PENERAPAN REGRESI *BIVARIATE POISSON INVERSE GAUSSIAN* DENGAN MENGGUNAKAN ALGORITMA *FISHER SCORING*

SKRIPSI

NURUL IKHSANI

H051181011

PROGRAM STUDI STATISTIKA DEPARTEMENSTATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2022

PENERAPAN REGRESI *BIVARIATE POISSON INVERSE GAUSSIAN* DENGAN MENGGUNAKAN ALGORITMA *FISHER SCORING*

SKRIPSI

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

NURUL IKHSANI

H051181011

PROGRAM STUDI STATISTIKA DEPARTEMENSTATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2022

LEMBAR PERNYATAAN KEOTENTIKAN

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

Penerapan Regresi *Bivariate Poisson Inverse Gaussian* Menggunakan *Algoritma Fisher Scoring*

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

Makassar, 19 Oktober 2022

SEPULUH RIBU RUPIAH 10000 TGL 20 METERAI TEMPEL 2FA59AKX062016479

Nurul Ikhsani

NIM H051181011

PENERAPAN REGRESI BIVARIATE POISSON INVERSE GAUSSIAN DENGAN MENGGUNAKAN ALGORITMA FISHER SCORING

Disetujui Oleh:

Pembimbing Utama

Anisa, S.Si., M.Si.

NIP. 197302271 199802 2 001

Pembimbing Pertama

Dr. Nirwan, M. Si.

NIP. 19630306 198702 1 002

Ketua Program Studi

Dr. Nurtiti Sunusi, S.Si., M.Si

NIP. 19720117 199703 2 002

Pada 19 Oktober 2022

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Nurul Ikhsani
NIM : H051181011
Program Studi : Statistika
Judul Skripsi : Penerapan Regresi *Bivariate Poisson Inverse Gaussian* Menggunakan *Algoritma Fisher Scoring*

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Anisa, S.Si., M.Si. (...............................)
2. Sekretaris : Dr. Nirwan, M. Si. (...............................)
3. Anggota : Andi Kresna Jaya, S.Si., M.Si. (...............................)
4. Anggota : Drs. Raupong, M. Si. (...............................)

Ditetapkan di : Makassar

Tanggal : 19 Oktober 2022

KATA PENGANTAR

Segala Puji dan Syukur kehadirat Allah SWT yang telah melimpahkan berkah, rahmat dan hidayah-Nya kepada penulis sehingga dapat menyelesaikan skripsi dengan judul "Penerapan Regresi *Bivariate Poisson Inverse Gaussian Menggunakan Algoritma Fisher Scoring*" yang disusun sebagai salah satu syarat akademik untuk meraih gelar sarjana pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin. Shalawat serta salam senantiasa tercurahkan kepada junjungan Nabi besar Muhammad SAW sebagai *uswatun hasanah* dalam meraih kesuksesan dunia dan akhirat.

Penulis menyadari bahwa penyelesaian skripsi ini tidak luput dari hambatan dan tantangan. Namun, berkat bantuan dan dorongan serta motivasi baik yang terkait secara langsung maupun tidak langsung dari berbagai pihak akhirnya kesulitan-kesulitan yang timbul dapat diatasi. Untuk itu suatu kewajiban kepada penulis untuk meyampaikan terima kasih yang tak terhingga kepada kedua orang tua penulis, **Ayahanda Muhammad Said** dan **Ibunda Suheni** atas segala doa, kasih sayang, pengertian, perhatian yang tulus, pengorbanan dan perjuangaan serta dukungan yang luar biasa yang telah diberikan selama ini, dan terima kasih selalu mendengarkan keluh kesah penulis. Terima kasih juga kepada keluarga besar penulis yang selalu mendoakan, serta memberikan motivasi dan berbagi banyak hal selama ini.

Keberhasilan penulisan skripsi ini tidak lepas dari bimbingan, pengarahan dan bantuan dari berbagai pihak baik berupa pikiran, motivasi, tenaga maupun doa karena itu perkenankan pula penulis untuk menyampaikan terima kasih dan penghargaan setinggi-tingginya kepada yang terhormat :

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Nurtiti Sunusi, S.Si, M.Si**., selaku Ketua Departemen Statistika serta segenap dosen pengajar dan staf Departemen Statistika yang telah membekali ilmu dan meluangkan waktunya untuk Penulis selama menjadi mahasiswa dilingkungan Universitas Hasanuddin.

4. **Ibu Anisa, S.Si, M.Si.**, selaku Pembimbing Utama dan **Bapak Dr. Nirwan, M.Si.**, selaku Penasehat Akademik dan Pembimbing pertama yang telah meluangkan waktu dan pemikirannya untuk membantu penulis menyeselasikan hambatan dan tantangan serta memberikan arahan, dorongan, motivasi dan semangat kepada Penulis dari awal hingga selesainya penulisan skripsi ini.
5. **Bapak Andi Kresna Jaya, S.Si, M.Si**., dan **Bapak Drs. Raupong, M. Si.**, selaku Tim Penguji atas segala saran dan pembelajaran yang membangun dalam penyempurnaan penyusunan skripsi ini serta telah meluangkan waktu dan pemikiran kepada Penulis dalam penyusunan skripsi ini.

