



TESIS - SS255401

PENAKSIRAN PARAMETER PADA MODEL *GENERALIZED POISSON SPATIAL AUTOREGRESSIVE*

(Studi Kasus: Pemodelan Jumlah Kasus Pneumonia pada Balita di Kabupaten Tuban)

JOSHUA CAPRI GUNAWAN SIHOMBING

NRP 6003232015

DOSEN PEMBIMBING:

Dr. Sutikno, S.Si, M.Si.

NIP. 19710313 199702 1 001

Dr. Achmad Choiruddin, S.Si, M.Sc.

NIP. 19911210 202406 1 001

Program Magister

Departemen Statistika

Fakultas Sains dan Analitika Data

Institut Teknologi Sepuluh Nopember

Surabaya

2026



TESIS - SS255401

PENAKSIRAN PARAMETER PADA MODEL *GENERALIZED POISSON SPATIAL AUTOREGRESSIVE*

(Studi Kasus: Pemodelan Jumlah Kasus Pneumonia pada Balita di Kabupaten Tuban)

JOSHUA CAPRI GUNAWAN SIHOMBING
NRP 6003232015

Dosen Pembimbing

Dr. Sutikno, S.Si, M.Si.

NIP. 19710313 199702 1 001

Dr. Achmad Choiruddin, S.Si, M.Sc.

NIP. 19911210 202406 1 001

Program Magister

Departemen Statistika

Fakultas Sains dan Analitika Data

Institut Teknologi Sepuluh Nopember

Surabaya

2025



THESIS- SS255401

**PARAMETER ESTIMATION OF *GENERALIZED POISSON*
*SPATIAL AUTOREGRESSIVE***
**(Case Study: Modeling Pneumonia Cases in Toddlers
in Tuban Regency)**

JOSHUA CAPRI GUNAWAN SIHOMBING
NRP 6003232015

Advisors

Dr. Sutikno, S.Si, M.Si.
NIP. 19710313 199702 1 001
Dr. Achmad Choiruddin, S.Si, M.Sc.
NIP. 19911210 202406 1 001

Master Program

Departement of Statistics
Faculty of Science and Data Analytics
Institut Teknologi Sepuluh Nopember
Surabaya
2025

LEMBAR PENGESAHAN TESIS

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Statistika (M.Stat)

di

Institut Teknologi Sepuluh Nopember

Oleh:

JOSHUA CAPRI GUNAWAN SIHOMBING

NRP: 6003232015

Tanggal Ujian: 9 Januari 2026

Periode Wisuda: April 2026

Disetujui oleh:
Pembimbing:

1. Dr. Sutikno, S.Si., M.Si.
NIP: 19710313 199702 1 001

.....

2. Dr. Achmad Choiruddin, S.Si., M.Sc.
NIP: 19911210 202406 1 001

.....

Penguji:

1. Santi Puteri Rahayu, S.Si., M.Si., Ph.D.
NIP: 19750115 199903 2 003

.....

2. Prof. Drs. Nur Iriawan, M.Ikom., Ph.D.
NIP: 19621015 198803 1 002

.....

Kepala Departemen Statistika
Fakultas Sains dan Analitika Data



Dr. rer.pol. Dedy Dwi Prastyo, S.Si., M.Si. ✓
NIP: 19831204 200812 1 002

**PENAKSIRAN PARAMETER PADA MODEL *GENERALIZED
POISSON SPATIAL AUTOREGRESSIVE***
**(Studi Kasus: Pemodelan Jumlah Kasus Pneumonia pada Balita di
Kabupaten Tuban)**

Nama Mahasiswa : Joshua Capri Gunawan Sihombing
NRP : 6003232015
Departemen : Statistika FSAD - ITS
Dosen Pembimbing : Dr. Sutikno, S.Si, M.Si.
Co-Pembimbing : Dr. Achmad Choiruddin, S.Si, M.Sc.

ABSTRAK

Jumlah kasus pneumonia pada balita di Kabupaten Tuban menyajikan dua tantangan data yang signifikan, yaitu overdispersi dan dependensi spasial. Penelitian ini bertujuan untuk mengembangkan dan menerapkan model *Generalized Poisson Spatial Autoregressive* (GPSAR) untuk mengatasi kedua masalah tersebut secara bersamaan. Estimasi parameter model dilakukan menggunakan metode *Maximum Likelihood Estimation* dengan prosedur iterasi Berndt-Hall-Hall-Hausman (BHHH). Hasil penelitian mengonfirmasi validitas dan keunggulan model spasial yang diajukan. Model GPSAR terbukti lebih unggul dibandingkan model non-spasial *Generalized Poisson Regression* (GPR) dalam hal kebaikan model (AICc: 1673,39 vs 1717,24). Secara statistik, parameter struktural untuk lag spasial dan parameter dispersi bernilai signifikan ($p < 0,001$), yang mengonfirmasi adanya pengelompokan spasial yang kuat dan fenomena overdispersi pada data. Tiga variabel prediktor ditemukan berpengaruh signifikan: Cakupan Air Bersih (X_4) yang menunjukkan efek protektif (negatif), serta Kelas Ibu Hamil (X_3) dan Imunisasi Campak (X_5) yang menunjukkan asosiasi positif yang berpotensi berkaitan dengan kualitas surveilans kesehatan.

Kata Kunci: *Data Cacah, Model SAR, Overdispersi, Pneumonia, Spatial Lag.*

(Halaman ini sengaja dikosongkan)

**PARAMETER ESTIMATION IN THE GENERALIZED POISSON
SPATIAL AUTOREGRESSIVE MODEL**
(Case Study: Modeling Pneumonia Cases in Toddlers in Tuban Regency)

Student Name : Joshua Capri Gunawan Sihombing
NRP : 6003232015
Department : Statistics FSAD - ITS
Advisor : Dr. Sutikno, S.Si, M.Si.
Co-Advisor : Dr. Achmad Choiruddin, S.Si, M.Sc.

ABSTRACT

The number of pneumonia cases in children under five in Tuban Regency presents two significant data challenges: overdispersion and spatial dependency. This study aims to develop and apply the Generalized Poisson Spatial Autoregressive (GPSAR) model to address both issues simultaneously. The model parameters were estimated using the Maximum Likelihood Estimation method with the Berndt-Hall-Hausman (BHHH) iteration procedure. The results confirm the validity and superiority of the spatial model. The GPSAR model outperformed the non-spatial Generalized Poisson Regression (GPR) model in terms of goodness-of-fit (AICc: 1673,39 vs. 1717,24). Statistically, the structural parameters for spatial lag and dispersion were highly significant ($p < 0,001$), confirming strong spatial clustering and overdispersion in the data. Three predictor variables were found to be statistically significant determinants: Clean Water Coverage (X_4), which showed a protective effect, as well as Prenatal Classes (X_3) and Measles Immunization (X_5), which showed positive associations potentially linked to healthcare surveillance quality.

Kata Kunci: *Count Data, SAR Model, Overdispersion, Pneumonia, Spatial Lag.*

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa atas segenap karunia-Nya yang tak terhitung jumlahnya, sehingga penulis dapat menyelesaikan laporan Tugas Akhir dengan judul **“Penaksiran Parameter Pada Model *Generalized Poisson Spatial Autoregressive* (Studi kasus: Pemodelan Kasus Pneumonia Pada Balita di Kabupaten Tuban)”** dengan baik dan lancar. Penyusunan laporan Tesis ini dapat diselesaikan bukan tanpa bantuan serta dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih dan penghargaan setinggi-tingginya kepada:

1. Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis mendapatkan kesempatan untuk melanjutkan studi pada jenjang Magister.
2. Kedua orang tua tercinta yaitu Ayahanda Parsaoran Sihombing dan Ibunda Valentina Pasaribu yang selalu memberikan kasih sayang, nasehat, dukungan dalam berbagai bentuk, serta doa yang tiada hentinya kepada penulis.
3. Bapak Dr. Sutikno, S.Si, M.Si. selaku dosen pembimbing yang telah yang telah meluangkan waktunya serta dengan sangat sabar memberikan bimbingan dan banyak memberi masukan kepada penulis.
4. Bapak Dr. Achmad Choiruddin, S.Si, M.Sc. selaku dosen wali dan Co-pembimbing atas bimbingan, arahan, saran, dukungan serta motivasi selama perkuliahan serta penyusunan Tesis ini.
5. Bapak/Ibu staf dan karyawan di Departemen Statistika ITS, terima kasih atas segala bantuan selama masa perkuliahan penulis.
6. Bapak dan Ibu dosen Departemen Statistika FSAD ITS yang telah memberikan ilmu yang bermanfaat dan barokah.
7. Kepada teman-teman semuanya.

Penulis menyadari bahwa dalam tugas akhir ini masih belum sempurna. Semoga ketidak sempurnaan itu menjadi celah untuk proses belajar dan pengembangan keilmuan kedepannya. Semoga Tesis ini bisa menjadi bahan referensi dan memberikan manfaat bagi pengembangan dunia statistika.

Surabaya, September 2026

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	i
ABSTRAK	iii
ABSTRACT	v
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xv
DAFTAR NOTASI.....	xvii
BAB I LATAR BELAKANG	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	6
1.5 Batasan Penelitian	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Regresi Poisson	7
2.1.1 Model Regresi Poisson	7
2.1.2 Estimasi Parameter.....	9
2.1.3 Pengujian Hipotesis	11
2.2 Pengujian <i>Equidispersion</i>	13
2.3 Distribusi Generalized Poisson	14
2.3.1 Generalized Poisson Regression (GPR).....	14
2.3.2 Penaksir Parameter.....	16
2.3.3 Pengujian Hipotesis	19
2.4 Analisis Spasial	20
2.5 Poisson Spatial Autoregressive	21
2.5.1 Estimasi Parameter.....	21
2.5.2 Pengujian Hipotesis	24

2.6 Pengujian Aspek Spasial.....	26
2.6.1 Uji Dependensi Spasial	26
2.7 Generalized Poisson Spatial Autoregressive	29
2.8 Pembobot Spasial.....	31
2.9 Kriteria Pemilihan Model Terbaik	32
2.10 Pneumonia Pada Balita	32
2.10.1 Gambaran Umum Pneumonia Pada Balita	32
2.10.2 Faktor-Faktor yang Memengaruhi Kasus Pnomonia Pada Balita	33
2.11 Hasil Penelitian Terdahulu.....	35
BAB III METODOLOGI	37
3.1 Langkah-langkah Penaksir Parameter GPSAR.....	37
3.2 Langkah-langkah Pemodelan Kasus Pneumonia pada Balita.....	39
3.2.1 Sumber Data	39
3.2.2 Spesifikasi Model, Variabel dan Struktur Data Penelitian.....	39
3.2.3 Langkah-langkah Analisis Pemodelan GPSAR	41
BAB IV HASIL DAN PEMBAHASAN.....	43
4.1 Penaksiran Parameter GPSAR.....	43
4.2 Penerapan Model GPSAR	57
4.2.1 Gambaran Umum Data Penelitian.....	57
4.2.2 Pola Hubungan Variabel Respon dan Variabel Prediktor	62
4.2.3 Deteksi Multikolinieritas	62
4.2.4 Pengujian Equidispersi	63
4.2.5 Pemodelan <i>Generalized Poisson Regression</i>	63
4.2.6 Pembobotan Spasial dan Pengujian Dependensi Spasial	64
4.2.7 Penaksir Parameter Generalized Poisson Spatial Autoregressive ..	65
4.2.8 Pemilihan Model Terbaik	68
BAB V KESIMPULAN DAN SARAN	69
5.1 Kesimpulan	69
5.2 Saran	70
DAFTAR PUSTAKA	73

LAMPIRAN.....	77
BIODATA PENULIS.....	89

DAFTAR GAMBAR

Gambar 1.1	Peta Jumlah Kasus Sebaran Pneumonia Kabupaten Tuban	3
Gambar 1.2	<i>Roadmap</i> Penelitian.....	5
Gambar 2.1	Penyebab Kasus Pneumonia pada Balita	33
Gambar 2.2	Kerangka Konsep Penelitian	35
Gambar 3.1	Wilayah Administrasi Tuban Berdasarkan Desa	39
Gambar 3.2	Diagram Alir Analisis Data.....	42
Gambar 4.1	Sebaran Kasus Jumlah Pneumonia.....	59
Gambar 4.2	Sebaran Kasus Variabel Prediktor.....	60

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

Tabel 2.1 Penelitian tentang Pneumonia pada Balita	34
Tabel 2.2 Pemetaan Hasil Penelitian Terdahulu	36
Tabel 3.1 Variabel Penelitian.....	40
Tabel 3.2 Struktur Data GPSAR.....	41
Tabel 4.1 Statistik Deskriptf.....	57
Tabel 4.2 Korelasi Variabel Respon dan Prediktor	62
Tabel 4.3 Nilai VIF Masing-masing Variabel Prediktor	62
Tabel 4.4 Nilai Taksiran Parameter GPR	63
Tabel 4.5 Pengujian <i>Lagrange Multiplier</i> (LM).....	64
Tabel 4.6 Nilai Taksiran Parameter GPSAR	65
Tabel 4.7 Perbandingan nilai AICc	68

(Halaman ini sengaja dikosongkan)

DAFTAR NOTASI

α	: Taraf signifikan
n	: Banyak unit observasi
q	: Banyak variabel prediktor
p	: Banyak variabel respon
ρ	: Efek spasial lag
\mathbf{y}	: Vektor variabel pengamatan
$\boldsymbol{\beta}$: Vektor parameter koefisien regresi berukuran $(p + 1) \times 1$
I	: Nilai Moran'I
Ω	: Ruang parameter dibawah populasi
ω	: Ruang parameter dibawah H_0
i	: Indeks unit penelitian, $i = 1, 2, \dots, n$
j	: Indeks variabel respon, $j = 1$
k	: Indeks variabel prediktor untuk observasi ke- i
\mathbf{x}_i	: Vektor variabel prediktor untuk observasi ke- i
\mathbf{W}	: Matriks pembobot spasial
ψ	: <i>Eigenvalue</i> dari matriks pembobot spasial
F	: Statistik uji pada uji serentak
G^2	: Statistik uji <i>Glejser</i>
\mathbf{H}	: Matriks Hessian
W_k	: Statistik uji pada uji parsial
$L(\cdot)$: Fungsi <i>likelihood</i>
$\boldsymbol{\varepsilon}$: Vektor residual
$\ell(\cdot)$: Fungsi <i>ln-likelihood</i>

BAB I

LATAR BELAKANG

1.1 Latar Belakang

Pneumonia merupakan infeksi saluran pernapasan akut yang menyerang paru-paru dan hingga saat ini masih tercatat sebagai salah satu penyebab utama kematian pada anak di bawah usia lima tahun (balita) secara global. Penyakit ini memiliki prevalensi yang tinggi karena sistem kekebalan tubuh balita yang belum terbentuk sempurna, terutama di negara berkembang dengan akses layanan kesehatan yang terbatas. Organisasi kesehatan dunia seperti WHO dan UNICEF bahkan telah mempublikasikan *The Integrated Global Action Plan for Prevention and Control Pneumonia and Diarrhoea* (GAPPD) sebagai respons terhadap tingginya mortalitas ini. Hal ini menegaskan bahwa penanganan pneumonia bukan sekadar masalah medis rutin, melainkan prioritas kesehatan masyarakat yang mendesak.

Di Indonesia, beban penyakit ini juga signifikan. Berdasarkan data Profil Kesehatan Indonesia tahun 2022, pneumonia menyumbang angka kematian balita terbesar, yaitu mencapai 12,5%. Kondisi ini tercermin pula di tingkat daerah, khususnya di Provinsi Jawa Timur dan Kabupaten Tuban. Meskipun data Badan Pusat Statistik (BPS) menunjukkan adanya penurunan kasus di Kabupaten Tuban dari 1.707 kasus pada tahun 2021 menjadi 1.485 kasus pada tahun 2022 (turun 3,03%), angka absolut ribuan kasus ini menunjukkan bahwa ancaman tersebut masih nyata. Fluktuasi kasus yang tinggi di berbagai desa mengindikasikan perlunya pemodelan statistik yang presisi untuk memetakan faktor risikonya.

Dalam analisis statistika, data jumlah kasus kejadian penyakit seperti pneumonia dikategorikan sebagai data *count*. Pendekatan standar untuk memodelkan hubungan antara variabel respon berupa data *count* dengan variabel prediktor adalah Regresi Poisson. Model ini dibangun di atas asumsi fundamental yang dikenal sebagai *equidispersion*, di mana nilai rata-rata (*mean*) dari variabel respon diasumsikan sama dengan nilai variansnya. Penggunaan model ini sangat

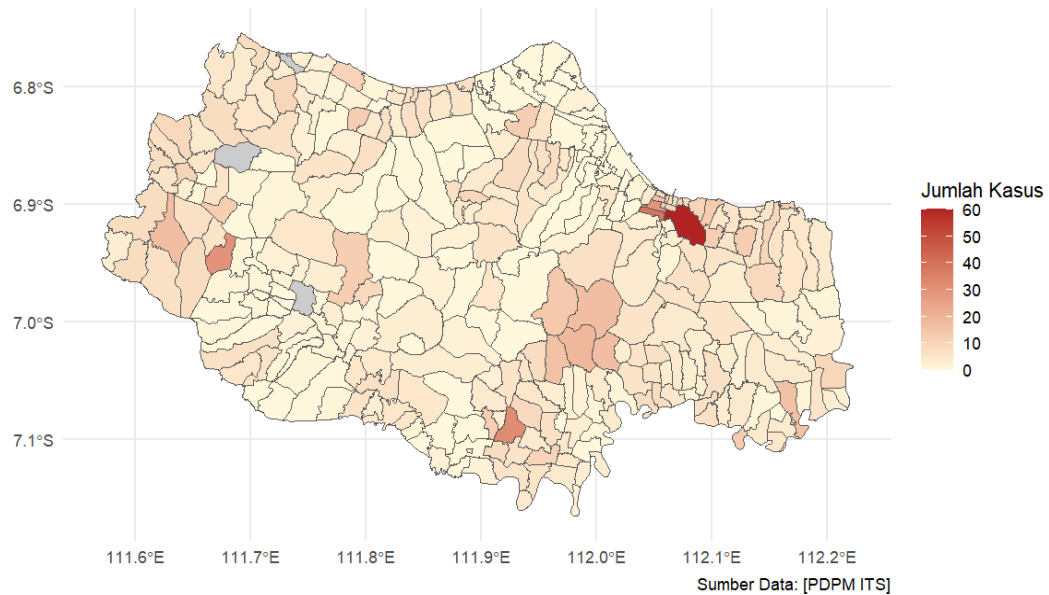
populer karena kesederhanaan interpretasinya dalam menjelaskan laju kejadian suatu peristiwa.

Namun, dalam penerapan pada data riil kesehatan, asumsi *equidispersion* tersebut sangat sering dilanggar. Pelanggaran ini terbukti secara empiris dalam data penelitian ini. Berdasarkan statistik deskriptif terhadap kasus pneumonia di 328 desa/kelurahan di Kabupaten Tuban, ditemukan karakteristik data yang ekstrem. Rata-rata jumlah kasus tercatat sebesar 4,527 kasus kejadian, sementara nilai variansnya melonjak hingga 60,29. Hal ini menghasilkan *Variance-to-Mean Ratio* sebesar 13,32, yang jauh lebih besar dari 1. Kondisi ini mengonfirmasi dugaan adanya *overdispersion*, yang jika dipaksakan menggunakan Regresi Poisson standar, akan menghasilkan kesimpulan yang tidak valid akibat *standard error* yang bias.

Untuk mengatasi keterbatasan Regresi Poisson dalam menangani *overdispersion*, diperlukan model alternatif yang lebih fleksibel. Salah satu pendekatan yang dapat digunakan adalah *Generalized Poisson Regression* (GPR). Keunggulan utama GPR dibandingkan model standar adalah kemampuannya menangani struktur data, baik yang mengalami *overdispersion* maupun *underdispersion*. Dengan menggunakan GPR, estimasi parameter menjadi lebih akurat karena model ini secara eksplisit mengakomodasi variabilitas data yang tinggi yang tidak tertangkap oleh model Poisson biasa.

Selain karakteristik distribusi data, analisis penyakit menular tidak dapat dilepaskan dari aspek lokasi atau geografis. Unit pengamatan dalam penelitian ini adalah desa/kelurahan, yang secara geografis saling berbatasan. Mengasumsikan bahwa kejadian penyakit di satu desa independen dengan desa tetangganya adalah asumsi yang tidak realistis. Hal ini sesuai dengan hukum pertama geografi Tobler yang menyatakan bahwa "segala sesuatu saling berhubungan, namun hal-hal yang dekat memiliki hubungan yang lebih erat daripada hal-hal yang jauh". Mengabaikan interaksi antarwilayah ini dapat menghilangkan informasi penting mengenai pola penyebaran penyakit.

Urgensi untuk memasukkan aspek spasial ke dalam model diperkuat oleh bukti visual dan statistik. Berdasarkan peta sebaran kasus pneumonia di Kabupaten Tuban pada Gambar 1.1 menunjukkan pola yang tidak acak, melainkan cenderung membentuk kluster atau pengelompokan di wilayah-wilayah tertentu. Pola ini menunjukkan adanya dependensi spasial yang jelas, di mana desa-desa yang berdekatan secara geografis cenderung memiliki jumlah kasus yang serupa.



Gambar 1.1 Sebaran Kasus Jumlah Pneumonia menurut desa/kelurahan di Kabupaten Tuban

Secara statistik, fenomena ini dikonfirmasi melalui Uji Moran's I yang menghasilkan $p\text{-value} < 0,001$. Hasil ini memberikan bukti mutlak bahwa terdapat dependensi spasial positif yang signifikan. Artinya, desa dengan kasus tinggi cenderung dikelilingi oleh desa dengan kasus tinggi pula, sehingga analisis spasial menjadi wajib dilakukan.

Meskipun metode spasial telah berkembang, terdapat kesenjangan dalam penerapannya untuk data spesifik ini. Penelitian terkini oleh Sutikno et al. (2025) di lokasi yang sama telah menggunakan *Multivariate Spatial Autoregressive* (MSAR) untuk menangani dependensi spasial. Namun, pendekatan tersebut dibangun untuk memodelkan persentase kasus. Asumsi normalitas tidak tepat digunakan untuk memodelkan data *count*.

Penerapan pemodelan SAR untuk kasus ril pada data *count* dilakukan oleh Rohimah (2015) untuk mengetahui faktor-faktor yang memengaruhi jumlah

penderita HIV di Provinsi Jawa Timur (Rohimah, 2015). Saat ini penerapan SAR untuk data *count* masih sebatas memodelkan dengan pendekatan distribusi Poisson yang mana artinya model belum mempertimbangkan kondisi saat asumsi *equidispersion* tidak terpenuhi. Oleh karena itu, pada penelitian ini dikembangkan model SAR untuk data *count* yang fleksibel terhadap asumsi *equidispersion* dengan data yang berdistribusi *Generalized Poisson*. Model ini selain mempertimbangkan aspek spasial juga mempertimbangkan aspek asumsi *equidispersion* yang tidak terpenuhi. Selain itu, penerapan model ini juga dilakukan pada kasus ril mengenai permasalahan kesehatan anak yaitu tentang penyakit pneumonia pada balita.

Berdasarkan kebutuhan untuk menangani *overdispersion* dan dependensi spasial secara simultan, penelitian ini bertujuan mengembangkan dan mengestimasi parameter model *Generalized Poisson Spatial Autoregressive* (GPSAR). Penelitian ini akan mengaplikasikan model tersebut untuk menentukan faktor-faktor yang memengaruhi kejadian pneumonia di Kabupaten Tuban secara presisi, melibatkan variabel prediktor seperti persentase pemberian ASI eksklusif, pemberian vitamin A, akses air bersih, serta balita yang mendapatkan imunisasi campak. Melalui pendekatan ini, diharapkan dihasilkan model yang tidak hanya valid secara statistik tetapi juga akurat dalam memberikan rekomendasi kebijakan kesehatan.

Penelitian ini mengkaji tentang penaksiran parameter model *Generalized Poisson Spatial Autoregressive* (GPSAR). Variabel respon yang digunakan dalam penelitian ini yaitu jumlah kasus pneumonia pada balita. Selanjutnya, penelitian ini juga menjelaskan tentang pengaplikasian model GPSAR pada variabel jumlah pneumonia pada balita di Kabupaten Tuban Tahun 2023.

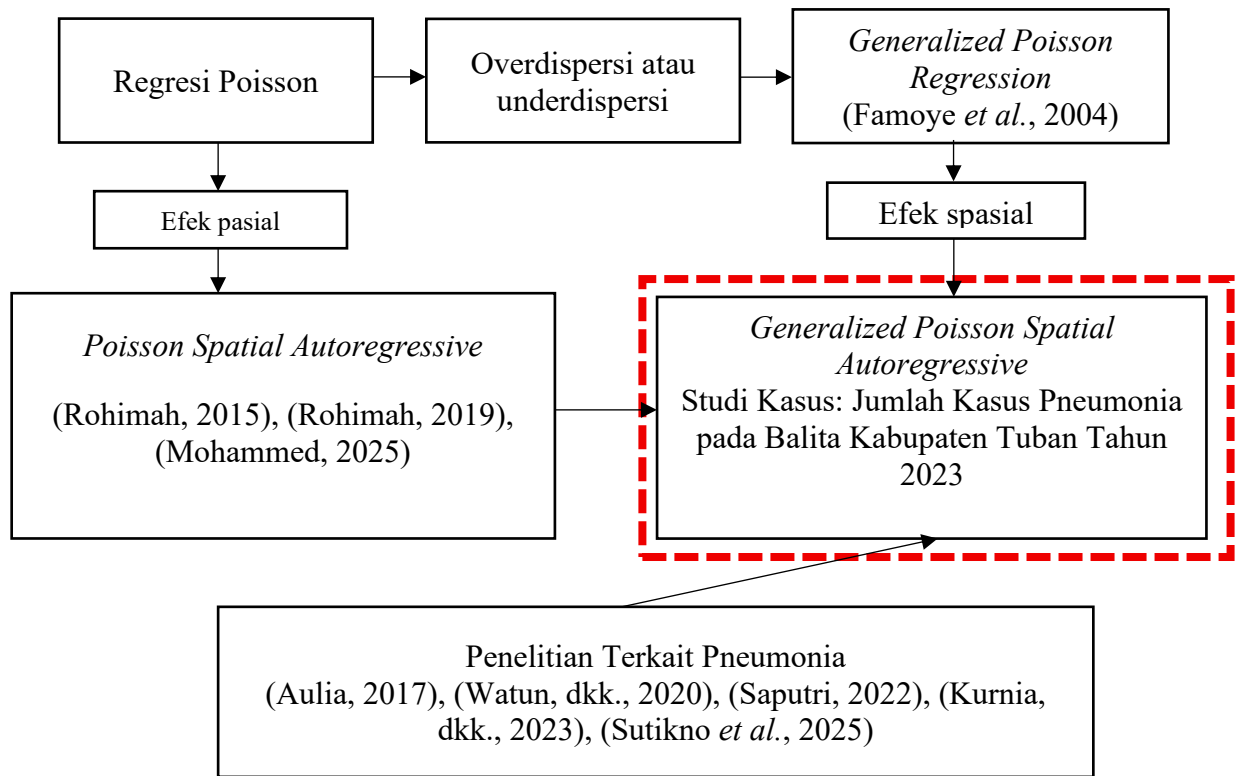
1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, data jumlah kasus pneumonia pada balita di Kabupaten Tuban tahun 2023 merupakan data *count* yang dihadapkan pada dua tantangan metodologis. Pertama, adanya bukti dependensi spasial antar desa (Uji Moran's I, $p\text{-value} < 0,001$). Kedua, adanya potensi ketidakterpenuhan asumsi *equidispersion*. Kedua kondisi tersebut tidak dapat ditangani secara memadai oleh model regresi Poisson standar. Oleh karena itu, penelitian ini mengusulkan pengembangan model *Generalized Poisson Spatial Autoregressive*

(GPSAR), yang mengintegrasikan kemampuan model *Generalized Poisson* (GP) dalam menangani permasalahan dispersi dengan kemampuan model *Spatial Autoregressive* (SAR) dalam mengakomodasi autokorelasi spasial, guna memperoleh hasil penaksiran dan pemodelan yang lebih akurat. Berdasarkan uraian tersebut, rumusan masalah dalam penelitian ini dirumuskan sebagai berikut:

1. Bagaimana bentuk penaksir parameter model GPSAR?
2. Bagaimana penerapan model GPSAR terhadap jumlah kasus pneumonia tiap desa di Kabupaten Tuban?

Roadmap penelitian disajikan pada Gambar 1.2.



Gambar 1.2 Roadmap Penelitian

Keterangan:

: yang diteliti

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan yang ingin dicapai pada penelitian ini sebagai berikut.

1. Mendapatkan bentuk penaksir parameter model GPSAR.
2. Mendapatkan model GPSAR terhadap jumlah kasus pneumonia tiap desa di Kabupaten Tuban.

1.4 Manfaat Penelitian

Manfaat penelitian ini terdiri atas dua bidang, yaitu keilmuan dan aplikasi. Manfaat keilmuan yang diperoleh adalah pengembangan ilmu statistika dengan dikembangkannya model GPSAR yaitu gabungan *Generalized Poisson Regression* dengan menambahkan efek spasial. Manfaat aplikasi adalah mendapatkan model GPSAR untuk menganalisis faktor-faktor yang mempengaruhi jumlah kasus pneumonia pada balita di Kabupaten Tuban. Hasil penelitian ini dapat dijadikan dasar pertimbangan dinas kesehatan di Kabupaten Tuban dalam upaya meningkatkan capaian penurunan jumlah pneumonia di Kabupaten Tuban.

1.5 Batasan Penelitian

Batasan penelitian diperlukan untuk mendapatkan hasil penelitian yang fokus dan maksimal, serta sesuai dengan rentang waktu yang telah direncanakan. Batasan-batasan penelitian diantaranya adalah sebagai berikut :

1. Metode pengujian hipotesis parameter yang digunakan adalah uji MLRT untuk pengujian hipotesis secara serentak dan uji *Wald* untuk pengujian hipotesis secara parsial.
2. Pengujian efek spasial menggunakan alat uji yang berbasis distribusi normal.
3. Pengujian aspek spasial yang di ujikan hanyalah dependensi spasial dengan *Moran's I*.

BAB II

TINJAUAN PUSTAKA

Bab ini menyajikan landasan teoretis yang menjadi dasar pengembangan model yang diusulkan. Pembahasan dimulai dari model fundamental untuk data *count*, yaitu Regresi Poisson, beserta asumsi *equidispersion* yang seringkali tidak terpenuhi. Sebagai solusi atas keterbatasan tersebut, akan diuraikan konsep GPR yang mampu menangani masalah *overdispersion*. Selanjutnya, akan dibahas keterbatasan model non-spasial dalam menganalisis data geografis, yang mengarah pada pengenalan model SAR untuk mengakomodasi dependensi spasial. Puncak dari tinjauan pustaka ini adalah sintesis dari konsep-konsep tersebut untuk membangun kerangka teoretis model GPSAR. Bab ini juga akan melengkapi pembahasan dengan konsep-konsep pendukung seperti pembobot spasial, pengujian asumsi, dan tinjauan penelitian terdahulu mengenai faktor-faktor yang memengaruhi kasus pneumonia pada balita.

