PENERAPAN *INVERSE GAUSSIAN HYBRID ESTIMATOR* (IGH) DALAM MENGATASI MULTIKOLINEARITAS PADA JUMLAH KASUS *TUBERCULOSIS* (TBC)

(Skripsi)

OLEH

GRACIA TRIFENA SINTAULI

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026

ABSTRACT

PENERAPAN *INVERSE GAUSSIAN HYBRID ESTIMATOR* (IGH) DALAM MENGATASI MULTIKOLINEARITAS PADA JUMLAH KASUS *TUBERCULOSIS* (TBC)

BY

GRACIA TRIFENA SINTAULI

The Inverse Gaussian Regression (IGR) model is one approach within the Generalized Linear Model (GLM) framework for modeling data with a positively skewed distribution. Parameter estimation is typically performed using the Inverse Gaussian Maximum Likelihood (IGML) method. However, under conditions of high multicollinearity, IGML becomes unstable due to increased coefficient variance, which leads to a higher MSE. This study compares IGML with the Inverse Gaussian Hybrid Estimator (IGH) in addressing multicollinearity in Tuberculosis cases across 28 districts/cities in West Java Province from 2022-2024. The analysis results indicate the presence of multicollinearity, characterized by high correlation values and large VIF values. The IGH method produces coefficient shrinkage, making the model more stable and superior to IGML.

Keywords: Inverse Gaussian Regression, IGML, IGH, Multicolinearitas, Tuberculosis, Mean Square Error.

ABSTRAK

PENERAPAN *INVERSE GAUSSIAN HYBRID ESTIMATOR* (IGH) DALAM MENGATASI MULTIKOLINEARITAS PADA JUMLAH KASUS *TUBERCULOSIS* (TBC)

Oleh

GRACIA TRIFENA SINTAULI

Inverse Gaussian Regression (IGR) merupakan salah satu pendekatan dalam *Generalized Linear Model* (GLM) yang digunakan untuk memodelkan data dengan distribusi menceng ke kanan (*positively skewed*). Estimasi parameter pada model IGR dilakukan menggunakan metode *Inverse Gaussian Maximum Likelihood* (IGML). Ketika terjadi multikolinearitas yang tinggi, estimator IGML menjadi tidak stabil karena variansi koefisien meningkat sehingga nilai MSE besar. Penelitian ini bertujuan untuk membandingkan kinerja metode IGML dengan *Inverse Gaussian Hybrid Estimator* (IGH) dalam mengatasi permasalahan multikolinearitas pada pemodelan Jumlah kasus *Tuberculosis* (TBC) di 28 kabupaten/kota Provinsi Jawa Barat tahun 2022–2024. Hasil analisis menunjukkan adanya multikolinearitas yang ditandai dengan nilai korelasi yang tinggi serta nilai VIF yang besar. Estimasi menggunakan metode IGH menghasilkan penyusutan koefisien (*shrinkage*) sehingga model menjadi lebih stabil. Dengan demikian, metode IGH merupakan estimator yang lebih baik dibandingkan IGML.

Kata Kunci: *Inverse Gaussian Regression*, IGML, IGH, Multikolinearitas, *Tuberculosis*, *Mean Square Error*.

PENERAPAN *INVERSE GAUSSIAN HYBRID ESTIMATOR* (IGH) DALAM MENGATASI MULTIKOLINEARITAS PADA JUMLAH KASUS *TUBERCULOSIS* (TBC)

Oleh

GRACIA TRIFENA SINTAULI

Skripsi

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026

Judul	: **PENERAPAN *INVERSE GAUSSIAN HYBRID ESTIMATOR* (IGH) DALAM MENGATASI MULTIKOLINEARITAS PADA JUMLAH KASUS *TUBERCULOSIS* (TBC)**
Nama Mahasiswa	: Gracia Trifena Sintauli
NPM	: **2217031172**
Program Studi	: **Matematika**
Fakultas	: **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Prof. Ir. Netti Herawati, M. Sc., Ph. D.
NIP. 196501251990032001

Misgiyati, S. Pd., M. Si.
NIP. 198509282023212032

2. Wakil Dekan Bidang Akademik dan Kerjasama,
FMIPA Universitas Lampung

Mulyono, S.Si., M.Si., Ph.D.
NIP. 197406112000031002

MENGESAHKAN

1. Tim Penguji

Ketua	: **Prof. Ir. Netti Herawati, M. Sc., Ph. D.**	....
Sekretaris	: **Misgiyati, S. Pd., M. Si.**	....
Penguji Bukan Pembimbing	: **Drs. Nusyirwan, M. Si.**	....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP.197110012005011002

Tanggal Lulus Ujian Skripsi: **12 Mei 2026**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama	: **Gracia Trifena Sintauli**
Nomor Pokok Mahasiswa	: **2217031172**
Jurusan	: **Matematika**
Judul Skrpsi	: **PENERAPAN *INVERSE GAUSSIAN HYBRID ESTIMATOR* (IGH) DALAM MENGATASI MULTIKOLINEARITAS PADA JUMLAH KASUS *TUBERCULOSIS* (TBC)**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 12 Mei 2026
Penulis

Gracia Trifena Sintauli
NPM. 2217031172

RIWAYAT HIDUP

Penulis Bernama lengkap Gracia Trifena Sintauli lahir di Jakarta pada tanggal 09 September 2002. Penulis lahir dari pasangan Bapak Bintang Simanjuntak dan Ibu Julita Saragih dan merupakan anak pertama dari dua bersaudara.

