

**IMPLEMENTASI MODEL REGRESI HURDLE POISSON
(STUDI KASUS: FAKTOR-FAKTOR BERPENGARUH
TERHADAP GEJALA TERPAPAR COVID-19 SAAT
PANDEMI DI PUSKESMAS BARANTI
KECAMATAN BARANTI)
SKRIPSI**



Oleh:

NURUL ANISA

H011181302

**PROGRAM STUDI MATEMATIKA
DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
OKTOBER 2022**

**IMPLEMENTASI MODEL REGRESI HURDLE POISSON
(STUDI KASUS: FAKTOR-FAKTOR BERPENGARUH
TERHADAP GEJALA TERPAPAR COVID-19 SAAT
PANDEMI DI PUSKESMAS BARANTI
KECAMATAN BARANTI)**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada
Program Studi Matematika Departemen Matematika dan Ilmu Pengetahuan**

Alam Universitas Hasanuddin

NURUL ANISA

H011181302

**PROGRAM STUDI MATEMATIKA
DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
OKTOBER 2022**

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Nurul Anisa

Nim : H011181302

Program Studi : Matematika

Jenjang : S1

Menyatakan dengan ini bahwa karya tulis saya berjudul

**Implementasi Model Regresi Hurdle Poisson
(Studi Kasus: Faktor-Faktor Berpengaruh Terhadap Gejala
Terpapar Covid-19 Saat Pandemi Di Puskesmas Baranti
Kecamatan Baranti)**

Adalah karya tulis saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain bahwa skripsi saya tulis ini benar-benar merupakan hasil karya saya sendiri. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 25 Oktober 2022

Yang menyatakan,


Nurul Anisa
H011181302

LEMBAR PENGESAHAN

IMPLEMENTASI MODEL REGRESI HURDLE POISSON (STUDI KASUS: FAKTOR-FAKTOR BERPENGARUH TERHADAP GEJALA TERPAPAR COVID-19 SAAT PANDEMI DI PUSKESMAS BARANTI KECAMATAN BARANTI)

Disusun dan diajukan oleh

NURUL ANISA

H011181302

Telah dipertahankan di tahapan Panitia Ujian yang dibentuk dalam rangka
Penyelesaian Studi Program Sarjana Program Studi Matematika Fakultas Matematika
dan Ilmu Pengetahuan Alam Universitas Hasanuddin

Pada tanggal 25 Oktober 2022

Dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,

Pembimbing Pertama,

Dr. Firman, S.Si., M.Si.
NIP. 196804292002121001

Jusmawati Massalessa, S.Si., M.Si.
NIP. 196806011995122001

Ketua Program Studi,

Prof. Dr. Nurdin, S.Si., M.Si.
NIP. 197008072000031002



KATA PENGANTAR

Tiada kata pembuka yang pantas diucapkan kecuali mengucapkan kata syukur kepada Allah SWT atas berkat dan karunia-Nya sehingga penulis bisa menyelesaikan skripsi yang berjudul **Implementasi Model Regresi Hurdle Poisson (Studi Kasus: Faktor-Faktor berpengaruh terhadap gejala terpapar Covid-19 saat pandemi di Puskesmas Baranti Kecamatan Baranti)**, sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin. Sholawat serta salam tidak lupa kita kirimkan kepada baginda Rasulullah SAW, nabi yang menjadi penuntun kita untuk menjadi lebih baik serta berakhlak mulia untuk menuju jannahnya, *amin ya robbal alamin*.

Penulis menyadari bahwa skripsi ini tidak mungkin selesai tanpa adanya dukungan, bantuan, bimbingan dan doa dari berbagai pihak selama masa pengerjaannya. Oleh karena itu, pada kesempatan ini dengan segala kerendahan hati penulis menyampaikan terima kasih yang setulus-tulusnya kepada:

1. Bapak **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.** selaku Rektor Universitas Hasanuddin beserta jajarannya.
2. Bapak **Dr. Eng. Amiruddin, M.Si.** selaku Dekan FMIPA Universitas Hasanuddin.
3. Bapak **Prof. Dr. Nurdin, S.Si., M.Si.** selaku Ketua Departemen Matematika beserta Bapak dan Ibu Dosen Departemen Matematika yang telah memberi banyak ilmu pengetahuan kepada penulis selama menjadi mahasiswa di Program Studi Matematika, serta para Staf Departemen Matematika yang telah membantu dan memudahkan penulis dalam administrasi.
4. Bapak **Dr. Firman, S.Si., M.Si.** selaku Pembimbing Utama yang telah menyediakan waktu, tenaga, pikiran untuk membimbing dan mengarahkan penulis dalam menyelesaikan skripsi ini.
5. Ibu **Jusmawati Massalesse, S.Si., M.Si.** selaku pembimbing pertama sekaligus Penasehat Akademik yang telah memberi waktu, tenaga, pikiran dan motivasi dalam perkuliahan hingga penyusunan skripsi ini.

