

**SKRIPSI**

**ANALISIS REGRESI POISSON, REGRESI BINOMIAL  
NEGATIF DAN REGRESI HURDLE BINOMIAL NEGATIF  
PADA DATA OVERDISPERSI**

(STUDI KASUS : JUMLAH KASUS KEMATIAN PENDERITA DEMAM BERDARAH  
DENGUE DI PROVINSI JAWA TIMUR TAHUN 2022)



STATE ISLAMIC UNIVERSITY  
**SUNAN KALIJAGA**  
YOGYAKARTA

**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
YOGYAKARTA**

**2024**

**ANALISIS REGRESI POISSON, REGRESI BINOMIAL  
NEGATIF DAN REGRESI HURDLE BINOMIAL NEGATIF  
PADA DATA OVERDISPERSI**

(STUDI KASUS : JUMLAH KASUS KEMATIAN PENDERITA DEMAM BERDARAH

DENGUE DI PROVINSI JAWA TIMUR TAHUN 2022)

Skripsi

Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh Derajat  
Sarjana Matematika  
Program Studi Matematika



SALMAN AL-FARISI

19106010037

STATE ISLAMIC UNIVERSITY  
PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
YOGYAKARTA

**2024**

**SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR**

Hal : Persetujuan Skripsi / Tugas Akhir

Lamp :

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

*Assalamu'alaikum wr. wb.*

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Salman Al-Farisi

NIM : 19106010037

Judul Skripsi : Analisis Regresi Poisson, Regresi Binomial Negatif dan Regresi Hurdle Binomial Negatif Pada Data Overdispersi

sudah dapat diajukan kembali kepada Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Matematika.

Dengan ini kami mengharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqasyahkan. Atas perhatiannya kami ucapan terima kasih.

*Wassalamu'alaikum wr. wb.*

STATE ISLAMIC UNIVERSITY  
**SUNAN KALIJAGA**  
YOGYAKARTA

Yogyakarta, 23 Juli 2024

Pembimbing I

Pembimbing II

Sri Utami Zuliana S.Si., M.Sc., Ph.D

NIP. 19741003 200003 2 002

Malahayati, S.Si., M.Sc

NIP. 19840412 201101 2 010



KEMENTERIAN AGAMA  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-1464/Un.02/DST/PP.00.9/08/2024

Tugas Akhir dengan judul : ANALISIS REGRESI POISSON, REGRESI BINOMIAL NEGATIF, DAN REGRESI HURDLE BINOMIAL NEGATIF PADA DATA OVERDISPERSI

yang dipersiapkan dan disusun oleh:

Nama : SALMAN AL-FARISI  
Nomor Induk Mahasiswa : 19106010037  
Telah diujikan pada : Jumat, 02 Agustus 2024  
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Sri Utami Zuliana, S.Si., M.Sc., Ph.D.  
SIGNED

Valid ID: 66bc097d25b10



Pengaji I

Malahayati, S.Si., M.Sc  
SIGNED

Valid ID: 66bae9b3e4bf9



Pengaji II

Mohammad Farhan Qudratullah, S.Si., M.Si  
SIGNED

Valid ID: 66b595625d513



Yogyakarta, 02 Agustus 2024

UIN Sunan Kalijaga  
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.  
SIGNED

Valid ID: 66c2de9a3e94f

## SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Salman Al-Farisi

NIM : 19106010037

Program Studi : Matematika

Fakultas : Sains dan Teknologi

Dengan ini menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sesungguhnya skripsi ini merupakan hasil pekerjaan penulis sendiri sepanjang pengetahuan penulis, bukan duplikasi atau saduran dari karya orang lain kecuali bagian tertentu yang penulis ambil sebagai bahan acuan. Apabila terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab penulis.

Yogyakarta, 26 Juli 2024



Salman Al-Farisi

STATE ISLAMIC UNIVERSITY  
**SUNAN KALIJAGA**  
YOGYAKARTA

## **PERSEMBAHAN**

KARYA SEDERHANAINI PENULIS PERSEMBAHKAN KEPADA:

KEDUA ORANG TUA, PAK DARSONO, BU YUNI DAN DIK NOURA YANG  
TELAH MEMBERIKAN DO'A, DAN DUKUNGAN YANG TAK TERHINGGA.  
KERABAT DEKAT, SAHABAT, TEMAN DAN SEMUA ORANG-ORANG YANG  
PERNAH SAYA KENAL.

ALMAMATER UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA YOGYAKARTA  
YANG TELAH BANYAK MENANAMKAN ILMU.



## MOTTO

*"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya..."*

(QS. Al Baqarah: 286)

*"Menuntut ilmu adalah taqwa, menyampaikan ilmu adalah ibadah, mengulang-ulang ilmu adalah zikir, mencari ilmu adalah jihad."*

(Imam Ghazali)



## PRAKATA

*Assalamu'alaikum Wr. Wb. Allhamdulillahirabbil'alamin*, puji syukur kehadirat Allah SWT yang telah melimpahkan nikmat, rahmat dan hidayah-Nya. Shalawat serta salam semoga selalu tercurahkan kepada junjungan kita, Nabi Muhammad SAW. Berkat rahmat Allah yang Maha Kuasa, penulis dapat menyelesaikan skripsi ini yang berjudul **"Analisis Regresi Poisson, Regresi Binomial Negatif dan Regresi Hurdle Binomial Negatif Pada Data Overdispersi"**.