Penulis pertama kali menempuh pendidikannya di TK Katolik Nusa Melati pada tahun 2007-2008, kemudian melanjutkan pendidikannya ke Sekolah Dasar di SD Katolik Nusa Melati pada tahun 2008–2014. Selanjutnya, penulis melanjutkan pendidikan menengah pertama di SMP Negeri 160 Jakarta pada tahun 2014–2017, lalu melanjutkan pendidikan menengah atas di SMK Prestasi Prima pada tahun 2017–2020.

Pada tahun 2022, penulis terdaftar sebagai mahasiswi Program Studi S1 Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung, yang berlokasi di Rajabasa, Kota Bandar Lampung melalui jalur SBMPTN. Selama menempuh pendidikannya, penulis pernah menjadi bagian dari Unit Kegiatan Mahasiswa (UKM-U) Sains dan Teknologi. Pada tahun 2025, penulis melaksanakan Kuliah Kerja Nyata (KKN) di Kelurahan Kaliawi, Kecamatan Tanjung Karang Pusat, Bandar Lampung sebagai bentuk pengabdian kepada masyarakat. Selain itu, penulis juga melaksanakan Kerja Praktik di Pemerintah Provinsi (Pemprov) Lampung sebagai sarana pengembangan wawasan dan pengalaman.

KATA INSPIRASI

"*To GOD Be The Glory*"

"Ombakku besar, perahuku kecil. Tapi, Tuhan Yesusku Terlebih besar"

"Dan apa saja yang kamu minta dalam doa dengan penuh kepercayaan, kamu akan menerimanya"

(Matius 21:22)

"Tetapi kamu ini, kuatkanlah hatimu, jangan lemah semangatmu, karena ada upah bagi usahamu"

(2 Tawarikh 15:7)

"Karena masa depan sungguh ada, dan harapanmu tidak akan hilang"

(Amsal 23:18)

"Aku tahu, bahwa Engkau sanggup melakukan segala sesuatu, dan tidak ada rencana-Mu yang gagal"

(Ayub 42:2)

"Jangan takut gagal, karena yang tidak pernah gagal hanyalah orang-orang yang tidak pernah melangkah"

(Buya Hamka)

PERSEMBAHAN

Segala puji dan syukur penulis panjatkan kepada Tuhan Yesus Kristus atas kasih, penyertaan, berkat dan anugerah-Nya sehingga skripsi ini dapat terselesaikan dengan baik. Penulis mempersembahkan karya sederhana ini kepada:

Orang Tua Tercinta

Terima kasih atas semua cinta, doa, kasih saying, perjuangan, dan pengorbanan yang selalu diberikan. Kalian adalah sumber semangat terbesar dalam setiap langkah perjuangan penulis hingga mampu menyelesaikan skripsi ini.

Diri Sendiri

Terima kasih Gracia Trifena Sintauli sudah mampu berusaha, bertahan, berjuang, dan tidak menyerah dalam setiap proses yang dilalui, mari berjuang untuk masa depan yang lebih baik.

Adik Tersayang

Terima kasih atas doa, dukungan, kebersamaan dan selalu menjadi penyemangat dalam setiap perjuangan.

Dosen Pembimbing dan Dosen Pembahas

Terima kasih atas bimbingan, arahan, ilmu dan motivasi yang diberikan selama penyusunan skripsi ini bagi penulis.

Almamater Tercinta, Universitas Lampung

SANWACANA

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena berkat limpahan rahmat serta karunia-Nya kepada penulis sehingga penulis dapat menyelesaikan skripsi yang berjudul "Penerapan *Inverse Gaussian Hybrid Estimator* (IGH) Dalam Mengatasi Multikolinearitas Pada Jumlah Kasus *Tuberculosis* (TBC)".

Dalam penyusunan skripsi ini tentunya tidak terlepas dari dukungan, bimbingan, serta masukan dari berbagai pihak yang sangat membantu. Untuk itu penulis mengucapkan terima kasih kepada:

1. Ibu Prof. Ir. Netti Herawati, M.Sc., Ph.D., selaku dosen pembimbing I yang telah memberikan bimbingan, arahan, dan masukan dalam proses penyusunan skripsi ini.
2. Ibu Misiyati, S.Pd., M.Si., selaku dosen pembimbing II yang telah memberikan bimbingan, saran, dan masukan dalam proses penyusunan skripsi ini.
3. Bapak Drs. Nusirwan, M.Si., selaku dosen penguji yang telah memberikan kritik dan saran, sehingga skripsi ini dapat terselesaikan.
4. Bapak Dr. Subian Saidi, S. Si., M.Si., selaku dosen pembimbing akademik yang senantiasa memotivasi dan membimbing selama menjalani perkuliahan.
5. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Bapak Dr. Eng. Heri Satria, S.Si., M.Si., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

7. Seluruh dosen dan staf Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
8. Kepada orang tua tercinta, Papa dan Mama. Terima kasih atas segala doa, cinta, dukungan, dan pengorbanan yang tak ternilai hingga penulis dapat menyelesaikan pendidikan ini.
9. Saudara kandung penulis, Adik Anggi Grasela Cikita. Terima kasih telah memberikan doa, dukungan, keceriaan dan kebersamaan yang menjadi penyemangat penulis.
10. Sahabat penulis Yohana, Astrid, Ayu, Mega, Winne, Yolan, Agustino, Rafael, Lusi, Elisabeth, Annisa, Enjel, Venny, Fiola, Adela, Fanny, Widi, Andika, Amar, Igun, Ka Ika, Agnes, Puji dan Amel terima kasih atas doa, dukungan, semangat, dan kebersamaan selama ini bersama penulis.
11. Teman – teman Jurusan Matematika FMIPA UNILA angkatan 2022.
12. Semua pihak yang telah membantu dalam penyelesaian skripsi ini yang tidak bisa penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan dan penulisan laporan ini masih jauh dari kesempurnaan, serta memiliki kesalahan didalamnya, akan tetapi penulis berharap dengan adanya skripsi ini dapat memberikan manfaat informasi bagi yang membacanya.