2.1 Regresi Poisson

Analisis regresi Poisson mengasumsikan bahwa distribusi variabel respon Y yang digunakan adalah berdistribusi Poisson. Fungsi kepadatan peluang dengan parameter μ dinyatakan sebagai berikut (Cameron & Trivedi, 2013):

$$P(y_i|\mu_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}; y_i = 0, 1, \dots; 0 \leq \mu \leq \infty \quad (2.1)$$

Distribusi Poisson memiliki *mean* yang sama dengan *variance*, yaitu $E(y_i) = V(y_i) = \mu_i$, kondisi ini disebut sebagai *equidispersion*. Pada praktiknya kondisi *equidispersion* tidak selalu terpenuhi, seringkali nilai *variance* lebih kecil dari *mean* (*underdispersion*) atau nilai *variance* lebih besar dari *mean* (*overdispersion*).

2.1.1 Model Regresi Poisson

Regresi Poisson merupakan salah satu pendekatan yang dapat digunakan untuk melakukan analisis data *count*, dimana model yang terbentuk merupakan model non linier (Cameron & Trivedi, 2013). Jika y_i adalah data dari banyaknya suatu kejadian yang terjadi dalam periode waktu dan/atau wilayah tertentu, maka

regresi Poisson mengasumsikan y_i merupakan variabel random berdistribusi Poisson.

Model regresi Poisson merupakan salah satu contoh *Generalized Linear Model* (GLM). *Generalized Linear Model* terdiri dari tiga komponen yaitu,

1. komponen random, komponen sistematis, dan *link function* (McCullagh & Nelder, 1989). Komponen random terdiri dari variabel respon \mathbf{y} dengan nilai observasi $(y_1, y_2, \dots, y_n)^T$ yang independen.
2. Komponen sistematis untuk kovariat x_{ij} dimana $j = 1, 2, \dots, p$ menghasilkan suatu kombinasi linier variabel prediktor atau *linear predictor* $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$. \mathbf{X} adalah matriks dari variabel prediktor dan $\boldsymbol{\beta}$ adalah vektor dari parameter model.
3. *Link function* merupakan komponen yang menghubungkan komponen random dengan komponen sistematis, $E(\mathbf{y}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$.

Cara mendapatkan *link function* untuk model regresi Poisson adalah sebagai berikut.

- a) Logaritamkan kedua ruas persamaan (2.1)

$$\log P(y_i|\mu_i) = \log \left(\frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \right) = -\mu_i + y_i \log \mu_i - \log y_i!$$

- b) Eksponensialkan kedua ruas persamaan yang didapatkan pada tahap (a)

$$\exp(\log P(y_i|\mu_i)) = \exp(-\mu_i + y_i \log \mu_i - \log y_i!)$$

$$P(y_i|\mu_i) = \exp(-\mu_i + y_i \log \mu_i - \log y_i!)$$

- c) Menyatakan kembali persamaan yang didapatkan pada tahap (b) dalam bentuk umum keluarga distribusi eksponensial. Bentuk umum keluarga distribusi eksponensial adalah sebagai berikut:

$$P(\mathbf{y}, \boldsymbol{\theta}, \phi) = \exp \left\{ \frac{\mathbf{y}\boldsymbol{\theta} - b(\boldsymbol{\theta})}{a(\phi)} + c(\mathbf{y}, \phi) \right\} \quad (2.2)$$

sehingga

$$P(y_i|\mu_i) = \exp(-\mu_i + y_i \log \mu_i - \log y_i!)$$

$$P(y_i|\mu_i) = \exp(y_i \log \mu_i - \mu_i - \log y_i!)$$

Jadi, *link function* untuk model regresi Poisson adalah $\log(\mu_i)$, sehingga model yang terbentuk adalah:

$$\text{Log}(\mu_i) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$$

$$\mu_i = \exp(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}), i = 1, 2, \dots, n$$

atau jika dinyatakan dalam bentuk vektor menjadi

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

$$\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \quad (2.3)$$

dengan,

$$\mathbf{x}_i_{(p+1) \times 1} = [1 \quad x_{1i} \quad \cdots \quad x_{pi}]^T$$

$$\boldsymbol{\beta}_{(p+1) \times 1} = [\beta_0 \quad \beta_1 \quad \cdots \quad \beta_p]^T$$

2.1.2 Estimasi Parameter

Salah satu metode yang dapat digunakan untuk menaksir parameter regresi Poisson adalah *Maximum Likelihood Estimation* (MLE). Adapun langkah – langkah dari penaksiran parameter Regresi Poisson adalah sebagai berikut:

1. Menyusun fungsi *likelihood* untuk regresi Poisson.

$$\begin{aligned} L(\mu_i) &= \prod_{i=1}^n P(y_i | \mu_i) \\ &= \prod_{i=1}^n \left(\frac{\mu_i^{y_i} \exp(-\mu_i)}{y_i!} \right) \\ &= \frac{\prod_{i=1}^n \mu_i^{y_i} \exp(-\mu_i)}{\prod_{i=1}^n y_i!} \end{aligned} \quad (2.4)$$

2. Menyusun fungsi *ln likelihood*.

$$\ell(\mu_i) = \sum_{i=1}^n y_i \ln(\mu_i) - \sum_{i=1}^n \mu_i - \sum_{i=1}^n \ln(y_i!) \quad (2.5)$$

Jika $\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$, maka fungsi *ln likelihood* yang terbentuk:

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n y_i \ln(\exp(\mathbf{x}_i^T \boldsymbol{\beta})) - \sum_{i=1}^n \exp(\mathbf{x}_i^T \boldsymbol{\beta}) - \sum_{i=1}^n \ln(y_i!)$$

$$= \sum_{i=1}^n [y_i(\mathbf{x}_i^T \boldsymbol{\beta}) - \exp(\mathbf{x}_i^T \boldsymbol{\beta}) - \ln(y_i!)] \quad (2.6)$$

3. Menurunkan fungsi *ln likelihood* terhadap parameter yang ditaksir, yaitu $\boldsymbol{\beta}$.

Kemudian, hasil penurunan tersebut disamakan dengan nol.

$$\begin{aligned} \frac{\partial(\ell(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}} &= \frac{\partial(\sum_{i=1}^n [y_i(\mathbf{x}_i^T \boldsymbol{\beta}) - \exp(\mathbf{x}_i^T \boldsymbol{\beta}) - \ln(y_i!)])}{\partial \boldsymbol{\beta}} \\ 0 &= \sum_{i=1}^n [y_i \mathbf{x}_i - \mathbf{x}_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})] \end{aligned} \quad (2.7)$$

Solusi yang didapatkan dari persamaan turunan pertama *ln likelihood* tidak memiliki bentuk *closed-form*, estimasi parameter harus dilakukan melalui prosedur iterasi numerik. Metode iterasi numerik yang digunakan dalam penelitian ini adalah algoritma *Berndt-Hall-Hall-Hausman* (BHHH). Metode BHHH dipilih karena metode ini memanfaatkan informasi dari gradien turunan pertama untuk mengaproksimasi matriks Hessian, sehingga menghindari turunan kedua yang kompleks. Adapun prosedur iterasi BHHH didefinisikan melalui langkah-langkah berikut.

Definisi:

Misalkan $\boldsymbol{\beta}_{((p+1) \times 1)} = [\beta_0 \beta_1 \dots \beta_p]^T$ adalah vektor parameter yang akan di estimasi, $\ell_i(\boldsymbol{\beta})$ adalah *ln likelihood* untuk observasi ke- i , dimana $i = 1, 2, \dots, n$. Fungsi *ln likelihood* total adalah $\sum_{i=1}^n \ell_i(\boldsymbol{\beta})$

Inisialisasi:

1. Menentukan nilai awal parameter $\boldsymbol{\beta}^{(0)}$. Nilai awal parameter menentukan kecepatan tercapainya konvergensi pada proses iterasi. Idealnya nilai awal parameter untuk model regresi Poisson adalah 0 untuk semua parameter.
2. Menentukan toleransi konvergensi $\epsilon > 0$ yang mana dalam penelitian ini 10^{-3} menjadi batas toleransi konvergensi.
3. Menentukan batas maksimum iterasi (m_{maks}). Misalkan m_{maks} yang ditetapkan adalah 1000. Ketika iterasi tidak mencapai konvergensi pada iterasi ke-1000 maka proses iterasi akan berhenti walaupun konvergensi belum tercapai.
4. Menentukan set iterasi awal adalah $k = 0$.

Prosedur Iterasi:

Langkah 1: Menghitung vektor gradien individu untuk setiap observasi $i = 1, 2, \dots, n$.

$$g_i(\boldsymbol{\beta}^{(k)}) = \frac{\partial \ell_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(k)}}$$

Langkah 2: Menghitung penjumlahan gradien seluruh sampel.

$$g(\boldsymbol{\beta}^{(k)}) = \sum_{i=1}^n g_i(\boldsymbol{\beta}^{(k)})$$

Langkah 3: Mengaproksimasi matriks Hessian menggunakan informasi vektor gradien individu. Misalkan H^* adalah aproksimasi matriks Hessian, maka

$$H^*(\boldsymbol{\beta}^{(k)}) = - \sum_{i=1}^n g_i(\boldsymbol{\beta}^{(k)}) g_i(\boldsymbol{\beta}^{(k)})^T$$

Langkah 4: Memperbaharui nilai parameter menggunakan arah BHHH.

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} - H^{*-1}(\boldsymbol{\beta}^{(k)}) g(\boldsymbol{\beta}^{(k)})$$

Langkah 5: Mengevaluasi konvergensi dengan melihat kondisi berikut.

$$\|\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\beta}^{(k)}\| \leq \epsilon$$

Jika iterasi BHHH belum mencapai kondisi konvergen, set $k = k + 1$ dan kembali ke langkah 1. Jika kondisi konvergen telah tercapai, maka $\boldsymbol{\beta}^{(k+1)}$ adalah estimator dari $\hat{\boldsymbol{\beta}}$.

2.1.3 Pengujian Hipotesis

Pengujian hipotesis dapat dilakukan dengan menggunakan metode MLRT. Pengujian hipotesis pada model regresi Poisson dilakukan dengan dua tahap yaitu pengujian hipotesis secara serentak dan pengujian hipotesis secara parsial. Untuk pengujian hipotesis secara serentak memiliki hipotesis sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{minimal ada satu } \beta_k \neq 0, k = 1, 2, \dots, p$$

Statistik uji yang digunakan adalah sebagai berikut:

$$G^2 = -2 \ln \left[\frac{L(\hat{\boldsymbol{\omega}})}{L(\hat{\boldsymbol{\Omega}})} \right] = -2 [\ell(\hat{\boldsymbol{\omega}}) - \ell(\hat{\boldsymbol{\Omega}})]$$

Pada persamaan diatas $\ell(\hat{\omega})$ yaitu fungsi *ln likelihood* dari himpunan parameter di bawah H_0 dan $\ell(\hat{\Omega})$ yaitu fungsi *ln likelihood* dari himpunan parameter dibawah populasi. Himpunan parameter di bawah H_0 adalah $\omega = \{\beta_{0\omega}\}$ dan himpunan parameter dibawah populasi adalah $\Omega = \{\beta_0, \beta_1, \dots, \beta_p\}$. Bentuk fungsi *likelihood* dan *ln likelihood* dari $\hat{\omega} = \{\hat{\beta}_0\}$ yaitu:

$$\begin{aligned}
 L(\hat{\omega}) &= \prod_{i=1}^n \left(\frac{\exp(\hat{\beta}_{0\omega})^{y_i} \exp(-\exp(\hat{\beta}_{0\omega}))}{y_i!} \right) \\
 &= \frac{\prod_{i=1}^n \exp(\hat{\beta}_{0\omega})^{y_i} \exp(-\exp(\hat{\beta}_{0\omega})^{y_i})}{\prod_{i=1}^n y_i!} \\
 \ell(\hat{\omega}) &= \sum_{i=1}^n y_i \ln(\exp(\hat{\beta}_{0\omega})) - \sum_{i=1}^n \exp(\hat{\beta}_{0\omega}) - \sum_{i=1}^n \ln(y_i!) \\
 &= \sum_{i=1}^n [y_i(\hat{\beta}_{0\omega}) - \exp(\hat{\beta}_{0\omega}) - \ln(y_i!)] \tag{2.8}
 \end{aligned}$$

Kemudian bentuk fungsi *likelihood* dan *ln likelihood* dari himpunan parameter di bawah $\hat{\Omega} = \{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p\}$ yaitu:

$$\begin{aligned}
 L(\hat{\Omega}) &= \prod_{i=1}^n \left(\frac{\exp(\mathbf{x}_i^T \hat{\beta})^{y_i} \exp(-\exp(\mathbf{x}_i^T \hat{\beta}))}{y_i!} \right) \\
 &= \frac{\prod_{i=1}^n \exp(\mathbf{x}_i^T \hat{\beta})^{y_i} \exp(-\exp(\mathbf{x}_i^T \hat{\beta})^{y_i})}{\prod_{i=1}^n y_i!} \\
 \ell(\hat{\Omega}) &= \sum_{i=1}^n y_i \ln(\exp(\mathbf{x}_i^T \hat{\beta})) - \sum_{i=1}^n \exp(\mathbf{x}_i^T \hat{\beta}) - \sum_{i=1}^n \ln(y_i!) \\
 &= \sum_{i=1}^n [y_i(\mathbf{x}_i^T \hat{\beta}) - \exp(\mathbf{x}_i^T \hat{\beta}) - \ln(y_i!)] \tag{2.9}
 \end{aligned}$$

Dari persamaan (2.8) dan (2.9) selanjutnya diperoleh nilai G dengan bentuk sebagai berikut:

$$\begin{aligned}
 G^2 &= -2 \ln \left[\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right] = -2 [\ell(\hat{\omega}) - \ell(\hat{\Omega})] \\
 G^2 &= -2 [\sum_{i=1}^n [y_i(\hat{\beta}_{0\omega}) - \exp(\hat{\beta}_{0\omega}) - \ln(y_i!)] - \sum_{i=1}^n [y_i(\mathbf{x}_i^T \hat{\beta}) - \exp(\mathbf{x}_i^T \hat{\beta}) - \ln(y_i!)]]
 \end{aligned}$$

$$G^2 = 2[\sum_{i=1}^n [y_i(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) - \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) - \ln(y_i!)] - \sum_{i=1}^n [y_i(\hat{\beta}_{0\omega}) - \exp(\hat{\beta}_{0\omega}) - \ln(y_i!)] \quad (2.10)$$

Nilai G^2 dibandingkan dengan distribusi $\chi_{(p)}^2$ dimana H_0 akan ditolak jika nilai dari statistik $G^2 > \chi_{(\alpha;p)}^2$ dengan α adalah taraf signifikan yang digunakan.

Kemudian jika didapat keputusan tolak H_0 maka selanjutnya dilakukan pengujian parsial terhadap masing-masing parameter yang bertujuan untuk mengetahui parameter mana yang berpengaruh signifikan terhadap variabel respon secara individual. Statistik uji yang digunakan untuk uji parsial yaitu uji *Wald* dimana saat dibawah H_0 maka β_k sama dengan nol. Hipotesis dalam pengujian parsial adalah sebagai berikut.

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0; k = 1, 2, \dots, k, \dots, p$$

Untuk ukuran sampel yang cukup besar, penaksir β_k terdistribusi mendekati distribusi normal dengan rata-rata β_k dan varians β_k adalah elemen diagonal dari $-\mathbf{H}^{*-1}(\hat{\boldsymbol{\beta}})$ sehingga diperoleh $se(\hat{\beta}_k) = \sqrt{\widehat{Var}(\hat{\beta}_k)}$. Statistik uji *Wald* di bawah H_0 adalah seperti pada persamaan (2.11).

$$W_k = \left(\frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \right)^2 \quad (2.11)$$

Nilai W_k dibandingkan dengan distribusi $\chi_{(1)}^2$ dimana H_0 akan ditolak jika nilai dari statistik $W_k > \chi_{(\alpha;1)}^2$ dengan α adalah taraf signifikan yang digunakan.

2.2 Pengujian *Equidispersion*

Pada regresi Poisson, terdapat kondisi yang perlu dipenuhi, yaitu kondisi *equidispersion*. Kondisi *equidispersion* dapat dipenuhi jika nilai variansi sama dengan nilai rata-rata. Salah satu metode pengujian yang dapat digunakan untuk mengetahui adanya kasus *overdispersion* adalah uji *Deviance*. Langkah-langkah pengujian adalah sebagai berikut:

$$H_0 : \frac{V(y)}{E(y)} = 1 \text{ (Equidispersion)}$$

$H_1: \frac{V(y)}{E(y)} \neq 1$ (Terjadi kasus *overdispersion* atau *underdispersion*)

Statistik uji (Hilbe, 2014):

$$D = 2 \sum_{i=1}^n \left(y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right) \quad (2.12)$$

dengan $\hat{\mu}_i = e^{x_i^T \hat{\beta}}$.

Tolak H_0 jika $D > \chi_{\alpha, n-(p+1)}^2$. Kasus *overdispersion* terjadi ketika nilai dari *deviance* (D) dibagi dengan derajat bebas lebih besar dari 1, sedangkan kasus *underdispersion* terjadi jika nilai dari *deviance* (D) dibagi dengan derajat bebas lebih kecil dari 1.

2.3 Distribusi Generalized Poisson

Generalized Poisson (GP) memiliki dua parameter, yaitu μ dan ϕ sebagai parameter dispersi. Jika $\phi = 0$, maka terdapat kondisi *equidispersion*. Jika $\phi < 0$, maka terdapat kasus *underdispersion* dan jika $\phi > 0$, maka terdapat kasus *overdispersion* (Famoye *et al.*, 2004). Misalkan $y_i \sim GP(\mu_i, \phi)$ dengan $i = 1, 2, \dots, n$ merupakan sampel random maka fungsi distribusi peluang dari distribusi *Generalized Poisson* (GP) sebagai berikut (Noriszura, *et al.*, 2005):

$$P(y_i | \mu_i, \phi) = \left[\frac{\mu_i}{1 + \phi \mu_i} \right]^{y_i} \frac{(1 + \phi y_i)^{y_i - 1}}{y_i!} \exp \left(-\frac{\mu_i(1 + \phi y_i)}{1 + \phi \mu_i} \right), y_i = 0, 1, 2, \dots \quad (2.13)$$

dimana rata – rata dan variansi dari y_i adalah $E(y_i) = \mu_i$ dan $V(y_i) = \mu_i(1 + \phi \mu_i)^2$.

2.3.1 Generalized Poisson Regression (GPR)

Pada regresi Poisson terdapat beberapa asumsi yang harus dipenuhi, yaitu variabel respon harus berdistribusi Poisson, tidak terjadi multikolinearitas antar variabel prediktor, serta memiliki rata – rata dan variansi yang sama dari variabel respon. Namun, pada umumnya sering terjadi kasus dimana rata-rata dari variabel respon lebih besar daripada variansinya atau rata-rata lebih kecil daripada variansinya. Jika terjadi *underdispersion* dan *overdispersion*, maka dapat menyebabkan regresi Poisson tidak dapat digunakan. Salah satu pendekatan yang

dapat digunakan sebagai salah satu alternatif untuk mengatasi kasus tersebut adalah *Generalized Poisson Regression* (GPR).

GPR adalah *Generalized Linear Model* (GLM), sehingga untuk mendapatkan *link function* dari model GPR dilakukan langkah – langkah berikut.

- a. Logaritman kedua ruas persamaan (2.9).

$$\begin{aligned}\log(f(y_i|\mu_i, \phi)) &= \log\left(\left[\frac{\mu_i}{1 + \phi\mu_i}\right]^{y_i} \frac{(1 + \phi y_i)^{y_i-1}}{y_i!} \exp\left(-\frac{\mu_i(1 + \phi y_i)}{1 + \phi\mu_i}\right)\right) \\ \log(f(y_i|\mu_i, \phi)) &= \log\left[\frac{\mu_i}{1 + \phi\mu_i}\right]^{y_i} + \log\frac{(1 + \phi y_i)^{y_i-1}}{y_i!} \\ &\quad + \log \exp\left(-\frac{\mu_i(1 + \phi y_i)}{1 + \phi\mu_i}\right) \\ &= y_i \log \mu_i - y_i \log(1 + \phi y_i) + (y_i - 1) \log(1 + \phi y_i) \\ &\quad - \log(y_i!) - \frac{\mu_i(1 + \phi y_i)}{1 + \phi\mu_i} \\ &= y_i \log \mu_i + \log\frac{(1 + \phi y_i)^{y_i-1}}{(1 + \phi y_i)^{y_i}} - \log(y_i!) - \frac{\mu_i(1 + \phi y_i)}{1 + \phi\mu_i}\end{aligned}$$

- b. Eksponensialkan kedua ruas persamaan yang didapatkan pada tahap (a)

$$\begin{aligned}\exp\{\log(f(y_i|\mu_i, \phi))\} &= \frac{1}{y_i!} \exp\left\{y_i \log \mu_i + \log\frac{(1 + \phi y_i)^{y_i-1}}{(1 + \phi y_i)^{y_i}} - \frac{\mu_i(1 + \phi y_i)}{1 + \phi\mu_i}\right\} \\ &= \frac{1}{y_i!} \frac{(1 + \phi y_i)^{y_i-1}}{(1 + \phi y_i)^{y_i}} \exp\left\{-\frac{\mu_i(1 + \phi y_i)}{1 + \phi\mu_i}\right\}\end{aligned}$$

- c. Menyatakan kembali persamaan yang didapatkan pada tahap (b) dalam bentuk umum keluarga distribusi eksponensial. Bentuk umum keluarga distribusi eksponensial adalah sebagai berikut:

$$P(\mathbf{y}, \boldsymbol{\theta}, \phi) = \exp\left\{\frac{\mathbf{y}\boldsymbol{\theta} - b(\boldsymbol{\theta})}{a(\phi)} + c(\mathbf{y}, \phi)\right\}$$

sehingga

$$P(y_i|\mu_i, \phi) = \frac{1}{y_i!} \frac{(1 + \phi y_i)^{y_i-1}}{(1 + \phi \mu_i)^{y_i}} \exp \left\{ y_i \log \mu_i - \frac{\mu_i(1 + \phi y_i)}{1 + \phi \mu_i} \right\}$$

Jadi, dapat diketahui bahwa *link function* untuk model GPR adalah $\log(\mu_i)$ sehingga bentuk umum dari model GPR adalah:

$$\log(\mu_i) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}$$

$$\mu_i = \exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}), i = 1, 2, \dots, n$$

atau

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

$$\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \quad (2.14)$$

Dengan,

$$\mathbf{x}_{i(p+1) \times 1} = [1 \quad x_{1i} \quad \dots \quad x_{pi}]^T$$

$$\boldsymbol{\beta}_{(p+1) \times 1} = [\beta_0 \quad \beta_1 \quad \dots \quad \beta_p]^T$$

2.3.2 Penaksir Parameter

Penaksiran parameter dari model GPR dapat dilakukan dengan *Maximum Likelihood Estimation* (MLE). Adapun langkah – langkah dari penaksiran parameter tersebut adalah sebagai berikut:

1. Menyusun fungsi *likelihood* untuk regresi Poisson.

$$\begin{aligned} L(\mu_i, \phi) &= \prod_{i=1}^n f(y_i|\mu_i, \phi) \\ &= \prod_{i=1}^n \left[\frac{\mu_i}{1 + \phi \mu_i} \right]^{y_i} \frac{(1 + \phi y_i)^{y_i-1}}{y_i!} \exp \left(-\frac{\mu_i(1 + \phi y_i)}{1 + \phi \mu_i} \right) \end{aligned} \quad (2.15)$$

2. Menyusun fungsi *ln likelihood*.

$$\ell(\mu_i, \phi) = \ln \left(\prod_{i=1}^n \left[\frac{\mu_i}{1 + \phi \mu_i} \right]^{y_i} \frac{(1 + \phi y_i)^{y_i-1}}{y_i!} \exp \left(-\frac{\mu_i(1 + \phi y_i)}{1 + \phi \mu_i} \right) \right)$$

Jika $\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$, maka fungsi *ln likelihood* yang terbentuk:

$$\begin{aligned}
\ell(\boldsymbol{\beta}, \phi) &= \ln \left(\prod_{i=1}^n \left[\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \phi \exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right]^{y_i} \frac{(1 + \phi y_i)^{y_i-1}}{y_i!} \right) \\
&= \sum_{i=1}^n \left[y_i \mathbf{x}_i^T \boldsymbol{\beta} - y_i \ln(1 + \phi \exp(\mathbf{x}_i^T \boldsymbol{\beta})) + (y_i - 1) \ln(1 + \phi y_i) \right. \\
&\quad \left. - \ln y_i! - \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta}) (1 + \phi y_i)}{1 + \phi \exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right]
\end{aligned} \tag{2.16}$$

3. Menurunkan fungsi *ln likelihood* terhadap parameter yang ditaksir, yaitu $\boldsymbol{\beta}$ dan ϕ . Kemudian, hasil penurunan tersebut disamakan dengan nol.

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \left[\begin{array}{c} \mathbf{x}_i^T y_i - \frac{y_i \phi \exp(\mathbf{x}_i^T \boldsymbol{\beta})}{(1 + \phi \exp(\mathbf{x}_i^T \boldsymbol{\beta}))} \\ -(1 + \phi y_i) \left(\begin{array}{c} \frac{\mathbf{x}_i^T \exp(\mathbf{x}_i^T \boldsymbol{\beta})}{(1 + \phi \exp(\mathbf{x}_i^T \boldsymbol{\beta}))} \\ - \frac{\phi \mathbf{x}_i^T \exp(2\mathbf{x}_i^T \boldsymbol{\beta})}{(1 + \phi \exp(\mathbf{x}_i^T \boldsymbol{\beta}))^2} \end{array} \right) \end{array} \right] \tag{2.17}$$

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \phi} = \sum_{i=1}^n \left[\begin{array}{c} \frac{y_i \phi \exp(\mathbf{x}_i^T \boldsymbol{\beta})}{(1 + \phi \exp(\mathbf{x}_i^T \boldsymbol{\beta}))} + \frac{y_i(y_i - 1)}{(1 + \phi y_i)} \\ - \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \left(\frac{y_i}{(1 + \phi \exp(\mathbf{x}_i^T \boldsymbol{\beta}))} - \frac{(1 + \phi y_i) \exp(\mathbf{x}_i^T \boldsymbol{\beta})}{(1 + \phi \exp(\mathbf{x}_i^T \boldsymbol{\beta}))^2} \right) \end{array} \right] \tag{2.18}$$

Solusi yang didapatkan dari persamaan turunan pertama *ln likelihood* tidak memiliki bentuk *closed-form*, sehingga estimasi parameter harus dilakukan melalui prosedur iterasi numerik. Metode iterasi numerik yang digunakan dalam penelitian ini adalah algoritma BHHH. Metode BHHH dipilih karena metode ini memanfaatkan informasi dari gradien turunan pertama untuk mengaproksimasi matriks Hessian, sehingga menghindari turunan kedua yang kompleks. Adapun prosedur iterasi BHHH didefinisikan melalui langkah-langkah berikut.

Definisi: Misalkan $\boldsymbol{\theta}_{((p+2) \times 1)} = [\beta_0 \ \beta_1 \ \dots \ \beta_p \ \phi]^T$ adalah vektor parameter yang akan di estimasi, $\ell_i(\boldsymbol{\theta})$ adalah *ln likelihood* untuk observasi ke- i , dimana $i = 1, 2, \dots, n$. Fungsi *ln likelihood* total adalah $\sum_{i=1}^n \ell_i(\boldsymbol{\theta})$

Inisialisasi:

1. Menentukan nilai awal parameter $\theta^{(0)}$. Nilai awal parameter menentukan kecepatan tercapainya konvergensi pada proses iterasi. Idealnya nilai awal parameter untuk model *Generalized Poisson Regression* adalah nilai hasil estimasi parameter regresi Poisson.
2. Menentukan toleransi konvergensi $\epsilon > 0$ yang mana dalam penelitian ini 10^{-6} menjadi batas toleransi konvergensi.
3. Menentukan batas maksimum iterasi (m_{maks}). Misalkan m_{maks} yang ditetapkan adalah 1000. Ketika iterasi tidak mencapai konvergensi pada iterasi ke-1000 maka proses iterasi akan berhenti walaupun konvergensi belum tercapai.
4. Menentukan set iterasi awal adalah $k = 0$.

Prosedur Iterasi:

Langkah 1: Menghitung vektor gradien individu untuk setiap observasi $i = 1, 2, \dots, n$.

$$g_i(\theta^{(k)}) = \left. \frac{\partial \ell_i(\theta)}{\partial \theta} \right|_{\theta=\theta^{(k)}}$$

Langkah 2: Menghitung penjumlahan gradien seluruh sampel.

$$g(\theta^{(k)}) = \sum_{i=1}^n g_i(\theta^{(k)})$$

Langkah 3: Mengaproksimasi matriks Hessian menggunakan informasi vektor gradien individu. Misalkan H^* adalah aproksimasi matriks Hessian, maka

$$H^*(\theta^{(k)}) = - \sum_{i=1}^n g_i(\theta^{(k)}) g_i(\theta^{(k)})^T$$

Langkah 4: Memperbaharui nilai parameter menggunakan arah BHHH.

$$\theta^{(k+1)} = \theta^{(k)} - H^{*-1}(\theta^{(k)}) g(\theta^{(k)})$$

Langkah 5: Mengevaluasi konvergensi dengan melihat kondisi berikut.

$$\|\theta^{(k+1)} - \theta^{(k)}\| \leq \epsilon$$

Jika iterasi BHHH belum mencapai kondisi konvergen, set $k = k + 1$ dan kembali ke langkah 1. Jika kondisi konvergen telah tercapai, maka $\theta^{(k+1)}$ adalah estimator dari $\hat{\theta}$.