Tidak lupa ucapan terima kasih yang mendalam Penulis sampaikan untuk seluruh pihak yang senantiasa memberikan bantuan berupa semangat ataupun dukungan yang tak ternilai besarnya, kepada :

1. Sahabat seperjuangan FaRIN**: Riskayani**, **Isra Rizka Utami**, **Andi Ummi Melin Aicha**, dan **Rahma Alifah**. Terima kasih atas semua bantuan, canda tawa, suka duka perkuliahan, dan kebersamaan yang telah lewati, serta telah mendengarkan keluh kesah Penulis selama menjalani masa kuliah sampai detik ini dan seterusnya.
2. Teman seperjuangan **Nur Hidayah. L** terima kasih atas segala motivasi dan bantuan yang telah diberikan selama penyusunan skripsi.
3. Teman seperjuangan **Statistika 2018**, terima kasih atas ilmu, kebersamaan, perjuangan, dukungan dan kenangan indah selama di Kampus Universitas Hasanuddin, kalian orang-orang hebat.
4. **Dian Islamia Muhtar** sebagai teman kos dahulu atas segala kebersamaan selama ini.
5. **Ridhah Wahyuni** dan **Evelin Bangkulu’** terima kasih atas waktu dan kebersamaan selama ini.
6. **Teman-teman KKN Tematik Gelombang 106**, terima kasih telah menjadi teman sekaligus keluarga selama sebulan lebih, banyak pembelajaran yang berharga dilalui bersama selama satu bulan lebih.

7. Semua pihak yang telah banyak berpartisipasi, baik secara langsung maupun tidak langsung yang tak sempat penulis sebutkan satu per satu. Terima kasih atas segala bantuan dan dukungannya.

Penulis menyadari bahwa masih banyak kekurangan dalam skripsi ini. Oleh karena itu, dengan segala kerendahan hati Penulis memohon maaf. Akhir kata semoga skripsi ini dapat memberikan manfaat bagi pembaca.

Makassar, 19 Oktober 2022

Nurul Ikhsani

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIK

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

Nama	: Nurul Ikhsani
NIM	: H05118101
Program Studi	: Statistika
Departemen	: Statistika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam
Jenis Karya	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-ekslusif** (***Non-exclusive Royalty- Free Right***) atas tugas akhir saya yang berjudul:

“Penerapan Regresi *Bivariate Poisson Inverse Gaussian Menggunakan Algoritma Fisher Scoring*”

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.
Dibuat di Makassar pada tanggal, 19 Oktober 2022

Yang menyatakan

(Nurul Ikhsani)

ABSTRAK

Regresi Poisson merupakan model regresi non-linier dengan variabel responnya berupa data diskrit dan berdistribusi Poisson. Salah satu asumsi yang harus dipenuhi dalam regresi Poisson adalah asumsi equidispersi, yaitu keadaan dengan mean dan variansi dari variabel responnya bernilai sama. Namun dalam aplikasinya asumsi tersebut kadang dilanggar karena nilai variansinya lebih besar daripada meannya yang disebut overdispersi, sehingga digunakan metode *Poisson Inverse Gaussian*. Pada kasus dengan dua variabel respon yang berkorelasi dan memerlukan estimasi bersama, digunakan model regresi *Bivariate Poisson*. Sehingga, untuk kasus yang mengalami overdispersi digunakan metode *Bivariate Poisson Inverse Gaussian* (BPIG). Distribusi BPIG merupakan model yang berdistribusi campuran antara distribusi Poisson dan *Inverse Gaussian* dengan dua variabel respon. Parameter model regresi BPIG diestimasi menggunakan metode *maximum likelihood estimation* (MLE) dengan algoritma *Fisher Scoring*. Penelitian ini diaplikasikan pada data jumlah kematian ibu dan neonatal di Sulawesi Selatan tahun 2019. Hasil yang diperoleh adalah variabel prediktor yang mempengaruhi jumlah kematian ibu dan neonatal di Sulawesi Selatan pada tahun 2019 yaitu pelayanan K4 ibu hamil (X_1), peserta KB aktif (X_2), penanganan komplikasi kebidanan (X_3), penanganan komplikasi neonatal (X_4) dan jumlah puskesmas (X_5).

Kata Kunci : *distribusi BPIG, MLE, Fisher Scoring, overdispersi, regresi BPIG*

ABSTRACT

Poisson regression is a non-linear regression model with response variables in the form of discrete data and Distributed by Poisson. One of the assumptions that must be met in Poisson regression is the assumption of equidistence, that is, the state with the mean and the variance of the response variable is equally valued. However, in its application, this assumption is sometimes violated because the variance value is greater than the mean called overdispersion, so the Poisson Inverse Gaussian method is used. In cases with two response variables that are correlated and require a shared estimate, the Bivariate Poisson regression model is used. Thus, for cases that experience overdispersion, the Bivariate Poisson Inverse Gaussian (BPIG) method is used. The BPIG distribution is a model that distributes a mixture of the Poisson Bivariate distribution and the Gaussian Inverse. The parameters of the BPIG regression model were estimated using the maximum likelihood estimation (MLE) method with the Fisher Scoring algorithm. This study was applied to data on the number of maternal and neonatal deaths in South Sulawesi in 2019. The results obtained are predictor variables that affect the number of maternal and neonatal deaths in South Sulawesi in 2019, namely K4 services for pregnant women (X_1), *active birth control participants*(X_2), *handling obstetric complications* (X_3), *handling neonatal complications* (X_4) *and the number of puskesmas* (X_5).

Keywords : *BPIG distribution, MLE, Fisher Scoring, overdispersion, BPIG regression*

DAFTAR ISI

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Data cacahan (*count data*) merupakan data yang menerangkan beberapa peristiwa yang terjadi pada suatu periode masa tertentu. Pemodelan data cacahan tidak bisa menggunakan regresi OLS (*Ordinary Least Square*), karena pemodelan data cacahan akan menyanggah dua asumsi yang disyaratkan dalam regresi OLS yaitu, error mengikuti distribusi normal dan mempunyai sifat homokedastisitas (Sofyan dkk., 2017). Oleh karena itu, digunakan *Generalized Linear Models* (GLMs). Hal ini dikarenakan, GLMs tidak mengharuskan variabel respon berdistribusi normal dan varians homogen (Jong & Heller, 2008).