6. **Prof. Dr. Syamsuddin Toaha, M.Sc** dan **Dr. Amran, S.Si., M.Si.** selaku dosen penguji yang telah memberi saran dan arahan kepada penulis untuk penyempurnaan skripsi ini.
7. Orang tua dan keluarga tercinta yang sudah memberi doa, dukungan serta semangat yang tinggi.
8. Teman-teman seperjuangan Matematika 2018 yang telah mengukir kisah perkuliahan yang indah dan menjadi saksi akan perjuangan berkuliahan.
9. UKM Koperasi Mahasiswa Unhas yang banyak membantu penulis dalam kisah perkuliahan hingga penyusunan skripsi.
10. Keluarga cemara Asrama X Putri Baranti yang menjadi tempat berkeluh kesah dan sebagai tempat pertama saat butuh pertolongan.
11. Sahabat ANSUNOSA (Andir Suma Novi Anisa) yang memberi semangat tinggi dalam menyelesaikan skripsi ini sekaligus menjadi saksi kehidupan penulis.
12. Keluarga besar KKN SIDRAP 3 yang menjadi penyemangat akan penyelesaian skripsi.

Akhir kata, semoga segala bentuk kebaikan yang telah diberikan bernilai ibadah dan mendapat balasan dari Allah SWT. Semoga skripsi ini memberi manfaat bagi pengembangan ilmu.

Makassar, 25 Oktober 2022

Nurul Anisa

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama : Nurul Anisa

NIM : H011181302

Program Studi : Matematika

Departemen : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Jenis Karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

Implementasi Model Regresi Hurdle Poisson (Studi Kasus: Faktor-Faktor Berpengaruh Terhadap Gejala Terpapar Covid-19 Saat Pandemi Di Puskesmas Baranti Kecamatan Baranti)

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 Lamasi pada tanggal 25 Oktober 2022

Yang menyatakan,



Nurul Anisa

ABSTRAK

Infeksi Covid-19 menjadi masalah kesehatan dunia. Kasus ini berawal ketika Badan Kesehatan Dunia menyebutkan adanya kasus kluster pneumonia dengan etiologi yang tidak jelas di kota Wuhan, provinsi Hubei, China. Kemudian WHO telah menetapkan sebagai darurat kesehatan masyarakat yang meresahkan dunia karena penambahan jumlah kasus Covid-19 dengan cepat. Hingga saat ini masih ada banyak penyebarannya diseluruh wilayah. Gejala terpapar Covid-19 merupakan hal yang sangat dihindari, maka dari itu kesehatan sangat perlu dijaga. Beberapa faktor perpengaruh terhadap gejala Covid-19 ialah adanya gejala batuk, faktor usia, pekerjaan dan pengonsumsi rokok. Perilaku hidup bersih yang kurang diperhatikan dapat menyebabkan kesehatan menurun, sehingga mudah terkena penyakit yang menyebabkan cepat terpapar virus. Terutama saat protokol kesehatan tidak diterapkan. Pada kasus ini merupakan data yang berbentuk diskrit namun tidak biner sehingga dapat dikatakan sebagai data cacahan yang memiliki banyak peluang dalam *Distribusi Poisson*. Kasus ini dapat dianalisis menggunakan metode *Regresi Poisson*. Namun metode *Regresi Poisson* tidak sinkron saat terjadi pelanggaran asumsi yaitu *Overdispersi* (nilai varian lebih besar dari rata-rata atau nilai deviance dibagi derajat bebas). Pada kasus ini mengalami *Overdispersi* karena pada variabel banyak data bernilai nol sehingga dilanjutkan menggunakan model *Regresi Hurdle Poisson*. Dan diperoleh faktor yang berpengaruh secara signifikan pada gejala terpapar Covid-19 adalah faktor Usia.

Kata kunci: Gejala Covid-19, *Multikolinearitas*, *Regresi Poisson*, *Overdispersi*, *Regresi Hurdle Poisson*.

Judul : Implementasi Model Regresi Hurdle Poisson (Studi Kasus:
Faktor-Faktor Berpengaruh Terhadap Gejala Terpapar Covid-19
Saat Pandemi Di Puskesmas Baranti Kecamatan Baranti)

Nama : Nurul Anisa

NIM : H011181302

Program Studi : Matematika

ABSTRACT

Covid-19 infection is a global health problem. This case began when the World Health Organization mentioned a cluster case of pneumonia with an unclear etiology in the city of Wuhan, Hubei province, China. Then WHO declared it a public health emergency that worried the world because of the rapid increase in the number of Covid-19 cases. Until now there are still many distributions throughout the region. Symptoms of Covid-19 are things that are highly avoided, therefore health is very important to maintain. Some factors that influence the symptoms of Covid-19 are cough symptoms, age, occupation and cigarette consumption. Clean living behavior that is not paid attention to can cause health to decline, making it easy to get diseases that cause rapid exposure to viruses. Especially when health protocols are not implemented. In this case, it is discrete data but not binary, so it can be said as count data that has many opportunities in the Poisson distribution. This case can be analyzed using the Poisson Regression method. However, the Poisson Regression method is not synchronous when there is a violation of the assumption, namely Overdispersion (variant value is greater than the average or deviance value divided by degrees of freedom). In this case, there is overdispersion because many variables have no value, so we continue to use the Hurdle Poisson Regression model. And the factors that have a significant effect on the symptoms of being exposed to Covid-19 are the age factor.