2.3.3 Pengujian Hipotesis

Pengujian hipotesis dapat dilakukan dengan menggunakan metode MLRT. Pengujian hipotesis pada model regresi Poisson dilakukan dengan dua tahap yaitu pengujian hipotesis secara serentak dan pengujian hipotesis secara parsial. Pengujian parameter regresi secara serentak dapat menggunakan *Likelihood ratio Test*. Untuk pengujian hipotesis secara serentak memiliki hipotesis sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{minimal ada satu } \beta_k \neq 0, k = 1, 2, \dots, p$$

Statistik uji yang digunakan adalah sebagai berikut:

$$G^2 = -2 \ln \left[\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right] = -2 [\ell(\hat{\omega}) - \ell(\hat{\Omega})]$$

Pada persamaan diatas $\ell(\hat{\omega})$ adalah fungsi *likelihood* dari himpunan parameter di bawah H_0 dan $\ell(\hat{\Omega})$ adalah fungsi *likelihood* dari himpunan parameter dibawah populasi. Himpunan parameter dibawah H_0 adalah $\omega = \{\beta_0, \phi\}$ dan himpunan parameter dibawah populasi adalah $\Omega = \{\beta_0, \beta_1, \dots, \beta_p, \phi\}$. Bentuk fungsi *likelihood* dan *ln likelihood* dari $\hat{\omega} = \{\hat{\beta}_{0\omega}, \hat{\phi}\}$ yaitu:

$$\begin{aligned} L(\hat{\omega}) &= \prod_{i=1}^n \left[\frac{\exp(\hat{\beta}_{0\omega})}{1 + \phi \exp(\hat{\beta}_{0\omega})} \right]^{y_i} \frac{(1 + \phi y_i)^{y_i-1}}{y_i!} \exp \left(- \frac{\exp(\hat{\beta}_{0\omega}) (1 + \phi y_i)}{1 + \phi \exp(\hat{\beta}_{0\omega})} \right) \\ \ell(\hat{\omega}) &= \sum_{i=1}^n \left[y_i \hat{\beta}_{0\omega} - y_i \ln(1 + \phi \exp(\hat{\beta}_{0\omega})) + (y_i - 1) \ln(1 + \phi y_i) \right. \\ &\quad \left. - \ln y_i! - \frac{\exp(\hat{\beta}_{0\omega}) (1 + \phi y_i)}{1 + \phi \exp(\hat{\beta}_{0\omega})} \right] \end{aligned} \quad (2.19)$$

Kemudian bentuk fungsi *likelihood* dan *ln likelihood* dari himpunan parameter di bawah $\hat{\Omega} = \{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p, \hat{\phi}\}$ yaitu:

$$\begin{aligned} L(\hat{\Omega}) &= \prod_{i=1}^n \left[\frac{\exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})}{1 + \phi \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})} \right]^{y_i} \frac{(1 + \phi y_i)^{y_i-1}}{y_i!} \exp \left(- \frac{\exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) (1 + \phi y_i)}{1 + \phi \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})} \right) \\ \ell(\hat{\Omega}) &= \sum_{i=1}^n \left[y_i \mathbf{x}_i^T \hat{\boldsymbol{\beta}} - y_i \ln(1 + \phi \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})) + (y_i - 1) \ln(1 + \phi y_i) \right. \\ &\quad \left. - \ln y_i! - \frac{\exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) (1 + \phi y_i)}{1 + \phi \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})} \right] \end{aligned} \quad (2.20)$$

Dari persamaan (2.19) dan (2.20) selanjutnya diperoleh nilai G dengan penjabaran sebagai berikut:

$$\begin{aligned}
 G &= -2 \ln \left[\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right] = -2 [\ell(\hat{\omega}) - \ell(\hat{\Omega})] \\
 &= 2 \left[\sum_{i=1}^n \left(\begin{aligned} &\left[y_i \mathbf{x}_i^T \hat{\beta} - y_i \ln(1 + \phi \exp(\mathbf{x}_i^T \hat{\beta})) + (y_i - 1) \ln(1 + \phi y_i) \right] \\ &- \ln y_i! - \frac{\exp(\mathbf{x}_i^T \hat{\beta}) (1 + \phi y_i)}{1 + \phi \exp(\mathbf{x}_i^T \hat{\beta})} \end{aligned} \right) \right. \\
 &\quad \left. - \left[y_i \hat{\beta}_0 - y_i \ln(1 + \phi \exp(\hat{\beta}_0)) + (y_i - 1) \ln(1 + \phi y_i) \right] \right. \\
 &\quad \left. - \ln y_i! - \frac{\exp(\hat{\beta}_0) (1 + \phi y_i)}{1 + \phi \exp(\hat{\beta}_0)} \right] \right] \quad (2.21)
 \end{aligned}$$

Tolak H_0 jika $G^2 > \chi^2_{(\alpha; p)}$.

Kemudian jika didapat keputusan tolak H_0 selanjutnya dilakukan pengujian parsial terhadap parameter yang bertujuan untuk mengetahui parameter mana saja yang berpengaruh signifikan terhadap variabel respon secara individual. Statistik uji yang digunakan untuk uji parsial yaitu uji *Wald* dimana saat dibawah H_0 maka β_k sama dengan nol. Hipotesis dalam pengujian parsial adalah sebagai berikut.

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0; k = 1, 2, \dots, k, \dots, p$$

Untuk ukuran sampel yang cukup besar, penaksir β_k terdistribusi mendekati distribusi normal dengan rata-rata β_k dan varians β_k adalah elemen diagonal dari $-\mathbf{H}^{*-1}(\hat{\beta})$ sehingga diperoleh $se(\hat{\beta}_k) = \sqrt{\widehat{var}(\hat{\beta}_k)}$. Statistik uji *Wald* di bawah H_0 adalah seperti pada persamaan (2.22).

$$W_k = \left(\frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \right)^2 \quad (2.22)$$

Nilai W_k dibandingkan dengan distribusi $\chi^2_{(1)}$ dimana H_0 akan ditolak jika nilai dari statistik $W_k > \chi^2_{(\alpha; 1)}$ dengan α adalah taraf signifikan yang digunakan.

2.4 Analisis Spasial

Spasial merupakan hal – hal yang berhubungan dengan aspek kewilayahan, sedangkan data spasial adalah data yang observasinya merupakan suatu lokasi atau

wilayah. Keterkaitan antar wilayah sering kali dipengaruhi oleh lokasi yang saling berdekatan, sehingga diperlukan analisis spasial untuk mengetahui pengaruh hubungan antara lokasi dengan respon.

2.5 Poisson Spatial Autoregressive

Pengembangan model spasial telah banyak dilakukan, salah satunya adalah pengembangan model spasial untuk data *count*. Jika fungsi *link Poisson Spatial Autoregressive* (PSAR) didefinisikan dalam persamaan (2.23), dan variabel respon berdistribusi Poisson, yaitu:

$$y_i \sim \text{Poisson}(\mu_i); i = 1, 2, \dots, n$$

Model PSAR yang standar menerapkan lag spasial pada prediktor linear, yaitu (Lambert et al. 2010):

$$\ln \mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sum_{j \neq i}^n \rho w_{ij} \ln \mu_j$$

Dalam notasi matriks, persamaan di atas menjadi:

$$\begin{aligned} \ln \boldsymbol{\mu} &= \mathbf{X}\boldsymbol{\beta} + \rho \mathbf{W} \ln \boldsymbol{\mu} \\ \ln \boldsymbol{\mu} - \rho \mathbf{W} \ln \boldsymbol{\mu} &= \mathbf{X}\boldsymbol{\beta} \\ (\mathbf{I} - \rho \mathbf{W}) \ln \boldsymbol{\mu} &= \mathbf{X}\boldsymbol{\beta} \\ \ln \boldsymbol{\mu} &= (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{X}\boldsymbol{\beta} \\ \ln \boldsymbol{\mu} &= \mathbf{A}^{-1} \mathbf{X}\boldsymbol{\beta} \\ \boldsymbol{\mu}^{PSAR} &= \exp(\mathbf{A}^{-1} \mathbf{X}\boldsymbol{\beta}) \end{aligned} \tag{2.23}$$

dengan $\mathbf{A} = (\mathbf{I} - \rho \mathbf{W})$ dan \mathbf{A}^{-1} adalah matriks invers dari matriks \mathbf{A} , \mathbf{X} adalah matriks prediktor berukuran $n \times (p + 1)$, $\boldsymbol{\beta}$ adalah parameter koefisien regresi berukuran $(p + 1) \times 1$, dan ρ adalah parameter koefisien spasial *lag* pada variabel respon, dan \mathbf{W} adalah matriks pembobot berukuran $(n \times n)$.

2.5.1 Estimasi Parameter

Salah satu metode yang dapat digunakan untuk menaksir parameter PSAR adalah *Maximum Likelihood Estimation* (MLE). Parameter yang di estimasi adalah $\boldsymbol{\theta}_{PSAR} = [\beta_0 \ \beta_1 \ \dots \ \beta_p \ \rho]^T$. Adapun langkah – langkah dari penaksiran parameter PSAR adalah sebagai berikut:

1. Menyusun fungsi *likelihood* untuk regresi *Poisson Spatial Autoregressive*.

$$\begin{aligned}
L(\mu_i^{PSAR}) &= \prod_{i=1}^n P(y_i | \mu_i^{PSAR}) \\
&= \prod_{i=1}^n \left(\frac{(\mu_i^{PSAR})^{y_i} \exp(-\mu_i^{PSAR})}{y_i!} \right) \\
&= \frac{\prod_{i=1}^n (\mu_i^{PSAR})^{y_i} \exp(-\mu_i^{PSAR})}{\prod_{i=1}^n y_i!}
\end{aligned} \tag{2.24}$$

2. Menyusun fungsi *ln likelihood*.

$$\ell(\mu_i^{PSAR}) = \sum_{i=1}^n y_i \ln(\mu_i^{PSAR}) - \sum_{i=1}^n \mu_i^{PSAR} - \sum_{i=1}^n \ln(y_i!) \tag{2.25}$$

Jika $\mu_i^{PSAR} = \exp(\mathbf{A}^{-1} \mathbf{X} \boldsymbol{\beta})_i$, dimana $(\mathbf{A}^{-1} \mathbf{X} \boldsymbol{\beta})_i$ adalah baris ke- i dari vektor $(\mathbf{A}^{-1} \mathbf{X} \boldsymbol{\beta})$ maka fungsi *ln likelihood* yang terbentuk:

$$\begin{aligned}
\ell(\boldsymbol{\theta}_{PSAR}) &= \sum_{i=1}^n y_i \ln((\mathbf{A}^{-1} \mathbf{X} \boldsymbol{\beta})_i) - \sum_{i=1}^n (\mathbf{A}^{-1} \mathbf{X} \boldsymbol{\beta})_i - \sum_{i=1}^n \ln(y_i!) \\
&= \sum_{i=1}^n [y_i (\mathbf{A}^{-1} \mathbf{X} \boldsymbol{\beta}) - \exp(\mathbf{A}^{-1} \mathbf{X} \boldsymbol{\beta}) - \ln(y_i!)]
\end{aligned} \tag{2.26}$$

3. Menurunkan fungsi *ln likelihood* terhadap parameter yang ditaksir, yaitu $\boldsymbol{\beta}$.

Kemudian, hasil penurunan tersebut disamakan dengan nol.

$$\begin{aligned}
\frac{\partial(\ell(\boldsymbol{\theta}_{PSAR}))}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n [y_i - \exp(\mathbf{A}^{-1} \mathbf{X} \boldsymbol{\beta})_i] (\mathbf{A}^{-1} \mathbf{X})^T \\
\frac{\partial(\ell(\boldsymbol{\theta}_{PSAR}))}{\partial \rho} &= \sum_{i=1}^n [y_i - \exp(\mathbf{A}^{-1} \mathbf{X} \boldsymbol{\beta})_i] (\mathbf{A}^{-1} \mathbf{W} \mathbf{A}^{-1} \mathbf{X} \boldsymbol{\beta})
\end{aligned} \tag{2.27}$$

Solusi yang didapatkan dari persamaan turunan pertama *ln likelihood* tidak memiliki bentuk *closed-form*, sehingga estimasi parameter harus dilakukan melalui prosedur iterasi numerik. Metode iterasi numerik yang digunakan dalam penelitian ini adalah BHHH. Metode BHHH dipilih karena metode ini memanfaatkan informasi dari gradien turunan pertama untuk mengaproksimasi matriks Hessian, sehingga menghindari turunan kedua yang kompleks. Adapun prosedur iterasi BHHH didefinisikan melalui langkah-langkah berikut.

Definisi:

Misalkan $\boldsymbol{\theta}_{PSAR} = [\beta_0 \beta_1 \dots \beta_p \rho]^T$ adalah vektor parameter yang akan di estimasi, $\ell_i(\boldsymbol{\theta}_{PSAR})$ adalah *ln likelihood* untuk observasi ke- i , dimana $i = 1, 2, \dots, n$. Fungsi *ln likelihood* total adalah $\sum_{i=1}^n \ell_i(\boldsymbol{\theta}_{PSAR})$

Inisialisasi:

1. Menentukan nilai awal parameter $\boldsymbol{\theta}_{PSAR}^{(0)}$ Nilai awal parameter menentukan kecepatan tercapainya konvergensi pada proses iterasi. Idealnya nilai awal parameter untuk model *Generalized Poisson Regression* adalah nilai hasil estimasi parameter regresi Poisson.
2. Menentukan toleransi konvergensi $\epsilon > 0$ yang mana dalam penelitian ini 10^{-6} menjadi batas toleransi konvergensi.
3. Menentukan batas maksimum iterasi (m_{maks}) . Misalkan m_{maks} yang ditetapkan adalah 1000. Ketika iterasi tidak mencapai konvergensi pada iterasi ke-1000 maka proses iterasi akan berhenti walaupun konvergensi belum tercapai.
4. Menentukan set iterasi awal adalah $k = 0$.

Prosedur Iterasi:

Langkah 1: Menghitung vektor gradien individu untuk setiap observasi $i = 1, 2, \dots, n$.

$$\boldsymbol{g}_i(\boldsymbol{\theta}^{(k)}) = \left. \frac{\partial \ell_i(\boldsymbol{\theta}_{PSAR})}{\partial \boldsymbol{\theta}_{PSAR}} \right|_{\boldsymbol{\theta}_{PSAR} = \boldsymbol{\theta}_{PSAR}^{(k)}}$$

Langkah 2: Menghitung penjumlahan gradien seluruh sampel.

$$\boldsymbol{g}(\boldsymbol{\theta}_{PSAR}^{(k)}) = \sum_{i=1}^n \boldsymbol{g}_i(\boldsymbol{\theta}_{PSAR}^{(k)})$$

Langkah 3: Mengaproksimasi matriks Hessian menggunakan informasi vektor gradien individu. Misalkan \boldsymbol{H}^* adalah aproksimasi matriks Hessian, maka

$$\boldsymbol{H}^*(\boldsymbol{\theta}_{PSAR}^{(k)}) = - \sum_{i=1}^n \boldsymbol{g}_i(\boldsymbol{\theta}_{PSAR}^{(k)}) \boldsymbol{g}_i(\boldsymbol{\theta}_{PSAR}^{(k)})^T$$

Langkah 4: Memperbaharui nilai parameter menggunakan arah BHHH.

$$\boldsymbol{\theta}_{PSAR}^{(k+1)} = \boldsymbol{\theta}_{PSAR}^{(k)} - \boldsymbol{H}^{*-1}(\boldsymbol{\theta}_{PSAR}^{(k)}) \boldsymbol{g}(\boldsymbol{\theta}_{PSAR}^{(k)})$$

Langkah 5: Mengevaluasi konvergensi dengan melihat kondisi berikut.

$$\left\| \boldsymbol{\theta}_{PSAR}^{(k+1)} - \boldsymbol{\theta}_{PSAR}^{(k)} \right\| \leq \epsilon$$

Jika iterasi BHHH belum mencapai kondisi konvergen, set $k = k + 1$ dan kembali ke langkah 1. Jika kondisi konvergen telah tercapai, maka $\boldsymbol{\theta}_{PSAR}^{(k+1)}$ adalah estimator dari $\hat{\boldsymbol{\theta}}_{PSAR}$.

2.5.2 Pengujian Hipotesis

Pengujian hipotesis dapat dilakukan dengan menggunakan metode MLRT. Pengujian hipotesis pada model *Poisson Spatial Autoregressive* dilakukan dengan dua tahap, yaitu pengujian hipotesis secara serentak dan pengujian hipotesis secara parsial. Untuk pengujian hipotesis secara serentak memiliki hipotesis sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{minimal ada satu } \beta_k \neq 0, k = 1, 2, \dots, p$$

Statistik uji yang digunakan adalah sebagai berikut:

$$G^2 = -2 \ln \left[\frac{L(\hat{\boldsymbol{\omega}})}{L(\hat{\boldsymbol{\Omega}})} \right] = -2 [\ell(\hat{\boldsymbol{\omega}}) - \ell(\hat{\boldsymbol{\Omega}})]$$

Pada persamaan diatas $\ell(\hat{\boldsymbol{\omega}})$ yaitu fungsi *ln likelihood* dari himpunan parameter di bawah H_0 dan $\ell(\hat{\boldsymbol{\Omega}})$ yaitu fungsi *ln likelihood* dari himpunan parameter dibawah populasi. Himpunan parameter di bawah H_0 adalah $\boldsymbol{\omega} = \{\beta_0, \rho_{\omega}\}$ dan himpunan parameter dibawah populasi adalah $\boldsymbol{\Omega} = \{\beta_0, \beta_1, \dots, \beta_p, \rho\}$.

Bentuk fungsi *likelihood* dari $\hat{\boldsymbol{\omega}}$ yaitu:

$$\begin{aligned} L(\hat{\boldsymbol{\omega}}) &= \prod_{i=1}^n \left(\frac{\exp(\hat{\beta}_{0\omega} + \sum_{j=1}^n w_{ij} y_j)^{y_i} \exp(\hat{\beta}_{0\omega} + \sum_{j=1}^n w_{ij} y_j)}{y_i!} \right) \\ &= \frac{\prod_{i=1}^n \exp(\hat{\beta}_{0\omega} + \sum_{j=1}^n w_{ij} y_j)^{y_i} \exp(\hat{\beta}_{0\omega} + \sum_{j=1}^n w_{ij} y_j)}{\prod_{i=1}^n y_i!} \end{aligned}$$

$$\ell(\hat{\boldsymbol{\omega}}) = \sum_{i=1}^n [y_i (\hat{\beta}_{0\omega} + \sum_{j=1}^n w_{ij} y_j) - \exp(\hat{\beta}_{0\omega} + \sum_{j=1}^n w_{ij} y_j) - \ln(y_i!)] \quad (2.28)$$

Kemudian bentuk fungsi *likelihood* dari himpunan parameter di bawah $\hat{\boldsymbol{\theta}} = \hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ yaitu:

$$L(\hat{\boldsymbol{\Omega}}) = \prod_{i=1}^n \left(\frac{\exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})^{y_i} \exp(-\exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}))}{y_i!} \right)$$

$$\begin{aligned}
&= \frac{\prod_{i=1}^n \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})^{y_i} \exp(-\exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})^{y_i})}{\prod_{i=1}^n y_i!} \\
\ell(\hat{\boldsymbol{\Omega}}) &= \sum_{i=1}^n y_i \ln(\exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})) - \sum_{i=1}^n \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) - \sum_{i=1}^n \ln(y_i!) \\
&= \sum_{i=1}^n [y_i(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) - \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) - \ln(y_i!)] \tag{2.29}
\end{aligned}$$

Dari persamaan (2.8) dan (2.9) selanjutnya diperoleh nilai G dengan bentuk sebagai berikut:

$$G^2 = -2 \ln \left[\frac{L(\hat{\boldsymbol{\omega}})}{L(\hat{\boldsymbol{\Omega}})} \right] = -2[\ell(\hat{\boldsymbol{\omega}}) - \ell(\hat{\boldsymbol{\Omega}})] \underset{n \rightarrow \infty}{\overset{H_0}{\sim}} \chi_{df}^2 \tag{2.30}$$

Nilai G^2 dibandingkan dengan distribusi $\chi_{(df)}^2$ dimana H_0 akan ditolak jika nilai dari statistik $G^2 > \chi_{(\alpha; df)}^2$ dengan α adalah taraf signifikan yang digunakan dan $df = n(\boldsymbol{\Omega}) - n(\boldsymbol{\omega}) = (p + 2) - 2 = p$.

Kemudian jika didapat keputusan tolak H_0 selanjutnya dilakukan pengujian parsial terhadap parameter yang bertujuan untuk mengetahui parameter mana saja yang berpengaruh signifikan terhadap variabel respon secara individual. Statistik uji yang digunakan untuk uji parsial yaitu uji *Wald* dimana saat dibawah H_0 maka β_k sama dengan nol. Hipotesis dalam pengujian parsial untuk β_k adalah sebagai berikut.

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0; k = 1, 2, \dots, k, \dots, p$$

Untuk ukuran sampel yang cukup besar, penaksir β_k terdistribusi mendekati distribusi normal dengan rata-rata β_k dan varians β_k adalah elemen diagonal dari $-\mathbf{H}^{*-1}(\hat{\boldsymbol{\beta}})$ yang bersesuaian dengan $\hat{\beta}_k$. Sehingga diperoleh $se(\hat{\beta}_k) = \sqrt{\widehat{Var}(\hat{\beta}_k)}$.

Statistik uji *Wald* di bawah H_0 adalah seperti pada persamaan (2.31).

$$W_k = \left(\frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \right)^2 \tag{2.31}$$

Nilai W_k dibandingkan dengan distribusi $\chi^2_{(1)}$ dimana H_0 akan ditolak jika nilai dari statistik $W_\rho > \chi^2_{(\alpha;1)}$ dengan α adalah taraf signifikan yang digunakan. Kemudian untuk uji parsial parameter ρ hipotesis dalam pengujian parsialnya adalah sebagai berikut.

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Statistik uji didapatkan dengan uji *Wald* yang dapat dinyatakan sebagai berikut.

$$W_\rho = \left(\frac{\hat{\rho}}{se(\hat{\rho})} \right)^2 \quad (2.32)$$

Nilai W_ρ dibandingkan dengan distribusi $\chi^2_{(1)}$ dimana H_0 akan ditolak jika nilai dari statistik $W_\rho > \chi^2_{(\alpha;1)}$ dengan α adalah taraf signifikan yang digunakan. Nilai $se(\hat{\rho})$ adalah $\sqrt{\widehat{Var}(\hat{\rho})}$ dan $\widehat{Var}(\hat{\rho})$ merupakan elemen diagonal utama dari $-H^{*-1}(\hat{\theta})$ yang bersesuaian dengan ρ .

2.6 Pengujian Aspek Spasial

Efek spasial dapat dikategorikan menjadi dua, yaitu dependensi spasial dan heterogenitas spasial. Dependensi spasial menunjukkan adanya ketergantungan atau keterkaitan antar lokasi pengamatan, dimana nilai pengamatan di suatu lokasi dipengaruhi oleh nilai pengamatan di lokasi lain yang letaknya berdekatan. Kemudian, heterogenitas spasial dapat menunjukkan adanya perbedaan karakteristik pada setiap lokasi pengamatan sehingga setiap lokasi dapat memiliki bentuk fungsi dan parameter yang berbeda – beda. Dalam penelitian ini kita akan fokus dengan menguji dependensi spasial yang diuji dengan menggunakan Moran's I dan Lagrange Multiplier (LM).

2.6.1 Uji Dependensi Spasial

Dependensi spasial dapat diketahui melalui *Moran's I*. Indeks *moran's I* digunakan untuk menguji dependensi spasial. Jika indeks *moran's I* bernilai positif maka terdapat dependensi spasial yang positif yang artinya wilayah dengan nilai pengamatan tinggi cenderung berdekatan dengan wilayah yang memiliki nilai pengamatan tinggi, begitu juga wilayah dengan nilai rendah cenderung berdekatan dengan wilayah nilai rendah. Selanjutnya, jika indeks *moran's I* bernilai negatif

maka terdapat dependensi spasial yang negatif yang artinya wilayah dengan nilai tinggi cenderung berdekatan dengan wilayah nilai rendah, dan sebaliknya. Metode ini tidak menguji autokorelasi pada data mentah, melainkan pada Deviance Residuals dari model *Generalized Linear Model* (GLM) global.

Pendekatan ini didasarkan pada sifat teoretis bahwa *deviance residuals* ($r_{\{i,d\}}$) dari model GLM (seperti Poisson atau *Generalized Poisson*) terdistribusi secara asimtotik normal (Z-score) ketika ukuran sampel cukup besar. Dengan demikian, residual ini memenuhi syarat distribusi untuk dijadikan input dalam statistik Moran's I standar. Statistik uji ini, yang dinotasikan sebagai $I_{\{DR\}}$, didefinisikan sebagai adaptasi dari formula Moran's I klasik dengan mensubstitusi variabel amatan dengan *deviance residuals* (Zhang & Lin, 2008):

$$I_{DR} = \frac{n}{S_0} \frac{(\sum_{i=1}^n \sum_{i^*=1}^n w_{ii^*} (r_{i,d} - \bar{r}_d)(r_{i^*,d} - \bar{r}_d))}{\sum_{i=1}^n (r_{i,d} - \bar{r}_d)^2} \quad (2.33)$$

dimana I_{DR} adalah adaptasi dari *Moran's I* yang menggunakan residual devians sebagai input, w_{ii^*} adalah elemen pembobot spasial antara lokasi i dan lokasi i^* . Biasanya $w_{ii^*} = 1$ jika i dan i^* bertetangga, dan 0 jika tidak (dengan $w_{ii} = 0$), $r_{i,d}$ adalah residual devians pada lokasi i , $r_{i^*,d}$ adalah residual devians pada lokasi i^* , \bar{r}_d adalah rata-rata residual devians, dan $S_0 = \sum_{i=1}^n \sum_{i^*=1}^n w_{ii^*}$.

Residual devians dapat dinyatakan sebagai $r_{i,d} = \text{sign}(n_i - \hat{n}_i) \sqrt{d_i^{GP}}$ dimana $\text{sign}(\cdot)$ adalah fungsi tanda yang didefinisikan sebagai $\text{sign}(a) = 1$ jika $a \geq 0$, dan $\text{sign}(a) = -1$ jika $a < 0$, $d_i^{GP} = 2 \left[n_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) + (y_i - 1) \log \left(\frac{y_i(1+\phi)}{\hat{\mu}_i + \phi y_i} \right) \right]$ untuk $y_i \neq 0$ dan $d_i^{GP} \approx 2[\hat{\mu}_i - \log(1 + \phi \hat{\mu}_i)]$ untuk $y_i = 0$.

Matriks pembobot spasial yang digunakan pada uji dependensi spasial adalah matriks pembobot dengan pendekatan *contiguity* (persinggungan). Terdapat tiga macam *contiguity* untuk membentuk matriks pembobot spasial, yaitu persinggungan sisi (*rook*), persinggungan sudut (*bishop*), dan persinggungan sisi dan sudut (*queen*). Elemen matriks pembobot spasial berdasarkan kriteria – kriteria tersebut adalah:

$$w_{ii^*} \begin{cases} 1, & \text{jika } i \text{ dan } i^* \text{ bersinggungan} \\ 0, & \text{jika } i \text{ dan } i^* \text{ tidak bersinggungan} \end{cases}$$

Hipotesis pengujian signifikansi untuk indeks *moran's I* adalah:

$H_0: I = 0$ (Tidak terdapat dependensi spasial)

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

dengan statistik uji:

$$Z_{I_{DR}} = \frac{I_{DR} - E(I_{DR})}{\sqrt{Var(I_{DR})}} \quad (2.34)$$

Adapun nilai harapan (ekspektasi) dari statistik *Moran's I*, dinotasikan sebagai $E(I_{DR}) = -\frac{1}{n-1}$, Sementara itu, variansnya, yaitu $V(I)$, diperoleh melalui $V(I_{DR}) = \frac{m[(m^2-3m+3)S_1-mS_2+3S_0^2]-b_2[(m^2-m)S_1-2mS_2+6S_0^2]}{(m-1)(m-2)(m-3)S_0^2} - [E(I_{DR})]^2$. Aspek ini bergantung pada tiga komponen utama yang didasarkan pada matriks pembobot spasial (w). Komponen pertama, S_1 , adalah setengah dari jumlah total kuadrat dari penjumlahan elemen matriks pembobot dan transposnya, atau dapat dinyatakan sebagai:

$$S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{i^*=1}^n (w_{ii^*} + w_{i^*i})^2$$

Komponen kedua, S_2 , adalah jumlah dari kuadrat penjumlahan setiap baris dan kolom matriks pembobot, yang dirumuskan sebagai

$$S_2 = \sum_{i=1}^n \left(\sum_{i^*=1}^n w_{ii^*} + \sum_{i^*=1}^n w_{i^*i} \right)^2$$

dan

$$b^2 = \frac{n \sum_{i=1}^n (r_{i,d} - \bar{r}_d)^4}{\left(\sum_{i=1}^n (r_{i,d} - \bar{r}_d)^2 \right)^2}$$

Tolak H_0 jika $|Z_{I(hitung)}| > Z_{\frac{\alpha}{2}}$, artinya terdapat dependensi spasial, sehingga dapat dilanjutkan pemodelan regresi spasial.

a. Lagrange Multiplier (LM)

Uji Lagrange Multiplier (LM) digunakan untuk menentukan ada atau tidaknya efek dependensi spasial, serta menentukan model spasial yang akan

dibentuk. Pengujian hipotesis dalam uji LM salah satunya adalah model *Spatial Autoregressive Model* (SAR).

$H_0 : \rho = 0$ (tidak ada efek dependensi spasial *autoregressive* pada model)

$H_1 : \rho \neq 0$ (terdapat efek spasial dependensi spasial *autoregressive*)

Statistik Uji :

$$LM_1 = \frac{\left(\frac{\mathbf{e}'\mathbf{W}\mathbf{y}}{S^2}\right)^2}{(nJ)} \quad (2.35)$$

Adapun nilai $nJ = T + \frac{(\mathbf{W}\mathbf{X}\boldsymbol{\beta})'\mathbf{M}(\mathbf{W}\mathbf{X}\boldsymbol{\beta})}{s^2}$, Dalam persamaan ini, komponen T merupakan nilai *trace* dari perkalian matriks pembobot spasial, yang dihitung dengan rumus $T = \text{tr}[(\mathbf{W} + \mathbf{W}')\mathbf{W}]$. Selanjutnya, komponen s^2 adalah varians galat yang didefinisikan sebagai $s^2 = \frac{\mathbf{e}'\mathbf{e}}{n}$, dan adapun matriks \mathbf{M} yang digunakan pada bagian pembilang adalah matriks proyeksi (matriks idempoten dan simetris) yang diperoleh dari $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Tolak H_0 jika $LM_1 > \chi^2_{(1;1-\alpha)}$.

2.7 Generalized Poisson Spatial Autoregressive

Model dari bentuk umum dari model *Generalized Poisson Spatial Autoregressive* didasarkan pada formulasi multiplikatif untuk menangani sifat simultan dari interaksi spasial dan masalah pelanggaran asumsi *equidispersion*. Model ini mengasumsikan bahwa nilai harapan di suatu lokasi (μ_i) bergantung pada kovariat di lokasi tersebut dan pada nilai harapan di lokasi-lokasi tetangganya (μ_j). Formulasi spasial multiplikatif untuk data cacah adalah sebagai berikut (Lambert et al. 2010):

$$E(y_i) \equiv \mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \prod_{j \neq i}^n [E(y_j)]^{\rho w_{ij}}$$

Karena $E(y_j) \equiv \mu_j$, maka berlaku

$$\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \prod_{j \neq i}^n \mu_j^{\rho w_{ij}}$$

Bentuk multiplikatif ini sulit untuk dioperasionalkan secara langsung dalam analisis dan estimasi. Untuk memudahkan penyelesaian aljabar, model tersebut perlu

ditransformasi ke dalam bentuk linear dengan menerapkan fungsi logaritma natural pada kedua ruas, sehingga menghasilkan model struktural berikut:

$$\ln(\mu_i) = \ln \left(\exp(\mathbf{x}_i^T \boldsymbol{\beta}) \prod_{j \neq i}^n \mu_j^{\rho w_{ij}} \right)$$

$$\ln(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \sum_{j \neq i}^n \rho w_{ij} \ln \mu_j$$

Untuk generalisasi model pada seluruh lokasi observasi, persamaan di atas diekspresikan dalam notasi matriks. Transformasi ini menghasilkan bentuk umum dari model GPSAR:

$$\begin{aligned} \ln \boldsymbol{\mu} &= \mathbf{X}\boldsymbol{\beta} + \rho \mathbf{W} \ln \boldsymbol{\mu} \\ \ln \boldsymbol{\mu} - \rho \mathbf{W} \ln \boldsymbol{\mu} &= \mathbf{X}\boldsymbol{\beta} \\ (\mathbf{I} - \rho \mathbf{W}) \ln \boldsymbol{\mu} &= \mathbf{X}\boldsymbol{\beta} \\ \ln \boldsymbol{\mu} &= (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{X}\boldsymbol{\beta} \\ \ln \boldsymbol{\mu} &= \mathbf{A}^{-1} \mathbf{X}\boldsymbol{\beta} \end{aligned}$$

dengan $\mathbf{A} = (\mathbf{I} - \rho \mathbf{W})$ dan \mathbf{A}^{-1} adalah matriks invers dari matriks \mathbf{A} . Dari penurunan di atas, diperoleh bentuk *reduced form* dari rata-rata kondisional untuk model GPSAR, yaitu:

$$\boldsymbol{\mu}^{GPSAR} = \exp(\mathbf{A}^{-1} \mathbf{X}\boldsymbol{\beta}) \quad (2.36)$$

$$\mu_i^{GPSAR} = \exp(\mathbf{A}^{-1} \mathbf{X}\boldsymbol{\beta})_i \quad (2.37)$$

Dimana $\boldsymbol{\mu}$ adalah vektor ekspektasi rata-rata berukuran $(n \times 1)$ dimana n adalah jumlah total lokasi atau area observasi, ρ adalah parameter koefisien spasial lag pada variabel respon, \mathbf{W} adalah matrik pembobot berukuran $(n \times n)$, $(\mathbf{A}^{-1} \mathbf{X}\boldsymbol{\beta})_i$ Baris ke- i dari vektor $(\mathbf{A}^{-1} \mathbf{X}\boldsymbol{\beta})$, \mathbf{X} merupakan matriks yang memuat nilai-nilai dari semua variabel prediktor yang berukuran $(n \times (p + 1))$ di mana k adalah jumlah variabel prediktor dan kolom tambahan adalah untuk *intercept*, $\boldsymbol{\beta}$ adalah vektor kolom yang berisi koefisien regresi yang akan diestimasi oleh model yang berukuran $((p + 1) \times 1)$.

$$y_{n \times 1} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}, \boldsymbol{\beta}_{(p+1) \times 1} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \mathbf{W}_{n \times n} = \begin{pmatrix} w_{11} & w_{12} & w_{13} & \cdots & w_{1n} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & w_{n3} & \cdots & w_{nn} \end{pmatrix}$$

$$\mathbf{X}_{n \times (p+1)} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

Persamaan (2.34) merepresentasikan model SAR yang telah diadaptasi untuk data count dengan asumsi distribusi *Generalized Poisson* dalam bentuk vektor dan persamaan (2.35) merepresentasikan model SAR yang telah diadaptasi untuk data count dengan asumsi distribusi *Generalized Poisson* dalam bentuk unit observasi.