Regresi Poisson adalah salah satu anggota famili dari *Generalized Linear Models* (GLMs) yang bersumber dari distribusi Poisson. Distribusi Poisson adalah distribusi diskrit dan nilai variabel random berupa nilai bulat positif sehingga menjadi opsi yang baik akan pemodelan data cacahan. Distribusi Poisson semata-mata ditentukan oleh satu parameter yang menerangkan baik mean maupun varians dari distribusi tersebut, sehingga dalam regresi Poisson terdapat asumsi yang harus terpenuhi yaitu mean dan varians variabel respon harus sama (*equidispersion*). Namun pada kenyataannya kerap ada pelanggaran asumsi tersebut dimana varians lebih kecil dari mean (*underdispersion*) atau varians lebih besar dari mean (*overdispersion*). Kebanyakan data cacahan diketahui memiliki kasus overdispersi (Purnamasari, 2016).

Apabila terdapat kasus overdispersi pada data, maka penerapan regresi Poisson kurang akurat untuk dianalisis, karena pada nilai *standard error* akan mengakibatkan terjadinya *underestimate* (lebih kecil dari nilai sebenarnya), akibatnya hasil kesimpulan yang ditemukan nanti menjadi tidak valid. Adapun langkah untuk mengatasi terjadinya tersebut adalah melalui memilih sejumlah pemodelan yang merupakan kombinasi dari distribusi Poisson dengan sejumlah distribusi baik diskrit maupun kontinu (*mixed Poisson distribution*). Salah satu *mixed Poisson distribution* yang kerap digunakan dalam penelitian untuk mengatasi kasus overdispersi adalah distribusi *Poisson Inverse Gaussian* (PIG). (Jamaluddin, 2019).

Distribusi *Poisson Inverse Gaussian* (PIG) pertama kali diperkenalkan oleh Holla pada tahun 1996 (Karlis & Xekalaki, 2005). Beberapa penelitian menggunakan distribusi PIG dikarenakan mempunyai fungsi *likelihood* yang *close form* dan perhitungannya yang lebih mudah (Herindrawati, 2017). Beberapa penelitian yang mengenai distribusi PIG diantaranya yaitu pada penelitian Zha, Lord, dan Zou (2014) pada kejadian jumlah kecelakaan motor yang terdapat di Texas dan Washington, menunjukkan bahwa pemodelan regresi PIG lebih baik dari model regresi binomial negatif untuk memodelkan kejadian jumlah kecelakaan motor di dua tempat tersebut. Tahun 2017 Andriana Yoshinta Herindrawati juga menggunakan model PIG pada kasus baru HIV di Provinsi Jawa Tengah tahun 2015 dengan kesimpulan bahwa faktor yang berpengaruh signifikan adalah PUS yang menggunakan kondom, rasio fasilitas Kesehatan, persentase daerah perkotaan, dan persentase penduduk usia 25-34 tahun.

Dalam metode PIG, terdapat tahap pengestimasian parameter menggunakan metode *Maximum Likelihood Estimation* (MLE) yang berfungsi untuk memaksimumkan fungsi *likelihood*. Dalam proses tersebut tidak semuanya bisa diselesaikan dengan cara analitik. Jika diperoleh bentuk implisit dan non linear, maka dapat diselesaikan dengan menggunakan algoritma *Fisher scoring*. Algoritma *Fisher scoring* merupakan salah satu bentuk perluasan dari algoritma *Newton Raphson*. Algoritma *Newton Raphson* menggunakan matriks Hessian yang elemennya ialah turunan kedua fungsi *likelihood*-nya. Hal ini yang mengakibatkan hasil iterasinya tidak selalu konvergen. Oleh karena itu, algoritma *Newton Raphson* diperluas dengan mengganti matriks Hessian menjadi matriks informasi (Resmiasih, 2019). Beberapa penelitian sebelumnya mengenai algoritma *Fisher scoring* diantaranya yaitu pada penelitian Schworer dan Hovey (2004) dengan membandingkan keunggulan *Newton Raphson* dan algoritma *Fisher Scoring* dalam MLE. Penelitian selanjutnya yaitu Purba (2018) dengan MLE berdasarkan algoritma *Newton Raphson*, *Fisher scoring*, dan algoritma Ekspektasi Maksimasi. Pada penelitian tersebut ditunjukkan bahwa algoritma *Fisher scoring* merupakan algoritma iterasi terbaik diantara ketiga algoritma yang diterapkan pada data kecelakaan tersebut.

Salah satu contoh data cacahan dalam lingkungan kesehatan adalah kasus kematian ibu dan kematian neonatal. Kematian ibu dan kematian neonatal merupakan dua hal yang saling berhubungan karena status gizi dan kesehatan ibu erat kaitannya dengan kesehatan bayi dalam kandungan. Menurut WHO (2010), kematian ibu merupakan kematian wanita dalam periode kehamilan, persalinan, dan dalam kurun waktu 42 hari (6 minggu) setelah pasca persalinan dengan penyebab langsung dan tidak langsung terhadap kehamilan. Menurut UNICEF (2018), lebih dari setengah kematian bayi, terjadi pada masa neonatus. Usia bayi 0-28 hari (masa neonatus) merupakan masa paling rentang untuk terkena berbagai masalah Kesehatan. Berdasarkan data Kemenkes (2019), 80% terjadi pada periode enam hari pertama kehidupan. Oleh karena itu, diperlukan penelitian tentang jumlah kasus kematian ibu dan kematian neonatal dengan faktor yang mempengaruhi keduanya dengan menggunakan metode *Bivariate Poisson Inverse Gaussian* (BPIG).