Key words: Covid-19 Symptoms, *Multikolinearitas*, *Regresi Poisson*, *Overdispersi*, *Regresi Hurdle Poisson*.

Title :Implementation of Hurdle Poisson Regression Model (Case Study: The Factors that Affected Covid-19 Phenomenon on Pandemic Era at Puskesmas Baranti Kecamatan Baranti)

Name : Nurul Anisa

Student ID : H011181302

Study Program : Math

DAFTAR ISI

HALAMAN SAMBUTAN	i
HALAMAN JUDUL.....	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN KEASLIAN	iv
KATA PENGANTAR	v
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR	vii
ABSTRAK	viii
<i>ABSTRACT</i>	ix
DAFTAR ISI	x
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH	4
1.3 TUJUAN PENELITIAN	4
1.4 MANFAAT PENELITIAN.....	4
1.5 BATASAN MASALAH.....	4
1.6 SISTEMATIKA PENULISAN.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Distribusi Poisson.....	6
2.2 Uji Multikolinieritas.....	7
2.3 Regresi Poisson.....	7
2.4 Overdispersi.....	8
2.5 Model Regresi Hurdle Poisson	9
2.6 Pengujian Parameter.....	13
2.6.1 Pengujian Serentak (<i>Uji Rasio Likelihood</i>)	13
2.6.2 Pengujian Parsial (<i>Uji Wald</i>)	14
BAB III METODOLOGI PENELITIAN	16
3.1 Pendekatan Penelitian	16
3.2 Populasi dan Sampel Penelitian.....	16
3.3 Variabel Penelitian	17

3.4 Analisis Data	17
3.4.2 Implementasi model Regresi Hurdle Poisson pada faktor-faktor berpengaruh terhadap gejala terpapar Covid-19 saat pandemi.....	17
3.5 Alur Kerja	19
BAB IV PEMBAHASAN.....	20
4.1.Deskriptif Data	20
4.2 Uji Multikolinearitas	21
4.3 Pemodelan faktor-faktor berpengaruh terhadap gejala terpapar Covid-19 saat masa pandemi	22
4.3.1. Pemodelan Regresi Poisson	22
4.3.2 Model Regresi Hurdle Poisson	24
BAB V KESIMPULAN DAN SARAN	30
5.1 KESIMPULAN.....	30
5.2 SARAN	30
DAFTAR PUSTAKA	32
LAMPIRAN	32

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Matematika merupakan salah satu keilmuan dasar yang digunakan dan dibutuhkan sebagai dasar pengembangan teknologi. Matematika juga sangat berkaitan dengan kehidupan sehari-hari. Maka dari itu pelajaran matematika diberikan dalam semua jenjang pendidikan (Cookson dan Stirk, 2019).

Infeksi Covid-19 menjadi masalah kesehatan dunia. Kasus ini berawal ketika Badan Kesehatan Dunia pada tanggal 31 Desember 2019 menyebutkan adanya kasus kluster *pneumonia* dengan etiologi yang tidak jelas di kota Wuhan, Provinsi Hubei, China. Virus *Corona* adalah *zoonosis* (ditularkan antara hewan dan manusia). Pada tanggal 30 Januari 2020 WHO telah menetapkan sebagai Darurat Kesehatan Masyarakat yang meresahkan dunia karena penambahan jumlah kasus Covid-19 dengan cepat. Pada 02 Maret 2020 presiden Jokowi mengumumkan kasus pertama virus *corona* di Indonesia. Hingga saat ini masih ada banyak penyebarannya di seluruh wilayah (Ruhyat, 2021).

Virus *corona* menyerang paru-paru dan perilaku yang melemahkan paru-paru membuat seseorang jadi lebih berisiko, salah satunya adalah merokok. Kebiasaan merokok pada masyarakat masih sukar untuk dihentikan, rokok mengandung lebih dari 4000 zat kimia beracun. Faktanya seorang perokok telah mengetahui akibat dari merokok, akan tetapi mereka tetap melakukan hal tersebut. Apa yang menjadi kesulitan seseorang untuk berhenti merokok? *Nicotine regulation model* menegaskan bahwa perokok dapat menjaga tingkat nikotin yang berada di dalam darah dan terhindar dari dampak pengaruh putus zat. Interaksi dari pengaruh nikotin pada otak yakni berefek pada psikologis, seperti turunnya kemampuan dalam mengontrol emosi dan mudah depresi, sehingga membuat para perokok selalu ingin merokok untuk menenangkan diri.

Menurut WHO, Indonesia merupakan negara ketiga dengan jumlah perokok terbesar di dunia setelah Cina dan India. Menurut Patanavanich, dkk., menunjukkan

terdapat hubungan antara merokok dengan perburukan Covid-19. Terdapat 9.025 pasien dan 878 diantaranya mengalami keterburukan Covid-19 (17,8% perokok dan 9,3% tidak merokok) bukti lainnya dalam (Francisco dan Alqahtani, 2020), pasien perokok aktif lebih besar memiliki resiko untuk komplikasi berat dan resiko meninggal karena Covid-19.