2.8 Pembobot Spasial

Matriks pembobot spasial (\mathbf{W}) diperoleh melalui informasi jarak dari ketetanggaan (*neighbourhood*) atau kedekatan antara satu wilayah dengan wilayah yang lain. Wilayah yang berdekatan cenderung akan memberikan pengaruh yang lebih besar dibandingkan wilayah yang lebih jauh jaraknya, hal ini merupakan hukum pertama geografi yang dinyatakan oleh Tobler (Anselin, 1988). Terdapat beberapa metode yang dapat digunakan dalam mendefinisikan hubungan persinggungan (*contiguity*) antar wilayah.

1. *Rook Contiguity* (persinggungan sisi); mendefinisikan $w_{ij} = 1$ untuk wilayah yang bersisian (*common side*) dengan wilayah yang menjadi perhatian, $w_{ij} = 0$ untuk wilayah lainnya.
2. *Bhisop Contiguity* (persinggungan sudut); mendefinisikan $w_{ij} = 1$ untuk wilayah yang titik sudutnya (*common vertex*) bertemu dengan sudut wilayah yang menjadi perhatian, $w_{ij} = 0$ untuk wilayah lainnya.
3. *Queen Contiguity* (persinggungan sisi-sudut); mendefinisikan $w_{ij} = 1$ untuk wilayah yang bersisian (*common side*) atau titik sudutnya (*common vertex*) bertemu dengan wilayah yang menjadi perhatian, $w_{ij} = 0$ untuk wilayah lainnya.

Dari berbagai macam metode dalam pembobotan, pembobotan *queen contiguity* merupakan pembobotan yang lebih kompleks atau lengkap. Dengan

mempertimbangkan karakteristik antar wilayah desa/kelurahan kabupaten Tuban yang tidak simetris maka penelitian ini digunakan pembobotan *queen contiguity*. Hal ini karena bentuk dari tiap daerah cocok dengan mekanisme pembobotan *queen contiguity*.

2.9 Kriteria Pemilihan Model Terbaik

Dalam pemilihan model regresi terbaik, akan digunakan AICc yang mempertimbangkan ukuran sampel dengan meningkatkan penalty relative untuk model yang kompleks dengan ukuran data yang kecil. Formula dari AICc adalah sebagai berikut (Burnham & Anderson, 2004):

$$AICc = -2 \log \left(L(\hat{\theta}) \right) + 2p + \frac{2p(p+1)}{n-p-1} \quad (2.38)$$

dimana p adalah jumlah parameter dan n adalah jumlah observasi. Model terbaik adalah model dengan nilai AICc terkecil (Akaike, 1998)

2.10 Pneumonia Pada Balita

Studi kasus dalam penelitian ini adalah kasus pneumonia pada balita. Oleh karena itu, pada subab ini akan dijelaskan tentang gambaran umum pneumonia pada balita dilanjutkan mengenai faktor-faktor yang diduga mempengaruhi kasus pneumonia pada balita.

2.10.1 Gambaran Umum Pneumonia Pada Balita

Balita dapat diartikan sebagai anak usia di bawah usia lima tahun. Periode ini merupakan periode penting dalam perkembangan dan pertumbuhan karena dapat menentukan keberhasilan pertumbuhan dan perkembangan di periode selanjutnya. Oleh karena itu, penyakit-penyakit yang diderita balita pada masa ini dapat menjadi penghambat dalam pertumbuhan dan perkembangan.

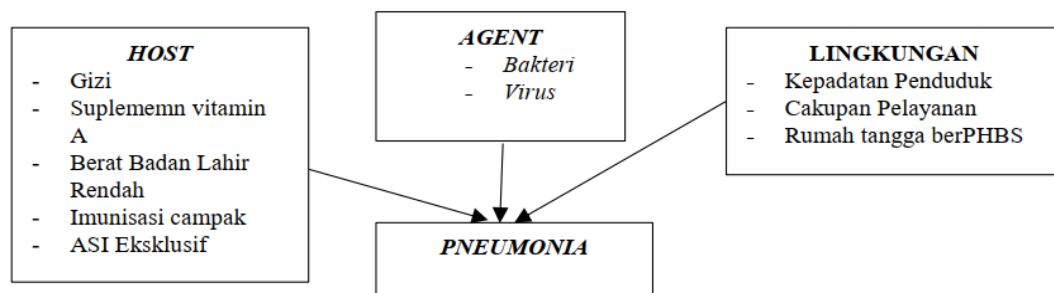
Pneumonia merupakan penyakit menular berupa peradangan pada daerah saluran pernapasan bagian bawah yaitu paru-paru. Pneumonia disebabkan oleh beberapa agen infeksi seperti virus, bakteri dan jamur. Agen infeksi pneumonia yang paling umum adalah bakteri *Streptococcus pneumoniae*, *Haemophilus influenzae* tipe b, dan virus pernapasan *syncytial* (WHO, 2022).

Gejala pneumonia pada anak diawali dengan infeksi saluran pernapasan atas (hidung dan tenggorokan). Selain itu, pneumonia pada anak juga ditandai dengan gejala diare apabila penyebabnya adalah bakteri (Tim Medis Siloam Hospitals, 2024).

2.10.2 Faktor-Faktor yang Diduga Memengaruhi Kasus Pnomonia Pada Balita

Pneumonia lebih sering terjadi pada balita. Hal ini dikarenakan sistem kekebalan tubuh pada balita masih lemah. Kerangka berpikir untuk menganalisis faktor-faktor yang diduga memengaruhi kasus pneumonia dari *the Epidemiologic Triangle through infectious disease* (Gordon, 1950)) seperti Gambar 2.1.

Dalam menentukan faktor-faktor yang mempengaruhi penyakit dilakukan pembagian menjadi 3 bagian, yaitu berdasarkan *host*, *agent* dan *environment*. Faktor *host* merupakan faktor-faktor dalam diri manusia yang mempengaruhi timbulnya penyakit. Faktor *agent* adalah suatu substansi tertentu yang kehadirannya dapat menimbulkan penyakit. Terakhir, faktor *environment* atau lingkungan yaitu faktor ekstrinsik yang menunjang terjadinya suatu penyakit. Faktor lingkungan mempengaruhi penularan, penyebaran dan perkembangan suatu *agent* penyebab penyakit. Penelitian ini hanya berfokus pada faktor *host* dan *environment*.



Gambar 2.1 Penyebab Kasus Pneumonia pada Balita

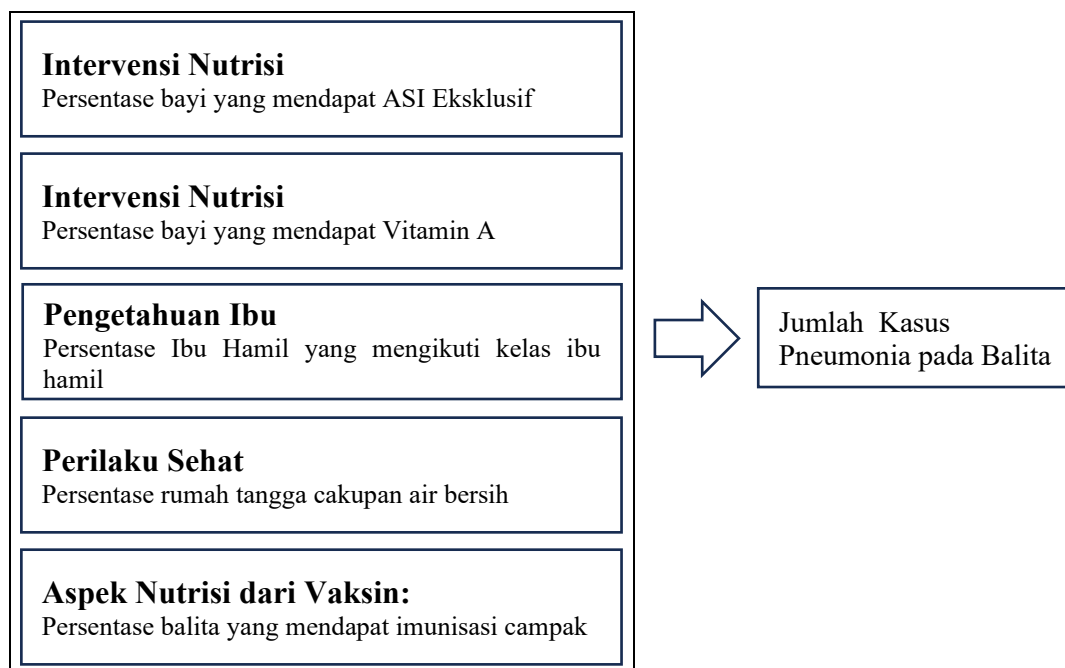
Penelitian terdahulu terkait metode analisis dan variabel penelitian ditunjukkan pada Tabel 2.1.

Tabel 2.1 Penelitian tentang Pneumonia pada Balita

No.	Penulis (Tahun)	Studi Kasus/ Metode	Variabel yang signifikan
1.	Wutun., dkk, 2020	<i>Risk Factors of Pneumonia in Children under Five in Lewoleba Hospital, East Nusa Tenggara, Indonesia</i>	1. Pemberian ASI Eksklusif 2. Riwayat Asma 3. Pengetahuan Ibu tentang penanggulangan pneumonia 4. Riwayat anggota keluarga yang merokok
2.	Saputri, 2022	Pemodelan Faktor-Faktor yang Mempengaruhi Kasus Pneumonia pada Balita di Provinsi Jawa Barat dengan Metode GWGPR	1. Persentase pemberian vitamin A 2. Persentase perilaku hidup bersih dan sehat
3.	Kurnia., dkk., 2023	<i>Analysis of Factors Affecting Penumonia in Toodlers Based on Theory Epidemiology Triad</i>	1. Berat badan lahir rendah 2. Status imunisasi dasar lengkap 3. Pemberian ASI Eksklusif 4. Status nutrisi 5. Jarak dari rumah ke pabrik/jalan besar 5. Paparan Asap Rokok
4.	Fachrunisah, 2024	Pemodelan Pneumonia pada Balita di Kabupaten Tuban Menggunakan <i>Bivariate Spatial Autoregressive Models</i>	Variabel Pneumonia: 1. Persentase ASI eksklusif 2. Persentasi imunisasi dasar lengkap Variabel Diare 1. Persentasi imunisasi dasar lengkap 2. Persentase air bersih

Referensi penelitian terdahulu terkait faktor-faktor yang diduga berpengaruh terhadap kasus pneumonia pada balita telah banyak yang sesuai dengan kerangka berpikir *the Epidemiologic Triangle*. Dari *host*, *agent* dan *environment*, penelitian terdahulu banyak terfokus pada faktor *host* dan *environment*. Faktor *host* terdiri dari riwayat asma, status nutrisi, berat badan lahir rendah, pemberian imunasi dasar lengkap, pemberian ASI Eksklusif, Pemberian Vitamin A, Pemberian Vaksin A dan D. Faktor *environment* dari penelitian terdahulu terdiri dari kepadatan penduduk, rumah tangga berperilaku bersih dan sehat, pengetahuan ibu, persentase rumah tangga dengan akses jamban sehat, adanya riwayat anggota keluarga yang merokok, bahkan jarak dari rumah ke jalan besar.

Berdasarkan ketersediaan data dan pertimbangan unit analisis yang berupa desa/kelurahan maka variabel-variabel penelitian yang akan dimasukkan dalam model GPSAR ditunjukkan pada Gambar 2.2.



Gambar 2.2 Kerangka Konsep Penelitian

2.11 Hasil Penelitian Terdahulu

Adapun untuk menunjang penelitian yang akan dilakukan maka diperoleh beberapa penelitian yang sesuai dan akan menjadi dasar acuan. Berangkat dari

penelitian tersebut perlu dikembangkan pada metode baru untuk memenuhi syarat kebaruan. Berikut hasil studi penelitian terdahulu pada Tabel 2.2.

Tabel 2.2 Pemetaan Hasil Penelitian Terdahulu

No.	Penulis dan Tahun	Judul	Metode	Hasil
1.	Lee, K. L., & Bell, D. R. (2009)	<i>A Spatial Negative Binomial Regression of Individual-level Count Data with Regional and Person-Specific Covariates</i>	Model hirarkis Bayesian yang memperluas regresi Binomial Negatif standar untuk memasukkan kovariat tingkat individu dan wilayah, serta autokorelasi spasial yang tidak teramati.	Mengembangkan model yang dapat memisahkan efek pada level yang berbeda dan menangani autokorelasi spasial. Menunjukkan bahwa penggunaan kovariat tingkat wilayah dalam model NBR standar dapat menghasilkan koefisien yang bias.
2.	Sellner, R., Fischer, M. M., & Koch, M. (2013)	<i>A Spatial Autoregressive Poisson Gravity Model</i>	Model SAR Poisson yang diestimasi menggunakan <i>2-stage nonlinear least squares</i> (2NLS) yang <i>heteroscedasticity-robust</i> .	Estimator 2NLS secara efektif mengontrol masalah overdispersi atau underdispersi yang sering muncul akibat adanya autokorelasi spasial. Model ini fleksibel untuk asumsi distribusi Poisson maupun Binomial Negatif.
3.	Mohebbi, M., et al. (2014)	<i>Disease Mapping and Regression with Count Data in the Presence of Overdispersion and Spatial Autocorrelation: A Bayesian Model Averaging Approach</i>	Model hirarkis Bayesian dengan distribusi <i>Generalized Poisson</i> dan Binomial Negatif untuk menangani overdispersi, dikombinasikan dengan efek acak spasial (CAR) dan <i>Bayesian Model Averaging</i> (BMA).	Menunjukkan bahwa strategi pemodelan yang menggunakan <i>Generalized Poisson</i> dan Binomial Negatif dengan autokorelasi spasial bekerja dengan baik dan memberikan dasar yang kuat untuk inferensi. BMA efektif dalam menangani ketidakpastian model.
4.	Glaser, S. (2017)	<i>A review of spatial econometric models for count data</i>	Tinjauan komparatif model ekonometrik spasial untuk data count, mengklasifikasikannya menjadi model <i>observation-driven</i> dan <i>parameter-driven</i> .	Menyoroti bahwa model SAR untuk data count kurang berkembang dibandingkan model untuk data kontinu karena non-linearitas. Membandingkan berbagai upaya untuk memasukkan efek spasial, mulai dari model auto-Poisson Besag hingga pendekatan yang lebih baru.
5.	N.A. Cruz J.D. Toloza-Delgado O.O. Melo (2024)	<i>Generalized spatial autoregressive model</i>	Mengusulkan model GSAR untuk analisis spasial pada data non-normal (seperti data count/binomial) dalam kerangka GLM.	Membuktikan melalui teori dan simulasi bahwa model GSAR lebih unggul (efisien, akurat, varians rendah) dibandingkan metode SAR yang ada, dan mendemonstrasikan aplikasinya pada data pemilu riil.

BAB III

METODOLOGI

Bab ini menyajikan alur kerja penelitian secara detail, mulai dari pengembangan teoretis hingga implementasi empiris. Metodologi diawali dengan penjabaran langkah-langkah untuk menaksir parameter model GPSAR yang diusulkan. Selanjutnya, akan dipaparkan secara rinci tahapan pemodelan pada data kasus pneumonia di Kabupaten Tuban, yang meliputi deskripsi sumber data, pendefinisian variabel, serta serangkaian analisis data yang akan dijalankan untuk menjawab tujuan penelitian.

3.1 Langkah-langkah Penaksir Parameter GPSAR

Penaksir parameter untuk model GPSAR diperoleh dengan mengikuti metode yang dijabarkan di bawah ini:

1. Menentukan model GPSAR.

Model GPSAR yang terbentuk mengikuti persamaan (2.36) atau (2.37).

2. Membentuk fungsi *likelihood* dari model GPSAR.
3. Menentukan fungsi *ln-likelihood* model GPSAR.
4. Memaksimumkan fungsi *ln-likelihood* model GPSAR dengan mendapatkan turunan pertama terhadap parameter β, ρ, ϕ kemudian disamadengankan nol.
5. Jika hasil langkah ke 4 tidak *closed form*, maka dibutuhkan iterasi numerik untuk mendapatkan estimasi parameternya. Salah satu prosedur iterasi numerik yang dapat digunakan adalah iterasi *Berndt Hall-Hall-Hausman* (BHHH) dengan langkah – langkah sebagai berikut:

Definisi

Misalkan $\theta = [\beta_0 \beta_1 \dots \beta_p \phi \rho]^T$ adalah vektor parameter yang akan di estimasi, $\ell_i(\theta)$ adalah *ln likelihood* untuk observasi ke- i , dimana $i = 1, 2, \dots, n$. Fungsi *ln likelihood* total adalah $\sum_{i=1}^n \ell_i(\theta)$

Inisialisasi:

1. Menentukan nilai awal parameter $\theta^{(0)}$. Idealnya nilai awal parameter untuk model regresi Poisson adalah 0 untuk semua parameter.

2. Menentukan toleransi konvergensi $\epsilon > 0$ yang mana dalam penelitian ini 10^{-6} menjadi batas toleransi konvergensi.
3. Menentukan batas maksimum iterasi (m_{maks}). Misalkan m_{maks} yang ditetapkan adalah 1000. Ketika iterasi tidak mencapai konvergensi pada iterasi ke-1000 maka proses iterasi akan berhenti walaupun konvergensi belum tercapai.
4. Menentukan set iterasi awal adalah $k = 0$.

Prosedur Iterasi:

Langkah 1: Menghitung vektor gradien individu untuk setiap observasi $i = 1, 2, \dots, n$.

$$g_i(\theta^{(k)}) = \left. \frac{\partial \ell_i(\theta)}{\partial \theta} \right|_{\theta=\theta^{(k)}}$$

Langkah 2: Menghitung penjumlahan gradien seluruh sampel.

$$g(\theta^{(k)}) = \sum_{i=1}^n g_i(\theta^{(k)})$$

Langkah 3: Mengaproksimasi matriks Hessian menggunakan informasi vektor gradien individu. Misalkan H^* adalah aproksimasi matriks Hessian, maka

$$H^*(\theta^{(k)}) = - \sum_{i=1}^n g_i(\theta^{(k)}) g_i(\theta^{(k)})^T$$

Langkah 4: Memperbaharui nilai parameter menggunakan arah BHHH.

$$\theta^{(k+1)} = \theta^{(k)} - H^{*-1}(\theta^{(k)}) g(\theta^{(k)})$$

Langkah 5: Mengevaluasi konvergensi dengan melihat kondisi berikut.

$$\|\theta^{(k+1)} - \theta^{(k)}\| \leq \epsilon$$

Jika belum konvergen, set $k = k + 1$ dan kembali ke langkah 1. Jika kondisi konvergen telah tercapai, maka $\theta^{(k+1)}$ adalah estimator dari $\hat{\theta}$.

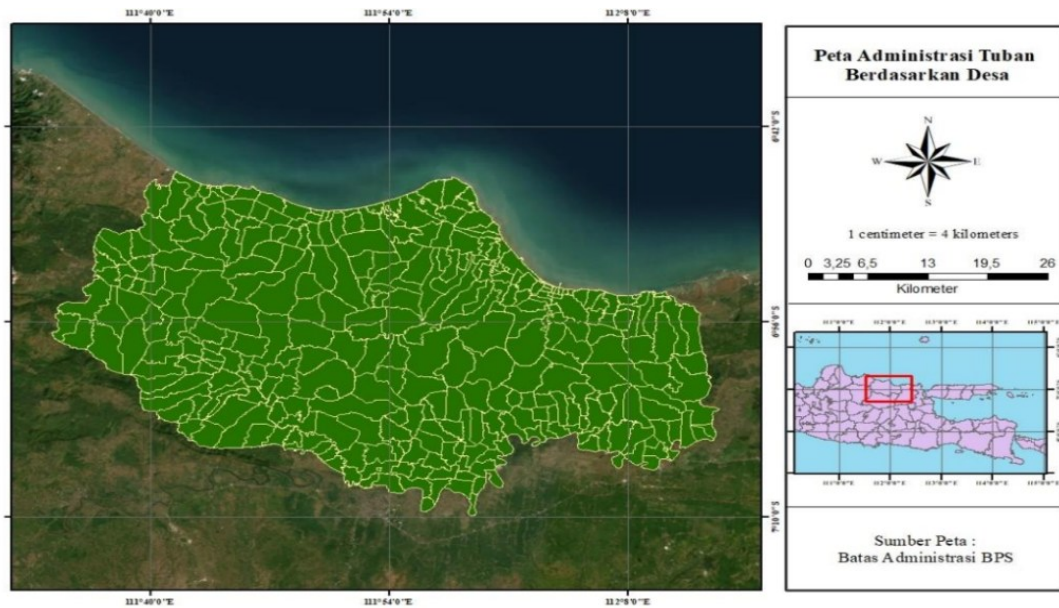
Setelah mendapatkan penaksir parameter maka selanjutnya adalah mendapatkan pengujian hipotesis parameter model GPSAR dilakukan secara serentak dan secara parsial.

3.2 Langkah-langkah Pemodelan Kasus Pneumonia pada Balita

Pada subbab ini, dijelaskan tentang sumber data jumlah kasus pneumonia pada balita, serta variabel dan struktur data yang digunakan dalam penelitian ini yang digunakan untuk melakukan pemodelan jumlah kasus pneumonia pada balita. Selanjutnya, akan dijelaskan mengenai langkah-langkah dalam pemodelan jumlah kasus pneumonia pada balita.

3.2.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Pemerintah Kabupaten Tuban dan pusat kajian Potensi Daerah dan Pemberdayaan Masyarakat (PDPM) ITS. Unit observasi dalam penelitian ini adalah 311 desa dan 17 kelurahan. Data yang digunakan merupakan data tahun 2023. Adapun peta Kabupaten Tuban tergambar pada Gambar 3.1.



Gambar 3.1 Wilayah Administrasi Tuban Berdasarkan Desa.

3.2.2 Spesifikasi Model, Variabel dan Struktur Data Penelitian

Model yang digunakan untuk menganalisis faktor-faktor yang berpengaruh terhadap jumlah kasus pneumonia pada balita adalah GPSAR. Spesifikasi fungsi *link* yang akan digunakan dalam penelitian ini adalah sama seperti pada persamaan 2.37 namun akan ditambahkan variabel eksposur sebagai berikut.

$$\mu_i^{GPSAR} = q_i \exp(\mathbf{A}^{-1} \mathbf{X} \boldsymbol{\beta})_i \quad (3.1)$$

dengan $A = (I - \rho W)$ dan A^{-1} adalah matriks invers dari matriks A , q_i adalah variabel eksposur yaitu jumlah balita.

Variabel yang digunakan dalam penelitian ini terdiri dari satu variabel respon (Y) dan lima variabel prediktor (X). Definisi operasional dan skala pengukuran untuk setiap variabel disajikan secara rinci pada Tabel 3.1.

Tabel 3.1 Variabel Penelitian

No.	Simbol	Variabel	Skala Pengukuran
1.	Y	Jumlah kasus pneumonia pada balita	Rasio
2.	X ₁	Persentase bayi yang mendapatkan ASI Eksklusif	Rasio
3.	X ₂	Persentase balita yang mendapatkan vitamin A	Rasio
4.	X ₃	Persentase ibu hamil yang mengikuti kelas ibu hamil	Rasio
5.	X ₄	Persentase rumah tangga dengan cakupan air bersih	Rasio
6.	X ₅	Persentase balita yang mendapatkan Imunisasi Campak	Rasio

Definisi operasional dari variabel penelitian yang digunakan adalah sebagai berikut:

1. Jumlah kasus pneumonia pada balita adalah jumlah balita yang menderita pneumonia (pneumonia).
2. Persentase bayi yang mendapatkan ASI eksklusif adalah jumlah bayi kurang dari 6 bulan yang mendapatkan ASI tanpa menambahkan dan/atau mengganti dengan makanan atau minuman lain kecuali obat, vitamin dan mineral dibandingkan dengan jumlah bayi lahir hidup dikalikan 100 persen.
3. Persentase balita yang mendapatkan imunisasi campak adalah perbandingan jumlah balita yang mendapatkan imunisasi campak dengan jumlah balita dikalikan 100 persen.
4. Persentase balita yang mendapatkan vitamin A adalah perbandingan jumlah balita yang mendapatkan vitamin A dengan jumlah balita tertimbang dikalikan 100 persen.
5. Persentase ibu hamil yang mengikuti kelas ibu hamil adalah perbandingan ibu hamil yang mengikuti kelas ibu hamil di puskesmas dengan jumlah ibu hamil dikalikan 100 persen.
6. Persentase rumah tangga dengan cakupan air bersih adalah perbandingan jumlah rumah tangga dengan sumber air minumnya menggunakan air bersih

dengan jumlah rumah tangga dikalikan 100 persen. Struktur data penelitian ini ditunjukkan pada Tabel 3.2.

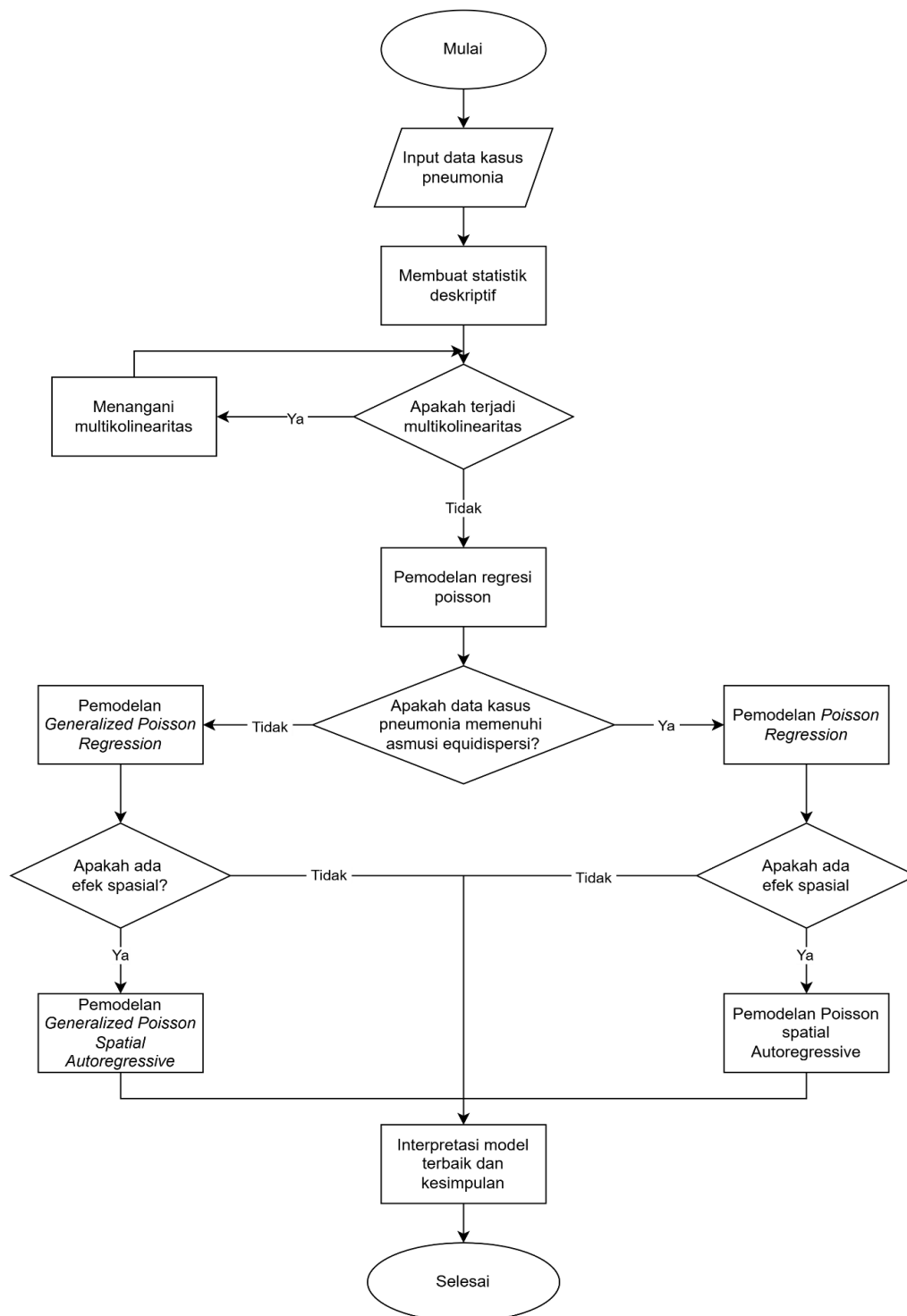
Tabel 3.2 Struktur Data GPSAR

Des/ Kel	Variabel Respon	Variabel Prediktor					Variabel Eksposure
		X_1	...	X_k	...	X_5	
1	y_1	$x_{1,1}$...	$x_{1,k}$...	$x_{1,5}$	q_1
2	y_2	$x_{2,1}$...	$x_{2,k}$...	$x_{2,5}$	q_2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
i	y_i	$x_{i,1}$...	$x_{i,k}$...	$x_{i,5}$	q_i
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
328	y_{328}	$x_{328,1}$...	$x_{328,k}$...	$x_{328,5}$	q_{328}

3.2.3 Langkah-langkah Analisis Pemodelan GPSAR

Untuk mengetahui faktor-faktor yang memengaruhi kasus pneumonia pada balita di Kabupaten Tuban, dilakukan pemodelan *Generalized Poisson Spatial Autoregressive* (GPSAR) melalui beberapa langkah berikut.