Berdasarkan uraian diatas, penulis tertarik untuk meneliti penerapan metode *Bivariate Poisson Inverse Gaussian* (BPIG) pada data yang mengalami overdispersi dengan menggunakan algoritma *Fisher scoring*. Oleh karena itu, penulis mengangkat judul “Penerapan Regresi *Bivariate Poisson Inverse Gaussian* dengan Menggunakan Algoritma *Fisher Scoring*”. Metode ini diaplikasikan pada data kasus kematian ibu dan kematian neonatal di Sulawesi Selatan pada tahun 2019.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, permasalahan yang dapat dirumuskan yaitu:

1. Bagaimana bentuk estimasi parameter model regresi BPIG dengan menggunakan algoritma *Fisher scoring*?
2. Bagaimana model regresi BPIG dengan menggunakan algoritma *Fisher scoring* pada data kasus kematian ibu dan kematian neonatal di Sulawesi Selatan tahun 2019?

1.3 Batasan masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penaksiran parameter model regresi BPIG menggunakan metode *Maximum Likelihood Estimation* (MLE) dengan algoritma *Fisher scoring.*
2. Studi kasus yang digunakan yaitu kasus kematian ibu dan kematian neonatal di Sulawesi Selatan tahun 2019. Jumlah kasus kematian ibu dan kematian neonatal di Sulawesi Selatan saling berkorelasi dan berdistribusi *Bivariate Poisson*.

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini yaitu

1. Memperoleh bentuk estimasi parameter model regresi BPIG dengan menggunakan algoritma *Fisher scoring*.
2. Memodelkan regresi BPIG dengan menggunakan algoritma *Fisher scoring* pada data kasus kematian ibu dan kematian neonatal di Sulawesi Selatan tahun 2019.

1.5 Manfaat Penelitian

Adapun manfaat penelitian ini yaitu

1. Penelitian ini dapat dijadikan sebagai bahan rujukan dalam pembelajaran tentang penerapan BPIG pada data yang mengalami overdispersi dengan menggunakan algoritma *Fisher scoring*.
2. Memberikan informasi untuk masyarakat dan dinas terkait mengenai faktor yang mempengaruhi kasus kematian ibu dan kematian neonatal di Sulawesi Selatan tahun 2019.

BAB II
TINJAUAN PUSTAKA

2.1 Regresi *Bivariate Poisson*

Pada dasarnya, terdapat dua jenis analisis regresi yaitu regresi linear dan nonlinear. Model regresi Poisson yaitu model yang termasuk dalam regresi nonlinear untuk data diskrit. Regresi poisson didasarkan pada penggunaan distribusi Poisson (Aulele, 2012). Jika variabel random diskrit Y berdistribusi Poisson dengan parameter μ $(\mu > 0)$, maka Y memiliki fungsi probabilitas dinyatakan dalam Persamaan (2.1) berikut (Arkandi & Winahju, 2015):

$$P(Y = y) = p(y) = \frac{e^{-\mu}\mu^{y}}{y!} \tag{2.1}$$

Regresi Poisson adalah penerapan dari *Generalized Linear Model* (GLM). GLM merupakan pengembangan dari model regresi umum untuk variabel respon yang tidak mengharuskan berdistribusi normal (Sari, 2018). Pada GLM, terdapat hubungan antara rata-rata dari variabel respon dengan variabel prediktor yang dihubungkan oleh suatu fungsi linear (Cahyandari, 2014), yaitu:

$$g(\mu_i) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{np}$$

$$g(\mu_i) = \boldsymbol{X}_i^T \boldsymbol{\beta}$$

dengan

g = fungsi penghubung

μ_i = rata-rata dari variabel respon yang berdistribusi Poisson

$\boldsymbol{X}_i^T = \left[1 \; X_{i1} \; X_{i2} \; \dots \; X_{ip}\right]_{1\times(p+1)}$ sebagai vektor variabel prediktor $k = 1,2,\dots,p$

pada pengamatan ke- $i = 1,2,\dots,n$

$\boldsymbol{\beta} = \left[\beta_0 \; \beta_1 \; \beta_2 \; \cdots \; \beta_p\right]_{1\times(p+1)}^T$ sebagai vektor koefisien regresi.

Regresi Poisson biasanya menggunakan fungsi penghubung log, karena rata-rata dari variabel responnya akan berbentuk fungsi eksponensial dan menjamin nilai variabel yang diduga dari variabel responnya akan bernilai non-negatif. Sehingga fungsi penghubung berbentuk sebagai berikut:

$g(\mu_i) = \ln \mu_i$

Maka hubungan rata-rata variabel respon dan variabel prediktor dengan fungsi penghubung seperti berikut:

$ln(\mu_i) = \boldsymbol{X}_i^T \boldsymbol{\beta}$

$\mu_i = exp(\boldsymbol{X}_i^T \boldsymbol{\beta})$

Oleh karena itu, model regresi Poisson dapat dituliskan sebagai berikut (Cahyandari, 2014):

$y_i = \mu_i + \varepsilon_i$

$y_i = exp(\boldsymbol{X}_i^T \boldsymbol{\beta}) + \varepsilon_i$

dengan ε_i sebagai galat.