Berdasarkan Kemenkes 2020, perokok lebih mudah terjangkit covid-19 dibandingkan orang yang tidak merokok, dengan resiko dua kali lebih tinggi terjangkit Covid-19 dibandingkan tidak perokok. Dr. Sumardi menjelaskan bahwa merokok dapat menyebabkan kerusakan saluran sistem pernapasan, sehingga mengakibatkan terjadinya Penyakit Paru Obstruktif Kronis (COPD).

Data perkembangan Covid-19 di Indonesia khususnya Sulawesi Selatan telah menyebar sampai 04 Februari 2022 terlapor 115 orang dirawat, 110.216 terkonfirmasi, 107.857 sembuh dan 2.244 meninggal. (Data Perkembangan COVID-19 di Indonesia - ANTARA News).

Perilaku hidup bersih yang kurang diperhatikan dapat menyebabkan kesehatan menurun, sehingga mudah mengalami demam dan batuk. Dengan gejala ini dapat menyebabkan penularan Covid-19 dikarenakan adanya droplet atau percikan ludah yang dapat menimbulkan virus (Dai,2020).

Aktifitas bekerja diluar rumah menjadi hal yang rutin dilakukan, dengan adanya pekerjaan dari kantor atau yang lain mengharuskan bertemu dengan orang banyak sehingga dapat menyebabkan cepatnya terpapar virus. Terutama saat protokol kesehatan tidak diterapkan (Winogroho, 2021).

Hasil pemeriksaan covid-19 dalam sehari merupakan salah satu kasus data cacahan dengan banyak nilai 0 (*excess zero*). Munculnya nilai 0 terjadi karena hasil pemeriksaan individu negatif (potensial pemerksaan positif). Pengamatan pada variabel respon berbentuk diskrit namun tidak biner sehingga dapat dikatakan sebagai data cacahan yang menyangkut banyaknya kejadian dalam distribusi *Poisson*. Distribusi *Poisson* memiliki asumsi kesamaan rata-rata dengan varian yang disebut *equidispersion*. Satu metode yang digunakan untuk menganalisis data cacahan adalah metode regresi *poisson*. Tetapi metode ini menjadi tidak sinkron apabila terdapat banyak data yang bernilai nol (Afi Nurani, Sutawanir Darwis, 2012).

Ada beberapa metode yang dapat digunakan untuk memodelkan variabel respon dengan *excess zero* yaitu *Zero Inflated Poisson* dan *Hurdle Poisson*. Pemodelan dengan *Zero Inflated Poisson* adalah model *mixture* dengan memodelkan observasi bernilai 0 dengan proses biner serta model independen Poisson untuk observasi bernilai positif. Model *Hurdle Poisson* diasumsikan menjadi dua proses independen fungsi dari variabel respon yaitu model biner untuk observasi bernilai nol serta *Truncated Poisson* untuk observasi bernilai positif. Kedua model ini sama-sama menggunakan metode maximum likelihood untuk memperoleh estimasi parameter yang dipergunakan.

Model *Regresi Hurdle Poisson* adalah model yang terbentuk secara terpisah sehingga lebih mudah dalam pengaplikasian dan interpretasinya. Model *Hurdle Poisson* adalah model untuk data cacahan (*count*) dengan menggunakan dua bagian pendekatan (*two part model*). Pertama ialah model untuk data biner yang bernilai nol atau positif. Kedua ialah *Truncated* model data bernilai positif lebih dari nol (Ruhyat, 2021).

Model *Regresi Hurdle Poisson* merupakan model yang dapat digunakan apabila terjadi *overdispersi* ketika data variabel respon banyak bernilai nol. Model *Regresi Hurdle Poisson* terbagi menjadi dua bagian yakni model data biner yang bernilai nol atau nilai positif yang ditaksir dengan menggunakan model logit dan model untuk data yang bernilai positif saja yang ditaksir dengan model *Truncated Poisson*. *Regresi Hurdle Poisson* kemudian diinterpretasikan dalam memodelkan kasus sehingga diperoleh hasil pengujian secara signifikan terhadap variabel (Julianda H dkk, 2019).

Berdasarkan uraian diatas peneliti bermaksud melakukan penelitian dengan judul “Implementasi Model Regresi Hurdle Poisson (Studi kasus: Faktor-Faktor berpengaruh terhadap gejala terpapar Covid-19 saat pandemi di Puskesmas Baranti Kecamatan Baranti).

1.2 RUMUSAN MASALAH

1. Bagaimana mengetahui variabel yang berpengaruh secara signifikan pada faktor-faktor berpengaruh terhadap gejala terpapar Covid-19 saat pandemi?
2. Bagaimana model *Regresi Hurdle Poisson* yang terjadi saat overdispersi pada faktor-faktor berpengaruh terhadap gejala terpapar Covid-19 saat pandemi?