Penulis ingin mengucapkan terima kasih kepada Ibu Sri Utami Zuliana, S.Si., M.Sc., Ph.D., selaku pembimbing I dan Ibu Malahayati, S.Si., M.Sc., selaku pembimbing II yang telah meluangkan waktu dan tenaga untuk memberikan saran dan bimbingan serta masukannya kepada penulis. Penulis menyadari bahwa proses penulisan skripsi ini terdapat banyak hambatan dan rintangan. Namun berkat adanya bimbingan, motivasi, dan dorongan dari berbagai pihak, *alhamdulillah* skripsi ini dapat terselesaikan. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Prof. Dr. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
2. Bapak Muchammad Abrori, S.Si.,M.Kom., selaku Ketua Program Studi Matematika.
3. Ibu Malahayati, S.Si.,M.Sc., selaku dosen pembimbing akademik dan dosen pembimbing skripsi II yang telah memberikan pengarahan kepada penulis

selama menempuh pendidikan di Universitas Islam Negeri Sunan Kalijaga Yogyakarta.

4. Ibu Sri Utami Zuliana, S.Si.,M.Sc.,Ph.D., selaku dosen pembimbing skripsi I yang telah meluangkan waktu, tenaga, pikiran dan kesabaran dalam membimbing penulis untuk menyusun skripsi ini hingga dapat terselesaikan dengan baik.
5. Seluruh dosen dan staf fakultas Sains dan Teknologi yang telah memberikan ilmu dan pelayanan yang terbaik kepada penulis.
6. Kedua orang tua penulis, Pak Darsono, dan Bu Yuni serta Dik noura yang telah banyak memberikan doa, kasih sayang, nasehat dan dukungannya baik berupa materi maupun non materi.
7. Keluarga besar mahasiswa Matematika UIN Sunan Kalijaga, khususnya angkatan 2019.
8. Teman-teman KKN 108 kelompok 78 Pule, Lum'ah, Fransiska, Diah, Idwar, La Afrina, Lathifah, Putri, Rizal, Umay dan Vendy atas kerjasamanya selama KKN di Gunung Kidul.
9. Terakhir semua pihak yang telah memberikan dorongan penulis hingga terselesaikan skripsi ini.

Penulis berharap skripsi ini dapat memberikan manfaat bagi penulis pada khususnya dan bagi pembaca pada umumnya. Penulis merasakan bahwa masih banyak kekurangan dalam penulisan skripsi ini, oleh karena itu perlu adanya kritik dan saran yang membangun untuk penulis demi mengembangkan penelitian.

Yogyakarta, 23 Juli 2024

Penulis

Salman Al-Farisi

NIM. 19106010037



## DAFTAR ISI

|  |              |
|--|--------------|
| <b>HALAMAN JUDUL</b> .....             | <b>i</b>     |
| <b>SURAT PERSETUJUAN</b> .....         | <b>ii</b>    |
| <b>SURAT PENGESAHAN</b> .....          | <b>iii</b>   |
| <b>SURAT PERNYATAAN KEASLIAN</b> ..... | <b>iv</b>    |
| <b>HALAMAN PERSEMBAHAN</b> .....       | <b>v</b>     |
| <b>HALAMAN MOTTO</b> .....             | <b>vi</b>    |
| <b>PRAKATA</b> .....                   | <b>vii</b>   |
| <b>DAFTAR ISI</b> .....                | <b>x</b>     |
| <b>DAFTAR TABEL</b> .....              | <b>xiii</b>  |
| <b>DAFTAR GAMBAR</b> .....             | <b>xiv</b>   |
| <b>DAFTAR LAMBANG</b> .....            | <b>xv</b>    |
| <b>INTISARI</b> .....                  | <b>xvii</b>  |
| <b>ABSTRACT</b> .....                  | <b>xviii</b> |
| <b>I PENDAHULUAN</b> .....             | <b>1</b>     |
| 1.1. Latar Belakang Masalah .....      | 1            |
| 1.2. Batasan Masalah .....             | 4            |
| 1.3. Rumusan Masalah .....             | 4            |
| 1.4. Tujuan Penelitian .....           | 4            |
| 1.5. Manfaat Penelitian .....          | 5            |
| 1.6. Tinjauan Pustaka .....            | 5            |
| 1.7. Sistematika Penulisan .....       | 7            |
| <b>II DASAR TEORI</b> .....            | <b>9</b>     |
| 2.1. Distribusi Normal .....           | 9            |