Asumsi yang dipenuhi pada regresi ini ialah rata-rata dan variansinya harus equidispersi. Equidispersi merupakan keadaan nilai rata-rata dan variansi sama besar (Adiatma dkk., 2021).

Distribusi *Bivariate Poisson* terjadi Ketika 3 variabel random berdistribusi Poisson dengan parameter μ_1, μ_2, dan μ_3. Misalkan, terdapat variabel random Y_1 dan Y_2 yang tercipta dari variabel Z_1, Z_2 dan Z_3 yaitu $Y_1 = Z_1 + Z_3$ dan $Y_2 = Z_2 + Z_3$. Secara bersama-sama variabel random Y_1 dan Y_2 berdistribusi *Bivariate Poisson* mempunyai fungsi peluang seperti berikut:

$$f(y_1, y_2 | \mu_1, \mu_2, \mu_3) = = e^{-(\mu_1 + \mu_2 + \mu_3)} \sum_{k=0}^{min(y_1,y_2)} \frac{{\mu_1}^k {\mu_2}^{(y_1-k)} {\mu_3}^{(y_2-k)}}{k!(y_1-k)!\,(y_2-k)!};\ y_i = 0,1,2,.. \quad (2.2)$$

dengan

y_1, y_2 = variabel respon

μ_1, μ_2, μ_3 = parameter rata-rata dari variabel Z_1, Z_2, dan Z_3

k = nilai yang mungkin untuk variabel Z_3

Parameter μ_3 adalah suatu ukuran untuk mengetahui besarnya korelasi antara Y_1 dan Y_2 . Jika μ_3 bernilai nol, maka Y_1 dan Y_2saling bebas dan bukan merupakan *Bivariate Poisson* (Karlis & Ntzoufras, 2005).

2.2 Regresi *Bivariate Poisson Inverse Gaussian*

Salah satu solusi memodelkan data cacahan yaitu dengan menggunakan distribusi *mixed Poisson*. Salah satu distibusi dari *mixed Poisson* yaitu distribusi *Poisson Inverse Gaussian* (PIG). Distribusi PIG merupakan gabungan dari distribusi Poisson dan *Inverse* Gaussian.

Misalkan Y adalah variabel respon yang berdistribusi *Poisson Inverse Gaussian*, maka fungsi kepadatan peluang bagi Y adalah

$P(Y = y|\mu, \tau) = \int_0^{\infty} f(y|\mu, v)g(v, \tau)\, dv$

dengan

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

$g(v, \tau) \quad = (2\pi v^3)^{-\frac{1}{2}} e^{-(v-1)^2/2\tau v}$

v = efek random yang berdistribusi *inverse* gaussian

Berdasarkan parameternya, distribusi PIG terdiri dari dua parameter,yaitu parameter rata-rata (μ) dan parameter dispersi (τ). Jika Y adalah variabel respon yang berdistribusi PIG, maka distribusi PIG dapat dinotasikan dengan Y$\sim PIG(\mu, \tau)$. Sehingga fungsi kepadatan peluang dapat ditulis pada Persamaan (2.3) berikut (Willmot, 1987):

$$P(y|\, \mu, \tau) = e^{\frac{1}{\tau}} K_s(z) \left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}} (1 + 2\tau\mu)^{-\frac{\left(y-\frac{1}{2}\right)}{2}} \frac{\mu^y}{y\,!}\ ; y = 0,1,2, \ldots \qquad (2.3)$$

dengan

$z \quad = \sqrt{\frac{1}{\tau^2} + \frac{2\mu}{\tau}}$

$s \quad = y - \frac{1}{2}$

$K_s(z) = K_{y-\frac{1}{2}} \sqrt{\frac{1}{\tau^2} + \frac{2\mu}{\tau}}$ sebagai fungsi Bessel modifikasi jenis ketiga.

Untuk

y = variabel respon

τ = parameter dispersi

μ = rata-rata.

Shoukri dkk., (2004) dalam publikasinya "*The Poisson Inverse Gaussian Regression Model in the Analysis of Clustered Counts Data*" mengatakan bahwa dengan memanfaatkan sifat-sifat fungsi Bessel maka diketahui bahwa:

$$M(y_i) = \frac{1}{\sqrt{1+2\tau\mu}} \frac{K_{y+\frac{1}{2}}(z)}{K_{y-\frac{1}{2}}(z)} \qquad (2.4)$$

Regresi PIG digunakan sebagai alternatif dari model regresi Poisson yang mengalami masalah overdispersi. Pada regresi PIG dimisalkan Y_i adalah variabel respon untuk pengamatan ke-i. Sehingga model regresi PIG dapat dituliskan seperti pada rumus berikut:

$$\mu_i = e^{\boldsymbol{X}_i^T \boldsymbol{\beta}} \text{ atau } ln(\mu_i) = \boldsymbol{X}_i^T \boldsymbol{\beta} \qquad (2.5)$$

dengan

$\boldsymbol{X}_i^T$: vektor variabel prediktor pengamatan ke-i

$\boldsymbol{\beta}$: vektor koefisien regresi

untuk $i = 1,2, \dots, n$ sebagai pengamatan.