1. Membuat analisis deskriptif dan membuat peta tematik variabel prediktor.
2. Mendeteksi multikolinearitas dan mengatasi jika terjadi multikolinearitas sampai tidak terjadi multikolinearitas.
3. Melakukan pengujian *equidispersion* dengan uji *Deviance* seperti pada persamaan (2.12).
4. Melakukan pemodelan data kasus pneumonia pada balita dengan *generalized poisson regression* jika asumsi *equidispersion* tidak terpenuhi.
5. Melakukan pembobotan spasial dengan *queen contiguity*.
6. Melakukan uji dependensi spasial dengan uji *Moran's I* seperti pada persamaan (2.34).
7. Memodelkan data kasus pneumonia pada balita dengan GPSAR dengan menentukan nilai penaksir parameter dan melakukan pengujian parameter secara serentak dan parsial.
8. Melakukan interpretasi model terbaik berdasarkan model dengan nilai uji AICc terkecil (2.38).
9. Menarik kesimpulan dan saran.
10. Diagram alir dari langkah analisis data pada penelitian ini adalah ditunjukkan pada Gambar 3.2.



Gambar 3.2 Diagram Alir Analisis Data.

BAB IV

HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang estimasi parameter model GPSAR, statistik uji untuk pengujian hipotesis model GPSAR dan pada bagian akhir akan dijelaskan penerapan model GPSAR beserta pengujian hipotesis parameternya pada jumlah kasus pneumonia pada balita di desa/kelurahan kabupaten Tuban.

4.1 Penaksiran Parameter GPSAR

Penaksiran parameter model GPSAR dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Tujuan utama dari estimasi ini adalah memperoleh vektor parameter $\boldsymbol{\theta} = [\boldsymbol{\beta}^T \phi \rho]^T$ yang memaksimumkan fungsi *likelihood*. Berbeda dengan regresi linear klasik, fungsi *ln-likelihood* pada model GPSAR bersifat non-linear dan kompleks, sehingga solusi analitik bentuk tertutup (*closed-form solution*) tidak dapat diperoleh. Oleh karena itu, proses optimasi dilakukan menggunakan pendekatan numerik iteratif dengan algoritma BHHH.

Algoritma BHHH dipilih karena efisiensi dalam mengaproksimasi matriks Hessian menggunakan penjumlahan *outer product* dari vektor gradien sehingga menghindari komputasi turunan parsial kedua yang rumit secara analitis dan berat secara komputasi. Selain itu, algoritma ini menjamin bahwa matriks informasi yang diestimasi selalu bersifat semidefinit positif, yang mendukung stabilitas konvergensi algoritma menuju nilai optimum global.

Langkah awal dalam prosedur ini adalah mendefinisikan fungsi distribusi peluang yang mendasari model GPSAR. Berdasarkan fungsi *link* logaritmik dan struktur dependensi spasial yang telah diuraikan pada persamaan (3.1), fungsi massa peluang distribusi GPSAR untuk variabel respon y_i didefinisikan sebagai berikut:

$$P(Y_i = y_i | x_i, \mathbf{W}, \boldsymbol{\theta}) = \left[\frac{\mu_i^*}{1 + \phi \mu_i^*} \right]^{y_i} \frac{(1 + \phi y_i)^{y_i - 1}}{y_i!} \exp \left(- \frac{\mu_i^* (1 + \phi y_i)}{1 + \phi \mu_i^*} \right) \quad (4.1)$$

dengan $\mu_i^{GPSAR} = q_i \exp([(I_n - \rho \mathbf{W})^{-1} \mathbf{X} \boldsymbol{\beta}]_i)$, rata-rata dan variansi dari y_i adalah $E(y_i) = \mu_i^{GPSAR}$ dan $V(y_i) = \mu_i(1 + \phi \mu_i^{GPSAR})^2$ yang merujuk pada spesifikasi model persamaan (3.1), di mana y_i merepresentasikan variabel respon pada lokasi ke- i untuk $i = 1, 2, \dots, n$, dan q_i merupakan variabel exposure, serta ϕ merupakan parameter dispersi yang mengontrol kondisi *overdispersion* atau *underdispersion* pada data.

Fungsi *likelihood* gabungan dibangun berdasarkan asumsi bahwa kejadian antar lokasi bersifat independen secara kondisional apabila parameter rata-rata μ^{GPSAR} telah diketahui. Oleh karena itu, fungsi *likelihood* $L(\boldsymbol{\theta})$ didefinisikan sebagai perkalian dari fungsi massa peluang seluruh pengamatan y_i untuk $i = 1, 2, \dots, n$. Persamaan fungsi *likelihood* dinyatakan sebagai berikut:

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^n P(Y_i = y_i | x_i, \mathbf{W}, \boldsymbol{\theta}) \\ &= \prod_{i=1}^n \left[\left(\frac{\mu_i^{GPSAR}}{1 + \phi \mu_i^{GPSAR}} \right)^{y_i} \frac{(1 + \phi y_i)^{y_i-1}}{y_i!} \exp \left(- \frac{\mu_i^{GPSAR} (1 + \phi y_i)}{1 + \phi \mu_i^{GPSAR}} \right) \right] \end{aligned} \quad (4.2)$$

Persamaan (4.2) menunjukkan bahwa probabilitas gabungan merupakan produk dari probabilitas marginal masing-masing lokasi. Untuk mempermudah proses penurunan rumus fungsi tersebut ditransformasikan ke dalam bentuk *ln-likelihood*.

$$\ell(\boldsymbol{\theta}) = \ln(L(\boldsymbol{\theta})) = \sum_{i=1}^n \ln P(Y_i = y_i | x_i, \mathbf{W}, \boldsymbol{\theta}) \quad (4.3)$$

Persamaan (4.2) dapat disederhanakan menjadi fungsi *ln-likelihood* sebagai berikut:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \left[y_i \ln \left(\frac{\mu_i^{GPSAR}}{1 + \phi \mu_i^{GPSAR}} \right) + (y_i - 1) \ln(1 + \phi y_i) - \frac{\mu_i^{GPSAR} (1 + \phi y_i)}{1 + \phi \mu_i^{GPSAR}} - \ln(y_i!) \right] \quad (4.4)$$

Estimasi parameter menggunakan algoritma BHHH mensyaratkan adanya informasi mengenai turunan parsial pertama dari fungsi *ln-likelihood* terhadap masing-masing parameter. Informasi yang didapatkan ini kemudian disusun dalam vektor gradien $\mathbf{g}(\boldsymbol{\theta})$. Oleh karena itu untuk mendapatkan turunan parsial pertama

dari fungsi *ln-likelihood* terhadap masing-masing parameter perlu dilakukan dengan menjabarkan fungsi *ln-likelihood* pada persamaan (4.4) dan menerapkan aturan rantai pada turunan fungsi *ln-likelihood* terhadap $\boldsymbol{\beta}$ dan ρ . Hal ini diperlukan karena parameter $\boldsymbol{\beta}$ dan ρ tidak berhubungan langsung dengan fungsi *ln-likelihood*, melainkan melalui perantara prediktor linear (η_i) yang telah mengalami transformasi spasial dan fungsi *link* eksponensial. Sedangkan turunan parsial fungsi *ln-likelihood* terhadap ϕ dilakukan dengan cara biasa. Sehingga akan didapatkan bentuk vektor gradien berikut.

$$\mathbf{g}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu_i^{GPSAR}} \cdot \frac{\partial \mu_i^{GPSAR}}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \boldsymbol{\beta}} \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho} = \sum_{i=1}^n \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu_i^{GPSAR}} \cdot \frac{\partial \mu_i^{GPSAR}}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \rho} \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \phi} \end{bmatrix} \quad (4.5)$$

dimana $\mu_i^{GPSAR} = q_i \exp([A^{-1}X\boldsymbol{\beta}]_i)$ dan η_i adalah elemen $ke - i$ dari vektor $A^{-1}X\boldsymbol{\beta}$.

Turunan parsial fungsi *ln-likelihood* terhadap $\boldsymbol{\beta}$ dan terhadap ρ sama-sama mengandung unsur turunan parsial fungsi *ln-likelihood* terhadap μ_i^{GPSAR} dan turunan parsial μ_i^{GPSAR} terhadap η_i pada proses penurunannya. Sehingga perlu dilakukan turunan parsial terlebih dahulu fungsi *ln-likelihood* terhadap μ_i^{GPSAR} dan turunan parsial μ_i^{GPSAR} terhadap η_i .

1. Turunan Parsial Fungsi *ln-likelihood* terhadap μ_i^{GPSAR}

Langkah pertama adalah menurunkan fungsi $\ell(\boldsymbol{\theta})$ terhadap μ_i^{GPSAR} . Untuk mempermudah proses diferensiasi, persamaan (4.4) dipecah menjadi suku-suku yang mengandung μ_i^{GPSAR} dimana suku-suku yang dimaksud adalah suku pertama dan suku ketiga. Berikut adalah langkah-langkah turunan parsial fungsi *ln-likelihood* terhadap μ_i^{GPSAR} secara lengkap.

a. Suku pertama

$$\frac{\partial}{\partial \mu_i^{GPSAR}} [y_i \ln(\mu_i^{GPSAR}) - y_i \ln(1 + \phi \mu_i^{GPSAR})] = \frac{y_i}{\mu_i^{GPSAR}} - \frac{y_i \phi}{1 + \phi \mu_i^{GPSAR}}$$

disamakan penyebutnya menjadi seperti berikut.

$$\begin{aligned} \frac{y_i}{\mu_i^{GPSAR}} - \frac{y_i \phi}{1 + \phi \mu_i^{GPSAR}} &= \frac{y_i(1 + \phi \mu_i^{GPSAR}) - y_i \phi \mu_i^{GPSAR}}{\mu_i^{GPSAR}(1 + \phi \mu_i^{GPSAR})} \\ &= \frac{y_i}{\mu_i^{GPSAR}(1 + \phi \mu_i^{GPSAR})} \end{aligned}$$

b. Suku ketiga

Bentuk dari suku ketiga merupakan bentuk fungsi rasional dimana kita dapat menggunakan aturan turunan pembagian yang dinyatakan sebagai berikut.

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$$

Misalkan $u = \mu_i^{GPSAR}(1 + \phi y_i)$ dan $v = 1 + \phi \mu_i^{GPSAR}$,

maka akan didapatkan bentuk

$$u' = (1 + \phi y_i) \text{ dan } v' = \phi$$

Sehingga

$$\begin{aligned} \frac{\partial}{\partial \mu_i^{GPSAR}} \left[\frac{\mu_i^{GPSAR}(1 + \phi y_i)}{1 + \phi \mu_i^{GPSAR}} \right] &= \frac{(1 + \phi y_i)(1 + \phi \mu_i^{GPSAR}) - \mu_i^{GPSAR}(1 + \phi y_i)(\phi)}{(1 + \phi \mu_i^{GPSAR})^2} \\ &= \frac{(1 + \phi y_i)[(1 + \phi \mu_i^{GPSAR}) - \phi \mu_i^{GPSAR}]}{(1 + \phi \mu_i^{GPSAR})^2} \\ &= \frac{1 + \phi y_i}{(1 + \phi \mu_i^{GPSAR})^2} \end{aligned}$$

Dengan menggabungkan hasil turunan Suku 1 dan negatif dari Suku 3, diperoleh:

$$\frac{\partial \ell(\theta)}{\partial \mu_i^{GPSAR}} = \frac{y_i}{\mu_i^{GPSAR}(1 + \phi \mu_i^{GPSAR})} - \frac{1 + \phi y_i}{(1 + \phi \mu_i^{GPSAR})^2}$$

$$\begin{aligned}
&= \frac{y_i(1 + \phi\mu_i^{GPSAR}) - \mu_i^{GPSAR}(1 + \phi y_i)}{\mu_i^{GPSAR}(1 + \phi\mu_i^{GPSAR})^2} \\
&= \frac{y_i + y_i\phi\mu_i^{GPSAR} - \mu_i^{GPSAR} - \mu_i^{GPSAR}\phi y_i}{\mu_i^{GPSAR}(1 + \phi\mu_i^{GPSAR})^2} \\
&= \frac{y_i - \mu_i^{GPSAR}}{\mu_i^{GPSAR}(1 + \phi\mu_i^{GPSAR})^2} \tag{4.6}
\end{aligned}$$

Setelah mendapatkan bentuk turunan parsial fungsi *ln-likelihood* terhadap μ_i^{GPSAR} selanjutnya adalah mencari turunan parsial μ_i^{GPSAR} terhadap η_i .

2. Turunan Parsial μ_i^{GPSAR} terhadap η_i

Mengingat fungsi *link* adalah $\mu_i^{GPSAR} = \exp(\eta_i)$, maka

$$\frac{\partial \mu_i^{GPSAR}}{\partial \eta_i} = \frac{\partial}{\partial \eta_i} (\exp(\eta_i)) = \exp(\eta_i) = \mu_i^{GPSAR} \tag{4.7}$$

Setelah kita mendapatkan turunan parsial fungsi *ln-likelihood* terhadap μ_i^{GPSAR} dan turunan parsial μ_i^{GPSAR} terhadap η_i , maka yang perlu dilakukan sekarang adalah mencari secara utuh turunan parsial fungsi *ln-likelihood* terhadap seluruh parameter.

1. Turunan Parsial terhadap Koefisien Regresi (β)

Penurunan gradien terhadap parameter koefisien regresi β dilakukan dengan menerapkan aturan rantai secara bertingkat seperti yang telah dijelaskan sebelumnya. Secara matematis, hubungan fungsional tersebut dapat diuraikan sebagai berikut:

$$\ell(\theta) \xrightarrow{\text{fungsi dari}} \mu_i^{GPSAR} \xrightarrow{\text{fungsi dari}} \eta_i \xrightarrow{\text{fungsi dari}} \beta$$

di mana prediktor linear spasial η_i didefinisikan dari persamaan struktur model:

$$\eta = (I_n - \rho W)^{-1} X\beta = A^{-1} X\beta$$

Berdasarkan alur dependensi di atas, turunan parsial fungsi *ln-likelihood* terhadap β diformulasikan sebagai penjumlahan dari perkalian tiga komponen turunan:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu_i^{GPSAR}} \cdot \frac{\partial \mu_i^{GPSAR}}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}$$

Untuk komponen satu dan komponen dua telah didapatkan sebelumnya yang mana hasilnya dinyatakan pada persamaan (4.5) dan (4.6). Sehingga selanjutnya adalah mencari komponen ketiga. Diketahui bahwa η_i adalah elemen $ke - i$ dari vektor $\mathbf{A}^{-1}\mathbf{X}\boldsymbol{\beta}$. Misalkan $\mathbf{Z} = \mathbf{A}^{-1}\mathbf{X}$ adalah matriks prediktor yang telah terboboti secara spasial, maka $\eta_i = \mathbf{z}_i^T \boldsymbol{\beta}$, di mana \mathbf{z}_i^T adalah baris $ke - i$ dari matriks \mathbf{Z} . Turunan skalar η_i terhadap vektor $\boldsymbol{\beta}$ adalah vektor baris tersebut yang ditransposkan:

$$\frac{\partial \eta_i}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}} (\mathbf{z}_i^T \boldsymbol{\beta}) = (\mathbf{z}_i^T)^T = (\mathbf{A}^{-1}\mathbf{X})_i^T$$

Selanjutnya, ketiga komponen tersebut disubstitusikan kembali ke dalam persamaan (4.6):

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n \left[\left(\frac{y_i - \mu_i^{GPSAR}}{\mu_i^{GPSAR} (1 + \phi \mu_i^{GPSAR})^2} \right) \cdot (\mu_i^{GPSAR}) \cdot (\mathbf{A}^{-1}\mathbf{X})_i^T \right] \\ &= \sum_{i=1}^n \left[\frac{y_i - \mu_i^{GPSAR}}{(1 + \phi \mu_i^{GPSAR})^2} (\mathbf{A}^{-1}\mathbf{X})_i^T \right] \end{aligned} \quad (4.8)$$

2. Turunan Parsial terhadap Parameter Spasial (ρ)

Penurunan gradien terhadap parameter dependensi spasial ρ dilakukan dengan menerapkan aturan rantai secara bertingkat yang telah dijelaskan sebelumnya. Secara matematis, hubungan fungsional tersebut dapat diuraikan sebagai berikut:

$$\ell(\boldsymbol{\theta}) \xrightarrow{\text{fungsi dari}} \mu_i^{GPSAR} \xrightarrow{\text{fungsi dari}} \eta_i \xrightarrow{\text{fungsi dari}} \rho$$

Di mana prediktor linear spasial $\boldsymbol{\eta}$ didefinisikan dari persamaan struktur model:

$$\boldsymbol{\eta} = (\mathbf{I}_n - \rho \mathbf{W})^{-1} \mathbf{X}\boldsymbol{\beta} = \mathbf{A}^{-1} \mathbf{X}\boldsymbol{\beta}$$

Berdasarkan alur dependensi di atas, turunan parsial terhadap ρ diformulasikan sebagai penjumlahan dari perkalian tiga komponen turunan:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho} = \sum_{i=1}^n \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu_i^{GPSAR}} \cdot \frac{\partial \mu_i^{GPSAR}}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \rho}$$

Untuk komponen satu dan komponen dua telah didapatkan sebelumnya yang mana hasilnya dinyatakan pada persamaan (4.5) dan (4.6). Sehingga selanjutnya adalah mencari komponen ketiga. Diketahui bahwa $\boldsymbol{\eta} = \mathbf{A}^{-1}\mathbf{X}\boldsymbol{\beta}$. Karena $\mathbf{X}\boldsymbol{\beta}$ tidak memuat ρ maka penurunan hanya dilakukan pada matriks invers \mathbf{A}^{-1} . Dengan demikian didapatkan turunannya dalam bentuk berikut.

$$\begin{aligned} \frac{\partial \boldsymbol{\eta}}{\partial \rho} &= \left(\frac{\partial}{\partial \rho} \mathbf{A}^{-1} \right) \mathbf{X}\boldsymbol{\beta} \\ &= (\mathbf{A}^{-1} \mathbf{W} \mathbf{A}^{-1}) \mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{A}^{-1} \mathbf{W} \underbrace{(\mathbf{A}^{-1} \mathbf{X}\boldsymbol{\beta})}_{\boldsymbol{\eta}} \\ &= \mathbf{A}^{-1} \mathbf{W} \boldsymbol{\eta} \end{aligned}$$

sehingga

$$\frac{\partial \eta_i}{\partial \rho} = [\mathbf{A}^{-1} \mathbf{W} \boldsymbol{\eta}]_i$$

Selanjutnya, ketiga komponen tersebut disubstitusikan kembali ke dalam persamaan utama:

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho} &= \sum_{i=1}^n \left[\left(\frac{y_i - \mu_i^{GPSAR}}{\mu_i^{GPSAR} (1 + \phi \mu_i^{GPSAR})^2} \right) \cdot (\mu_i^{GPSAR}) \cdot [\mathbf{A}^{-1} \mathbf{W} \boldsymbol{\eta}]_i \right] \\ &= \sum_{i=1}^n \left[\frac{y_i - \mu_i^{GPSAR}}{(1 + \phi \mu_i^{GPSAR})^2} [\mathbf{A}^{-1} \mathbf{W} \boldsymbol{\eta}]_i \right] \end{aligned} \quad (4.9)$$

3. Turunan Parsial terhadap Parameter Dispersi (ϕ)

Berbeda dengan parameter regresi dan spasial yang berada di dalam fungsi μ_i^{GPSAR} , parameter dispersi ϕ muncul secara eksplisit dalam fungsi distribusi peluang *Generalized Poisson* untuk mengontrol varians data. Oleh karena itu, penurunan gradien terhadap ϕ dilakukan langsung pada fungsi *ln-likelihood*

persamaan (4.4) tanpa melalui aturan rantai yang kompleks terhadap η . Fungsi *ln-likelihood* yang memuat komponen ϕ adalah seperti pada persamaan (4.4). Turunan parsial pertama terhadap ϕ diperoleh dengan memecah dan menurunkan setiap suku sebagai berikut:

$$\frac{\partial}{\partial \phi} \left[y_i \ln(\mu_i^{GPSAR}) - y_i \ln(1 + \phi \mu_i^{GPSAR}) + (y_i - 1) \ln(1 + \phi y_i) - \frac{\mu_i^{GPSAR}(1 + \phi y_i)}{1 + \phi \mu_i^{GPSAR}} - \ln(y_i!) \right]$$

Turunan Suku Logaritma Pertama

$$\frac{\partial}{\partial \phi} [y_i \ln(\mu_i^{GPSAR}) - y_i \ln(1 + \phi \mu_i^{GPSAR})] = -\frac{y_i \mu_i^{GPSAR}}{1 + \phi \mu_i^{GPSAR}}$$

Turunan Suku Logaritma Kedua

$$\frac{\partial}{\partial \phi} [(y_i - 1) \ln(1 + \phi y_i)] = \frac{y_i(y_i - 1)}{1 + \phi y_i}$$

Turunan Suku Rasio (Pecahan) Menggunakan aturan pembagian $\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$ di mana $u = \mu_i^{GPSAR} + \mu_i^{GPSAR} \phi y_i$ dan $v = 1 + \phi \mu_i^{GPSAR}$, sehingga

$$\frac{\partial}{\partial \phi} \left[\frac{\mu_i^{GPSAR}(1 + \phi y_i)}{1 + \phi \mu_i^{GPSAR}} \right] = \frac{\mu_i^{GPSAR}(y_i - \mu_i^{GPSAR})}{(1 + \phi \mu_i^{GPSAR})^2}$$

Setelah penyederhanaan aljabar, suku ini menjadi:

Dengan menggabungkan ketiga komponen tersebut, diperoleh gradien akhir untuk parameter dispersi:

$$\frac{\partial \ell(\theta)}{\partial \phi} = \sum_{i=1}^n \left[-\frac{y_i \mu_i^{GPSAR}}{1 + \phi \mu_i^{GPSAR}} + \frac{y_i(y_i - 1)}{1 + \phi y_i} + \frac{\mu_i^{GPSAR}(y_i - \mu_i^{GPSAR})}{(1 + \phi \mu_i^{GPSAR})^2} \right] \quad (4.10)$$

Berdasarkan turunan parsial yang telah diperoleh untuk masing-masing parameter pada persamaan (4.7), (4.8), dan (4.9), didapati bahwa bentuk tersebut tidak *closed form*. Kemudian vektor unit gradien dapat disusun komponennya menjadi seperti berikut.

$$\mathbf{g}_i(\boldsymbol{\theta}) = \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{y_i - \mu_i^{GPSAR}}{(1 + \phi \mu_i^{GPSAR})^2} (\mathbf{A}^{-1} \mathbf{X})_i^T \\ \frac{y_i - \mu_i^{GPSAR}}{(1 + \phi \mu_i^{GPSAR})^2} [\mathbf{A}^{-1} \mathbf{W} \boldsymbol{\eta}]_i \\ \frac{y_i(y_i - 1)}{1 + \phi y_i} - \frac{y_i \mu_i^{GPSAR}}{1 + \phi \mu_i^{GPSAR}} - \frac{\mu_i^{GPSAR}(y_i - \mu_i^{GPSAR})}{(1 + \phi \mu_i^{GPSAR})^2} \end{bmatrix} \quad (4.11)$$

dan

$$\mathbf{g}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \sum_{i=1}^n \frac{y_i - \mu_i^{GPSAR}}{(1 + \phi \mu_i^{GPSAR})^2} (\mathbf{A}^{-1} \mathbf{X})_i^T \\ \sum_{i=1}^n \frac{y_i - \mu_i^{GPSAR}}{(1 + \phi \mu_i^{GPSAR})^2} [\mathbf{A}^{-1} \mathbf{W} \boldsymbol{\eta}]_i \\ \sum_{i=1}^n \frac{y_i(y_i - 1)}{1 + \phi y_i} - \frac{y_i \mu_i^{GPSAR}}{1 + \phi \mu_i^{GPSAR}} - \frac{\mu_i^{GPSAR}(y_i - \mu_i^{GPSAR})}{(1 + \phi \mu_i^{GPSAR})^2} \end{bmatrix} \quad (4.12)$$

Estimasi parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \rho, \phi)^T$ dilakukan melalui prosedur iteratif metode BHHH untuk memaksimumkan fungsi *ln-likelihood*. Berikut adalah langkah-langkah algoritma estimasi:

Definisi

Misalkan $\boldsymbol{\theta} = [\beta_0 \ \beta_1 \ \dots \ \beta_p \ \phi \ \rho]^T$ adalah vektor parameter yang akan di estimasi, $\ell_i(\boldsymbol{\theta})$ adalah *ln likelihood* untuk observasi ke- i , dimana $i = 1, 2, \dots, n$. Fungsi *ln likelihood* total adalah $\sum_{i=1}^n \ell_i(\boldsymbol{\theta})$

Inisialisasi:

1. Menentukan nilai awal parameter $\boldsymbol{\theta}^{(0)}$. Nilai awal parameter menentukan kecepatan tercapainya konvergensi pada proses iterasi. Idealnya nilai awal parameter untuk model GPSAR adalah nilai estimasi parameter dari model GPR.
2. Menentukan toleransi konvergensi $\epsilon > 0$ yang mana dalam penelitian ini 10^{-6} menjadi batas toleransi konvergensi.
3. Menentukan batas maksimum iterasi (m_{maks}). Misalkan m_{maks} yang ditetapkan adalah 1000. Ketika iterasi tidak mencapai konvergensi pada iterasi ke-1000 maka proses iterasi akan berhenti walaupun konvergensi belum tercapai.

4. Menentukan set iterasi awal adalah $k = 0$.

Prosedur Iterasi:

Langkah 1: Pada setiap iterasi ke- k , langkah pertama adalah menghitung nilai numerik dari gradien individu $g_i(\theta)$ untuk setiap unit observasi $i = 1, 2, \dots, n$. Perhitungan ini dilakukan dengan mensubstitusi nilai estimasi parameter $\theta^{(k)}$ ke dalam formulasi vektor gradien individu yang telah dijabarkan pada persamaan (4.11). Komponen ini nantinya akan digunakan untuk menyusun aproksimasi matriks Hessian.

Langkah 2: Menyusun vektor gradien $g(\theta^{(k)})$ yang dibentuk dengan mengagregasi seluruh hasil evaluasi gradien individu dari langkah sebelumnya. Proses ini merujuk pada definisi vektor gradien total yang telah dinyatakan dalam persamaan (4.12)

Langkah 3: Mengaproksimasi matriks Hessian menggunakan informasi vektor gradien individu. Misalkan H^* adalah aproksimasi matriks Hessian, maka

$$H^*(\theta^{(k)}) = - \sum_{i=1}^n g_i(\theta^{(k)}) g_i(\theta^{(k)})^T$$

Langkah 4: Memperbaharui nilai parameter menggunakan arah BHHH.

$$\theta^{(k+1)} = \theta^{(k)} - H^{*-1}(\theta^{(k)}) g(\theta^{(k)})$$

Langkah 5: Mengevaluasi konvergensi dengan melihat kondisi berikut.

$$\|\theta^{(k+1)} - \theta^{(k)}\| \leq \epsilon$$

Jika belum konvergen, set $k = k + 1$ dan kembali ke langkah 1. Jika kondisi konvergen telah tercapai, maka $\theta^{(k+1)}$ ditetapkan sebagai estimator $\hat{\theta}$.

Setelah didapatkan bentuk penaksir parameter selanjutnya menentukan pengujian hipotesis parameter model GPSAR yang dilakukan melalui dua tahap, yaitu uji serentak menggunakan metode *Maximum Likelihood Ratio Test* (MLRT) dan uji parsial menggunakan uji *Wald*. Uji serentak bertujuan untuk mengetahui signifikansi pengaruh variabel prediktor terhadap variabel respon secara bersama. Prinsip MLRT didasarkan pada perbandingan antara nilai maksimum fungsi *ln-likelihood* pada ruang parameter populasi (Ω_{GPSAR}) dengan nilai maksimum pada ruang parameter di bawah H_0 (ω_{GPSAR}) yang tidak melibatkan variabel prediktor.

Adapun hipotesis untuk pengujian serentak model GPSAR dapat dinyatakan sebagai berikut

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = \dots = \beta_p = 0$$

$$H_1: \text{Minimal ada satu } \beta_k \neq 0, k = 1, 2, \dots, p$$

Fungsi *likelihood* untuk ruang parameter $\boldsymbol{\Omega}_{GPSAR}$ telah diuraikan pada persamaan (4.2), di mana penaksir parameter yang memaksimalkan fungsi tersebut $\hat{\boldsymbol{\Omega}}_{GPSAR} = \{\hat{\boldsymbol{\beta}}, \hat{\rho}, \hat{\phi}\}$ disubstitusikan untuk menghasilkan nilai maksimum fungsi *ln likelihood* seperti yang dapat dinyatakan pada Persamaan (4.10).

$$\ell(\hat{\boldsymbol{\Omega}}_{GPSAR}) = \ell(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\mu}_i^{GPSAR}}{1 + \hat{\phi} \hat{\mu}_i^{GPSAR}} \right) + (y_i - 1) \ln(1 + \hat{\phi} y_i) \right. \\ \left. - \frac{\hat{\mu}_i^{GPSAR} (1 + \phi y_i)}{1 + \hat{\phi} \hat{\mu}_i^{GPSAR}} - \ln(y_i!) \right] \quad (4.13)$$

Sebaliknya, himpunan parameter di bawah H_0 ($\boldsymbol{\omega}_{GPSAR}$) dibentuk dengan menerapkan restriksi bahwa seluruh koefisien variabel prediktor bernilai nol. Fungsi *likelihood* untuk kondisi restriksi ini dinyatakan dalam Persamaan (4.11).

$$L(\boldsymbol{\omega}_{GPSAR}) = \prod_{i=1}^n \left[\frac{\left(\frac{\mu_{i\omega}^{GPSAR}}{1 + \phi_{\omega} \mu_{i\omega}^{GPSAR}} \right)^{y_i} (1 + \phi_{\omega} y_i)^{y_i-1}}{y_i!} \right. \\ \left. \exp \left(- \frac{\mu_{i\omega}^{GPSAR} (1 + \phi_{\omega} y_i)}{1 + \phi_{\omega} \mu_{i\omega}^{GPSAR}} \right) \right] \quad (4.14)$$

dimana $\mu_{i\omega}^{GPSAR} = \exp(\mathbf{A}_{\omega}^{-1} \mathbf{1} \beta_{0\omega})_i$, $\mathbf{A}_{\omega} = (\mathbf{I} - \rho_{\omega} \mathbf{W})$, \mathbf{A}_{ω}^{-1} adalah matriks invers dari \mathbf{A}_{ω} , dan $\mathbf{1} = [1 \ 1 \ \dots \ 1]^T$ berukuran $n \times 1$. Sama halnya dengan kondisi populasi, penaksir parameter di bawah H_0 diperoleh dengan memaksimumkan fungsi pada persamaan (4.11). Hasil tersebut didapatkan setelah melakukan penaksiran parameter di bawah hipotesis nol dengan menggunakan metode MLE dan iterasi BHHH yang dijelaskan lebih lanjut pada lampiran 3. Penaksir parameter yang memaksimalkan fungsi tersebut $\hat{\boldsymbol{\omega}}_{GPSAR} = \{\hat{\beta}_{0\omega}, \hat{\rho}_{\omega}, \hat{\phi}_{\omega}\}$ disubstitusikan untuk menghasilkan nilai maksimum fungsi *ln likelihood* seperti yang dapat dinyatakan pada Persamaan (4.12).

$$\ell(\hat{\omega}_{GPSAR}) = \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\mu}_{i\omega}^{GPSAR}}{1 + \hat{\phi}_{\omega} \hat{\mu}_{i\omega}^{GPSAR}} \right) + (y_i - 1) \ln(1 + \hat{\phi}_{\omega} y_i) \right. \\ \left. - \frac{\hat{\mu}_{i\omega}^{GPSAR} (1 + \hat{\phi}_{\omega} y_i)}{1 + \hat{\phi}_{\omega} \hat{\mu}_{i\omega}^{GPSAR}} - \ln(y_i!) \right] \quad (4.15)$$

dimana $\hat{\mu}_{i\omega}^{GPSAR} = \exp(\hat{A}_{\omega}^{-1} \mathbf{1} \hat{\beta}_{0\omega})_i$, $\hat{A}_{\omega} = (I - \hat{\rho}_{\omega} W)$, \hat{A}_{ω}^{-1} adalah matriks invers dari \hat{A}_{ω} , dan $\mathbf{1} = [1 \ 1 \dots 1]^T$ berukuran $n \times 1$.