|  |           |
|--|-----------|
| 2.2. Analisis Regresi . . . . .  | 11        |
| 2.3. Regresi Poisson . . . . .   | 13        |
| 2.4. Estimasi Parameter Regresi Poisson . . . . .  | 17        |
| 2.5. Data Overdispersi . . . . .   | 19        |
| 2.6. Regresi Binomial Negatif . . . . .  | 21        |
| 2.7. Estimasi Parameter Regresi Binomial Negatif . . . . .   | 23        |
| 2.8. Regresi Hurdle Binomial Negatif . . . . .   | 25        |
| 2.9. Estimasi Parameter Regresi Hurdle Binomial Negatif . . . . .  | 27        |
| 2.10. Uji Hipotesis Parameter Regresi Poisson, Regresi Binomial Negatif<br>dan Regresi Hurdle Binomial Negatif . . . . . | 29        |
| 2.11. Pemilihan Model Terbaik . . . . .  | 31        |
| 2.12. Faktor-faktor yang Mempengaruhi Jumlah Kasus Kematian Pende-<br>rita Demam Berdarah Dengue (DBD) . . . . .         | 31        |
| <b>III METODE PENELITIAN . . . . .</b>   | <b>35</b> |
| 3.1. Jenis dan Sumber Data . . . . .   | 35        |
| 3.2. Variabel Penelitian . . . . .   | 35        |
| 3.3. Langkah-langkah Analisis . . . . .  | 35        |
| 3.4. Software yang Digunakan . . . . .   | 36        |
| 3.5. Flowchart Tahapan Penelitian . . . . .  | 37        |
| <b>IV HASIL DAN PEMBAHASAN . . . . .</b>   | <b>38</b> |
| 4.1. Statistik Deskriptif . . . . .  | 38        |
| 4.2. Uji Distribusi Poisson . . . . .  | 39        |
| 4.3. Uji Multikolinearitas . . . . .   | 40        |
| 4.4. Pemodelan Regresi Poisson . . . . .   | 41        |
| 4.5. Uji Overdispersi . . . . .  | 44        |
| 4.6. Pemodelan Regresi Binomial Negatif . . . . .  | 45        |

|  |           |
|--|-----------|
| 4.7. Pemodelan Regresi Hurdle Binomial Negatif . . . . . | 47        |
| 4.8. Pemilihan Model Terbaik . . . . .                   | 51        |
| 4.9. Interpretasi Model . . . . .                        | 51        |
| <b>V PENUTUP . . . . .</b>                               | <b>54</b> |
| 5.1. Kesimpulan . . . . .                                | 54        |
| 5.2. Saran . . . . .                                     | 55        |
| <b>LAMPIRAN . . . . .</b>                                | <b>60</b> |
| <b>A Data Penelitian . . . . .</b>                       | <b>60</b> |
| <b>B Tabel Distribusi T . . . . .</b>                    | <b>63</b> |
| <b>C Tabel Kolmogorov-Smirnov . . . . .</b>              | <b>64</b> |
| <b>D Tabel Chi Square . . . . .</b>                      | <b>65</b> |
| <b>E Source Code Program R . . . . .</b>                 | <b>66</b> |
| <b>Curriculum Vitae . . . . .</b>                        | <b>74</b> |



## **DAFTAR TABEL**

|     |  |    |
|-----|--|----|
| 4.1 | Statistik Deskriptif . . . . .                                     | 38 |
| 4.2 | Nilai <i>Variance Inflation Factor</i> (VIF) . . . . .             | 41 |
| 4.3 | Hasil Estimasi Parameter Regresi Poisson . . . . .                 | 42 |
| 4.4 | Hasil Estimasi Parameter Regresi Binomial Negatif . . . . .        | 45 |
| 4.5 | Hasil Estimasi Parameter Regresi Hurdle Binomial Negatif . . . . . | 47 |
| 4.6 | Nilai AIC . . . . .  | 51 |
| 4.7 | Perbandingan Pengaruh Variabel Prediktor . . . . .                 | 52 |



## **DAFTAR GAMBAR**

|     |                                     |    |
|-----|-------------------------------------|----|
| 2.1 | Kurva Distribusi Normal . . . . .   | 10 |
| 2.2 | Kurva Distribusi Poisson . . . . .  | 15 |
| 2.3 | Kurva Distribusi Binomial . . . . . | 22 |
| 3.1 | Flowchart Penelitian . . . . .      | 37 |