1.3 TUJUAN PENELITIAN

Tujuan dari penelitian tugas akhir ini antara lain:

1. Mendapatkan variabel yang berpengaruh secara signifikan pada faktor-faktor berpengaruh terhadap gejala terpapar Covid-19 saat pandemi.
2. Mendapatkan Model *Regresi Hurdle Poisson* saat terjadi *overdispersi* pada faktor-faktor berpengaruh terhadap gejala terpapar Covid-19 saat pandemi.

1.4 MANFAAT PENELITIAN

Manfaat yang dapat diambil dari penelitian ini adalah:

1. Bagi penulis
Menambah wawasan penulis terkait metode *Regresi Hurdle Poisson* pada faktor-faktor berpengaruh terhadap gejala terpapar Covid-19 saat pandemi.
2. Bagi mahasiswa matematika
Menambah referensi dalam penerapan ilmu matematika dalam bidang kesehatan.

1.5 BATASAN MASALAH

Data yang digunakan dalam penelitian ini adalah data faktor-faktor berpengaruh terhadap gejala terpapar Covid-19 saat pandemi periode Januari-Juni 2021 di Puskesmas Baranti, Kecamatan Baranti.

1.6 SISTEMATIKA PENULISAN

Pada penelitian ini menggunakan sistem penulisan yang terdiri dari 5 BAB dimana rinciannya akan dijelaskan sebagai berikut:

BAB I : PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, dan batasan penelitian.

BAB II : TINJAUAN PUSTAKA

Pada bab ini menjelaskan tentang teori Distribusi *Poisson*, *Multikolinieritas*, Regresi *Poisson*, *Overdispersi*, model *Regresi Hurdle Poisson*.

BAB III : METODOLOGI PENELITIAN

Pada bab ini membahas tentang metodologi penelitian dan alur kerja.

BAB IV : HASIL DAN PEMBAHASAN

Pada bab ini berisi tentang hasil dan pembahasan yang diperoleh dari masalah penelitian.

BAB V : PENUTUP

Pada bab ini membahas tentang kesimpulan dari hasil dan pembahasan serta saran untuk peneliti selanjutnya.

BAB II

TINJAUAN PUSTAKA

Pada bab ini akan dijelaskan beberapa teori berdasarkan penelitian diantaranya sebagai berikut:

2.1 Distribusi Poisson

Distribusi *Poisson* ialah suatu distribusi yang digunakan untuk kejadian yang memiliki probabilitas kejadiannya kecil, pada kejadian tersebut bergantung pada interval waktu tertentu atau suatu daerah tertentu yang menghasilkan pengamatan berupa variabel diskrit.

Distribusi *Poisson* sering digunakan untuk menentukan peluang sebuah peristiwa yang ada dalam area kesempatan tertentu diharapkan terjadinya sangat jarang (Sudjana, 2016).

Distribusi *Poisson* mempunyai ciri-ciri sebagai berikut :

- a. Banyaknya hasil percobaan yang terjadi dalam suatu interval waktu atau wilayah tertentu, tidak bergantung pada banyaknya hasil percobaan yang terjadi pada interval waktu atau daerah lain yang terpisah.
- b. Probabilitas terjadinya hasil percobaan suatu interval waktu yang singkat atau dalam suatu daerah yang kecil, sebanding dengan panjang interval waktu atau besarnya daerah tersebut dan tidak bergantung pada banyaknya hasil percobaan diluar interval waktu atau daerah tersebut.
- c. Probabilitas lebih dari satu hasil percobaan yang terjadi dalam interval waktu yang singkat atau dalam daerah yang kecil dapat diabaikan.

Rumus fungsi peluang *Distribusi Poisson* adalah :

$$p(x) = P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (2.1)$$

Dengan:

$x = 0,1,2,3 \dots$

$e =$ bilangan konstan (2,7183)

$\lambda =$ rata-rata terjadinya suatu peristiwa (Sudjana,2016)

2.2 Uji Multikolinieritas

Multikolinieritas adalah suatu kondisi dimana terjadi korelasi antara variabel bebas dengan variabel terikat. Besaran yang dapat digunakan untuk mendeteksi adanya multikolinieritas adalah faktor inflasi ragam (*Variance Inflation Factor/VIF*). VIF digunakan sebagai kriteria untuk mendeteksi multikolinieritas pada regresi linear yang melibatkan lebih dari dua variabel bebas. VIF didefinisikan sebagai berikut:

$$VIF = \frac{1}{1 - R_i^2} \quad (2.2)$$

dengan R_i^2 adalah koefisien determinasi antara variabel prediktor. Nilai VIF lebih besar dari 10 menunjukkan multikolinieritas (Sriningsih dkk, 2018).

2.3 Regresi Poisson

Regresi Poisson digunakan untuk menganalisis hubungan antara variabel respon yang berupa data diskrit dengan satu atau lebih variabel prediktor dimana nilai rata-rata dan variansnya diasumsikan sama (*equidispersion*) dalam hal ini respon data tersebut berdistribusi *Poisson* dengan parameter μ (Sari & Sihombing, 2021). Parameter μ ini sangat bergantung pada beberapa unit tertentu atau periode dari waktu, jarak, luas area, volume, dan sebagainya. Distribusi ini kemudian digunakan untuk memodelkan suatu peristiwa yang keberadaannya relatif jarang atau langka untuk terjadi pada satuan unit tertentu (Susilawati dkk, 2022).