Bandar Lampung, 12 Mei 2026
Penulis

Gracia Trifena Sintauli

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Model *Inverse Gaussian Regression* (IGR) adalah model yang ketika variabel responnya condong positif atau positif *skewed*. *Skewed* adalah suatu derajat atau ukuran dari ketidaksimetrisan suatu distribusi data atau sebagai penyimpangan kesimetrisan dari suatu distribusi. Ketika data berdistribusi secara simetris, maka kemiringan sama dengan nol. Distribusi *Inverse Gaussian* ini pertama kali diperkenalkan oleh Erwin Schrödinger (1915) dalam konteks waktu lintasan partikel dan kemudian dikembangkan dalam kerangka regresi oleh John Nelder dan Robert Wedderburn (1972) melalui konsep *Generalized Linear Model* (GLM). Dalam GLM, distribusi *Inverse Gaussian* termasuk dalam keluarga eksponensial dimana varians meningkat secara kubik terhadap rata-rata.

Regresi *Inverse Gaussian* merupakan salah satu pendekatan yang sesuai untuk memodelkan data kontinu yang tidak berdistribusi normal dan memiliki tingkat kemencengan tinggi. Oleh karena itu, model ini banyak digunakan dalam berbagai bidang seperti kesehatan, ekonomi, dan teknik, khususnya ketika data menunjukkan variansi yang tidak konstan serta pola distribusi yang tidak simetris.

Estimasi parameter dalam regresi *Inverse Gaussian* umumnya menggunakan metode *Maximum Likelihood* (ML). Estimator ini dikenal sebagai *Inverse Gaussian Maximum Likelihood* (IGML). Estimator ini memiliki sifat konsisten, tidak bias secara asimtotik, dan efisien jika asumsi model terpenuhi (Greene, 2018). Namun, ketika terdapat multikolinearitas tinggi antar variabel independen, estimator ML menjadi tidak stabil, varians koefisien meningkat. Koefisien regresi menjadi tidak stabil dan nilai *Mean Squared Error* (MSE) membesar (Gujarati & Porter, 2009).

Secara matematis, multikolinearitas menyebabkan matriks desain menjadi mendekati singular sehingga invers matriks menjadi tidak stabil dan menghasilkan variansi estimator yang sangat besar (Montgomery, *et al*., 2021). Kondisi ini menunjukkan bahwa meskipun metode IGML optimal secara teoritis, performanya dapat menurun secara signifikan pada kondisi data riil yang memiliki korelasi tinggi antar variabel independen. Dengan demikian, diperlukan pendekatan alternatif yang mampu mengatasi ketidakstabilan tersebut tanpa mengorbankan kemampuan model dalam menjelaskan data.

Inverse Gaussian Hybrid Estimator (IGH), yaitu estimator yang memodifikasi matriks dengan dua parameter *shrinkage* untuk mengendalikan nilai eigen kecil yang menyebabkan instabilitas. Keunggulan IGH umumnya dilakukan melalui perbandingan MSE, penurunan *trace* matriks varians, efek *shrinkage* pada norma koefisien. Pendekatan berbasis *shrinkage* seperti ini telah terbukti mampu mengurangi variansi estimator dan meningkatkan kestabilan model pada kondisi multikolinearitas tinggi (Hoerl & Kennard, 1970).

Salah satu fenomena yang memenuhi karakteristik tersebut adalah jumlah kasus *Tuberculosis* (TBC). Menurut *World Health Organization* (WHO), TBC masih menjadi salah satu penyakit menular dengan beban tinggi di berbagai negara berkembang. Jumlah kasus TBC umumnya dipengaruhi oleh faktor demografis dan sosial ekonomi seperti kepadatan penduduk, jumlah penduduk, luas wilayah, ketersediaan fasilitas kesehatan, tingkat kemiskinan, dan laju pertumbuhan penduduk. Variabel-variabel tersebut saling berkorelasi dan berpotensi menimbulkan multikolinearitas, misalnya hubungan antara jumlah penduduk dan fasilitas kesehatan atau antara kepadatan penduduk dan luas wilayah. Selain itu, distribusi jumlah kasus TBC cenderung tidak simetris dan meningkat tajam pada wilayah tertentu sehingga sesuai dimodelkan menggunakan distribusi *Inverse*

Gaussian. Penelitian ini menerapkan *Inverse Gaussian Regression* (IGR) dengan estimasi IGML serta membandingkannya dengan *Inverse Gaussian Hybrid Estimator* (IGH) dalam pemodelan jumlah kasus TBC. Penelitian ini bertujuan untuk mengevaluasi apakah IGH mampu menghasilkan estimasi yang lebih stabil dan memiliki MSE lebih kecil dibandingkan IGML pada kondisi multikolinearitas tinggi dalam data riil kasus TBC.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Membandingkan kinerja *Inverse Gaussian Maximum Likelihood* (IGML) dan *Inverse Gaussian Hybrid Estimator* (IGH) berdasarkan nilai *Mean Squared Error* (MSE), stabilitas estimator, serta efek *shrinkage*.
2. Mengevaluasi faktor-faktor yang memengaruhi kasus *Tuberculosis* (TBC) di Jawa Barat.