Jika terdapat dua variabel random Y_1 dan Y_2 yang berdistribusi Poisson dan tidak saling bebas, yang mempunyai rata-rata $v\mu_1$ dan $v\mu_2$. Variabel *v* merupakan variabel random yang berdistribusi *Inverse Gaussian*. Hal tersebut menunjukkan bahwa Y_1 dan Y_2 berdistribusi *mixed Poisson*, yaitu *Bivariate Poisson Inverse Gaussian* (BPIG). Berdasarkan persamaan (2.4), distribusi BPIG memiliki fungsi kepadatan gabungan sebagai berikut (Mardalena dkk., 2021):

$$P(y_j | j = 1,2) = e^{\frac{1}{\tau}} K_s(z) \left(\frac{2}{\pi\tau}\right)^{\frac{1}{2}} \left(1 + 2\tau \sum_{j=1}^{2} \mu_j\right)^{-\frac{\left(2\sum_{j=1}^{2} y_j - 1\right)}{4}} \prod_{j=1}^{2} \frac{{\mu_j}^{y_j}}{y_j\,!} \qquad (2.6)$$

dengan

$$z = \sqrt{\frac{1}{\tau^2} + \frac{2(\mu_1 + \mu_2)}{\tau}}$$

$$s = y_1 + y_2 - \frac{1}{2}$$

$K_s(z) = K_{y_1+y_2-\frac{1}{2}}\left(\frac{1}{\tau}\sqrt{2\tau(\mu_1+\mu_2)+1}\right)$ sebagai fungsi Bessel modifikasi jenis ketiga

Model regresi *Bivariate Poisson Inverse Gaussian* (BPIG) memiliki dua variabel respon yang saling berkorelasi. Misalkan y_{ij} sebagai variabel respon untuk pengamatan ke-i dan variabel respon ke-j dengan sampel random $Y_{1i}Y_{2i} \sim BPIG(\mu_{ij}, \tau)$ dimana $i = 1,2,..,n$ dan $j = 1,2,$ .Fungsi penghubung log natural (ln) diperlukan dalam pemodelan BPIG. Fungsi penghubung ln digunakan untuk menghubungkan parameter μ_{ij} dengan variabel penjelas. Sehingga model regresi BPIG sebagai berikut (Mardalena dkk., 2021):

$$ln\left(\mu_{ij}\right) = \boldsymbol{X}_i^T\boldsymbol{\beta}_j + \varepsilon_{ij}$$

$$\mu_{ij} = exp\left(\boldsymbol{X}_i^T\boldsymbol{\beta}_j + \varepsilon_{ij}\right) \tag{2.7}$$

dengan

$\boldsymbol{X}_i^T = \left[1\ X_{i1}\ X_{i2}\ \dots\ X_{ip}\right]_{1\times(p+1)}$ sebagai vektor variabel prediktor $k = 1,2,\dots,p$ pada pengamatan ke- $i = 1,2,\dots,n$

$\boldsymbol{\beta}_j = \left[\beta_{j0}\ \beta_{j1}\ \beta_{j2}\ \cdots\ \beta_{jp}\right]_{1\times(p+1)}^T$ sebagai vektor koefisien regresi

ε_{ij} = galat

2.3 Metode *Maximum Likelihood Estimation*

Metode *Maximum Likelihood Estimation* (MLE) digunakan untuk mengestimasi parameter suatu model yang fungsi probabilitasnya diketahui. Misalkan $Y_1, Y_2, \dots, Y_n$ adalah sampel random dan berdistribusi Poisson, maka fungsi probabilitas bersama dari $Y_1, Y_2, \dots, Y_n$ dapat dinyatakan sebagai berikut.

$$f(y_1, y_2, \dots, y_n|\ \boldsymbol{\theta}) = f(y_1|\theta).f(y_2|\theta)\dots f(y_n|\theta) = f(y_i|\theta)$$

dengan θ adalah parameter yang tidak diketahui. Fungsi *likelihood* didefinisikan sebagai fungsi probabilitas bersama dari $Y_1, Y_2, \dots, Y_n$ yang dimisalkan sebagai fungsi dari θ yang dituliskan pada Persamaan (2.8) berikut (Hoog dkk., 2019).

$$L(\theta|\ y) = f(y_1, y_2, \dots, y_n|\ \theta) = f(y_1|\theta).f(y_2|\theta)\dots f(y_n|\theta) = \prod_{i=1}^{n} f(y_i|\theta) \tag{2.8}$$

Penaksir dari maksimum *likelihood* $\hat{\theta}$ yakni nilai θ yang memaksimumkan fngsi *likelihood* $L(\theta|\ y)$. Akan tetapi, lebih mudah bekerja dengan logaritma natural dari fungsi *likelihood*, yaitu $l(\theta|\, y) = L(\theta|\, y)$ sehingga dapat dituliskan pada Persamaan (2.9) berikut.

$$l(\theta|\, y) = L(\theta|\, y) = \ln\{\prod_{i=1}^{n} f(y_i|\theta)\} = \sum_{i=1}^{n} \ln f(y_i|\theta) \quad (2.9)$$

Untuk mendapatkan nilai θ yang memaksimumkan fungsi $l(\theta|\, y)$, maka $l(\theta|\, y)$ diturunkan terhadap θ dan kemudian menyamakannya dengan nol seperti pada Persamaan (2.10) berikut (Hogg dkk., 2019).

$$\frac{\partial\, l(\theta|\, y)}{\partial \theta} = 0 \quad (2.10)$$

2.4 Algoritma *Fisher Scoring*

Algoritma *Fisher Scoring* merupakan salah satu bentuk perluasan dari metode Newton Raphson yang digunakan dalam statistik untuk menyelesaikan persamaan *Maximum Likelihood*. Algoritma *Fisher scoring* serupa dengan algoritma *Newton Raphson*, perbedaanya adalah *Fisher scoring* memerlukan matriks informasi. Matriks informasi tersebut adalah negatif dari nilai ekspektasi dari matriks turunan kedua fungsi yang akan dimaksimumkan sedangkan algoritma newton raphson memerlukan matriks turunan kedua dari nilai yang diamati. Persamaan iterasi *Fisher Scoring* sebagai berikut:

$$\hat{\boldsymbol{\theta}}_{(t+1)} = \hat{\boldsymbol{\theta}}_{(t)} + \boldsymbol{I}^{-1}\left(\hat{\theta}_{(t)}\right)\boldsymbol{D}\left(\hat{\theta}_{(t)}\right) \quad (2.11)$$

Untuk $I_{(t)}$ sebagai taksiran ke-t dari matriks informasi yang diamati. Matriks informasi dalam penulisan ini yaitu $\boldsymbol{I}\left(\hat{\theta}_{(t)}\right) = -E[H\left(\hat{\theta}_{(t)}\right)]$ (Purba, 2018).