Model Regresi Poisson dinyatakan sebagai berikut:

$$y_i = f(\mu_i) + \varepsilon_i ; i = 1, 2, 3, \dots, n \quad (2.3)$$

dimana y_i adalah jumlah kejadian dan μ_i adalah *mean* jumlah kejadian dimana μ_i diasumsikan tidak berubah dari data ke data (Susilawati, 2022). *Regresi Poisson* menggunakan *General Linear Model* (GLM), terdapat sebuah fungsi g yang linear menghubungkan *mean* dari variabel respon dengan variabel prediktor, yaitu:

$$g(\mu_i) = \ln \mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} \quad (2.4)$$

Fungsi g merupakan fungsi penghubung sedangkan hubungan antara mean dan prediktor linear sebagai berikut:

$$\mu_i = g^{-1}(\mu_i) = g^{-1}(x_i^T \beta) \quad (2.5)$$

Model regresi poisson, biasanya berfungsi penghubung yang digunakan adalah fungsi penghubung log, karena mean dari variabel responnya akan berbentuk fungsi eksponensial dan menjamin bahwa nilai variabel yang ditaksir dari variabel respon akan bernilai non-negatif. Fungsi penghubung log sebagai berikut:

$$g(\mu_i) = \ln \mu_i = x_i^T \beta \quad (2.6)$$

Hubungan antara mean variabel respon dan prediktor linear dengan menggunakan fungsi log adalah:

$$\ln \mu_i = x_i^T \beta \quad (2.7)$$

Sehingga fungsi penghubung untuk model regresi poisson sebagai berikut:

$$\begin{aligned} \ln \mu_i &= x_i^T \beta = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \\ \mu_i &= \exp(x_i^T \beta) = \exp \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \end{aligned} \quad (2.8)$$

2.4 Overdispersi

Overdispersi adalah kondisi dimana varians lebih besar nilai rata-rata. Penyebab dari overdispersi yang sering terjadi pada regresi poisson adalah peluang bernilai nol yang berlebih pada peubah Y. Pada Regresi Poisson disyaratkan keadaan yang equidispersi yaitu nilai mean dan varians dari variabel respon sama (Pradawati dkk, 2013). Sebelum membuat model regresi Hurdle Poisson, akan dilihat terlebih dahulu apakah data tersebut mengalami overdispersi atau tidak. Cara untuk mendeteksi *Overdispersi* pada model *Regresi Poisson* adalah dengan menggunakan statistik *deviance*, dimana nilai statistik *deviance* yang dibagi dengan derajat bebasnya. Nilai hasil bagi tersebut harus mendekati 1, jika hasil bagi tersebut lebih dari 1 maka terdapat *Overdispersi*. Langkah-langkah yang akan dilakukan pengujian tersebut adalah sebagai berikut (Julianda H dkk, 2019).

1. Perumusan Hipotesis

$H_0: \hat{\phi} = 0$ (Tidak terdapat *Overdispersi* pada model regresi *Poisson*)

$H_1: \hat{\phi} > 0$ (Terdapat *Overdispersi* pada model regresi *Poisson*)

2. Statistik Uji

Statistik uji yang digunakan adalah:

$$\hat{\phi} = \frac{\text{nilai deviance}}{n - p}$$

$$D = 2 \sum_{i=1}^n \left\{ y_i \log \frac{y_i}{\mu_i} - (y_i - \mu_i) \right\} \quad (2.9)$$

Dengan:

D = nilai deviance

$\hat{\phi}$ = Parameter disperse

y_i = nilai variabel respon dari pengamatan ke- i

μ_i = taksiran rata-rata kanyak kasus ke- i pada model *regresi Poisson*

n = banyaknya pengamatan

p = banyaknya parameter

3. Kriteria Pengujian

Dengan mengambil taraf nyata sebesar α , maka H_0 ditolak jika $\hat{\phi} > 1$.

4. Kesimpulan

Interpretasi ditolak atau diterimanya H_0 .

2.5 Model Regresi Hurdle Poisson

Model Hurdle Poisson merupakan model untuk sebuah data cacahan dengan menggunakan dua bagian pendekatan (two part model). Model Hurdle Poisson dapat mengatasi permasalahan excess zero yaitu dengan membagi dua model kedalam dua bagian yaitu pada model biner yang memiliki nilai nol akan ditaksir dengan model logit, dan untuk model yang datanya bernilai positif akan ditaksir dengan model Truncated Poisson (Alwani & Achmad, 2017).

Penggunaan model *Hurdle* yang paling penting adalah ketika datanya bersifat diskrit (data cacahan). Salah satu pendekatan untuk mengatasi masalah *overdispersi* adalah model *Hurdle*. Overdispersi terjadi ketika nilai varian lebih besar dari nilai mean. Salah satu penyebab terjadi *overdispensi* yaitu terjadi terdapat banyak nilai nol (*excess zero*) pada variabel respon.