Penaksir parameter $\hat{\Omega}_{GPSAR} = \hat{\beta}, \hat{\rho}, \hat{\phi}$ dan $\hat{\omega}_{GPSAR} = \hat{\beta}_{0\omega}, \hat{\rho}_{\omega}, \hat{\phi}_{\omega}$ selanjutnya disubstitusikan pada masing-masing fungsi *likelihood* menjadi $L(\hat{\Omega}_{GPSAR})$ dan $L(\hat{\omega}_{GPSAR})$ sehingga *likelihood ratio test* dihitung dengan rumus pada persamaan (4.13).

$$LR = \frac{L(\hat{\omega}_{GPSAR})}{L(\hat{\Omega}_{GPSAR})} < LR_0 \quad (4.16)$$

Statistik uji untuk pengujian hipotesis parameter secara serentak diperoleh pada persamaan (4.14).

$$\begin{aligned} G_{GPSAR}^2 &= -\ln(LR)^2 \\ G_{GPSAR}^2 &= -2 \ln(LR) \\ G_{GPSAR}^2 &= -2 \ln \frac{L(\hat{\omega}_{GPSAR})}{L(\hat{\Omega}_{GPSAR})} \\ G_{GPSAR}^2 &= -2 \left(\ell(\hat{\omega}_{GPSAR}) - \ell(\hat{\Omega}_{GPSAR}) \right) \end{aligned} \quad (4.17)$$

dimana $\ell(\hat{\omega}_{GPSAR})$ Daerah kritis pengujian hipotesis adalah sebagai berikut.

$$\begin{aligned} \alpha &= P(LR < LR_0), 0 < LR_0 \leq 1 \\ &= P(\ln(LR^2) < \ln(LR_0^2)) \\ &= P(\ln LR^2 < \ln LR_0^2) \\ &= P(G_{GPSAR}^2 > \chi_{(\alpha, df)}^2) \end{aligned} \quad (4.18)$$

G_{GPSAR}^2 mengikuti distribusi *chi-square* untuk $n \rightarrow \infty$ sehingga daerah penolakan H_0 adalah $G_{GPSAR}^2 > \chi_{(\alpha, df)}^2$ atau *p-value* $< \alpha$ dengan derajat bebas

(*df*) merupakan jumlah parameter di bawah populasi dikurangi dengan jumlah parameter di bawah H_0 .

$$\begin{aligned} df &= n(\hat{\boldsymbol{\Omega}}_{GPSAR}) - n(\hat{\boldsymbol{\omega}}_{GPSAR}) \\ &= [(p+1) + 1 + 1] - [1 + 1 + 1] = p \end{aligned}$$

Apabila pengujian hipotesis parameter secara serentak telah dilakukan dan diperoleh keputusan tolak H_0 , maka langkah selanjutnya adalah pengujian hipotesis parameter secara parsial. Pengujian hipotesis parameter secara parsial dalam hal ini bertujuan untuk mengetahui variabel prediktor dan parameter struktural (ρ, ϕ) mana yang berpengaruh signifikan terhadap variabel respon secara individu.

Pengujian hipotesis parameter secara parsial menggunakan uji *Wald* karena uji *Wald* mampu menguji signifikansi parameter individu dalam model yang kompleks dan bersesuaian dengan asumsi distribusi normal. Pengujian hipotesis secara parsial terhadap parameter ρ terlebih dahulu dilakukan dengan hipotesis sebagai berikut.

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Statistik uji parsial untuk $n \rightarrow \infty$ disusun berdasarkan sifat asimtotik penaksir $\hat{\rho}$. Penaksir ini berdistribusi normal dengan rata-rata ρ , sedangkan variansnya merupakan diagonal utama utama dari $-\mathbf{H}^{*-1}(\hat{\boldsymbol{\theta}})$ pada posisi yang bersesuaian dengan ρ . Oleh karena itu, diperoleh statistik uji sebagai berikut.

$$\frac{\hat{\rho} - \rho}{se(\hat{\rho})} \sim N(0,1)$$

dimana $se(\hat{\rho})$ merupakan *standart error* $\hat{\rho}$ yang diperoleh dari $\sqrt{\widehat{Var}(\hat{\rho})}$. Dengan demikian, statistik uji *Walds* di bawah H_0 dapat dinyatakan dalam persamaan (4.19).

$$W_\rho = \left(\frac{\hat{\rho}}{se(\hat{\rho})} \right)^2 \sim \chi^2_{(\alpha,1)} \quad (4.19)$$

Kriteria pengambilan keputusan untuk statistik uji pada persamaan (4.19) adalah H_0 ditolak jika nilai $W_\rho > \chi^2_{(\alpha,1)}$.

Selanjutnya pengujian signifikansi parameter β_k secara parsial. Hipotesis yang digunakan untuk menguji parameter β_k secara parsial adalah sebagai berikut.

$$H_0: \beta_k = 0$$

$$H_1: \beta_k \neq 0$$

Statistik uji parsial untuk $n \rightarrow \infty$ disusun berdasarkan sifat asimtotik penaksir $\hat{\beta}_k$. Penaksir ini berdistribusi normal dengan rata-rata β_k , sedangkan variansnya merupakan diagonal utama utama dari $-\mathbf{H}^{*-1}(\hat{\boldsymbol{\theta}})$ pada posisi yang bersesuaian dengan β_k . Oleh karena itu, diperoleh statistik uji sebagai berikut.

$$\frac{\hat{\beta}_k - \beta_k}{se(\hat{\beta}_k)} \sim N(0,1)$$

dimana $se(\hat{\beta}_k)$ merupakan *standart error* $\hat{\beta}_k$ yang diperoleh dari $\sqrt{\widehat{Var}(\hat{\beta}_k)}$.

Dengan demikian, statistik uji *Walds* di bawah H_0 dapat dinyatakan dalam persamaan (4.20).

$$W_{\beta_k} = \left(\frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \right)^2 \sim \chi^2_{(\alpha,1)} \quad (4.20)$$

Kriteria pengambilan keputusan untuk statistik uji *Wald* pada persamaan (4.20) adalah H_0 ditolak jika nilai $W_{\beta_k} > \chi^2_{(\alpha,1)}$.

Selanjutnya pengujian signifikansi parameter ϕ secara parsial. Hipotesis yang digunakan untuk menguji parameter secara parsial adalah sebagai berikut.

$$H_0: \phi = 0$$

$$H_1: \phi \neq 0$$

Statistik uji parsial untuk $n \rightarrow \infty$ disusun berdasarkan sifat asimtotik penaksir $\hat{\phi}$. Penaksir ini berdistribusi normal dengan rata-rata ϕ , sedangkan variansnya merupakan diagonal utama utama dari $-\mathbf{H}^{*-1}(\hat{\boldsymbol{\theta}})$ pada posisi yang bersesuaian dengan ϕ . Oleh karena itu, diperoleh statistik uji sebagai berikut.

$$\frac{\hat{\phi} - \phi}{se(\hat{\phi})} \sim N(0,1)$$

dimana $se(\hat{\phi})$ merupakan *standart error* ϕ yang diperoleh dari $\sqrt{\widehat{Var}(\hat{\phi})}$. Dengan demikian, statistik uji *Walds* di bawah H_0 dapat dinyatakan dalam persamaan (4.21).

$$W_{\phi} = \left(\frac{\hat{\phi}}{se(\hat{\phi})} \right)^2 \sim \chi^2_{(\alpha,1)} \quad (4.21)$$

Kriteria pengambilan keputusan untuk statistik uji *Wald* persamaan (4.21) adalah H_0 ditolak jika nilai $W_{\phi} > \chi^2_{(\alpha,1)}$.

4.2 Penerapan Model GPSAR Terhadap Jumlah Kasus Pneumonia Pada Balita di Kabupaten Tuban

Pada subbab ini akan dijelaskan terlebih dahulu mengenai deskripsi variabel yang digunakan dalam penelitian ini dan dilanjutkan dengan pemodelan GPSAR pada jumlah kasus pneumonia pada balita di Kabupaten Tuban.

4.2.1 Gambaran Umum Data Penelitian

Statistik deskriptif memberikan gambaran umum mengenai karakteristik data yang digunakan dalam penelitian, meliputi nilai rata-rata (*mean*), standar deviasi, nilai minimum, dan nilai maksimum. Ringkasan statistik untuk variabel respon dan kelima variabel prediktor disajikan pada Tabel 4.1.

Tabel 4.1 Statistik Deskriptif

Variabel	Rata-rata	Standar Deviasi	Minimum	Maksimum
[Y1] Pneumonia	4,53	7,78	0	115
[X1] Asi Eksklusif (%)	6,54	7,23	0,00	64,10
[X2] Vitamin A (%)	110,69	71,25	0,00	813,04
[X3] Kelas Ibu Hamil (%)	63,83	57,69	0,00	500,00
[X4] Cakupan Air Bersih	97,17	6,09	61,11	100,00
[X5] Imunisasi Campak	21,86	8,61	6,56	89,74

Berdasarkan hasil analisis statistik deskriptif yang disajikan pada Tabel 4.1, data menunjukkan bahwa rata-rata kejadian pneumonia di setiap desa/kelurahan adalah sekitar 4 hingga 5 kasus. Namun, angka rata-rata ini tidak mencerminkan kondisi seragam di seluruh wilayah. Rentang data menunjukkan variasi yang sangat lebar, di mana terdapat wilayah yang mencatatkan 0 kasus, sementara wilayah lain mencatat hingga 115 kasus. Hal ini mengindikasikan bahwa kejadian pneumonia

terpusat secara signifikan pada lokasi-lokasi tertentu saja, tidak menyebar merata di seluruh kabupaten.

Pada aspek pencegahan dan gizi, terdapat beberapa catatan penting. ASI menjadi makanan tunggal terbaik karena mengandung nutrisi lengkap dan antibodi untuk mendukung pertumbuhan, perkembangan otak, dan kekebalan tubuh bayi. Namun menurut data yang terjadi di kabupaten Tuban tingkat partisipasi pemberian ASI Eksklusif tercatat rendah, dengan rata-rata capaian wilayah hanya sebesar 6,54%. Angka ini menunjukkan bahwa praktik pemberian ASI eksklusif belum menjadi budaya yang dominan di sebagian besar wilayah.

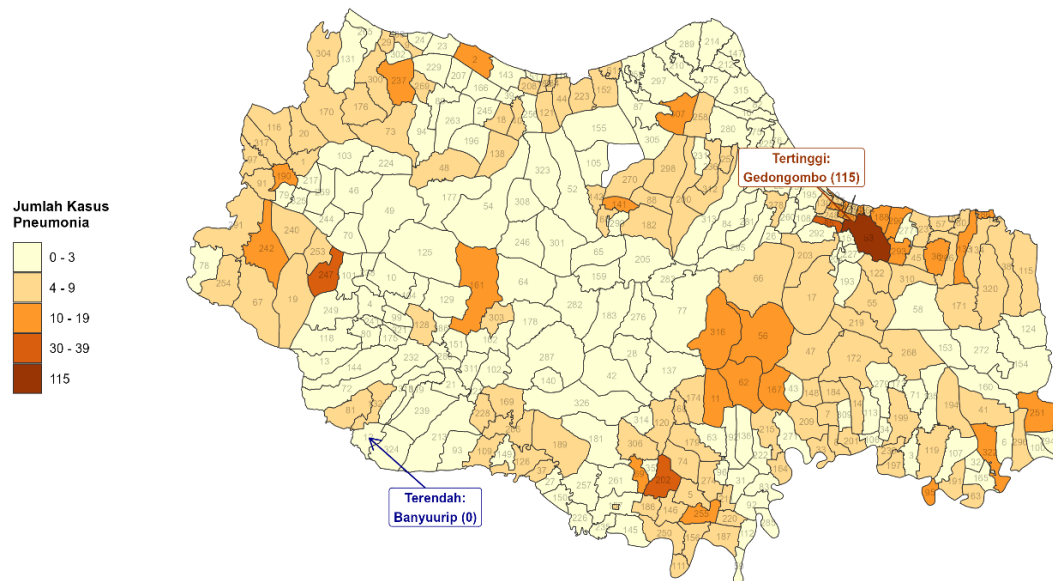
Selanjutnya data menunjukkan rata-rata capaian di atas 100% untuk pemberian Vitamin A (110,69%) dan angka yang bervariasi luas pada Kelas Ibu Hamil (63,83%). Nilai maksimum yang tinggi pada kedua variabel ini (mencapai 813,04% untuk Vitamin A dan 500% untuk Kelas Ibu Hamil) kemungkinan disebabkan oleh jumlah pelayanan yang melebihi estimasi target penduduk asli, misalnya karena adanya pelayanan bagi warga pendatang atau perbedaan basis data sasaran. Tingginya cakupan Vitamin A di data ini adalah modal positif, artinya secara programatik, distribusi suplemen sudah berjalan baik untuk menjaga struktur jaringan paru balita. Kemudian variasi angka partisipasi dalam data ini menyiratkan bahwa kesiapan ibu dalam merawat kesehatan bayi belum merata.

Secara rata-rata, 97,17% wilayah telah terlayani air bersih. Meskipun demikian, perhatian tetap diperlukan bagi wilayah dengan cakupan terendah yang berada di angka 61,11% untuk memastikan pemerataan akses sanitasi dasar. Ketersediaan air bersih berkorelasi langsung dengan Higiene dan Sanitasi (PHBS). Meskipun pneumonia adalah penyakit pernapasan, kebiasaan mencuci tangan ibu dan pengasuh dengan air bersih terbukti efektif memutus rantai penularan virus dan bakteri. Tingginya akses air bersih di Tuban adalah fondasi yang kuat untuk pencegahan penyakit infeksi secara umum.

Terdapat ketimpangan yang cukup tajam dalam pelaksanaan imunisasi campak. Rata-rata cakupan imunisasi berada di angka 21,86%. Namun, kesenjangan antarwilayah sangat terlihat, di mana wilayah dengan kinerja terendah hanya mencakup 6,56% sasaran, sedangkan wilayah dengan kinerja terbaik mampu menjangkau hingga 89,74% sasaran. Campak dikenal dapat melemahkan sistem

imun anak selama beberapa minggu hingga bulan setelah infeksi, yang sering kali berujung pada komplikasi berat berupa pneumonia. Rendahnya cakupan imunisasi campak di data ini (21,86%) membuka celah risiko besar yang mana banyak balita yang tidak terlindungi dari virus campak, yang secara tidak langsung meningkatkan risiko mereka terkena pneumonia berat sebagai komplikasi ikutan.

Selanjutnya untuk memberikan gambaran yang lebih komprehensif mengenai lokasi sebaran data dan mengidentifikasi wilayah dengan kecenderungan nilai ekstrem, berikut ditampilkan peta sebaran bagi setiap variabel penelitian.

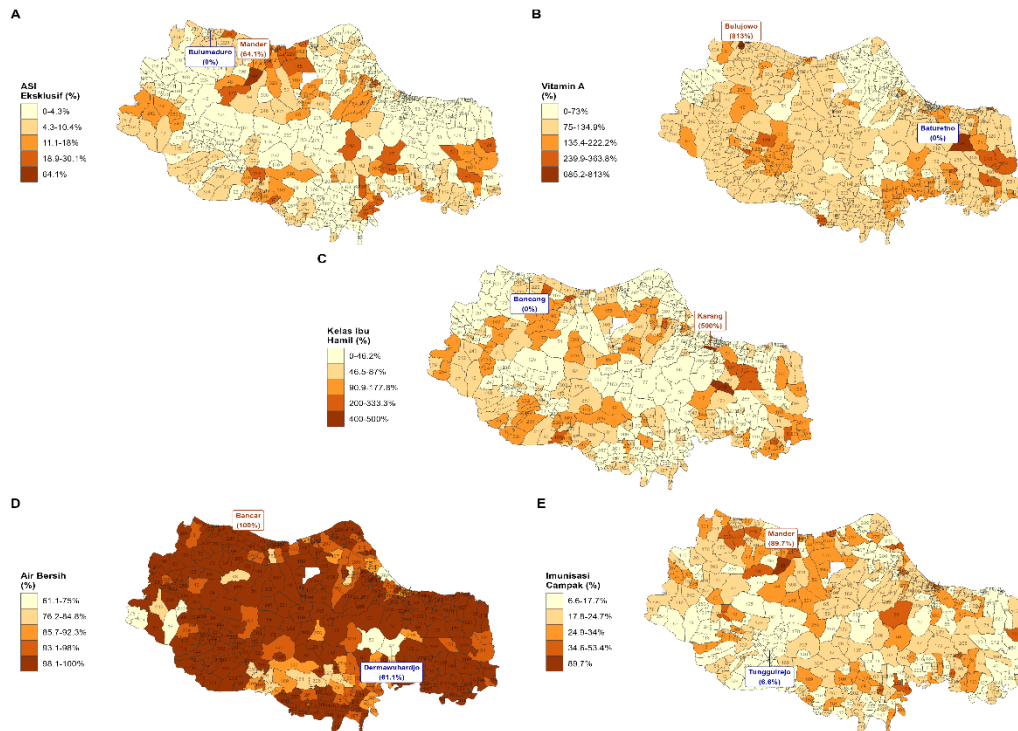


Gambar 4.1 Sebaran Kasus Jumlah Pneumonia menurut desa/kelurahan di Kabupaten Tuban Tahun 2023

Perbedaan jumlah kasus yang sangat mencolok antar desa/kelurahan seperti yang disajikan pada Gambar 4.1. Di satu sisi, terdapat desa yang tidak memiliki kasus sama sekali atau nol kasus, seperti Desa Banyuurip. Di sisi lain, terdapat desa dengan jumlah kasus yang sangat tinggi, yaitu Desa Gedongombo dengan 115 kasus. Kondisi ini menunjukkan bahwa beban penyakit pneumonia tidak tersebar merata di semua wilayah. Ada desa tertentu yang menanggung jumlah kasus jauh lebih besar dibandingkan rata-rata wilayah lainnya. Dalam istilah statistik, variasi data yang besar ini disebut sebagai overdispersi.

Selain ketimpangan jumlah kasus, peta juga memperlihatkan pola penyebaran yang mengelompok. Desa-desa dengan kasus tinggi cenderung berkumpul di

wilayah bagian timur Kabupaten Tuban. Wilayah yang berwarna coklat tampak saling berdekatan satu sama lain. Sebaliknya, wilayah bagian barat didominasi oleh warna kuning muda yang menandakan kasus rendah. Pola ini mengindikasikan bahwa kejadian pneumonia di satu desa berkaitan dengan kondisi di desa tetangganya. Keterkaitan antarwilayah yang berdekatan ini merupakan tanda adanya autokorelasi spasial.



Gambar 4.2 Sebaran Kasus Persentase ASI Eksklusif (A), Vitamin A (B), Kelas Ibu Hamil (C), Air Bersih (D), dan Imunisasi Campak (E) Menurut Desa/kelurahan di Kabupaten Tuban Tahun 2023

Peta sebaran spasial di Kabupaten Tuban menyingkap sebuah dikotomi tajam antara kemapanan infrastruktur fisik dan kerapuhan partisipasi kesehatan masyarakat. Akses Air Bersih (Peta D) berdiri sebagai satu-satunya variabel yang menunjukkan keberhasilan kawasan yang solid; blok warna gelap yang mendominasi hampir seluruh peta menegaskan bahwa sanitasi dasar telah menjadi "tulang punggung" yang kokoh dan merata, tidak lagi menjadi pembeda risiko yang signifikan antarwilayah. Kontras yang sangat mencolok justru terlihat pada ASI Eksklusif (Peta A), yang tampil sebagai "sisi terlemah" dalam pertahanan kesehatan balita. Alih-alih merata, peta ASI didominasi oleh lautan warna pucat dengan

capaian rendah yang nyaris seragam di seluruh kabupaten, menyisakan hanya segelintir desa sebagai "pulau keberhasilan" yang terisolasi. Ini menandakan bahwa intervensi yang menuntut perubahan perilaku budaya (menyusui) jauh lebih sulit ditembus dibandingkan sekadar membangun instalasi pipa air.

Di sisi lain, peta perlindungan medis dan edukasi (Vitamin A, Kelas Ibu Hamil, dan Imunisasi) memperlihatkan wajah ketimpangan yang lebih kaotis dan sporadis. Tidak seperti pola air bersih yang stabil, sebaran Vitamin A dan partisipasi Kelas Ibu Hamil tampak acak desa dengan capaian ekstrem tinggi bisa bersebelahan langsung dengan desa yang nol capaiannya mencerminkan ketidakkonsistenan performa pelayanan di tingkat mikro. Lebih mengkhawatirkan lagi adalah pola pada Imunisasi Campak (Peta E); meskipun sekilas terlihat lebih baik dari ASI, keberadaan "kantong-kantong" wilayah berwarna terang yang menyebar di tengah wilayah bergambar gelap menciptakan celah kerentanan yang berbahaya. Dalam epidemiologi, lubang-lubang kecil dalam cakupan imunisasi ini sudah cukup untuk menjadi pintu masuk wabah yang fatal. Secara keseluruhan, lanskap spasial Tuban menceritakan satu simpulan tegas yaitu risiko pneumonia di wilayah ini tidak lagi dipicu oleh lingkungan fisik yang buruk, melainkan oleh kegagalan sistemik dalam pemerataan perilaku sehat dan perlindungan biologis pada balita.

Kemudian kondisi tumpang tindih dari peta-peta ini memperlihatkan adanya akumulasi risiko yang serius pada wilayah tertentu. Desa-desa yang konsisten berwarna terang di berbagai peta misalnya desa yang cakupan ASI-nya rendah sekaligus imunisasinya minim pada dasarnya membentuk zona kerentanan ganda. Di wilayah ini, balita menghadapi ancaman dari dua arah sekaligus, yaitu pertahanan internal tubuh yang lemah karena kurangnya asupan alamiah, dan hilangnya perlindungan eksternal karena tidak terbentuknya kekebalan kelompok di lingkungan mereka. Realitas spasial ini menegaskan bahwa pekerjaan rumah Kabupaten Tuban saat ini bukan lagi sekadar pembangunan fisik, melainkan pembangunan sosial. Oleh karena itu tantangannya adalah bagaimana memastikan standar kesehatan di satu desa dapat "menular" ke desa tetangganya, menghapus batas tegas antara wilayah yang sehat dan yang rentan demi keselamatan balita secara menyeluruh.

4.2.2 Pola Hubungan Variabel Respon dan Variabel Prediktor

Sebelum analisis lebih lanjut dilakukan, pola hubungan antara variabel respon dan variabel prediktor perlu diketahui. Pola hubungan yang mungkin terjadi adalah hubungan positif, Negatif, dan tidak berhubungan. Hubungan variabel respon dengan variabel prediktor ditunjukkan dengan nilai korelasi. Nilai korelasi antara variabel respon dengan variabel prediktor ditunjukkan pada Tabel 4.2.

Tabel 4.2 Korelasi Variabel Respon dan Variabel Prediktor

Variabel	Pneumonia	Asi Eks	Vit A	Kls Ibu Hamil	Air Bersih	Campak
Pneumonia	1,00					
Asi Eks	-0,04	1,00				
Vit A	-0,02	0,03	1,00			
Kls Ibu Hamil	0,12	0,11	0,14	1,00		
Air Bersih	-0,06	-0,12	0,11	-0,04	1,00	
Im Campak	0,03	0,41	-0,07	-0,07	-0,03	1,00

Berdasarkan Tabel 4.2, hasil uji korelasi menunjukkan nilai koefisien korelasi antara variabel respon (Pneumonia) dengan variabel-variabel prediktor tergolong sangat lemah (mendekati 0), baik yang bernilai positif maupun negatif. Sebagai contoh, korelasi ASI Eksklusif tercatat sebesar -0,04 dan Vitamin A sebesar -0,02.

Lemahnya korelasi linier global ini tidak serta-merta menyimpulkan bahwa tidak ada hubungan kausalitas. Justru, hal ini memberikan indikasi awal bahwa hubungan antara faktor risiko dan kejadian Pneumonia bersifat kompleks dan tidak dapat ditangkap sepenuhnya hanya dengan pendekatan linier sederhana. Kemungkinan besar, hubungan tersebut tertutupi oleh variasi spasial (efek lokasi) atau adanya hubungan non-linier akibat distribusi data yang *overdispersed*. Temuan ini semakin memperkuat urgensi penggunaan metode multivariat yang lebih *robust* seperti regresi spasial untuk menyingkap pola hubungan yang sebenarnya.

4.2.3 Deteksi Multikolinieritas

Pemeriksaan multikolinearitas dilakukan untuk memenuhi asumsi pemodelan regresi, yaitu antar variabel prediktor harus independen. Hasil perhitungan nilai VIF pada variabel prediktor disajikan dalam Tabel 4.3.

Tabel 4.3 Nilai VIF Masing-masing Variabel Prediktor

Variabel	VIF
[X1] Asi Eksklusif (%)	1,25
[X2] Vitamin A (%)	1,04
[X3] Kelas Ibu Hamil (%)	1,05
[X4] Cakupan Air Bersih	1,03
[X5] Imunisasi Campak	1,22

Seluruh variabel prediktor memiliki nilai VIF yang sangat rendah, berkisar antara 1,03 hingga 1,25. Nilai ini berada jauh di bawah ambang batas kritis yang umumnya ditetapkan dalam literatur statistik, yaitu $VIF < 10$ (atau ambang batas yang lebih ketat $VIF < 5$).

Rendahnya nilai VIF ini mengonfirmasi bahwa tidak ada lagi masalah multikolinearitas dalam model. Dengan demikian, estimasi koefisien regresi yang dihasilkan dapat dipastikan bersifat stabil, presisi, dan tidak bias akibat inflasi varians, sehingga signifikansi pengaruh masing-masing variabel prediktor terhadap kejadian Pneumonia dapat diinterpretasikan secara valid.

4.2.4 Pengujian Equidisersi

Berdasarkan hasil analisis, diperoleh nilai statistik *Deviance* sebesar 1.601,227 dengan derajat bebas (*df*) 322. Nilai ini terbukti melampaui nilai kritis pada tabel distribusi Chi-Square (χ^2) untuk $df \approx 320$ pada taraf nyata 5%, yaitu sebesar 362,72.

4.2.5 Pemodelan *Generalized Poisson Regression*

Setelah dikonfirmasi adanya overdispersi pada data, artinya pemodelan dengan regresi Poisson tidak disarankan, sehingga pemodelan dilanjutkan menggunakan pendekatan *Generalized Poisson Regression* (GPR). Tabel 4.4 menyajikan hasil estimasi parameter model GPR yang mencakup nilai koefisien, *standard error*, serta nilai signifikansi statistik (*p – value*) untuk setiap variabel.

Tabel 4.4 Nilai Taksiran Parameter GPR				
Parameter	Nilai Taksiran	Standar Error	Z	<i>p-value</i>
$[\beta_0]$ Intercept	-0,6636	0,9042	-0,7339	0,4629
$[\phi]$ Dispersi	-6,1351	0,0899	-68,2117	0,0001
$[\beta_1]$ Asi Eksklusif	0,0210	0,0084	2,5175	0,0118
$[\beta_2]$ Vitamin A	-0,0013	0,0009	-1,4660	0,1426
$[\beta_3]$ Kelas Ibu Hamil	0,0011	0,0012	0,9771	0,3285
$[\beta_4]$ Cakupan Air Bersih	-0,0302	0,0091	-3,3050	0,0009
$[\beta_5]$ Imunisasi Campak	0,0072	0,0070	1,0301	0,3029

Setelah data selesai dimodelkan dengan model GPR maka selanjutnya adalah menguji residual dari model untuk menguji dependensi spasial.

4.2.6 Pembobotan Spasial dan Pengujian Dependensi Spasial

Pemodelan GPSAR untuk memodelkan jumlah kasus pneumonia pada balita di Kabupaten Tuban menggunakan matriks pembobot spasial *queen contiguity*. Hal tersebut berdasarkan bentuk wilayah Kabupaten Tuban yang tidak simetris. Matriks pembobot spasial dibuat berdasarkan persinggungan antar desa/kelurahan yang ditampilkan pada Lampiran 2.

Berdasarkan hasil analisis, diperoleh nilai indeks *Moran's I* teramati sebesar 0,278, yang lebih besar dari nilai ekspektasinya ($E[I] = -0,003$). Nilai positif ini mengindikasikan adanya pola autokorelasi spasial positif, di mana wilayah-wilayah dengan karakteristik sisaan serupa cenderung saling berkelompok (*clustering*). Signifikansi pola ini dikonfirmasi oleh nilai statistik uji $Z_{hitung}(8,29) > Z_{tabel}(1,96)$ dengan nilai *p-value* yang sangat kecil, yaitu $2,2 \times 10^{-16}$. Mengingat nilai *p-value* jauh di bawah taraf nyata 5%, maka diputuskan untuk menolak Hipotesis Nol yang menyatakan tidak ada autokorelasi spasial. Keberadaan dependensi spasial yang signifikan ini menegaskan bahwa model regresi klasik seperti *Generalized Poisson Regression* tidak memadai untuk memodelkan data jumlah kasus Pneumonia di wilayah penelitian, sehingga diperlukan pendekatan pemodelan spasial untuk menangkap efek ketergantungan antarwilayah tersebut secara eksplisit. Untuk menentukan spesifikasi model spasial yang paling tepat dilakukan pengujian *Lagrange Multiplier* (LM) sesuai dengan prosedur keputusan Anselin (1988). Berikut adalah hasil pengujian *Lagrange Multiplier* yang disajikan pada tabel 4.5.

Tabel 4.5 Pengujian *Lagrange Multiplier* (LM)

Uji Diagnostik	Nilai Statistik	df	<i>p-value</i>	Keterangan
LM Error (SEM)	26,748	1	$2,32 \times 10^{-7}$	Signifikan
LM Lag (SAR)	29,730	1	$4,97 \times 10^{-8}$	Signifikan
Robust LM Error	3,187	1	0,074	Tidak Signifikan
Robust LM Lag	6,169	1	0,013	Signifikan
SARMA	32,918	2	$7,12 \times 10^{-8}$	Signifikan

Pada tahap awal, uji LM standar menunjukkan bahwa kedua spesifikasi, baik *LM Lag* maupun *LM Error*, bernilai signifikan pada taraf nyata 5% ($p < 0,001$). Kondisi signifikansi ganda ini mengindikasikan perlunya pemeriksaan lanjutan menggunakan uji LM Robust untuk mengidentifikasi efek spasial yang lebih

dominan. Berdasarkan hasil uji Robust, terlihat perbedaan yang menentukan statistik Robust *LM Lag* tetap menunjukkan signifikansi statistik dengan nilai p sebesar 0,013 ($< 0,05$), sedangkan statistik Robust *LM Error* menjadi tidak signifikan dengan nilai p sebesar 0,074 ($> 0,05$).