1.3 Manfaat Penelitian

Adapun manfaat yang dapat diambil dari penelitian ini adalah untuk memperluas wawasan bagi penulis dan bagi pembaca mengenai penerapan *Inverse Gaussian Maximum Likelihood* (IGML) dan *Inverse Gaussian Hybrid* (IGH) dalam mengatasi permasalahan multikolinearitas pada jumlah kasus *Tuberculosis* (TBC).

II. TINJAUAN PUSTAKA

2.1 *Generalized Linear Model* (GLM)

Generalized Linear Model digunakan untuk menilai dan mengukur hubungan antara peubah respons dengan peubah penjelas. Menurut Jong & Heller (2008), *Generalized Linear Model* merupakan perluasan dari proses pemodelan linear yang mengizinkan penentuan model dari suatu data dengan peubah acak tidak harus menyebar normal, asalkan sebaran tersebut termasuk dalam distribusi keluarga eksponensial, antara lain yaitu *Inverse Gaussian*, *Gamma*, dan *Poisson*. Uji hipotesis yang diterapkan pada *Generalized Linear Model* tidak memerlukan asumsi kenormalan dari peubah respons ataupun kehomogenan ragam. Sehingga GLM tak hanya dapat digunakan pada peubah respons yang berdistribusi normal tetapi juga untuk peubah respons yang berdistribusi lain dan ragam yang tidak konstan atau homogen. Model umum *Generalized Linear Model* sebagai berikut:

$$Y_i \sim \text{Distribusi Keluarga Eksponensial}, E(Y_i) = \mu_i, g(\mu_i) = \eta_i = \boldsymbol{x}_i^T \boldsymbol{\beta} \qquad (2.1)$$

dan Y_i termasuk dalam distribusi keluarga eksponensial sebagai berikut:

$$Y_i = X_i\beta + \varepsilon \qquad (2.2)$$

dengan:

Y_i : peubah respons,

X_i : peubah penjelas,

β : parameter,

ε : galat.

Misalkan terdapat vektor pengamatan y yang mempunyai sejumlah n pengamatan yaitu $y_1, y_2, \ldots, y_n$ yang menyebar normal dengan nilai harapan μ. Kemudian dalam kasus model linear klasik dapat dirumuskan (Mc Cullagh & Nelder, 1989).

$$\mu = \sum_{j=1}^{k} x_j \beta_j \tag{2.3}$$

Jika nilai harapan μ dinyatakan dalam bentuk matriks maka:

$$\mu = \boldsymbol{X}\beta \tag{2.4}$$

dengan:

μ : vektor nilai harapan berukuran $n \times 1$, dengan $\mu_i = E(Y \mid X = x_i), x_i = (1, x_{i1}, \ldots, x_{ik})$,

X : matriks berukuran $n \times k$,

β : vektor parameter berukuran $k \times 1$.

Misalkan Yadalah peubah respons, maka bentuk GLM sebagai berikut:

$$f(y) = c(y, \phi)\exp\left(\frac{y\theta - b(\theta)}{a(\phi)}\right), g(\mu_i) = \boldsymbol{X}_i^T\boldsymbol{\beta} \tag{2.5}$$

Pada persamaan $f(y)$ merupakan fungsi kepadatan peluang peubah respons pada distribusi keluarga eksponensial. Sedangkan $g(\mu_i)$ adalah fungsi hubung yang menggambarkan hubungan antara penduga linear terhadap nilai harapan μ_i.

2.2 Distribusi *Inverse Gaussian*

Distribusi *Inverse Gaussian* merupakan distribusi kontinu dengan fungsi kepadatan sama dengan distribusi gamma tetapi dengan kemencengan lebih besar dan keruncingan yang tajam (Jong & Heller, 2008). Misalkan y merupakan suatu peubah acak yang mengikuti distribusi *Inverse Gaussian* dengan fungsi kepadatan peluang sebagai berikut:

$$f(y_i) = \frac{1}{\sqrt{2\pi\phi y_i^3}} \exp\left(-\frac{(y_i-\mu_i)^2}{2\phi\mu_i^2 y_i}\right), y_i > 0 \qquad (2.6).$$

Sehingga fungsi kepadatan peluangnya dapat ditulis dalam bentuk keluarga eksponensial sebagai:

$$f(y) = c(y, \phi)\exp\left(\frac{y\theta - a(\theta)}{\phi}\right) \qquad (2.7).$$

Dengan rata-rata dan variansi dari distribusi *Inverse Gaussian* adalah sebagai berikut:

$$E(y) = \mu, Var(Y) = \phi\mu^3 \qquad (2.8)$$

dan ϕ adalah parameter dispersi.

Uji kesesuaian distribusi dapat dilakukan dengan menggunakan *Uji Kolmogorov–Smirnov*. Dengan hipotesis sebagai berikut:

a. Hipotesis:

   $H_0 : F(x) = F_0(x)$ (Data distribusi *Inverse Gaussian*).

   $H_1 : F(x) \neq F_0(x)$ (Data tidak distribusi *Inverse Gaussian*).

b. Taraf Signifikansi : $\alpha = 5\% = 0{,}05$

c. Statistik Uji

$$D = \sup_x \mid F_n(x) - F(x) \mid \qquad (2.9).$$

   Keterangan:

   $F_n(x)$: fungsi distribusi empiris.