Model *Hurdle* mampu mengatasi kasus-kasus *excess zeros* dengan cara membagi dua model ke dalam dua bagian yaitu:

1. Model untuk data biner yang bernilai nol (*zero count*) atau nilai positif (*positive counts*), dimana model tersebut ditaksir dengan model logit.
2. Model untuk data yang bernilai positif (*positive counts*) saja, dimana model tersebut ditaksir dengan model *Truncated Poisson*.

Misalkan $k_1(0)$ adalah peluang ketika variabel respon (Y) bernilai sama dengan 0 dan $k_1(y), y = 1, 2, 3, \dots$ adalah sebuah fungsi peluang ketika variabel respon (Y) bernilai positif (Saffari dkk, 2012). Sehingga fungsi peluang dari model *Hurdle* adalah sebagai berikut:

$$P(Y_i = y) = \begin{cases} k_1(0) & , y = 0 \\ (1 - k_1(0))k_2(y) & , y = 1, 2, \dots \end{cases} \quad (2.10)$$

Kedua model *Hurdle* yang terbentuk pada fungsi peluang untuk nilai non-negatif. Misalkan f_1 dan f_2 dengan melihat bentuk umum *Hurdle* pada persamaan 2.10, maka

$$k_1(0) = f_1 = 0 \text{ dan } k_2(y) = \frac{f_2(y)}{(1 - f_2(0))}. \quad (2.11)$$

Pada kasus k_2 akan dilakukan normalitas, karena f_2 berlaku untuk nilai non negatif ($y = 0, 1, 2, \dots$) sedangkan k_2 berlaku untuk nilai positif ($y = 1, 2, \dots$). Ini berarti perlu dilakukan untuk menghitung fungsi peluang f_2 .

$$f_2(0) + f_2(1) + f_2(2) + \dots = 1, \quad y = 0, 1, 2, \dots$$

$$f_2(1) + f_2(2) + \dots = 1 - f_2(0), \quad y = 1, 2, 3, \dots$$

$$\frac{f_2(1)}{1 - f_2(0)} + \frac{f_2(2)}{1 - f_2(0)} + \dots = 1, \quad y = 1, 2, 3, \dots$$

$$\frac{f_2(y)}{1 - f_2(0)}, \quad y = 1, 2, \dots$$

Maka distribusi peluang dari model *Hurdle* adalah sebagai berikut:

$$P(y = 0) = f(y = 0) = f_1(0)$$

$$P(y = y) = f(y = y) = \frac{1 - f_1(0)}{1 - f_2(0)} f_2(y) = \theta f_2(y), y = 1, 2, 3, \dots \quad (2.12)$$

Berdasarkan persamaan 2.12 maka dapat ditulis distribusi peluang dari model *Hurdle* sebagai berikut:

$$P(Y_i = y) = \begin{cases} f_1(0), & y = 0 \\ \theta f_2(y), & y = 1, 2, 3, \dots \end{cases} \quad (2.13)$$

Model *Hurdle Poisson* digunakan untuk menjelaskan hubungan antara variabel respon yang berupa data cacahan dengan variabel prediktor yang berupa kontinu, diskrit atau campuran antara keduanya (Faidah & Pontoh, 2015). Misalkan $Y_i, i = 1, 2, 3, \dots, n$ adalah sebuah variabel acak non negatif, dan misalkan $Y_i = 0$ adalah observasi dengan frekuensi nilai 0 yang terlalu banyak, sehingga tidak bisa ditangani dengan menggunakan model Regresi Poisson biasa. Model Regresi *Hurdle Poisson* dengan variabel respon $Y_i, (i=1, 2, 3, \dots, n)$ memiliki distribusi sebagai berikut (Saffari dkk, 2012).

$$P(Y_i = y_i) = \begin{cases} 1 - \pi_i, & y_i = 0 \\ (\pi_i) \frac{e^{-\mu_i} \mu_i^{y_i}}{(1 - \pi_i) y_i!}, & y_i > 0 \end{cases} \quad (2.14)$$

dengan $i = 1, 2, 3 \dots n$; $0 < 1 - \pi_i < 1$ dan. Pada persamaan 2.14 selanjutnya terbagi kedalam model, yaitu:

1. Model Logit

Model logit adalah model untuk data biner yang bernilai nol atau nilai positif. Model logit dapat dituliskan sebagai berikut:

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \sum_{j=1}^p z_{ij} \delta_j \quad (2.15)$$

Dengan:

p = jumlah variabel independen

z = vektor kovariat pada variabel prediktor

$z_i = [z_{i1}, z_{i2}, \dots, z_{ip}]$

δ adalah vektor kolom parameter koefisien regresi untuk model logit

$\delta = [\delta_1 \delta_2 \dots \delta_p]^T$

2. Model *Truncated Poisson*

Model *Truncated Poisson* adalah model untuk data bernilai positif saja. Model *Truncated Poisson* dapat ditulis sebagai berikut:

$$\log(\mu_i) = \sum_{j=1}^p x_{ij}\beta_j$$

atau

$$\mu_i = \exp\left(\sum_{j=1}^p x_{ij}\beta_j\right) \quad (2.16)$$

Dengan:

p = jumlah variabel independen

x = vektor kovariat pada variabel prediktor

$x_i = [x_{i1}, x_{i2}, \dots, x_{ip}]$

β adalah vektor kolom parameter koefisien regresi untuk model *truncated poisson*