2.5 Pengujian Hipotesis Model Regresi *Bivariate Poisson Inverse Gaussian* (BPIG)

Pengujian hipotesis regresi BPIG menggunakan metode *Maximum Likelihood Ratio Test* (MLRT). Sebelum melakukan pengujian model regresi BPIG, dilakukan pengujian asumsi yang harus dipenuhi pada model regresi terlebih dahulu. Pengujian asumsi dalam metode BPIG terdiri atas empat pengujian, yaitu:

2.5.1 Uji Asumsi Distribusi *Bivariate Poisson*

Pengujian asumsi distribusi *Bivariate Poisson* dapat dilakukan dengan pendekatan *index of dispersion test* (I_B). Pengujian ini berfungsi untuk mengetahui variabel respon mengikuti distribusi *Bivariate Poisson* atau tidak. Hipotesis yang digunakan yaitu:

$H_0 : F(y_1, y_2) = F_0(y_1, y_2) = 0$ (Y_1 dan Y_2 mengikuti distribusi *Bivariate Poisson*)

$H_1 : F(y_1, y_2) \neq F_0(y_1, y_2) \neq 0$ (Y_1 dan Y_2 tidak mengikuti distribusi *Bivariate Poisson*)

Statistik uji yang digunakan untuk pengujian hipotesis sebagai berikut:

$$I_B = \frac{n\left(\bar{Y}_2 S_{Y_1}^2 - 2m_{11}^2 + \bar{Y}_1 S_{Y_2}^2\right)}{\left(\bar{Y}_1 \bar{Y}_2 - m_{11}^2\right)} \tag{2.12}$$

dengan

$$S_{Y_1}^2 = \frac{\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)^2}{n}\ ;\ S_{Y_2}^2 = \frac{\sum_{i=1}^{n}(Y_{2i} - \bar{Y}_2)^2}{n} \text{ dan } m_{11} = \frac{\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)(Y_{2i} - \bar{Y}_2)}{n}$$

untuk $n =$ jumlah data pada variabel respon

Kriteria penolakan H_0 jika $I_B > \mathcal{X}^2_{(\alpha;2n-3)}$ (Best, 1999).

2.5.2 Uji Korelasi Variabel Respon

Salah satu syarat pada regresi BPIG yaitu adanya hubungan linier antar variabel respon. Pengujian korelasi untuk variabel respon menggunakan uji Korelasi *Pearson* dengan hipotesis sebagai berikut:

$H_0 :\ \rho_{y_1 y_2} = 0$ (Y_1 dan Y_2 tidak berkorelasi)

$H_1 :\ \rho_{y_1 y_2} \neq 0$ (Y_1 dan Y_2 berkorelasi)

Statistik uji yang digunakan untuk uji korelasi adalah sebagai berikut:

$$r = \frac{\sum_{i=1}^{n}(y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2)}{\sum_{i=1}^{n}(y_{1i} - \bar{y}_1)^2 \sum_{i=1}^{n}(y_{2i} - \bar{y}_2)^2}$$

dengan kriteria penolakan, H_0 ditolak jika $|p_{value}| < \alpha$ (Putri dkk., 2019).

Koefisien korelasi Koefisien korelasi dapat menunjukkan dua hubungan, yaitu positif dan negatif. Nilai positif dan negatif dikarenakan nilai koefisien korelasi berkisar antara -1 hingga 1. Apabila nilai korelasi mendekati 1, baik positif maupun negatif berarti kedua variabel memiliki hubungan yang erat secara linier. Nilai korelasi yang positif menunjukkan adanya hubungan berbanding lurus pada dua variabel tersebut, sedangkan nilai korelasi yang negatif menunjukkan hubungan yang berbanding terbalik (Putri dkk, 2019).

2.5.3 Uji Multikolinearitas

Suatu model regresi yang mengalami multikolinieritas apabila terdapat hubungan linier tinggi diantara sejumlah atau semua variabel prediktor dari suatu model regresi. Kasus multikolinieritas bisa diketahui dengan melihat nilai VIF (*Variance Inflation Factors*). Hipotesis yang digunakan sebagai berikut:

H_0 : Tidak terjadi multikolinearitas

H_1 : Terjadi multikolinearitas

Statistik uji yang digunakan untuk pengujian hipotesis sebagai berikut:

$VIF = \frac{1}{1-R^2}$

dengan R^2 adalah kofisien determinasi antar variabel prediktor satu dengan yang lainnya. Kriteria penolakan, H_0 ditolak jika $VIF > 10$. (Putri dkk., 2019).

2.5.4 Uji Overdispersi

Pada model regresi Poisson terdapat sejumlah asumsi yang harus dipenuhi. Salah satunya adalah asumsi kesamaan antara mean dan variansi yang disebut equidispersi. Akan tetapi, dalam analisis data statistik sering dijumpai kondisi data dengan variansinya lebih kecil dari rataan ataupun sebaliknya. Kondisi ini disebut dengan underdispersi (*underdispersion*) atau overdispersi (*overdispersion*) (Kusuma dkk., 2013).