## DAFTAR LAMBANG

$x$  : Variabel prediktor/independen

$y$  : Variabel respon/dependen

! : Faktorial

$\beta$  : Parameter regresi

$f$  : Fungsi

$g$  : Fungsi penghubung

$\ln$  : Logaritma natural

$\sum$  : Penjumlahan total

$e$  : Basis logaritma natural

$\exp$  : Eksponensial

$L$  : *Likelihood*

$\infty$  : Tak terhingga

$\sigma^2$  : Variansi

$\pi$  : Rasio keliling lingkaran

$\mu$  : Rata-rata

$\bar{\mu}$  : Rata-rata

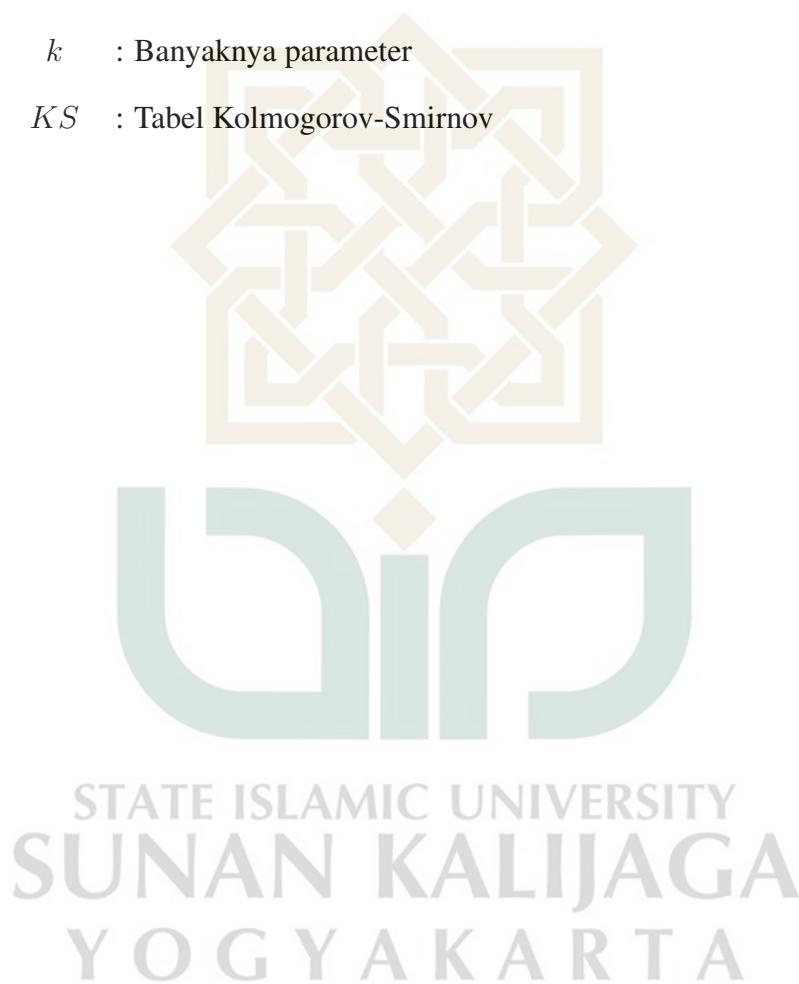
$\sigma$  : Simpangan baku

$R^2$  : Koefisien determinasi

$\log$  : Logaritma

$\prod$  : Perkalian total

- $G^2$  : Deviance  
 $\alpha$  : Parameter dispersi  
 $\Gamma$  : Gamma  
 $w_i$  : Peluang keadaan nol  
logit : Fungsi model regresi logistik  
 $k$  : Banyaknya parameter  
 $KS$  : Tabel Kolmogorov-Smirnov



## INTISARI

# ANALISIS REGRESI POISSON, REGRESI BINOMIAL NEGATIF DAN REGRESI HURDLE BINOMIAL NEGATIF PADA DATA OVERDISPERSI

(STUDI KASUS : JUMLAH KASUS KEMATIAN PENDERITA DEMAM BERDARAH

DENGUE DI PROVINSI JAWA TIMUR TAHUN 2022)

Oleh

SALMAN AL-FARISI

19106010037

Regresi Poisson bertujuan untuk memodelkan hubungan variabel prediktor dengan variabel respon yang harus berupa data diskrit. Asumsi yang harus terpenuhi adalah equidispersi yaitu nilai variansi variabel respon sama dengan rata-ratanya. Namun, seringkali asumsi ini dilanggar yaitu nilai variansi lebih besar dari rata-ratanya, hal ini disebut overdispersi. Untuk mengatasinya, dapat dilakukan dengan regresi Binomial Negatif dan Hurdle Binomial Negatif (HBN). Estimasi parameter ditentukan menggunakan pendekatan *Maximum Likelihood Estimation* (MLE). Model terbaik dipilih menggunakan metode *Akaike Information Criterion* (AIC). Tiga analisis regresi tersebut diaplikasikan pada data jumlah kematian penderita demam berdarah dengue di Provinsi Jawa Timur tahun 2022 dengan beberapa faktor yang mungkin mempengaruhinya. Berdasarkan hasil perhitungan AIC, model yang memiliki kemampuan prediksi terbaik adalah regresi Binomial Negatif sebesar 191,94. Model regresi Binomial Negatif yang diperoleh adalah  $\mu_i = \exp(6.47067 + 0,25272X_1 - 0,04879X_5 - 0,10942X_6)$

**Kata Kunci:** Regresi Poisson, Binomial Negatif, HBN, Overdispersi, MLE, AIC

## ABSTRACT

# POISSON REGRESSION ANALYSIS, NEGATIVE BINOMIAL REGRESSION AND NEGATIVE HURDLE BINOMIAL REGRESSION ON OVERDISPERSED DATA

(CASE STUDY: DENGUE FEVER DEATH TOLL IN EAST JAVA PROVINCE 2022)

By

SALMAN AL-FARISI

19106010037

The Poisson regression aims to illustrates the connection between the predicting variable and the response variable which has to be a discreet data. The required assumption is equidispersion, which is the response variable variant that is equal to the mean. Yet, this assumption is often violated, in which the variant value is larger than the mean, this is called overdispersion. To solve this, the Binomial Negative regression and the Hurdle Negative Binomial (HNB) can be done. The parameter is estimated using the Maximum Likelihood Estimation (MLE) approach. The best model is chosen through the Akaike Information Criterion (AIC) method. All three regression analyses are applied to the Dengue Fever Death Toll in East Java Province 2022 with several factors that may affect it. Based on the AIC calculations, the model that has the best predicting ability is the Negative Binomial regression which is 191,94. The Negative Binomial regression model obtained is  $\mu_i = \exp(6.47067 + 0,25272X_1 - 0.04879X_5 - 0.10942X_6)$

**Keyword:** Poisson regression, Negative Binomial, HNB, Overdispersion, MLE, AIC

## BAB I

### PENDAHULUAN

#### 1.1. Latar Belakang Masalah

Analisis regresi adalah metode untuk menentukan hubungan antara satu variabel respon dengan satu atau beberapa variabel prediktor. Analisis regresi sering diterapkan untuk memprediksi/memperkirakan suatu peristiwa atau fenomena tertentu. Data yang digunakan yaitu data diskrit atau data kontinu. Data diskrit yaitu data yang bisa dihitung atau dicacah. Sedangkan, data kontinu yaitu data yang nilainya berada pada interval tertentu (Supandi, 2020).