   $F(x)$: fungsi distribusi *Inverse Gaussian*.

d. Daerah Kritis:

   Tolak H_0 jika $D_{hit} > D_{(\alpha,n)}$ atau p-value $< \alpha$.

   Tidak tolak H_0 jika $D_{hit} < D_{(\alpha,n)}$ atau p-value $> \alpha$.

e. Kesimpulan.

2.3 Estimasi *Maximum Likelihood pada Model Inverse Gaussian* (IGML)

Maximum Likelihood merupakan metode estimasi parameter yang diperkenalkan oleh Fisher (1922) dan menjadi dasar utama dalam pemodelan statistik modern, seperti *Generalized Linear Model* (GLM). Prinsip dasar MLE adalah menentukan nilai parameter yang memaksimumkan peluang (*likelihood*) dari sampel yang diamati. Jika $Y_1, Y_2, \dots, Y_n$ merupakan sampel acak dari distribusi dengan fungsi kepadatan peluang $f(y;\theta)$. Maka fungsi likelihood didefinisikan sebagai berikut:

$$L(\theta) = \prod_{i=1}^{n} f(y_i;\theta) \quad (2.10).$$

Untuk mempermudah proses optimasi, digunakan fungsi *log-likelihood* yang diperoleh:

$$\ell(\theta) = \sum_{i=1}^{n} \ln\ f(y_i;\theta) \quad (2.11).$$

Estimator *Maximum Likelihood* diperoleh dengan mencari nilai parameter yang memaksimumkan fungsi log-likelihood $\frac{\partial \ell(\theta)}{\partial \theta} = 0$. Distribusi *Inverse Gaussian* dengan parameter rata-rata $\mu_i$dan parameter dispersi ϕ. Fungsi densitas peluang distribusi *Inverse Gaussian* seperti pada persamaan (2.6) maka fungsi *log-likelihood* untuk n pengamatan dapat dituliskan sebagai:

$$\ell(\beta,\phi) = -\frac{1}{2}\sum_{i=1}^{n}\left[\ln(2\pi\phi y_i^3) + \frac{(y_i-\mu_i)^2}{\phi\mu_i^2 y_i}\right] \quad (2.12)$$

dengan $\mu_i = g^{-1}(\boldsymbol{X}_i^T\boldsymbol{\beta})$. Karena bentuk turunan terhadap β tidak memiliki solusi eksplisit tertutup, maka estimasi dilakukan secara numerik. Dalam GLM, MLE diselesaikan menggunakan algoritma *Fisher Scoring* atau IRLS dengan rumus sebagai berikut:

$$\beta^{(r+1)} = \beta^{(r)} + \boldsymbol{I}^{-1}(\beta^{(r)})\mathbf{S}(\beta^{(r)}) \quad (2.13).$$

Dalam bentuk matriks, estimator IGML dapat dituliskan sebagai berikut:

$$\hat{\beta}_{IGML} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z} \tag{2.14}$$

dengan:

X : matriks desain berukuran $n \times p$,

X^T : transpose dari matriks X,

U : matriks bobot berukuran $n \times n$,

z : variabel respon.

Pada distribusi *Inverse Gaussian*, fungsi varians diberikan oleh $Var(Y_i) = \phi\mu_i^3$. Sehingga matriks bobot menjadi:

$$\boldsymbol{W}_i = \frac{1}{Var(Y_i)}\left(\frac{d\mu_i}{d\eta_i}\right)^2 = \frac{1}{\phi\mu_i^3}\left(\frac{d\mu_i}{d\eta_i}\right)^2 \qquad (2.15).$$

Apabila digunakan fungsi link kanonik $g(\mu) = 1/\mu^2$, dengan menggunakan fungsi link kanonik, turunan $\frac{d\mu_i}{d\eta_i}$ maka bentuk bobot menjadi $W_i = \mu_i^3$. Parameter dispersi ϕ ditaksir menggunakan pendekatan metode momen berbasis residual, yaitu:

$$\hat{\phi} = \frac{1}{n-p}\sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^3} \qquad (2.16).$$

Dengan kovarians estimator IGML adalah $Var(\hat{\beta}_{IGML}) = \phi(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$. Persamaan tersebut menunjukkan bahwa kestabilan estimator sangat bergantung pada sifat matriks $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}$. Apabila matriks tersebut mendekati singular akibat multikolinearitas, maka nilai invers menjadi sangat besar dan variansi estimator meningkat secara signifikan (Gujarati & Porter, 2009). Kondisi inilah yang kemudian mendorong pengembangan estimator berbasis shrinkage seperti *ridge* dan *hybrid estimator*.

2.4 Multikolinearitas

Dalam model *Generalized Linear Model* (GLM), termasuk *Inverse Gaussian Regression*, kovarians parameter diberikan oleh $Var(\hat{\beta}) = \phi(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$ Multikolinearitas adalah keadaan dua atau lebih variabel prediktor yang mempunyai korelasi yang kuat sehingga menimbulkan masalah dalam mengestimasi koefisien regresi. Tidak terjadinya multikolinieritas dapat meningkatkan presisi dari estimasi koefisien regresi dan uji statistik menjadi lebih andal. Besaran (*quality*) yang dapat digunakan untuk mendeteksi adanya multikolinearitas adalah faktor inflasi ragam (*Variance Inflation Factor*/VIF). VIF digunakan sebagai kriteria untuk mendeteksi multikolinearitas pada regresi linier yang melibatkan lebih dari dua variabel bebas. Nilai VIF lebih besar dari 10 mengidentifikasi adanya masalah multikolinearitas (Stock & Watson, 2018). Rumus untuk mencari nilai VIF adalah sebagai berikut:

$$VIF = \frac{1}{(1 - R_i^2)} \quad (2.17).$$

R_i^2 adalah koefisien determinasi antara X_j dengan variable bebas lainnya pada persamaan/model dugaan, sedangkan $j = 1,2,\dots,p$. Multikolinearitas juga dapat dideteksi melalui eigenvalue matriks $\boldsymbol{X}^T\boldsymbol{X}$. Jika terdapat eigenvalue sangat kecil (mendekati nol), maka terjadi near multicollinearity. Condition number didefinisikan sebagai $\kappa = \frac{\lambda_{max}}{\lambda_{min}}$.