$\beta = [\beta_1 \beta_2 \dots \beta_p]^T$

Model peluang *Hurdle poisson* yang terbentuk dari kombinasi model logit untuk observasi bernilai 0 dan model *Truncated Poisson* untuk data yang bernilai positif adalah:

$$P(Y_i = y_i) = \begin{cases} \frac{1}{1 + \exp(\sum_{j=1}^p z_{ij}\delta_j)} & , \text{jika } y_i = 0 \\ \left[\frac{\exp(\sum_{j=1}^p z_{ij}\delta_i)}{1 + \exp(\sum_{j=1}^p z_{ij}\delta_i)} \right] \left[\frac{[\exp(\sum_{j=1}^p x_{ij}\beta_j)]}{\exp(\exp(\sum_{j=1}^p x_{ij}\beta_j)) - 1} \right] |y_i| & , \text{jika } y_i > 0 \end{cases} \quad (2.17)$$

Penaksiran parameter $\delta_j\beta$ pada model regresi hurdle poisson dilakukan dengan menggunakan metode peluang Maksimum. Berikut fungsi peluang dari model regresi Hurdle Poisson.

$$L(\delta, \beta) = \prod P(Y_i = y_i)$$

$$L(\delta, \beta) = I_{y_i=0} \frac{1}{1 + \exp \sum_{j=1}^p z_{ij} \delta_j} I_{y_i>0} \left[\frac{\exp \sum_{j=1}^p z_{ij} \delta_j}{1 + \exp \sum_{j=1}^p z_{ij} \delta_j} \right] \left[\frac{[\exp \sum_{j=1}^p x_{ij} \beta_j]^{y_i}}{[\exp(\exp(\sum_{j=1}^p x_{ij} \beta_j))] y_i!} \right]$$

Metode peluang maksimum tidak menghasikan taksiran parameter secara eksplisit, tetapi memunculkan persamaan non linear yang kompleks dan cukup rumit sehingga diperlukan algoritma khusus untuk mendapatkan taksiran parameter. Algoritma yang digunakan adalah *Broyden-Fletcher-Goldfarb-Shanno*. Algoritma ini merupakan algoritma yang dapat menyelesaikan optimasi non linear tanpa kendala. Metode BFGH merupakan pengembangan dari metode Newton. Metode Newton menggunakan turunan kedua untuk menghitung perubahan nilai parameter untuk setiap iterasi kemudian dimodifikasi dengan penambahan *step legh* untuk memberikan jaminan konvergen apabila turunan kedua dibutuhkan perhitungan kompleks.

2.6 Pengujian Parameter

2.6.1 Pengujian Serentak (*Uji Rasio Likelihood*)

Pengujian keseluruhan model regresi hurdle poisson dilakukan dengan menggunakan *Likelihood Ratio Test* (LRT). Langkah-langkah hipotesis parameter dapat dituliskan sebagai berikut:

Hipotesis:

$H_0: \delta_1 = \delta_2 = \dots = \delta_k = \beta_1 = \beta_2 = \dots = \beta_k = 0$ (semua variabel prediktor tidak berpengaruh)

$H_1: \text{minimal ada satu } \delta_j \neq 0 \text{ atau } \beta_j \neq 0, j = 1, 2, \dots, k$ (paling sedikit ada satu variabel prediktor dalam model yang berpengaruh)

Uji statistik:

$$G^2 = -2(\ln L_0 - \ln L_1)$$

Dengan:

L_0 = nilai likelihood dari model tanpa variabel prediktor

L_1 = nilai likelihood dari model dengan variabel prediktor

Uji kriteria:

H_0 ditolak jika $G^2 > x_{k;\alpha}^2$ atau p - value $< \alpha$ dengan k merupakan banyaknya parameter dibawah H_0 .

2.6.2 Pengujian Parsial (Uji Wald)

a. Pengujian signifikansi parameter model logit $\pi_i = x_i^T \delta$ dengan $i = 1, \dots, n$

Hipotesis:

$$H_0: \delta_j = 0$$

$$H_1: \delta_j \neq 0, \text{ dengan } j = 1, 2, \dots, k$$

Uji statistik:

$$T_{hitung} = \frac{\hat{\delta}_j}{SE(\hat{\delta}_j)}$$

Uji kriteria:

$$H_0 \text{ ditolak jika } |T_{hitung}| > T_{1-\alpha/2} \text{ atau } p\text{-value} < \alpha$$

b. Pengujian signifikansi parameter model $\ln(\mu_i) = x_i^T \beta$ dengan $i = 1, \dots, n$

Hipotesis:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, k$$

Uji statistik:

$$T_{hitung} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

Uji kriteria:

H_0 ditolak jika $|T_{hitung}| > T_{1-\alpha/2}$ atau $p - value < \alpha$