Analisis regresi dibagi menjadi tiga macam, yakni regresi parametrik, non parametrik, dan semi parametrik (komponen analisis regresi parametrik serta non parametrik). Perbedaan utama analisis regresi parametrik dengan nonparametrik adalah data analisis regresi parametrik memerlukan asumsi terbatas dan mengikuti prinsip distribusi normal, jika data tidak memenuhi asumsi tersebut maka dapat menggunakan analisis regresi nonparametrik.

Data variabel acak kontinu, umumnya digunakan pada analisis regresi untuk menganalisis data. Namun ada juga data yang dianalisis berbentuk data diskrit. Model regresi Poisson merupakan salah satu model regresi yang menggunakan data diskrit. Model ini pertama kali diperkenalkan oleh Simeon-Denis Poisson (Poisson, 1837). Tujuannya untuk menganalisis hubungan antara variabel respon yang berbasis data diskrit dengan satu atau beberapa variabel prediktor.

Regresi Poisson termasuk regresi parametrik dan non linear yang variabel

responnya diasumsikan memenuhi asumsi equidispersi (Ruliana et al., 2016). Equidispersi yaitu nilai rata-rata variabel respon sama dengan nilai variansinya. Namun, asumsi ini terkadang menyimpang saat menganalisis data diskrit dengan analisis regresi Poisson. Penyimpangan tersebut terletak pada nilai variansi variabel respon lebih besar dari rata-ratanya. Hal ini sering disebut sebagai overdispersi.

Jika terjadi overdispersi pada data diskrit dan tetap menggunakan model regresi Poisson maka menyebabkan pentaksiran parameter pada *standar error* terlalu kecil sehingga menjadi *underestimate* (McCullagh, 1989). Hal ini dapat menyebabkan interpretasi model menjadi tidak valid. Untuk mengetahui apakah terjadi overdispersi atau tidak pada data, yaitu dengan cara membagi nilai *deviance* dengan derajat bebasnya. Apabila hasil bagi bernilai lebih dari satu, maka terindikasi adanya overdispersi.

Solusi alternatif untuk menangani masalah data overdispersi yaitu dengan menggunakan regresi Binomial Negatif dan Hurdle Binomial Negatif (HBN). Regresi Binomial Negatif dan HBN merupakan alternatif untuk menangani overdispersi pada data diskrit yang tidak memenuhi asumsi kesamaan antara nilai rata-rata dengan variansi dari variabel respon. Regresi Binomial Negatif dan HBN mengasumsikan bahwa variabel respon berdistribusi Binomial Negatif. Regresi Binomial Negatif pertama kali dikembangkan oleh John Nelder dan R.W.M. Wedderburn (Nelder & Wedderburn, 1972). Sedangkan untuk regresi HBN pertama kali diperkenalkan oleh John Muhally (Mullahy, 1986).

Regresi Poisson, Binomial Negatif dan HBN dapat diaplikasikan pada beberapa kasus dalam berbagai bidang, seperti kesehatan, iklim, pendidikan dan sebagainya. Salah satunya dalam bidang kesehatan terkait kasus jumlah penduduk terinfeksi penyakit demam berdarah *dengue* (DBD). Demam berdarah adalah suatu

penyakit infeksi virus yang ditularkan melalui nyamuk dan dapat berakibat fatal, sehingga perlu diwaspadai dan diatasi segera. Dalam beberapa dekade terakhir, insidensi penyakit dengue telah meningkat secara signifikan di seluruh dunia. World Health Organization (WHO) telah menempatkan DBD sebagai salah satu ancaman kesehatan global, bersama dengan 9 penyakit lainnya yang memerlukan perhatian khusus.

Dalam kurun waktu tahun 1968 hingga 2009, WHO merekam Indonesia merupakan negara tertinggi dengan jumlah kasus DBD di Asia Tenggara. Pada tahun 1968 DBD pertama kali ditemukan di Surabaya, 58 orang terinfeksi dan 24 orang meninggal dunia (Sukohar, 2014). Pada tahun 2021, jumlah kasus DBD di Indonesia mengalami penurunan signifikan sebesar 32,12 persen dibandingkan tahun sebelumnya menjadi 73.518 kasus. Akan tetapi pada tahun 2022 mengalami peningkatan drastis sebesar 94,8 persen menjadi 143.184 kasus. Provinsi Jawa Timur berada di urutan tertinggi nomor 2 dengan jumlah penderita DBD tertinggi tahun 2022.

Analisis regresi akan diaplikasikan pada data jumlah kasus kematian penderita DBD di Provinsi Jawa Timur tahun 2022 beserta beberapa faktor yang dipercirakan dapat mempengaruhinya. Setelah hasil model didapat, akan dilakukan evaluasi model untuk mencari model yang terbaik. Salah satu cara untuk mengevaluasi kualitas suatu model yaitu dengan *Akaike Information Criterion* (AIC) yaitu suatu metode matematis untuk mengevaluasi seberapa baik suatu model yang cocok dengan data yang diamati. AIC digunakan untuk membandingkan kualitas beberapa model yang dihasilkan dari data yang sama. Berdasarkan uraian diatas penulis mengambil judul “*Analisis Regresi Poisson, Binomial Negatif dan Hurdle Binomial Negatif Pada Data Overdispersi*” dengan studi jumlah kasus kematian penderita

DBD di Provinsi Jawa Timur 2022.