Dalam model *Inverse Gaussian Regression*, estimator IGML diperoleh melalui:

$$\hat{\beta}_{IGML} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z} \quad (2.18).$$

Menurut Amin, *et al*. (2020) menunjukkan bahwa dalam *Inverse Gaussian Regression*, multikolinearitas menyebabkan peningkatan *Mean Square Error* (MSE) estimator MLE, sehingga diperlukan estimator berbasis shrinkage seperti ridge atau generalized ridge.

2.5 Regresi *Ridge*

Ridge Regression diperkenalkan oleh Hoerl dan Kennard (1970) sebagai metode estimasi untuk mengatasi multikolinearitas pada regresi linear. $\hat{\beta}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$. Jika $\boldsymbol{X}^T\boldsymbol{X}$ singular akibat multikolinearitas, maka variansi estimator menjadi besar. *Ridge regression* memperbaiki masalah ini dengan menambahkan parameter penalti λpada diagonal matriks:

$$\hat{\beta}_{ridge} = (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{\lambda I})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \qquad (2.19).$$

Dengan $\lambda \geq 0$adalah parameter *shrinkage* dan $\boldsymbol{I}$ adalah matriks identitas
Ridge regression dilakukan untuk meminimumkan fungsi objektif:

$$\min_{\beta}\{(\boldsymbol{Y} - \boldsymbol{X\beta})^T(\boldsymbol{Y} - \boldsymbol{X\beta}) + \boldsymbol{\lambda\beta}^T\boldsymbol{\beta}\} \qquad (2.20).$$

Dengan komponen Error kuadrat (*goodness of fit*) dan penalti besar koefisien
Jika $\lambda = 0$ maka kembali ke OLS dan jika λ mendekati ∞ maka koefisien mendekati nol. Ridge menghasilkan estimator bias $E(\hat{\beta}_{ridge}) \neq \beta$

2.6 Distribusi *Inverse Gaussian Hybrid Estimator* (IGH)

Inverse Gaussian Hybrid Estimator (IGH) merupakan pengembangan dari estimator *Maximum Likelihood* pada model regresi *Inverse Gaussian* yang dikombinasikan dengan pendekatan *shrinkage* seperti pada *Ridge Regression*. Konsep dasar estimator hybrid adalah menggabungkan keunggulan estimator tidak bias berbasis likelihood dengan stabilitas estimator berbasis penalti untuk mengatasi multikolinearitas (Hoerl & Kennard, 1970). Pendekatan hybrid memperkenalkan parameter *shrinkage* k dan parameter penyesuaian bias d. Estimator IGH didefinisikan sebagai berikut:

$$\hat{\beta}_{IGH} = (\boldsymbol{X^T W X} + \boldsymbol{kI})^{-1}(\boldsymbol{X^T W X} - \boldsymbol{kI})(\boldsymbol{X^T W X} + \boldsymbol{k}(\boldsymbol{1} + \boldsymbol{d})\boldsymbol{I})^{-1}\boldsymbol{X^T W X}\hat{\beta}_{IGML} \quad (2.21)$$

dengan:

$k > 0$: parameter *shrinkage*,

d : parameter penyesuaian bias,

I : matriks identitas.

Penambahan parameter k bertujuan memperbaiki kondisi numerik matriks sehingga variansi estimator dapat ditekan, sedangkan parameter d digunakan untuk mengontrol tingkat bias agar tidak terlalu besar (Kibria, 2003). Parameter k ditentukan berdasarkan:

$$k = min\left(\frac{\phi}{\beta_j^2}\right) \quad (2.22).$$

Sedangkan parameter $d$ditentukan menggunakan:

$$d = min\left(\frac{\beta_j^2}{\beta_j^2 + \frac{\phi}{q_j}}\right) \quad (2.23)$$

dengan q_j merupakan eigenvalue dari matriks $\boldsymbol{X^T W X}$. IGH menghasilkan estimator yang bersifat bias namun memiliki variansi yang lebih kecil dibandingkan IGML dalam kondisi multikolinearitas tinggi. Dengan demikian, IGH sering menghasilkan *Mean Squared Error* (MSE) yang lebih kecil dibandingkan estimator *Maximum Likelihood* murni.

2.7 Pengujian Parameter

Pengujian parameter dilakukan untuk mengetahui apakah variabel independen berpengaruh signifikan terhadap variabel respon, baik secara simultan maupun parsial. Dalam kerangka *Generalized Linear Model*, pengujian parameter dilakukan menggunakan pendekatan *Likelihood Ratio Test* dan *Wald Test* (McCullagh & Nelder, 1989).