Ketika uji Robust *LM Lag* signifikan sementara Robust *LM Error* tidak signifikan, maka spesifikasi model yang paling sesuai untuk data ini adalah *Spatial Autoregressive* (SAR). Keputusan statistik ini sejalan dengan kerangka teoritis penyakit menular, di mana kejadian penyakit di suatu wilayah dipengaruhi secara langsung oleh kejadian penyakit di wilayah sekitarnya, bukan sekadar oleh kemiripan karakteristik error yang tidak teramati. Oleh karena itu, analisis selanjutnya akan menggunakan model *Generalized Poisson Spatial Autoregressive* yang mampu fleksibel terhadap asumsi *equidispersion* dan asumsi independen antar tiap pengamatan secara simultan.

4.2.7 Penaksir Parameter Generalized Poisson Spatial Autoregressive

Penaksiran parameter GPSAR dilakukan menggunakan metode MLE dan dilanjutkan dengan iterasi numerik BHHH karena tidak didapatkan bentuk yang *closed form* seperti yang telah diuraikan langkah-langkahnya pada Subbab 3.3. Setelah proses iterasi konvergen, diperoleh nilai penaksir $\hat{\beta}$, $\hat{\phi}$, dan $\hat{\rho}$. Hasil penaksiran parameter untuk model GPSAR pada data kasus ini ditunjukkan pada Tabel 4.6.

Tabel 4.6 Nilai Taksiran Parameter GPSAR

Parameter	Nilai Taksiran	Standar Error	Z	<i>p-value</i>
$[\beta_0]$ Intercept	$6,095 \times 10^{-1}$	$1,663 \times 10^{-1}$	2,692	0,0002
$[\beta_1]$ Asi Eksklusif	$-2,413 \times 10^{-3}$	$2,307 \times 10^{-3}$	-0,953	0,341
$[\beta_2]$ Vitamin A	$-6,110 \times 10^{-5}$	$1,294 \times 10^{-4}$	-0,085	0,932
$[\beta_3]$ Kelas Ibu Hamil	$1,432 \times 10^{-3}$	$4,976 \times 10^{-4}$	2,757	0,004
$[\beta_4]$ Cakupan Air Bersih	$-9,101 \times 10^{-3}$	$2,796 \times 10^{-3}$	-3,003	0,001
$[\beta_5]$ Imunisasi Campak	$7,170 \times 10^{-3}$	$2,851 \times 10^{-3}$	2,335	0,011
$[\rho]$ Lag	$9,899 \times 10^{-1}$	$1,895 \times 10^{-2}$	37,912	$2,2 \times 10^{-16}$
$[\phi]$ Dispersi	$2,991 \times 10^{-1}$	$2,392 \times 10^{-2}$	11,442	$2,2 \times 10^{-16}$

Hasil pengujian menunjukkan bahwa parameter dependensi spasial (ρ) memiliki nilai estimasi sebesar 0,989 dengan $p\text{-value} < 0,001$. Hal ini menunjukkan bahwa parameter ρ berpengaruh signifikan terhadap model. Nilai positif pada ρ mengindikasikan adanya autokorelasi spasial positif, di mana jumlah kasus pada

suatu lokasi berhubungan lurus dengan jumlah kasus pada lokasi-lokasi tetangganya. Selain itu, parameter dispersi (ϕ) memiliki nilai estimasi 0,299 dan bernilai signifikan, yang mengonfirmasi adanya *overdispersion* pada data sehingga penggunaan distribusi *Generalized Poisson* lebih tepat dibandingkan distribusi Poisson standar.

Secara parsial, variabel prediktor yang berpengaruh signifikan terhadap variabel respon pada taraf nyata 5% adalah Persentase ibu hamil yang mengikuti kelas ibu hamil (X_4), Persentase rumah tangga dengan cakupan air bersih (X_5), dan Persentase imunisasi campak (X_6).

Persamaan model GPSAR yang terbentuk berdasarkan nilai estimasi parameter adalah sebagai berikut:

$$\ln(\hat{\mu}_i^{GPSAR}) = \ln q_i + 0,9899 \sum_{j \neq i} w_{ij} (\ln \hat{\mu}_j^{GPSAR} - \ln q_j) + 0,6095 - 2,413X_{1i} \\ - 6,110 \times 10^{-5}X_{2i} + 0,0014X_{3i} - 0,0091X_{4i} + 0,0072X_{5i}$$

Dengan mengambil contoh Desa Sidomukti, model GPSAR desa Sidomukti dengan kode nomor desa 87 dapat dinyatakan sebagai berikut.

$$\ln(\hat{\mu}_{87}^{GPSAR}) = \ln q_i + 0,9899 \left[\frac{1}{6} \sum_j (\ln \hat{\mu}_j^{GPSAR} - \ln q_j) \right] + 0,601 - 2,41X_{1,87} \\ - 6,110 \times 10^{-5}X_{2,87} + 0,001X_{3,87} - 0,0091X_{4,87} + 0,007X_{5,87}$$

dengan $j \in \{55, 56, 85, 86, 88, 93\}$.

Model GPSAR pada Desa Sidomukti di atas dapat diinterpretasikan sebagai berikut.

1. Pengaruh spasial lag terbukti menjadi faktor signifikan secara statistik. Hal ini mengonfirmasi bahwa laju kejadian Pneumonia di Desa Sidomukti mengikuti pola wabah di desa-desa tetangganya seperti Jamprong, Bendonglateng, Sidohasri, Tawaran, Jombok, dan Karangtengah. Secara riil, dampaknya adalah jika rata-rata risiko di desa tetangga mengalami kenaikan sebesar 0,10 satuan, kasus di Sidomukti diproyeksikan turut melonjak sebesar 10,4% dengan asumsi faktor lain tetap. Apabila akumulasi risiko gabungan dari keenam desa tetangga meningkat satu unit, rata-rata kejadian di Sidomukti akan mengalami kenaikan drastis hingga 17,9%. Artinya, kondisi Desa Sidomukti tidak berdiri sendiri.

Kegagalan pengendalian wabah di desa tetangga akan secara langsung menggagalkan upaya kesehatan di Sidomukti.

2. Variabel Cakupan Kelas Ibu Hamil menunjukkan pengaruh positif yang signifikan terhadap jumlah kasus. Interpretasi nilai estimasi tersebut menunjukkan bahwa setiap kenaikan cakupan Kelas Ibu Hamil sebesar 1%, estimasi rata-rata kejadian Pneumonia di Desa Sidomukti meningkat sebesar 0,14%, dengan asumsi variabel lain konstan. Meskipun efek marginalnya relatif kecil yaitu sebesar 0,14%, signifikansi statistik ini mengindikasikan bahwa intensitas partisipasi dalam program kesehatan berkorelasi dengan tingkat pelaporan kasus. Hal ini kemungkinan merefleksikan fungsi deteksi dini yang mana semakin aktif kegiatan kelas ibu hamil, semakin tinggi probabilitas kasus Pneumonia yang teridentifikasi dan tercatat, bukan berarti kegiatan tersebut menyebabkan penyakit.
3. Variabel Cakupan Air Bersih terbukti memiliki pengaruh negatif yang signifikan terhadap laju kejadian Pneumonia. Estimasi ini mengonfirmasi bahwa perbaikan akses sanitasi dasar berfungsi sebagai faktor protektif yang efektif dalam mereduksi risiko penyakit. Secara kuantitatif, setiap peningkatan tingkat cakupan air bersih sebesar 1% di Desa Sidomukti berasosiasi dengan penurunan estimasi rata-rata kejadian Pneumonia sebesar 0,91%. Dalam simulasi peningkatan kapasitas layanan yang lebih luas sebesar 10%, model memprediksi penurunan rata-rata kasus yang substansial hingga 8,7%. Temuan ini sejalan dengan hipotesis epidemiologi lingkungan, di mana ketersediaan air bersih secara signifikan memutus rantai transmisi patogen penyebab Pneumonia.
4. Variabel Cakupan Imunisasi Campak menunjukkan asosiasi positif yang signifikan terhadap estimasi kejadian Pneumonia. Secara kuantitatif, setiap peningkatan cakupan imunisasi sebesar 1% berimplikasi pada kenaikan estimasi rata-rata kasus sebesar 0,72%. Dalam simulasi peningkatan cakupan yang lebih besar yaitu 10%, model memprediksi peningkatan angka pelaporan kasus sebesar 7,43%. Temuan arah koefisien yang positif ini perlu diinterpretasikan sebagai indikator kinerja surveilans, bukan hubungan kausalitas patologis. Korelasi positif ini mengindikasikan bahwa wilayah dengan cakupan imunisasi yang tinggi cenderung memiliki akses dan sistem pelaporan kesehatan yang

lebih baik, sehingga kapasitas deteksi dan pencatatan kasus menjadi lebih tinggi dibandingkan wilayah dengan cakupan layanan kesehatan yang rendah.

5. Estimasi parameter dispersi diperoleh sebesar $\hat{\phi} = 0,2991$ dengan tingkat signifikansi yang tinggi ($p < 0,001$). Signifikansi statistik pada parameter ini mengonfirmasi adanya fenomena overdispersi pada data, di mana variabilitas jumlah kasus Pneumonia antar wilayah secara nyata melebihi nilai rata-ratanya, sehingga melanggar asumsi ekidispersi yang disyaratkan oleh regresi Poisson standar. Temuan ini memberikan justifikasi empiris yang kuat bahwa penerapan model GPSAR adalah langkah metodologis yang tepat dibandingkan model Poisson klasik. Pengakomodasian parameter dispersi ini krusial untuk mengoreksi bias pada estimasi standard error, sehingga menghasilkan kesimpulan inferensi (uji signifikansi) yang lebih valid.

4.2.8 Pemilihan Model Terbaik

Pemilihan model terbaik antara GPR dan GPSAR dilakukan dengan membandingkan nilai AICc. Hasil perbandingan nilai AICc dan RMSE kedua model ditunjukkan pada Tabel 4.7.

Tabel 4.7 Perbandingan nilai AICc

Model	AICc
<i>Generalized Poisson Regression</i>	1717,24
<i>Generalized Poisson Spatial Autoregressive</i>	1673,39

Berdasarkan tabel tersebut, diperoleh nilai AICc untuk model GPSAR sebesar 1673,39, nilai ini lebih kecil dibandingkan dengan nilai AICc pada model GPR Non-Spasial sebesar 1717,24. Nilai AICc yang lebih kecil mengindikasikan bahwa model GPSAR memiliki keseimbangan yang lebih baik antara kompleksitas model.

BAB V

KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan yang dapat diambil berdasarkan hasil dan pembahasan yang diperoleh. Beberapa saran juga disampaikan sebagai rekomendasi untuk penelitian selanjutnya.

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan, diperoleh kesimpulan sebagai berikut.

1. Penaksir parameter model GPSAR dilakukan dengan metode *Maximum Likelihood Estimation* adalah sebagai berikut.

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n \left[\left(\frac{y_i - \mu_i^{GPSAR}}{(1 + \phi \mu_i^{GPSAR})^2} \right) (\mathbf{A}_{(n \times n)}^{-1} \mathbf{X}_{(n \times (p+1))})_i^T \right] \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \rho} &= \sum_{i=1}^n \left[\frac{y_i - \mu_i^{GPSAR}}{(1 + \phi \mu_i^{GPSAR})^2} [\mathbf{A}_{(n \times n)}^{-1} \mathbf{W}_{(n \times n)} \boldsymbol{\eta}_{(n \times 1)}]_i \right] \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \phi} &= \sum_{i=1}^n \left[-\frac{y_i \mu_i^{GPSAR}}{1 + \phi \mu_i^{GPSAR}} + \frac{y_i(y_i - 1)}{1 + \phi y_i} + \frac{\mu_i^{GPSAR}(y_i - \mu_i^{GPSAR})}{(1 + \phi \mu_i^{GPSAR})^2} \right]\end{aligned}$$

Hasil penaksiran parameter model GPSAR dengan pendekatan MLE tidak *closed form*, sehingga didekati secara numerik dengan iterasi BHHH.

2. Model GPSAR menghasilkan variabel-variabel yang signifikan terhadap jumlah kasus pneumonia pada balita di 328 desa/kelurahan di Kabupaten Tuban Tahun 2023, antara lain persentase ibu hamil yang mengikuti kelas ibu hamil, persentase rumah tangga dengan cakupan air bersih, dan persentase balita yang mendapatkan imunisasi campak. Berdasarkan nilai AICc model GPSAR memiliki nilai AICc yang lebih kecil (1673,39) daripada model *Generalized Poisson Regression* (GPR) yang memiliki nilai AICc sebesar 1717,24. Hal tersebut membuktikan bahwa model GPSAR lebih baik daripada model GPR dalam memodelkan jumlah kasus pneumonia pada balita di 328 desa/kelurahan di Kabupaten Tuban Tahun 2023. Selain itu, parameter lag spasial (ρ) dan parameter dispersi (ϕ) yang sangat signifikan menunjukkan

adanya dependensi spasial yang kuat dan fenomena overdispersi pada data, yang berhasil diakomodasi oleh model GPSAR.

5.2 Saran

Saran yang dapat diberikan untuk penelitian selanjutnya adalah sebagai berikut.

1. Penelitian selanjutnya sangat disarankan untuk beralih dari penggunaan data agregat level desa ke data berbasis titik lokasi individu, yaitu menggunakan alamat rumah pasien yang telah di-*geocode*. Penggunaan data titik ini memungkinkan analisis dilakukan dengan resolusi spasial yang lebih halus serta menghindari bias asumsi bahwa risiko penyakit seragam dalam satu wilayah administrasi (*homogeneous risk*). Sebagaimana ditunjukkan oleh Choiruddin et al. (2023), penggunaan data titik dalam analisis COVID-19 terbukti mampu mengungkap pola dispersi dan faktor risiko pada skala mikro dengan akurasi yang lebih tinggi.
2. Mengingat model GPSAR yang dikembangkan saat ini masih terbatas pada analisis *cross-sectional*, penelitian selanjutnya perlu memperluas cakupan model ke ranah spasio-temporal. Pengembangan ini bertujuan untuk menganalisis dinamika penyebaran penyakit dari tahun ke tahun, sehingga tren perubahan pola penyakit dapat dipahami secara lebih komprehensif daripada sekadar memodelkan data agregat spasial panel tradisional.
3. Berdasarkan hasil uji diagnostik *Lagrange Multiplier* (LM) pada penelitian ini, terindikasi adanya dependensi spasial yang tidak hanya terjadi pada lag variabel respon (*Spatial Lag*), tetapi juga berpotensi terjadi pada sisaan (*Spatial Error*). Hal ini dibuktikan dengan nilai statistik uji SARMA yang signifikan secara statistik. Oleh karena itu, penelitian selanjutnya disarankan untuk mengembangkan model *Generalized Poisson Spatial Autoregressive Moving Average* (GPSARMA). Pengembangan model ini diharapkan mampu mengakomodasi struktur hubungan spasial yang lebih kompleks, yaitu gabungan antara autokorelasi pada variabel dependen dan korelasi pada error model secara bersamaan.

4. Karena keterbatasan uji *Lagrange Multiplier* (LM) pada penelitian ini, kedepannya dapat mengembangkan statistik uji yang robust untuk data count.

DAFTAR PUSTAKA

- Adrian, T., & Brunnermeier, M. K. (2011). *CoVaR*. National Bureau of Economic Research. <https://doi.org/10.3386/w17454>
- Aina, S. Q., Fauziyah, M., & Nurmayanti, W. P. (2025). Comparison of Generalized Poisson Regression and Negative Binomial Regression Models Based on Akaike Information Criterion Values. *Statistika*, 25(1), 86–101. <https://doi.org/10.29313/statistika.v25i1.5402>
- Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199–213). Springer. https://doi.org/10.1007/978-1-4612-1694-0_15
- Alvionita, V., Sulfatimah, S., Astuti, A., & Nurfitri, N. (2022). Hubungan Status Gizi Dan Status Imunisasi Dengan Kejadian Pneumonia Pada Bayi: Hubungan Status Gizi Dan Status Imunisasi Dengan Kejadian Pneumonia Pada Bayi. *Ahmar Metastasis Health Journal*, 1(4), 137–143. <https://doi.org/10.53770/amhj.v1i4.92>
- Anselin, L. (1988). *Spatial econometrics: methods and models* (Vol. 4). Springer Science & Business Media. <https://doi.org/10.1007/978-94-015-7799-1>
- Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3), 247–267. [https://doi.org/10.1016/S0169-5150\(02\)00077-4](https://doi.org/10.1016/S0169-5150(02)00077-4)
- Aprilia, R., & Faisal, F. (2024). Tinjauan Literatur: Faktor Risiko dan Epidemiologi Pneumonia pada Balita. *Scientific Journal*, 3(3), 166–173. <https://doi.org/10.56260/sciena.v3i3.144>
- Aulia, I., Chamid, M. S., & Andari, S. (2017). Pemodelan Pneumonia pada Balita di Surabaya Menggunakan Spatial Autoregressive Models. *Jurnal Sains Dan Seni ITS*, 6(1), D82–D88. <https://doi.org/10.12962/j23373520.v6i1.22437>
- Beenstock, M., & Felsenstein, D. (2019). *The econometric analysis of non-stationary spatial panel data*. Springer. <https://doi.org/10.1007/978-3-030-03614-0>
- Berliana, S. M., Purihadi, Sutikno, & Rahayu, S. P. (2020). Parameter estimation and hypothesis testing of geographically weighted multivariate generalized poisson regression. *Mathematics*, 8(9), 1523. <https://doi.org/10.3390/math8091523>
- Bhati, A. S. (2005). Robust Spatial Analysis of Rare Crimes: An Information-Theoretic Approach. *Sociological Methodology*, 35(1), 227–289. <https://doi.org/10.1111/j.0081-1750.2006.00169.x>
- Bivand, R. S., & Wong, D. W. S. (2018). Comparing implementations of global and local indicators of spatial association. *Test*, 27(3), 716–748. <https://doi.org/10.1007/s11749-018-0599-x>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Issue 53). Cambridge university press. <https://doi.org/10.1017/CBO9780511814365>

- Chen, J., Liu, L., Xiao, L., Xu, C., & Long, D. (2020). Integrative analysis of spatial heterogeneity and overdispersion of crime with a geographically weighted negative binomial model. *ISPRS International Journal of Geo-Information*, 9(1), 60. <https://doi.org/10.3390/ijgi9010060>
- Choiruddin, A., Hannanu, F. F., Mateu, J., & Fitriyanah, V. (2023). COVID-19 transmission risk in Surabaya and Sidoarjo: an inhomogeneous marked Poisson point process approach. *Stochastic Environmental Research and Risk Assessment*, 37(6), 2271–2282. <https://doi.org/10.1007/s00477-023-02393-5>
- Cruz, N. A., Toloza-Delgado, J. D., & Melo, O. O. (2024). Generalized spatial autoregressive model. *ArXiv Preprint ArXiv:2412.00945*. <https://doi.org/10.48550/arXiv.2412.00945>
- Davis, D. E., & Schluter, G. E. (2005). Labor-force heterogeneity as a source of agglomeration economies in an empirical analysis of county-level determinants of food plant entry. *Journal of Agricultural and Resource Economics*, 480–501. <https://doi.org/10.22004/ag.econ.30975>
- Edi, Y. S. (2012). *Quasi-Maximum Likelihood untuk Regresi Panel Spasial*. Institut Teknologi Sepuluh Nopember. <https://doi.org/10.1063/1.5139745>
- El-Osta, H. S., Mishra, A. K., & Morehart, M. J. (2007). Determinants of economic well-being among US farm operator households. *Agricultural Economics*, 36(3), 291–304. <https://doi.org/10.1111/j.1574-0862.2007.00207.x>
- Farisa, F. A., Salby, S. N. H., Rahman, A. A., & Purhadi, P. (2023). Modeling the Number of Pneumonia in Toddlers in East Java Province in 2021 with Generalized Poisson Regression. *Inferensi*, 6(2), 91–96. <https://doi.org/10.12962/j27213862.v6i2.15339>
- Fingleton, B., & Le Gallo, J. (2008). Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties. *Papers in Regional Science*, 87(3), 319–340. <https://doi.org/10.1111/j.1435-5957.2008.00187.x>
- Fleming, M. M. (2004). Techniques for estimating spatially dependent discrete choice models. In *Advances in spatial econometrics: methodology, tools and applications* (pp. 145–168). Springer. https://doi.org/10.1007/978-3-662-05617-2_7
- Flores, O., Rossi, V., & Mortier, F. (2009). Autocorrelation offsets zero-inflation in models of tropical saplings density. *Ecological Modelling*, 220(15), 1797–1809. <https://doi.org/10.1016/j.ecolmodel.2009.01.030> Get rights and content
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2009). Geographically weighted regression. *The Sage Handbook of Spatial Analysis*, 1, 243–254. <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>
- Fotopoulos, G., & Louri, H. (2000). Location and survival of new entry. *Small Business Economics*, 14(4), 311–321. <https://doi.org/10.1023/A:1008180522759>
- Glaser, S. (2017). *A review of spatial econometric models for count data*. <https://hdl.handle.net/10419/168040>
- Goulias, K. G. (2002). Statistical and Econometric Data Analysis. *Transportation Systems Planning: Methods and Applications*.
- Griffith, D. A. (2002). A spatial filtering specification for the auto-Poisson model. *Statistics & Probability Letters*, 58(3), 245–251. <https://doi.org/10.1016/>

S0167-7152(02)00099-8

- Haining, R., Law, J., & Griffith, D. (2009). Modelling small area counts in the presence of overdispersion and spatial autocorrelation. *Computational Statistics & Data Analysis*, 53(8), 2923–2937. <https://doi.org/10.1016/j.csda.2008.08.014>
- Hays, J., & Franzese Robert, J. (2009). *A comparison of the small-sample properties of several estimators for spatial-lag count models*. <https://api.semanticscholar.org/CorpusID:34907698>
- Hoaglin, D. C., & Welsch, R. E. (1978). The Hat Matrix in Regression and ANOVA. *The American Statistician*, 32(1), 17–22. <https://doi.org/10.1080/00031305.1978.10479237>
- Islam, M. N., & Islam, N. (2003). Retrospective study of 273 deaths due to poisoning at Sir Salimullah Medical College from 1988 to 1997. *Legal Medicine*, 5, S129–S131. [https://doi.org/10.1016/S1344-6223\(02\)00128-1](https://doi.org/10.1016/S1344-6223(02)00128-1)
- Kaiser, M. S., & Cressie, N. (1997). Modeling Poisson variables with positive spatial dependence. *Statistics & Probability Letters*, 35(4), 423–432. [https://doi.org/10.1016/S0167-7152\(97\)00041-2](https://doi.org/10.1016/S0167-7152(97)00041-2)
- Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40(2), 509–533. <https://doi.org/10.1111/1468-2354.00027>
- Klier, T., & McMillen, D. P. (2008). Clustering of auto supplier plants in the United States: generalized method of moments spatial logit for large samples. *Journal of Business & Economic Statistics*, 26(4), 460–471. <https://doi.org/10.1198/073500107000000188>
- Lambert, D. M., McNamara, K. T., & Garrett, M. I. (2006). An application of spatial Poisson models to manufacturing investment location analysis. *Journal of Agricultural and Applied Economics*, 38(1), 105–121. <https://doi.org/10.1017/S1074070800022100>
- Lee, K. L., & Bell, D. R. (2009). A spatial negative binomial regression of individual-level count data with regional and person-specific covariates. *Information and Software Technology (March)*. Doi, 10.
- LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis*, 32(1), 19–35.
- LeSage, J. P., Fischer, M. M., & Scherngell, T. (2007). Knowledge spillovers across Europe: Evidence from a Poisson spatial interaction model with spatial effects. *Papers in Regional Science*, 86(3), 393–422. <https://doi.org/10.1111/j.1538-4632.2000.tb00413.x>
- LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781420064254>
- McMillen, D. P. (2004). *Geographically weighted regression: the analysis of spatially varying relationships*. JSTOR. https://doi.org/10.1111/j.0002-9092.2004.600_2
- Mohammed, J. M. (2025). Spatial Regression Analysis Using Poisson Regression: Applications in Studying Traffic Accidents. *European Journal of Applied Science, Engineering and Technology*, 3(3), 268–275. [https://doi.org/10.59324/ejaset.2025.3\(3\).18](https://doi.org/10.59324/ejaset.2025.3(3).18)
- Mohebbi, M., Wolfe, R., & Forbes, A. (2014). Disease mapping and regression with

- count data in the presence of overdispersion and spatial autocorrelation: a Bayesian model averaging approach. *International Journal of Environmental Research and Public Health*, 11(1), 883–902. <https://doi.org/10.3390/ijerph110100883>
- Murphy, K. M., & Topel, R. H. (2002). Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics*, 20(1), 88–97. <https://doi.org/10.1080/07350015.1985.10509471>
- Pinkse, J., & Slade, M. E. (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics*, 85(1), 125–154. [https://doi.org/10.1016/S0304-4076\(97\)00097-3](https://doi.org/10.1016/S0304-4076(97)00097-3)
- Rohimah, S. R. (2015). Model Spasial Autotoregresif Poisson Untuk Mendeteksi Faktor-faktor Yang Berpengaruh Terhadap Jumlah Penderita HIV Di Provinsi Jawa Timur. *Indonesian Journal of Mathematics and Natural Sciences*, 38(2), 169–175. <https://doi.org/10.15294/ijmns.v38i2.5822>
- Rohimah, S. R., & Riyantobi, A. M. (2019). Model Spasial Otoresif Poisson untuk Mendeteksi Faktor-Faktor yang Berpengaruh terhadap Jumlah Penderita HIV AIDS di Jakarta Timur. *Jurnal Statistika Dan Aplikasinya (JSA) Vol, 3*, 35–44. <https://doi.org/10.21009/jsa.03205>
- Rumeileh, S. A. (2024). *Global, regional, and national incidence and mortality burden of non-COVID-19 lower respiratory infections and aetiologies, 1990–2021: a systematic analysis from the Global Burden of Disease Study 2021*. [https://doi.org/10.1016/S1473-3099\(24\)00176-2](https://doi.org/10.1016/S1473-3099(24)00176-2).
- Sellner, R., Fischer, M. M., & Koch, M. (2013). A spatial autoregressive Poisson gravity model. *Geographical Analysis*, 45(2), 180–201. <https://doi.org/10.1111/gean.12007>
- Sutikno, Purhadi, Fachrunisah, & Cahyoko, F. D. (2025). Estimation of parameters and hypothesis testing of multivariate spatial autoregressive model. *MethodsX*, 14, 103294. <https://doi.org/10.1016/j.mex.2025.103294>
- Tyagi, K., Rane, C., & Manry, M. (2022). Regression analysis. In *Artificial intelligence and machine learning for EDGE computing* (pp. 53–63). Elsevier. <https://doi.org/10.1016/B978-0-12-824054-0.00007-1>
- Ver Hoef, J. M., & Boveng, P. L. (2007). Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology*, 88(11), 2766–2772. <https://doi.org/10.1890/07-0043.1>
- Zhang, T., & Lin, G. (2008). Identification of local clusters for count data: a model-based Moran's I test. *Journal of Applied Statistics*, 35(3), 293–306. <https://doi.org/10.1080/02664760701833248>

LAMPIRAN

Lampiran 1. Data Penelitian

Kode	KECAMATAN	DESA	Y_1	X_1	X_2	X_3	X_4	X_5	q_i
1	BANCAR	BANCAR	11	4.31	132.76	58.82	100.00	12.07	116
2	BANCAR	BANJARJO	4	4.91	99.43	22.73	100.00	22.26	265
3	BANCAR	BOGOREJO	1	21.43	89.71	64.52	100.00	21.43	84
4	BANCAR	BONCONG	3	5.08	101.12	0.00	100.00	30.51	59
5	BANCAR	BULU MEDURO	3	0.00	15.24	0.00	100.00	18.75	64
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
111	MERAKURAK	BOGOREJO	2	9.30	94.05	62.50	100.00	20.93	43
112	MERAKURAK	BOREHBANGLE	5	4.55	17.28	200.00	95.74	22.73	66
113	MERAKURAK	KAPU	0	14.63	100.00	31.25	100.00	22.76	123
114	MERAKURAK	MANDIREJO	0	10.43	67.72	45.45	100.00	12.88	163
115	MERAKURAK	PONGPONGAN	6	2.27	21.50	52.63	93.06	22.16	176
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
324	WIDANG	SIMOREJO	10	2.78	106.01	116.28	100.00	16.67	144
325	WIDANG	SUMBEREJO	0	12.20	265.52	35.71	100.00	13.01	123
326	WIDANG	TEGALREJO	2	4.69	106.87	128.21	100.00	19.53	128
327	WIDANG	TEGALSARI	6	3.28	121.48	121.95	100.00	14.75	122
328	WIDANG	WIDANG	16	3.18	103.74	52.63	100.00	16.61	283

Lampiran 2. Persinggungan antara 328 desa/kelurahan di Kabupaten Tuban

Kode	KECAMATAN	DESA	Count	Persinggungan
1	BANCAR	BANCAR	5	23 143 166 207 229
2	BANCAR	BANJARJO	7	23 24 29 30 229 237 302
3	BANCAR	BOGOREJO	5	2 9 24 207 229
4	BANCAR	BONCONG	3	9 23 30
5	BANCAR	BULUJOWO	5	9 30 265 300 302
...
111	MERAKURAK	BOREHBANGLE	3	236 280 312
112	MERAKURAK	KAPU	5	139 218 281 295 313
113	MERAKURAK	MANDIREJO	8	22 84 218 233 273 278 281 313
114	MERAKURAK	PONGPONGAN	8	88 182 231 236 258 298 312 313
115	MERAKURAK	SAMBONGGEDE	6	84 139 225 267 273 313
...
322	WIDANG	SIMOREJO	4	154 160 294 296
323	WIDANG	SUMBEREJO	6	58 153 154 160 171 320
324	WIDANG	TEGALREJO	3	100 251 296
325	WIDANG	TEGALSARI	6	6 41 100 160 251 294
326	WIDANG	WIDANG	6	6 32 41 107 165 194

Lampiran 3. Turunan fungsi *likelihood* di bawah H_0

Pada pengujian hipotesis pengaruh variabel prediktor, didefinisikan H_0 sebagai kondisi di mana tidak terdapat pengaruh variabel prediktor terhadap variabel respon. Model yang terbentuk di bawah H_0 adalah model yang hanya memuat parameter intersep $\beta_{0\omega}$, parameter dependensi spasial (ρ_ω), dan parameter dispersi (ϕ_ω).

1. Spesifikasi Model di Bawah H_0

Fungsi link yang menghubungkan rata-rata respon $\mu_{i\omega}^{GPSAR}$ dengan komponen sistematik model tanpa kovariat didefinisikan sebagai:

$$\ln(\mu_{i\omega}^{GPSAR}) = \beta_{0\omega} + \rho_\omega \sum_{j=1}^n w_{ij} y_j$$

Sehingga, nilai harapan (rata-rata) untuk observasi $ke - i$ adalah:

$$\mu_{i\omega}^{GPSAR} = \exp \left(\beta_{0\omega} + \rho \sum_{j=1}^n w_{ij} y_j \right)$$

di mana \mathbf{w}_i^T adalah baris $ke - i$ dari matriks pembobot spasial \mathbf{W} , dan \mathbf{y} adalah vektor variabel respon. Parameter yang akan diestimasi adalah $\boldsymbol{\theta}_\omega = [\beta_{0\omega}, \rho_\omega, \phi_\omega]^T$.