### **1.2. Batasan Masalah**

Dalam penelitian ini, beberapa batasan masalah yang digunakan adalah:

1. Pemilihan model terbaik dilakukan berdasarkan nilai *Akaike Information Criterion* (AIC) yang terkecil.
2. Software yang digunakan dalam penelitian ini adalah bahasa pemrograman R, dengan versi yang digunakan adalah 4.3.2.

### **1.3. Rumusan Masalah**

Beberapa rumusan masalah yang relevan dapat diidentifikasi sebagai berikut:

1. Bagaimana tahapan analisis regresi Poisson pada kasus kematian penderita DBD?
2. Bagaimana tahapan analisis regresi Binomial Negatif pada kasus kematian penderita DBD?
3. Bagaimana tahapan analisis regresi HBN pada kasus kematian penderita DBD?
4. Bagaimana menentukan model terbaik antara regresi Poisson, Binomial Negatif dan HBN dengan menggunakan optimasi AIC?
5. Bagaimana faktor-faktor yang mempengaruhi kasus kematian penderita DBD?

### **1.4. Tujuan Penelitian**

Tujuan penelitian ini adalah sebagai berikut:

1. Untuk mengkaji tahapan analisis regresi Poisson pada kasus kematian penderita DBD.
2. Untuk mengkaji tahapan analisis regresi Binomial Negatif pada kasus kematian penderita DBD.
3. Untuk mengkaji tahapan analisis regresi HBN pada kasus kematian penderita DBD.
4. Untuk mengetahui penentuan model terbaik antara regresi Poisson, Binomial Negatif dan HBN dengan menggunakan optimasi AIC.
5. Untuk mengetahui faktor-faktor yang mempengaruhi kasus kematian penderita DBD.

### **1.5. Manfaat Penelitian**

Dalam penyusunan penelitian ini, beberapa manfaat yang diharapkan adalah sebagai berikut:

1. Penelitian ini bertujuan untuk memberikan pemahaman tentang analisis regresi Poisson, Binomial Negatif, dan Hurdle Binomial Negatif serta cara menentukan model terbaik.
2. Penelitian ini dapat dijadikan sebagai referensi dan bahan pengembangan untuk penelitian selanjutnya bagi pembaca.

### **1.6. Tinjauan Pustaka**

Penelitian (Budiharti, 2021) mempelajari tentang metode *flexibly shaped spatial scan statistic* untuk mendeteksi *hotspot* jumlah penderita kusta di Jawa Barat, serta regresi Binomial Negatif dan Poisson untuk memodelkan jumlah penderita

kusta dan beberapa faktor yang berpengaruh. Hasil penelitian menunjukkan bahwa nilai AIC terbaik adalah regresi Binomial Negatif dan menghasilkan 2 *hotspot* kusta.

Penelitian (Arum et al., 2021) bertujuan untuk mengetahui faktor-faktor yang mempengaruhi jumlah publikasi ilmiah yang dihasilkan oleh mahasiswa Ph.D. Biokimia. Dalam penelitian ini, metode k-fold cross validation digunakan untuk menentukan model terbaik. Regresi Poisson dan regresi Binomial Negatif digunakan untuk memodelkan jumlah publikasi ilmiah dan faktor-faktor yang mempengaruhinya. Hasil penelitian menunjukkan bahwa model regresi Binomial Negatif dengan variabel independen jenis kelamin, status pernikahan, banyaknya anak dibawah 5 tahun, prestise, dan banyaknya artikel oleh mentor dalam 3 tahun terakhir adalah model terbaik yang didapatkan.

Penelitian (Ulfa et al., 2021) menggunakan data dari Dinas Kesehatan terkait profil kesehatan tahun 2017 dan Badan Pusat Statistik (BPS) terkait publikasi daerah dalam angka tahun 2018 untuk semua provinsi di Pulau Jawa. Metode yang digunakan adalah regresi Poisson dan regresi Binomial Negatif untuk memodelkan hubungan jumlah kasus baru kusta dan beberapa faktor yang mempengaruhinya. Hasil penelitian menunjukkan bahwa tiga faktor yang mempengaruhi jumlah kasus kusta adalah jumlah penduduk, persentase balita yang diimunisasi BCG, dan persentase penduduk dengan akses air bersih berkelanjutan

Penelitian (Zubedi et al., 2021) memodelkan kasus *stunting* pada balita di kota Gorontalo tahun 2018. Sebelumnya, penelitian ini melakukan uji asumsi untuk memastikan tidak terjadi multikolinearitas dan overdispersi. Metode yang digunakan adalah regresi Poisson dan regresi Binomial Negatif untuk mengetahui faktor-faktor yang signifikan mempengaruhi kasus *stunting* pada balita di kota Gorontalo.

Hasil penelitian menunjukkan bahwa dua faktor signifikan yang mempengaruhi kasus *stunting* di Gorontalo tahun 2018 adalah balita yang mendapatkan ASI eksklusif dan imunisasi dasar lengkap.