2.7.1. Pengujian Parameter Secara Simultan

Pengujian simultan bertujuan untuk mengetahui apakah seluruh variabel independen secara bersama-sama berpengaruh signifikan terhadap variabel respon.

a. Hipotesis:

   $H_0: \beta_1 = \beta_2 = \cdots = \beta_j = 0$ (tidak terdapat variabel independen yang berpengaruh secara signifikan terhadap model).

   $H_1: \beta_j \neq 0$ (terdapat minimal satu variabel independen yang berpengaruh secara signifikan terhadap model).

b. Taraf Signifikansi : $\alpha = 5\% = 0{,}05$

c. Statistik Uji

$$G = -2\ln(L_0 - L_1) \tag{2.24}$$

   dengan:

   L_0 : model tanpa melibatkan seluruh variabel independen.

   L_1 : model dengan melibatkan seluruh variabel independen.

d. Daerah Kritis:

   Jika $G > \chi^2_{(\alpha \cdot k)}$ atau nilai *P-value* < α, maka tolak H_0.

   Jika $G < \chi^2_{(\alpha \cdot k)}$ atau nilai *P-value* > α, maka tidak terima H_0.

e. Kesimpulan.

2.7.2. Pengujian Parameter Secara Parsial

Pengujian parsial dilakukan untuk mengetahui pengaruh masing-masing variabel independen terhadap variabel respon.

a. Hipotesis:

$H_0: \beta_j = 0$: variabel bebas tidak berpengaruh secara signifikan terhadap variabel terikat.

$H_1: \beta_j \neq 0$: variabel bebas berpengaruh secara signifikan terhadap variabel terikat.

b. Taraf Signifikansi : $\alpha = 5\% = 0{,}05$

c. Statistik Uji

$$Z_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \tag{2.25}$$

dengan:

$SE(\hat{\beta}_j)$: dugaan standar error dari β_j

d. Daerah Kritis

Jika $Z > \chi^2_{(0,05;1)}$ atau nilai *p-value* < α, maka tolak H_0.

Jika $Z < \chi^2_{(0,05;1)}$atau nilai *p-value* > α, maka tidak terima H_0.

e. Kesimpulan.

2.8 Jumlah kasus *Tuberculosis* (TBC)

Tuberculosis (TB) adalah suatu penyakit infeksi menular yang disebabkan oleh *Mycobacterium Tuberculosis* yang merupakan suatu patogen yang berbentuk basil (batang) dan dikenal dengan BTA (Basil Tahan Asam). Sebagian besar bakteri ini menyebar melalui udara dan umumnya menyerang organ paru tepatnya pada parenkim paru yang dapat mengakibatkan *Tuberculosis* (TB) paru. Penyakit ini dapat menimbulkan banyak komplikasi berbahaya yang bahkan dapat berakibat pada kematian apabila tidak diobati atau pencegahannya tidak tuntas.

Penyakit *Tuberculosis* (TB) masih menjadi salah satu masalah kesehatan di dunia walaupun sudah ribuan tahun lamanya sejak pertama kali penyakit ini muncul. *Tuberculosis* merupakan salah satu dari 10 penyakit menular dan menjadi urutan pertama penyebab kematian terbesar di dunia dan berada pada urutan ketiga untuk penyakit penyebab kematian setelah penyakit *kardiovaskuler* dan infeksi saluran pernafasan akut (ISPA) pada semua golongan umur menyebabkan kematian yang lebih besar dibandingkan HIV/AIDS setiap tahunnya. Penyakit ini masih menjadi perhatian dunia karena hingga saat ini belum ada negara yang bebas dari kasus *Tuberculosis*. Angka kematian akibat penyakit ini masih tinggi bahkan di negara berkembang karena merupakan penyakit yang menular dengan sangat cepat.

Menurut *Global Programme on Tuberculosis and Lung Health* 2024 menyebutkan bahwa Indonesia berada di posisi kedua dengan jumlah penderita TB paru terbanyak di dunia setelah India, yang diikuti oleh China dan dilanjutkan dengan negara lainnya. Kasus terbanyak dilaporkan dari provinsi yang memiliki jumlah penduduk yang tinggi seperti Jawa Barat, Jawa Timur, dan Jawa Tengah. Maka dari itu, terlepas dari kemajuan yang telah dicapai Indonesia, jumlah kasus *tuberkulosis* di Indonesia masih menjadi tantangan besar yang dihadapi dan memerlukan perhatian semua pihak.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada Semester Ganjil Tahun Ajaran 2025/2026 di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil melalui publikasi resmi Badan Pusat Statistik Provinsi Jawa Barat (https://jabar.bps.go.id/id). Data mencakup 28 kabupaten/kota di Provinsi Jawa Barat tahun 2022-2024 yang terdiri dari variabel jumlah kasus TBC (Y), kepadatan penduduk (X_1), jumlah penduduk (X_2), luas wilayah (X_3), jumlah fasilitas kesehatan (X_4), persentase penduduk miskin (X_5), dan laju pertumbuhan penduduk (X_6). Seluruh variabel ini dipilih karena memiliki relevansi epidemiologis dalam menggambarkan faktor-faktor sosial demografis yang memengaruhi penyebaran TBC di suatu wilayah.