Fungsi *log-likelihood* (ℓ) untuk satu observasi $ke - i$ adalah sebagai berikut:

$$\ell_{i\omega} = y_i \ln \left(\frac{\mu_{i\omega}^{GPSAR}}{1 + \phi_\omega \mu_{i\omega}^{GPSAR}} \right) + (y_i - 1) \ln(1 + \phi_\omega y_i) - \frac{\mu_{i\omega}^{GPSAR} (1 + \phi_\omega y_i)}{1 + \phi_\omega \mu_{i\omega}^{GPSAR}} - \ln(y_i!)$$

Total *log-likelihood* adalah

$$\begin{aligned} \ell(\omega) &= \sum_{i=1}^n \ell_{i\omega} = \sum_{i=1}^n y_i \ln \left(\frac{\mu_{i\omega}^{GPSAR}}{1 + \phi_\omega \mu_{i\omega}^{GPSAR}} \right) + (y_i - 1) \ln(1 + \phi_\omega y_i) - \frac{\mu_{i\omega}^{GPSAR} (1 + \phi_\omega y_i)}{1 + \phi_\omega \mu_{i\omega}^{GPSAR}} \\ &\quad - \ln(y_i!) \end{aligned}$$

Selanjutnya akan dicari turunan $\ell_{i\omega}(\boldsymbol{\theta}_\omega)$ terhadap $\mu_{i\omega}^{GPSAR}$.

$$\begin{aligned}
\frac{\partial \ell_{i\omega}(\theta_\omega)}{\partial \mu_{i\omega}^{GPSAR}} &= \frac{\partial}{\partial \mu_{i\omega}^{GPSAR}} \left(y_i \ln \left(\frac{\mu_{i\omega}^{GPSAR}}{1 + \phi_\omega \mu_{i\omega}^{GPSAR}} \right) + (y_i - 1) \ln(1 + \phi_\omega y_i) \right. \\
&\quad \left. - \frac{\mu_{i\omega}^{GPSAR}(1 + \phi_\omega y_i)}{1 + \phi_\omega \mu_{i\omega}^{GPSAR}} - \ln(y_i!) \right) \\
&= \left(\frac{y_i}{\mu_{i\omega}^{GPSAR}} - \frac{y_i \phi_\omega}{1 + \phi_\omega \mu_{i\omega}^{GPSAR}} \right) - \frac{1 + \phi_\omega y_i}{(1 + \phi_\omega \mu_{i\omega}^{GPSAR})^2} \\
&= \frac{y_i(1 + \phi_\omega \mu_{i\omega}^{GPSAR}) - y_i \phi_\omega \mu_{i\omega}^{GPSAR}}{\mu_{i\omega}^{GPSAR}(1 + \phi_\omega \mu_{i\omega}^{GPSAR})} - \frac{1 + \phi_\omega y_i}{(1 + \phi_\omega \mu_{i\omega}^{GPSAR})^2} \\
&= \frac{y_i}{\mu_{i\omega}^{GPSAR}(1 + \phi_\omega \mu_{i\omega}^{GPSAR})} - \frac{1 + \phi_\omega y_i}{(1 + \phi_\omega \mu_{i\omega}^{GPSAR})^2} \\
&= \frac{y_i(1 + \phi_\omega \mu_{i\omega}) - \mu_{i\omega}(1 + \phi_\omega y_i)}{\mu_{i\omega}^{GPSAR}(1 + \phi_\omega \mu_{i\omega}^{GPSAR})^2} \\
&= \frac{y_i + y_i \phi_\omega \mu_{i\omega}^{GPSAR} - \mu_{i\omega}^{GPSAR} - \mu_{i\omega}^{GPSAR} \phi_\omega y_i}{\mu_{i\omega}^{GPSAR}(1 + \phi_\omega \mu_{i\omega}^{GPSAR})^2} \\
&= \frac{y_i - \mu_{i\omega}^{GPSAR}}{\mu_{i\omega}^{GPSAR}(1 + \phi_\omega \mu_{i\omega}^{GPSAR})^2}
\end{aligned}$$

Selanjutnya adalah estimasi parameter regresi ($\beta_{0\omega}$) dan Spasial (ρ_ω). Akan digunakan aturan rantai untuk melakukan penurunan terhadap kedua parameter tersebut secara parsial dengan cara sebagai berikut.

$$\frac{\partial \ell(\omega_{GPSAR})}{\partial \theta} = \sum_{i=1}^n \frac{\partial \ell(\omega_{GPSAR})}{\partial \mu_{i\omega}^{GPSAR}} \cdot \frac{\partial \mu_{i\omega}^{GPSAR}}{\partial \eta_{i\omega}} \cdot \frac{\partial \eta_{i\omega}}{\partial \theta_\omega}$$

2. Turunan parsial terhadap $\beta_{0\omega}$.

Diketahui $\mu_{i\omega}^{GPSAR} = \exp(\eta_{i\omega})$, maka $\frac{\partial \mu_{i\omega}^{GPSAR}}{\partial \eta_{i\omega}} = \exp(\eta_{i\omega}) = \mu_{i\omega}^{GPSAR}$ Selanjutnya $\frac{\partial \eta_i}{\partial \beta_0} = 1$

Dengan demikian didapatkan bentuk berikut.

$$\frac{\partial \ell(\omega_{GPSAR})}{\partial \beta_{0\omega}} = \sum_{i=1}^n \left(\frac{y_i - \mu_{i\omega}^{GPSAR}}{\mu_{i\omega}^{GPSAR}(1 + \phi_\omega \mu_{i\omega}^{GPSAR})^2} \right) (\mu_{i\omega}^{GPSAR})(1) = \sum_{i=1}^n \frac{y_i - \mu_{i\omega}^{GPSAR}}{(1 + \phi_\omega \mu_{i\omega}^{GPSAR})^2}$$

3. Turunan parsial terhadap ρ_ω .

$$\frac{\partial \eta_{i\omega}}{\partial \rho_\omega} = \sum_{j=1}^n w_{ij} y_j$$

Dengan demikian didapatkan bentuk berikut.

$$\begin{aligned} \frac{\partial \ell(\omega_{GPSAR})}{\partial \rho_\omega} &= \sum_{i=1}^n \left[\left(\frac{y_i - \mu_{i\omega}^{GPSAR}}{\mu_{i\omega}^{GPSAR} (1 + \phi_\omega \mu_{i\omega}^{GPSAR})^2} \right) (\mu_{i\omega}^{GPSAR}) \left(\sum_{j=1}^n w_{ij} y_j \right) \right] \\ &= \sum_{i=1}^n \left[\left(\frac{y_i - \mu_{i\omega}^{GPSAR}}{(1 + \phi_\omega \mu_{i\omega}^{GPSAR})^2} \right) \left(\sum_{j=1}^n w_{ij} y_j \right) \right] \end{aligned}$$

4. Turunan parsial terhadap ϕ_ω .

Turunan $\partial \ell(\omega_{GPSAR})$ terhadap ϕ_ω dilakukan secara langsung.

$$\begin{aligned} \frac{\partial \ell_{i\omega}}{\partial \phi_\omega} &= \frac{\partial}{\partial \phi_\omega} \left(y_i \ln \left(\frac{\mu_{i\omega}^{GPSAR}}{1 + \phi_\omega \mu_{i\omega}^{GPSAR}} \right) + (y_i - 1) \ln(1 + \phi_\omega y_i) \right. \\ &\quad \left. - \frac{\mu_{i\omega}^{GPSAR} (1 + \phi_\omega y_i)}{1 + \phi_\omega \mu_{i\omega}^{GPSAR}} - \ln(y_i!) \right) \\ &= -\frac{y_i \mu_{i\omega}^{GPSAR}}{1 + \phi_\omega \mu_{i\omega}^{GPSAR}} + \frac{(y_i - 1) y_i}{1 + \phi_\omega y_i} - \frac{\mu_{i\omega}^{GPSAR} (y_i - \mu_{i\omega}^{GPSAR})}{(1 + \phi_\omega \mu_{i\omega}^{GPSAR})^2} \\ &= \frac{(y_i - 1) y_i}{1 + \phi_\omega y_i} - \frac{y_i \mu_{i\omega}^{GPSAR}}{1 + \phi_\omega \mu_{i\omega}^{GPSAR}} - \frac{\mu_{i\omega}^{GPSAR} (y_i - \mu_{i\omega}^{GPSAR})}{(1 + \phi_\omega \mu_{i\omega}^{GPSAR})^2} \end{aligned}$$

Sehingga didapatkan bentuk

$$\frac{\partial \ell(\omega_{GPSAR})}{\partial \phi_\omega} = \sum_{i=1}^n \frac{y_i (y_i - 1)}{1 + \phi_\omega y_i} - \frac{y_i \mu_{i\omega}^{GPSAR}}{1 + \phi_\omega \mu_{i\omega}^{GPSAR}} - \frac{\mu_{i\omega}^{GPSAR} (y_i - \mu_{i\omega}^{GPSAR})}{(1 + \phi_\omega \mu_{i\omega}^{GPSAR})^2}$$

Solusi numerik $\hat{\omega} = \{\hat{\beta}_{0\omega}, \hat{\rho}_\omega, \hat{\phi}_\omega\}$ dapat diperoleh dengan menyelesaikan sistem persamaan turunan parsial setiap parameter dengan iterasi BHHH.

Lampiran 4. Kode R

Berikut adalah kode beserta hasilnya.

```
> # --- 0. PREPARASI & LIBRARY ---
> rm(list = ls())
> if(!require(pacman)) install.packages("pacman")
> pacman::p_load(sf, readxl, janitor, dplyr, spdep, maxLik, Matrix, car, VGAM,
+               ggplot2, scales, patchwork, ggcorrplot, classInt, tidyr, ggrepel, writexl,
vcd)
> # SETUP OUTPUT
> output_folder <- "C:/Users/Parsaroan Sihombing/Desktop/"
> cat("\n[SYSTEM] Output folder set to:", output_folder, "\n")

[SYSTEM] Output folder set to: C:/Users/Parsaroan Sihombing/Desktop/
> # --- LOAD DATA & CLEANING ---
> cat("\n[INIT] Memuat Data...\n")

[INIT] Memuat Data...
> # 1. Load Peta
> path_peta <- file.choose() # Pilih .shp
> peta_indonesia <- st_read(path_peta, quiet = TRUE)
> peta_tuban <- peta_indonesia %>%
+   mutate(kunci_join = paste(toupper(ADM3_EN), toupper(ADM4_EN), sep = "_")) %>%
+   dplyr::select(ADM3_EN, ADM4_EN, kunci_join, geometry)
> # 2. Load Excel
> path_excel <- file.choose() # Pilih Excel
> data_statistik <- read_excel(path_excel)
> data_siap <- clean_names(data_statistik[, 1:11])
> # Rename & Cleaning
> colnames(data_siap)[4:9] <- c("asi_eksklusif", "vit_a", "idl",
+                               "kelas_ibuhamil", "air_bersih", "lm_campak")
> colnames(data_siap)[11] <- "jumlah_balita"
> data_siap <- data_siap %>%
+   mutate(
+     lm_campak = as.numeric(lm_campak),
+     desa_koreksi = case_when(
+       toupper(desa) == "BANJARJO" ~ "Banjarejo",
+       toupper(desa) == "BULU MEDURO" ~ "Bulumeduro",
+       toupper(desa) == "KEDUNGMAKAN" ~ "Kedungmakam",
+       toupper(desa) == "LAJU KIDUL" ~ "Lajo Kidul",
+       toupper(desa) == "SUGIWARAS" ~ "Sugihwaras",
+       TRUE ~ desa
+     ),
+     kunci_join = paste(toupper(kecamatan), toupper(desa_koreksi), sep = "_")
+   )
> # 3. Merge & Filter
> data_gabung <- left_join(peta_tuban, data_siap, by = "kunci_join")
> vars_model <- c("pneumonia", "asi_eksklusif", "vit_a", "kelas_ibuhamil",
+               "air_bersih", "lm_campak", "jumlah_balita")
> data_final <- data_gabung %>%
+   filter(!is.na(pneumonia) & jumlah_balita > 0) %>%
+   filter(complete.cases(across(all_of(vars_model)))) %>%
+   mutate(ID_Map = row_number()) %>%
+   mutate(
+     lon = st_coordinates(st_centroid(geometry))[1],
+     lat = st_coordinates(st_centroid(geometry))[2]
+   )
> # Buat Bobot Spasial (Digunakan di Tahap 7 & 8)
> nb_queen <- poly2nb(data_final, queen = TRUE)
> listw_queen <- nb2listw(nb_queen, style = "W", zero.policy = TRUE)
> W_matrix <- listw2mat(listw_queen)
> data_tabular <- st_drop_geometry(data_final)
> cat(">> Data Siap. Jumlah Observasi:", nrow(data_final), "\n")
>> Data Siap. Jumlah Observasi: 328
> # =====
> # 1. UJI DISTRIBUSI POISSON
> # =====
> cat("\n[1] Uji Distribusi Poisson (Chi-Square Goodness of Fit)...\n")
```

```

[1] Uji Distribusi Poisson (Chi-Square Goodness of Fit)...
> # Menggunakan library vcd
> gf_pois <- goodfit(data_final$pneumonia, type = "poisson", method = "ML")
> summary_gf <- summary(gf_pois)

      Goodness-of-fit test for poisson distribution

              X^2 df      P(> X^2)
Likelihood Ratio 1294.265 20 5.003258e-262
> cat("--- Hasil Uji Goodness of Fit ---\n")
--- Hasil Uji Goodness of Fit ---
> print(summary_gf)

              X^2 df      P(> X^2)
Likelihood Ratio 1294.265 20 5.003258e-262
> p_val_dist <- summary_gf[1,3] # Ambil P-value dari matriks
> if(p_val_dist < 0.05) {
+   cat(">> KESIMPULAN: Tolak H0 (Data TIDAK berdistribusi Poisson). Lanjut ke GPR.\n")
+ } else {
+   cat(">> KESIMPULAN: Gagal Tolak H0 (Data berdistribusi Poisson).\n")
+ }
>> KESIMPULAN: Tolak H0 (Data TIDAK berdistribusi Poisson). Lanjut ke GPR.
> # =====
> # 2. STATISTIK DESKRIPTIF
> # =====
> cat("\n[2] Statistik Deskriptif...\n")

[2] Statistik Deskriptif...
> tabel_deskriptif <- data_tabular %>%
+   dplyr::select(all_of(vars_model[1:6])) %>%
+   pivot_longer(cols = everything(), names_to = "Variabel", values_to = "Nilai") %>%
+   group_by(Variabel) %>%
+   summarise(
+     Mean = mean(Nilai), SD = sd(Nilai), Min = min(Nilai), Max = max(Nilai)
+   ) %>% mutate(across(where(is.numeric), ~round(., 2)))
> print(as.data.frame(tabel_deskriptif))
  Variabel  Mean   SD   Min   Max
1   air_bersih 97.17  6.09 61.11 100.00
2  asi_eksklusif  6.54  7.23  0.00  64.10
3 kelas_ibu_hamil 63.83 57.69  0.00 500.00
4   lm_campak  21.86  8.61  6.56  89.74
5   pneumonia   4.53  7.78  0.00 115.00
6      vit_a 110.69 71.25  0.00 813.04
> write.csv(tabel_deskriptif, paste0(output_folder, "1_Tabel_Deskriptif.csv"), row.names = FALSE)
> # =====
> # 3. UJI KORELASI
> # =====
> cat("\n[3] Uji Korelasi Pearson...\n")

[3] Uji Korelasi Pearson...
> data_num <- data_tabular %>% dplyr::select(all_of(vars_model[1:6]))
> # Hitung Matriks Angka
> matriks_r <- cor(data_num, method = "pearson")
> matriks_p <- cor_pmat(data_num)
> # Tampilkan Tabel di Console
> cat("--- Matriks Korelasi & Signifikansi ---\n")
--- Matriks Korelasi & Signifikansi ---
> format_korelasi <- function(r, p) {
+   r_txt <- sprintf("%.3f", r)
+   bintang <- case_when(p < 0.001 ~ "****", p < 0.01 ~ "***", p < 0.05 ~ "**", TRUE ~ "")
+   paste0(r_txt, bintang)
+ }
> matriks_tampil <- matrix(mapply(format_korelasi, matriks_r, matriks_p),
+   nrow=nrow(matriks_r), ncol=ncol(matriks_r),
+   dimnames=dimnames(matriks_r))
> matriks_tampil[upper.tri(matriks_tampil)] <- ""
> print(as.data.frame(matriks_tampil))
      pneumonia asi_eksklusif vit_a kelas_ibu_hamil air_bersih lm_campak
pneumonia      1.000***

```

```

asi_eksklusif      -0.044      1.000***
vit_a              -0.015        0.029 1.000***
kelas_ibuhamil    0.123*        0.112*  0.138*      1.000***
air_bersih        -0.063        -0.124*  0.112*      -0.039  1.000***
lm_campak         0.033         0.406*** -0.065        -0.070  -0.016  1.000***
> write.csv(matriks_tampil, paste0(output_folder, "2_Tabel_Korelasi.csv"))
> # =====
> # 4. UJI MULTIKOLINEARITAS
> # =====
> cat("\n[4] Uji Multikolinearitas (VIF)...\n")

[4] Uji Multikolinearitas (VIF)...
> # Model OLS Dummy hanya untuk VIF
> model_ols_dummy <- lm(pneumonia ~ asi_eksklusif + vit_a + kelas_ibuhamil + air_bersih +
lm_campak,
+                      data = data_num)
> nilai_vif <- vif(model_ols_dummy)
> tabel_vif <- data.frame(Variabel = names(nilai_vif), VIF = as.numeric(nilai_vif))
> print(tabel_vif)
  Variabel      VIF
1  asi_eksklusif 1.249157
2      vit_a    1.040354
3 kelas_ibuhamil 1.047661
4   air_bersih  1.033258
5    lm_campak  1.224585
> if(any(tabel_vif$VIF > 10)) cat("[WARNING] Ada VIF > 10!\n") else cat("[OK] Tidak ada
Multikolinearitas.\n")
[OK] Tidak ada Multikolinearitas.
> write.csv(tabel_vif, paste0(output_folder, "3_Tabel_VIF.csv"), row.names = FALSE)
> # =====
> # 5. UJI EQUIDISPERSI
> # =====
> # Ref: Hilbe (2014), Persamaan 2.12
> cat("\n[5] Uji Equidispersi (Metode Deviance Manual)...\n")

[5] Uji Equidispersi (Metode Deviance Manual)...
> # A. Fit Model Poisson (untuk mendapatkan mu_hat)
> model_pois <- glm(pneumonia ~ asi_eksklusif + vit_a + kelas_ibuhamil + air_bersih +
lm_campak,
+                 family = poisson(), data = data_tabular, offset =
log(data_tabular$jumlah_balita))
> y_i <- data_tabular$pneumonia
> mu_hat <- fitted(model_pois)
> n <- nrow(data_tabular)
> p <- 5 # Jumlah prediktor (asi, vit, ibu, air, campak)
> # B. Hitung Deviance (D) sesuai Rumus (2.12)
> #   D = 2 * Sigma [ y * ln(y/mu) - (y - mu) ]
> #   Note: Jika y=0, limit y*ln(y) = 0.
> term_1 <- ifelse(y_i == 0, 0, y_i * log(y_i / mu_hat))
> term_2 <- (y_i - mu_hat)
> D_hitung <- 2 * sum(term_1 - term_2)
> # C. Derajat Bebas (df) = n - (p + 1)
> df_model <- n - (p + 1)
> # D. Nilai Kritis Chi-Square (Tabel)
> alpha <- 0.05
> chi_tabel <- qchisq(1 - alpha, df = df_model)
> # E. Rasio Dispersi (Phi)
> disp_ratio <- D_hitung / df_model
> cat("--- Hasil Uji Deviance (Hilbe, 2014) ---\n")
--- Hasil Uji Deviance (Hilbe, 2014) ---
> cat(">> D Hitung (Manual)      :", round(D_hitung, 4), "\n")
>> D Hitung (Manual)      : 1601.227
> cat(">> Derajat Bebas (df)    :", df_model, "\n")
>> Derajat Bebas (df)    : 322
> cat(">> Chi-Square Tabel      :", round(chi_tabel, 4), "\n")
>> Chi-Square Tabel      : 364.8474
> cat(">> Rasio (D/df)          :", round(disp_ratio, 4), "\n")
>> Rasio (D/df)          : 4.9728
> # F. Keputusan Hipotesis
> if(D_hitung > chi_tabel) {

```

```

+   cat("\n[KESIMPULAN] TOLAK H0.\n")
+   if(dispratio > 1) {
+     cat(">> Terjadi OVERDISPERSI (Rasio > 1).\n")
+     cat(">> REKOMENDASI: Wajib menggunakan GPR/GPSAR.\n")
+   } else {
+     cat(">> Terjadi UNDERDISPERSI (Rasio < 1).\n")
+   }
+ } else {
+   cat("\n[KESIMPULAN] GAGAL TOLAK H0 (Equidispersion Terpenuhi).\n")
+ }

[KESIMPULAN] TOLAK H0.
>> Terjadi OVERDISPERSI (Rasio > 1).
>> REKOMENDASI: Wajib menggunakan GPR/GPSAR.
> # =====
> # 6. PEMODELAN GPR (Generalized Poisson Regression)
> # =====
> cat("\n[6] Estimasi Model GPR...\n")

[6] Estimasi Model GPR...
> model_gpr <- vglm(pneumonia ~ asi_eksklusif + vit_a + kelas_ibu_hamil + air_bersih +
+   lm_campak,
+   family = genpoisson2, data = data_tabular, offset =
+   log(data_tabular$jumlah_balita))
> coef_gpr <- coef(summary(model_gpr))
> print(coef_gpr)
              Estimate Std. Error      z value      Pr(>|z|)
(Intercept):1 -0.663692702 0.9042569377 -0.7339647 0.4629702321
(Intercept):2 -6.135168006 0.0899430395 -68.2117042 0.0000000000
asi_eksklusif  0.021063036 0.0083664722  2.5175529 0.0118173231
vit_a          -0.001374385 0.0009374845 -1.4660355 0.1426386180
kelas_ibu_hamil 0.001146330 0.0011731891  0.9771058 0.3285167921
air_bersih     -0.030273936 0.0091599445 -3.3050348 0.0009496459
lm_campak      0.007201301 0.0069906716  1.0301301 0.3029489230
> write.csv(coef_gpr, paste0(output_folder, "4_Hasil_GPR.csv"))
> # =====
> # 7. UJI DEPENDENSI SPASIAL (Pada Residual GPR)
> # =====
> cat("\n[7] Uji Dependensi Spasial (Moran's I pada Residual GPR)...\n")

[7] Uji Dependensi Spasial (Moran's I pada Residual GPR)...
> # Hitung Residual Deviance Manual untuk GPR
> y_obs <- data_final$pneumonia
> mu_fit <- fitted(model_gpr)
> phi_hat <- exp(coef(model_gpr, matrix = TRUE)[1, 2])
> dev_i <- numeric(length(y_obs))
> idx_pos <- y_obs > 0
> # Rumus Deviance GPR
> dev_i[idx_pos] <- 2 * (y_obs[idx_pos] * log(y_obs[idx_pos]/mu_fit[idx_pos]) -
+   (y_obs[idx_pos]-mu_fit[idx_pos]) +
+   (y_obs[idx_pos]-1) *
+   log((y_obs[idx_pos]*(1+phi_hat))/(mu_fit[idx_pos]+phi_hat*y_obs[idx_pos])))
> # Kasus y=0
> if(any(!idx_pos)) {
+   mu0 <- mu_fit[!idx_pos]
+   dev_i[!idx_pos] <- 2 * (mu0 - log(1 + phi_hat * mu0))
+ }
> dev_i[dev_i < 0] <- 0
> resid_gpr_manual <- sign(y_obs - mu_fit) * sqrt(dev_i)
> # Uji Moran
> moran_res <- moran.test(resid_gpr_manual, listw_queen, randomisation=TRUE,
+   alternative="two.sided")
> print(moran_res)

Moran I test under randomisation

data: resid_gpr_manual
weights: listw_queen

Moran I statistic standard deviate = 8.2911, p-value < 2.2e-16

```

```

alternative hypothesis: two.sided
sample estimates:
Moran I statistic      Expectation      Variance
      0.278136667      -0.003058104      0.001150244

> # =====
> # 8. PEMODELAN GPSAR (Estimasi BHHH)
> # =====
> cat("\n[8] Estimasi Model GPSAR...\n")

[8] Estimasi Model GPSAR...
> # Persiapan Matriks
> X_GPSAR <- model.matrix(~ asi_eksklusif + vit_a + kelas_ibu_hamil + air_bersih +
lm_campak, data = data_final)
> y <- data_final$pneumonia
> offset_vec <- data_final$jumlah_balita
> # Fungsi LogLik GPSAR
> logL_gpsar <- function(theta, y, X, W, offset_vec) {
+   n <- length(y); p <- ncol(X)
+   beta <- theta[1:p]; rho <- theta[p+1]; phi <- theta[p+2]
+   if(is.na(rho) || abs(rho) >= 0.99) return(rep(-1e10, n))
+
+   A <- diag(n) - rho * W
+   eta <- tryCatch(as.vector(solve(A) %*% X %*% beta), error=function(e) NULL)
+   if(is.null(eta)) return(rep(-1e10, n))
+
+   mu <- offset_vec * exp(eta)
+   term1 <- y * log(mu/(1+phi*mu)); term2 <- (y-1)*log(1+phi*y)
+   term3 <- -lgamma(y+1); term4 <- -(mu*(1+phi*y))/(1+phi*mu)
+   term2[y==0] <- 0
+   res <- term1 + term2 + term3 + term4
+   res[is.na(res)] <- -1e10
+   return(res)
+ }
> # Fungsi Gradien
> grad_gpsar <- function(theta, y, X, W, offset_vec) {
+   n <- length(y); p <- ncol(X)
+   beta <- theta[1:p]; rho <- theta[p+1]; phi <- theta[p+2]
+   A <- diag(n) - rho * W; A_inv <- tryCatch(solve(A), error=function(e) diag(n))
+   eta <- A_inv %*% X %*% beta; mu <- offset_vec * exp(eta)
+   common <- (y - mu) / (1 + phi * mu)^2
+
+   grad_beta <- as.vector(common) * (A_inv %*% X)
+   grad_rho <- as.vector(common) * as.vector(A_inv %*% W %*% eta)
+
+   t1 <- -(y*mu)/(1+phi*mu); t2 <- (y*(y-1))/(1+phi*y); t3 <- -(mu*(y-mu))/(1+phi*mu)^2
+   t2[y==0] <- 0; grad_phi <- t1 + t2 + t3
+   return(cbind(grad_beta, grad_rho, grad_phi))
+ }
> # Eksekusi
> start_val <- c(coef(model_pois), 0.3, 0.1)
> model_gpsar <- maxLik(logLik = logL_gpsar, grad = grad_gpsar, method = "BHHH",
+   start = start_val, y=y, X=X_GPSAR, W=W_matrix,
offset_vec=offset_vec,
+   control=list(iterlim=5000, tol=1e-6, printLevel=0))
> summ_gpsar <- summary(model_gpsar)
> print(summ_gpsar)
-----
Maximum Likelihood estimation
BHHH maximisation, 38 iterations
Return code 8: successive function values within relative tolerance limit (reltol)
Log-Likelihood: -828.4704
8 free parameters
Estimates:

```

	Estimate	Std. error	t value	Pr(> t)
(Intercept)	0.6094878	0.1663890	3.663	0.000249 ***
asi_eksklusif	-0.0024128	0.0023075	-1.046	0.295737
vit_a	0.0000611	0.0001294	0.472	0.636804
kelas_ibu_hamil	0.0014320	0.0004976	2.878	0.004004 **
air_bersih	-0.0091008	0.0027963	-3.255	0.001136 **

```

lm_campak      0.0071706  0.0028505   2.516 0.011885 *
               0.9899999  0.0189537  52.232 < 2e-16 ***
               0.2990521  0.0239279  12.498 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
-----
> est_gpsar <- coef(summ_gpsar)
> z_val <- est_gpsar[,1]/est_gpsar[,2]
> p_val <- 2*pnorm(abs(z_val), lower.tail=FALSE)
> tab_gpsar <- cbind(est_gpsar[,1:2], z_value=z_val, p_value=p_val)
> write.csv(tab_gpsar, paste0(output_folder, "5_Hasil_GPSAR.csv"))
> # =====
> # 9. KEBAIKAN MODEL (Goodness of Fit)
> # =====
> cat("\n[9] Perbandingan Kebaikan Model (GPR vs GPSAR)...\n")

[9] Perbandingan Kebaikan Model (GPR vs GPSAR)...
> calc_metrik <- function(logLik, n_par, n, y, pred) {
+   aic <- 2*n_par - 2*logLik
+   aicc <- aic + (2*n_par*(n_par+1))/(n-n_par-1)
+   rmse <- sqrt(mean((y-pred)^2))
+   return(c(AICc=aicc, RMSE=rmse))
+ }
> # Metrik GPR
> m_gpr <- calc_metrik(logLik(model_gpr), length(coef(model_gpr)), length(y), y,
fitted(model_gpr))
> # Metrik GPSAR
> if(model_gpsar$code %in% c(0,1,2,8)) {
+   cf <- coef(model_gpsar); rho_h <- cf[ncol(X_GPSAR)+1]
+   eta_h <- solve(diag(length(y)) - rho_h * W_matrix) %*% X_GPSAR %*% cf[1:ncol(X_GPSAR)]
+   mu_gpsar <- offset_vec * exp(eta_h)
+   m_gpsar <- calc_metrik(logLik(model_gpsar), length(cf), length(y), y, mu_gpsar)
+ } else { m_gpsar <- c(NA, NA) }
> tab_compare <- data.frame(Indikator=c("AICc", "RMSE"), GPR=m_gpr, GPSAR=m_gpsar)
> print(tab_compare)
      Indikator      GPR      GPSAR
AICc      AICc 1717.243627 1673.392207
RMSE      RMSE   7.117935   6.406422

```


BIODATA PENULIS



Penulis bernama Joshua Capri Gunawan Sihombing, lahir pada 2 Maret 2000 di Balikpapan. Penulis merupakan anak pertama dari empat bersaudara dari pasangan Parsaoran Sihombing dan Rosalin Valentina Pasaribu. Penulis memulai jenjang pendidikan formal tingkat SMA di SMA Negeri 3 Unggulan Tenggarong dan lulus pada tahun 2018. Penulis kemudian melanjutkan pendidikan pada jenjang sarjana (S1) di Program Studi Pendidikan Matematika, Universitas Mulawarman dan menyelesaikannya pada tahun 2023. Selama menempuh pendidikan hingga saat ini, penulis aktif mendedikasikan diri di dunia pendidikan sebagai pengajar matematika dengan pengalaman selama enam tahun. Penulis melanjutkan pendidikan Magister (S2) di Departemen Statistika, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember (ITS) mulai tahun 2024. Selama masa studi magister, penulis memfokuskan minat penelitiannya pada bidang pemodelan statistik spasial, khususnya *Generalized Poisson Spatial Autoregressive*. Salah satu karya tulis penulis terkait topik ini juga telah dipublikasikan pada jurnal ilmiah CAUCHY. Pembaca yang ingin memberikan kritik, saran, dan pertanyaan mengenai penelitian ini dapat menghubungi penulis melalui email: joe.cgsihombing12@gmail.com.