Overdispersi bisa menimbulkan perkiraan parameter yang diperoleh tidak efisien. Penggunaan yang tidak benar pada model regresi Poisson (yang mengalami overdispersi) bisa berakibat fatal dalam interpretasi model, terutama pada perkiraan parameter model karena bisa menaksir *standard error* yang terlalu rendah dan bisa menyampaikan kesimpulan yang salah tentang signifikan atau tidaknya parameter regresi yang terlibat (Darnah, 2011). Dalam mendeteksi kasus overdispresi dalam data, dapat dilihat dari hasil uji *Deviance*. Adapun hipotesis yang digunakan sebagai berikut:

$H_0 : \phi = 1$ Tidak terjadi overdispersi

$H_1 : \phi > 1$ Terjadi overdispersi

Statistik uji sebagai berikut:

$\phi = \frac{D^2}{db}$

dengan

$D^2 = 2\sum_{i=1}^{n}\left(y_i \ln\left(\frac{y_i}{\mu_i}\right)\right)$

$db = n - p$

$\mu_i =$ penduga bagi respon rata-rata ke-i

$n =$ banyaknya parameter termasuk konstanta

$p =$ banyaknya pengamatan

dengan kriteria penolakan, H_0 ditolak jika nilai $\phi > 1$ (Kusuma dkk., 2013).

2.5.5 Uji Secara Serentak Parameter Model

Pengujian secara serentak parameter model regresi digunakan untuk mengetahui bahwa seluruh variabel prediktor tidak berpengaruh terhadap variabel respon atau terdapat minimal salah satu variabel prediktor yang berpengaruh terhadap variabel respon. Parameter yang diuji pada pengujian ini mencakup seluruh parameter β secara bersama-sama dengan hipotesis sebagai berikut:

H_0 : $\beta_{jk} = 0$ untuk $j = 1,2$; $k = 1,2,\dots,p$

H_1 : setidaknya ada satu $\beta_{jk} \neq 0$ untuk $j = 1,2$; $k = 1,2,\dots,p$

Statistik uji yang digunakan untuk pengujian hipotesis sebagai berikut:

$$G^2 = -2\,ln\left(\frac{L(\widehat{\omega})}{L(\widehat{\Omega})}\right) = 2\left[ln\left(L(\widehat{\Omega})\right) - ln\left(L(\widehat{\omega})\right)\right] \tag{2.13}$$

Keterangan:

$L(\widehat{\omega}) =$ nilai *maximum likelihood* untuk model lengkap yang tidak menyertakan variabel prediktor

$L(\widehat{\Omega}) =$ nilai *maximum likelihood* untuk model lengkap yang menyertakan variabel prediktor

Dengan kriteria penolakan H_0 jika nilai $G^2 > \chi^2_{\alpha;dk}$ dengan dk merupakan derajat kebebasan yakni parameter dibawah populasi dikurangi parameter dibawah H_0, sehingga variabel prediktor secara bersama-sama memiliki pengaruh terhadap variabel respon (Herindrawati, 2017).

2.5.6 Uji Secara Parsial Satu Satu

Pengujian secara parsial digunakan untuk mengetahui parameter yang memberikan pengaruh yang signifikan terhadap model yang juga berarti digunakan untuk mengetahui variabel prediktor yang berpengaruh terhadap variabel respon. Parameter yang diuji pada pengujian ini yaitu parameter β . Statistik uji yang digunakan dalam pengujian ini yaitu menggunakan uji wald.

Hipotesis untuk menguji signifikansi parameter β sebagai berikut:

H_0 : $\beta_{jk} = 0$ untuk $j = 1,2$; $k = 1,2, \ldots, p$

H_1 : setidaknya ada satu $\beta_{jk} \neq 0$ untuk $j = 1,2$; $k = 1,2, \ldots, p$

Statistik uji yang digunakan untuk pengujian hipotesis sebagai berikut:

$$W(\hat{\beta}_{jp}) = \left(\frac{\hat{\beta}_{jp}}{se(\hat{\beta}_{jp})}\right)^2$$

Kriteria penolakan H_0 jika $W(\hat{\beta}_{jp}) > \chi^2_{\alpha;1}$ artinya variabel prediktor memiliki pengaruh signifikan terhadap variabel respon Y_j (Herindrawati, 2017).

2.6 Kematian Ibu dan Neonatal

Kematian ibu adalah kematian perempuan pada saat hamil atau kematian dalam kurun waktu 42 hari sejak terminasi kehamilan tanpa memandang lamanya kehamilan, yaitu kematian yang disebabkan karena kehamilannya atau penanganannya, tetapi bukan karena sebab-sebab lain seperti kecelakaan dan terjatuh. Angka Kematian Ibu (AKI) dihitung per 100.000 kelahiran hidup pada tahun tertentu (Dinkes, 2018).

Kematian neonatal adalah banyaknya kematian bayi yang terjadi pada bulan pertama setelah dilahirkan dan umumnya disebabkan oleh faktor yang dibawa anak sejak lahir yang diperoleh dari orang tua pada saat konsepsi atau selama kehamilan. Kematian neonatal dapat disebabkan oleh dua faktor, yaitu faktor ibu antara lain pelayanan kesehatan ibu hamil, infeksi ibu hamil, gizi ibu hamil, dan karakteristik dari ibu hamil (umur, paritas, dan jarak kehamilan) serta faktor janin antara lain bayi berat badan lahir rendah (BBLR), asfiksia, dan pneumonia (Dinkes, 2018).