3.3 Metode Penelitian

Pada penelitian ini menganalisis serta menerapkan *Inverse Gaussian Maximum Likelihood Estimator* (IGML) dan *Inverse Gaussian Hybrid Estimator* (IGH) untuk mengatasi permasalahan multikolinearitas pada pemodelan jumlah kasus *Tuberculosis* (TBC) di Provinsi Jawa Barat pada tahun 2022-2024. Adapun langkah-langkah pada penelitian ini adalah sebagai berikut:

1. Melakukan analisis deskriptif terhadap data untuk menjelaskan karakteristik setiap variabel dari data penelitian dengan menggunakan histogram dan boxplot.
2. Melakukan Uji distribusi dengan menggunakan uji *Kolmogorov–Smirnov*
3. Melakukan estimasi parameter menggunakan IGML
   a. Menghitung estimasi IGML dengan $\hat{\beta}_{IGML} = (\boldsymbol{X^T W X})^{-\mathbf{1}} \boldsymbol{X^T W z}$
   b. Menghitung parameter dispersi dengan $\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^{n} \frac{(Y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^3}$
4. Melakukan uji multikolinearitas.
   a. Matriks korelasi
   b. *Variance Inflation Factor* (VIF)
5. Melakukan uji parameter *Inverse Gaussian Hybrid Estimator* (IGH) dengan menentukan *shrinkage* k dan d dengan $k = \min\left(\frac{\phi}{\beta_j^2}\right)$ dan $d = min\left(\frac{\beta_j^2}{\beta_j^2 + \frac{\phi}{q_j}}\right)$

   dengan $q_j$adalah eigenvalue dari matriks $\boldsymbol{X^T U X}$
6. Melakukan estimasi parameter menggunakan IGH dengan
   $$\hat{\beta}_{IGH} = (\boldsymbol{X^T W X} + \boldsymbol{kI})^{-\mathbf{1}}(\boldsymbol{X^T W X} - \boldsymbol{kI})(\boldsymbol{X^T W X} + \boldsymbol{k}(\mathbf{1} + \boldsymbol{d})\boldsymbol{I})^{-\mathbf{1}} \boldsymbol{X^T W X} \hat{\beta}_{IGML}$$
7. Melakukan evaluasi dan perbandingan model dengan *Mean Squared Error* (MSE) dan *Root Mean Squared Error* (RMSE).
8. Melakukan uji parameter model
   a. Uji simultan dengan uji *Likelihood Ratio*.
   b. Uji parsial dengan uji *Wald*:

V. KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan yaitu dengan memodelkan jumlah kasus *Tuberculosis* (TBC) menggunakan pendekatan *Inverse Gaussian Regression* pada data kabupaten/kota tahun pengamatan, maka dapat disimpulkan:

1. Berdasarkan hasil analisis yang dilakukan, metode *Inverse Gaussian Hybrid Estimator* (IGH) menunjukkan kinerja yang lebih baik dibandingkan dengan *Inverse Gaussian Maximum Likelihood* (IGML), yang ditunjukkan oleh nilai *Mean Squared Error* (MSE) yang lebih rendah, sehingga menghasilkan estimasi yang lebih akurat dan stabil dalam mengatasi multikolinearitas.

2. Model regresi *Inverse Gaussian* diperoleh adalah sebagai berikut:

$$\log(\mu) = 10.16554 + 0.51046X_1 + 4.59483X_2 + 0.07363X_3 + 4.02780X_4 + 0.08774X_5 + 0.04555X_6$$

   Model tersebut menunjukkan bahwa hubungan antara variabel kepadatan penduduk (X_1), jumlah penduduk (X_2), luas wilayah (X_3), jumlah fasilitas kesehatan (X_4), persentase penduduk miskin (X_5), dan laju pertumbuhan penduduk (X_6) memiliki pengaruh terhadap jumlah kasus TBC, otomatis dengan regresi yang bernilai positif artinya bahwa peningkatan pada variabel tersebut cenderung diikuti oleh peningkatan jumlah kasus TBC, dengan asumsi variabel lain konstan.

DAFTAR PUSTAKA

Amin, M., Lukman, A. F., Ayinde, K., & Ogundimu, E. O. 2020. On the performance of some biased estimators in inverse Gaussian regression model. *Journal of Statistical Computation and Simulation*. **90**(12): 2141–2160.

Badan Pusat Statistik Provinsi Jawa Barat. 2024. *Statistik Kesehatan Jawa Barat*. BPS Jawa Barat.

Fisher, R. A. 1992. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons, New York.

Greene, W. H. 2018. *Basic Econometrics*. 5th Edition. McGraw-Hill International, New York.

Gujarati, D. N., & Porter, D. C. 2009. *Basic Econometrics*. 5th Edition. McGraw-Hill International, New York.

Hoerl, A. E., & Kennard, R. W. 1970. Ridge Regression: Biased estimation for nonorthogonal problems. *Technometrics*. **12**(1): 55–67.

Jong, P. de, & Heller, G. Z. 2008. *Generalized Linear Models for Insurance Data.* Cambridge University, Cambridge.

Kibria, B. M. G. 2003. Performance of some new ridge regression estimators. *Communications in Statistics – Simulation and Computation*. **32**(2): 419–435.

McCullagh, P., & Nelder, J. A. 1989. *Generalized Linear Models*. 2nd Edition. Chapman & Hall, London.

Montgomery, D. C., Peck, E. A., & Vining, G. G. 2021. *Introduction to Linear Regression Analysis.* 6th Edition. John Wiley & Sons, New York.

Schrödinger, E. 1915. Zur Theorie der Fall- und Steigversuche an Teilchen mit Brownscher Bewegung. *Physikalische Zeitschrift*. **16**(1): 289–295.

Stock, J. H., & Watson, M. W. 2018. *Introduction to Econometrics*. 4th Edition. Pearson, Boston.

World Health Organization 2024. *Global Tuberculosis Report 2024*. WHO.