Penelitian (Rahayu et al., 2018) tentang membandingkan regresi Zero Inflated Negative Binomial (ZINB) dan regresi Hurdle Negative Binomial (HNB) pada data overdispersi. Data yang digunakan yaitu jumlah kasus difteri dengan tujuan untuk mengetahui faktor-faktor yang mempengaruhinya. Hasil penelitian menunjukkan bahwa model ZINB memiliki nilai AIC sebesar 123,40 lebih kecil dibandingkan model HNB. Sehingga model ZINB merupakan model yang terbaik untuk mengatasi overdispersi pada data kasus difteri di Indonesia.

## 1.7. Sistematika Penulisan

Sistematika penulisan dalam penelitian ini sebagai berikut:

- BAB I** : Bab ini membahas tentang latar belakang masalah, batasan masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, tinjauan pustaka, dan sistematika penulisan.
- BAB II** : Bab ini membahas tentang analisis regresi Poisson, regresi Binomial Negatif serta hipotesisnya
- BAB III** : Membahas mengenai metode penelitian yang meliputi jenis data, sumber data, variabel penelitian serta langkah-langkah analisis.
- BAB IV** : Bab ini mencakup analisis data dan pembahasan rinci tentang masalah yang diteliti.
- BAB V** : Bab ini memuat kesimpulan permasalahan yang dibahas dan saran penulis kepada penelitian mendatang.

| No | Peneliti (Tahun)      | Judul Penelitian  | Perbedaan  | Persamaan  |
|----|-----------------------|---|--|--|
| 1  | (Budiharti, 2021)     | Pemodelan dan Pemetaan Jumlah Penderita Kusta di Jawa Barat dengan Regresi Binomial Negatif dan <i>Flexibly shaped spatial scan statistic</i> | Mengkaji tentang regresi Hurdle Binomial Negatif dan studi kasus jumlah kematian penderita DBD   | Mengkaji tentang regresi Poisson, Binomial Negatif dan <i>Akaike Information Criterion</i> |
| 2  | (Arum et al., 2021)   | Aplikasi <i>K-fold Cross Validation</i> Dalam Penentuan Model Regresi Binomial Negatif Terbaik  | Mengkaji tentang regresi Hurdle Binomial Negatif, metode <i>Akaike Information Criterion</i> dan studi kasus jumlah kematian penderita DBD | Mengkaji tentang regresi Poisson dan Binomial Negatif                                      |
| 3  | (Ulfah et al., 2021)  | Penanganan Overdispersi pada Model Regresi Poisson dengan Binomial Negatif untuk Jumlah Kasus Baru Kusta di Jawa                              | Mengkaji tentang regresi Hurdle Binomial Negatif dan studi kasus jumlah kematian penderita DBD   | Mengkaji tentang regresi Poisson, Binomial Negatif dan <i>Akaike Information Criterion</i> |
| 4  | (Zubedi et al., 2021) | Analisis Faktor-Faktor yang Mempengaruhi <i>stunting</i> pada balita di Kota Gorontalo Menggunakan Regresi Binomial Negatif                   | Mengkaji tentang regresi Hurdle Binomial Negatif dan studi kasus jumlah kematian penderita DBD   | Mengkaji tentang regresi Poisson, Binomial Negatif dan <i>Akaike Information Criterion</i> |
| 5  | (Rahayu et al., 2018) | Perbandingan Regresi Zero Inflated Negatif Binomial dan Regresi Hurdle Negatif Binomial pada Data Overdispersi                                | Mengkaji tentang regresi Binomial Negatif dan studi kasus jumlah kematian penderita DBD  | Mengkaji regresi Poisson, Hurdle Binomial Negatif dan <i>Akaike Information Criterion</i>  |

## BAB V

### PENUTUP

#### 5.1. Kesimpulan

Berdasarkan hasil studi literatur yang telah dipaparkan oleh penulis tentang analisis regresi Poisson, Binomial Negatif dan HBN pada data yang sudah terlampir pada penelitian ini, maka dapat diambil kesimpulan sebagai berikut:

1. Data jumlah kematian penderita DBD di provinsi Jawa Timur tahun 2022 termasuk data diskrit dan berdistribusi poisson. Terjadi multikolinearitas pada variabel ( $X_2$ ), maka dari itu perlu mengeliminasi variabel tersebut. Selanjutnya menentukan hasil estimasi parameter menggunakan metode MLE. Ada 5 variabel yang berpengaruh terhadap jumlah kematian penderita DBD.
2. Pada data jumlah kasus kematian penderita DBD terjadi overdispersi.Untuk mengatasi data overdispersi perlu menggunakan regresi Binomial Negatif. Parameter yang didapat diestimasikan menggunakan metode MLE. Terdapat 4 variabel yang berpengaruh terhadap jumlah kematian penderita DBD.
3. Regresi Hurdle Binomial Negatif juga dapat mengatasi data overdispersi. Pe-naksiran parameter regresi HBN menggunakan metode MLE. Pada model Hurdle dan *Truncated* masing-masing terdapat 5 dan 1 variabel yang ber-pengaruh terhadap jumlah kematian penderita DBD.
4. Model yang memiliki nilai AIC terkecil adalah regresi Binomial Negatif. Se-hingga dapat disimpulkan bahwa model regresi Binomial Negatif memiliki

kemampuan prediksi terbaik berdasarkan data yang digunakan.

5. Faktor yang berpengaruh pada model regresi Binomial Negatif yaitu persentase kepadatan penduduk ( $X_1$ ), rumah tangga menggunakan air sumur terlindungi ( $X_4$ ) dan memiliki akses sanitasi layak ( $X_5$ ) serta persentase penduduk miskin ( $X_6$ )

## 5.2. Saran

Adapun saran dari penelitian ini sebagai berikut:

1. Penelitian ini menggunakan data overdispersi pada variabel respon. Data yang mengalami overdispersi dapat diatasi dengan menggunakan regresi Binomial Negatif dan HBN. Penulis berharap adanya variasi metode yang digunakan untuk mengatasi data overdispersi dalam pengembangan penelitian selanjutnya.
2. Metode yang digunakan dalam menentukan model terbaik antara regresi Poisson, Binomial Negatif dan HBN pada penelitian ini dengan menggunakan optimisasi AIC. Optimisasi ini merupakan salah satu dari berbagai optimisasi yang dapat digunakan. Penulis berharap adanya variasi optimisasi yang digunakan untuk penentuan model terbaik pada pengembangan penelitian mendatang.

## DAFTAR PUSTAKA

- Akaike, H. (1998). A bayesian analysis of the minimum aic procedure. In *Selected Papers of Hirotugu Akaike*, pages 275–280. Springer.
- Arum, G. P., Widyaningsih, Y., & Prawira, K. (2021). Aplikasi k-fold cross validation dalam penentuan model regresi binomial negatif terbaik. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 15(2):315–322.
- Budiharti, L. M. (2021). Pemodelan dan pemetaan jumlah penderita kusta di jawa barat dengan regresi binomial negatif dan flexibly shaped spatial scan statistic. *Jurnal Riset Statistika*, pages 99–106.
- Candra, A. (2010). Demam berdarah dengue: Epidemiologi, patogenesis, dan faktor risiko penularan. *ASPIRATOR-Journal of Vector-borne Disease Studies*, 2(2).
- Darnah, D. (2010). Menentukan model terbaik dalam regresi poisson dengan menggunakan koefisien determinasi. *Jurnal Matematika, Statistika dan Komputasi*, 6(2):59–71.
- Faraway, J. J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. CRC press.
- Fisher, R. A. (1970). Statistical methods for research workers. In *Breakthroughs in statistics: Methodology and distribution*, pages 66–70. Springer.
- Irianto, A. (2015). *Statistik Konsep Dasar, Aplikasi, dan Pengembangannya*. Pre-nadamedia Grup.

Joseph, H. (2011). *Negative Binomial Regression, 2th Edition*. Cambridge University Press.

Martin, P. (2022). *Regression models for categorical and count data*. Sage.

McCullagh, P. (1989). *Generalized Linear Models, 2th Edition*. London: Chapman and Hall.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of econometrics*, 33(3):341–365.

Nelder, J. A. & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 135(3):370–384.

Nur, I. M., Yuniarti, D., & Hayati, M. N. (2017). Penerapan generalized poisson regression untuk mengatasi overdispersi pada regresi poisson. *Eksponensial*, 7(1):59–66.

Poisson, S.-D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile: précédées des règles générales du calcul des probabilités*. Bachelier.

Purwoko, S. (2018). Indikator air layak minum dan sanitasi layak dalam mendukung upaya kesehatan lingkungan di rumah tinggal. In *Proceeding National Seminar Germas 2018*, volume 1.

Qudratullah, M. F. (2017). *Statistik Nonparametrik Terapan: Teori, Contoh Kasus, dan Aplikasi dengan IBM SPSS*. Yogyakarta: Penerbit Andi.

Qudratullah, M. F., Zuliana, S. U., & Supandi, E. D. (2008). *Metode Statistika Bidang Akademik*.

- Rahayu, R. L., Asrof, A., Rustiana, S., Puspitasari, W., & Suparman, Y. (2018). Perbandingan regresi zero inflated negatif binomial dan regresi hurdle negatif binomial pada data overdispersi (studi kasus: Kejadian difteri di indonesia). *Jurnal Sains Matematika dan Statistika*, 4(1):16–25.
- Raymond P, C. S. C. (2002). Numerical methods for engineers.
- Ruliana, R., Hendikawati, P., & Agoestanto, A. (2016). Pemodelan generalized poisson regression (gpr) untuk mengatasi pelanggaran equidispersi pada regresi poisson kasus campak di kota semarang tahun 2013. *Unnes Journal of Mathematics*, 5(1):39–46.
- Shewhart, W. A. & WILKS, S. S. (2006). Regression analysis by example.
- Sukohar, A. (2014). Demam berdarah dengue (dbd). *Jurnal Medula*, 2(02).
- Supandi, E. D. (2020). Statistika dan terapannya.
- Ulfa, Y. A., Soleh, A. M., & Sartono, B. (2021). Handling of overdispersion in the poisson regression model with negative binomial for the number of new cases of leprosy in java: Penanganan overdispersi pada model regresi poisson dengan binomial negatif untuk jumlah kasus baru kusta di jawa. *Indonesian Journal of Statistics and Its Applications*, 5(1):1–13.
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2010). *Probability and Statistics for Engineers and Scientists, 9th Edition*. Pearson Education.
- Yosepana, S. (2009). *Belajar Efektif Geografi Untuk Siswa SMA/MA Kelas XI IPS*. PT Intimedia Cipta Nusantara.

Zubedi, F., Aliu, M. A., Rahim, Y., & Oroh, F. A. (2021). Analisis faktor-faktor yang mempengaruhi stunting pada balita di kota gorontalo menggunakan regresi binomial negatif. *JAMBURA Journal of probability and statistics*, 2(1):48–55.

