

TESIS
PEMODELAN REGRESI CONWAY MAXWELL POISSON DAN
REGRESI ZERO INFLATED CONWAY MAXWELL POISSON PADA
KASUS PENYAKIT DIFTERI TERHADAP MORTALITAS PASIEN DI
INDONESIA

MODELING OF CONWAY MAXWELL POISSON REGRESSION AND
ZERO INFLATED CONWAY MAXWELL POISSON REGRESSION IN
DIPHThERIA CASES ON PATIENT MORTALITY IN INDONESIA

A. EKA HERMIA FITRIANINGSY
H 062 202 007



PROGRAM STUDI MAGISTER STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2023

**PEMODELAN REGRESI *CONWAY MAXWELL POISSON* DAN REGRESI
ZERO INFLATED CONWAY MAXWELL POISSON PADA KASUS
PENYAKIT DIFTERI TERHADAP MORTALITAS PASIEN DI INDONESIA**

Tesis

sebagai salah satu syarat untuk mencapai gelar magister

Program Studi Magister Statistika

Disusun dan diajukan oleh

A. EKA HERMIA FITRIANINGSY

H062202007

Kepada

**PROGRAM STUDI MAGISTER STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2023**

TESIS

PEMODELAN REGRESI *CONWAY MAXWELL POISSON* DAN REGRESI
ZERO INFLATED CONWAY MAXWELL POISSON PADA KASUS
PENYAKIT DIFTERI TERHADAP MORTALITAS PASIEN DI INDONESIA

A. EKA HERMIA FITRIANINGSY

H062202007

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka
Penyelesaian Program Studi Magister Statistika Fakultas Matematika dan Ilmu
Pengetahuan Alam Universitas Hasanuddin
pada tanggal Februari 2023
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping



Dr. Nurtiti Sunusi, S.Si M.Si.
NIP. 1972017 199703 2 002

Dr. Erna Tri Herdiani, S.Si M.Si.
NIP. 19750429 200003 2 001

Ketua Program Studi
Magister Statistika,

Dekan Fakultas Matematika dan
Ilmu Pengetahuan Alam
Universitas Hasanuddin



Dr. Dr. Georgina M. Tinungki, M.Si
NIP. 19620926 198702 2 001



Dr. Eng. Amiruddin, M.Si
NIP. 19720515 199702 1 002

PERNYATAAN KEASLIAN TESIS DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, tesis berjudul Pemodelan Regresi Conway Maxwell Poisson Dan Regresi Zero Inflated Conway Maxwell Poisson Pada Kasus Penyakit Difteri Terhadap Mortalitas Pasien Di Indonesia adalah benar karya saya dengan arahan dari komisi pembimbing (Dr. Nurtiti Sunusi, S.Si., M.Si dan Dr. Erna Tri Herdiani, S.Si., M.Si.). Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari isi tesis ini akan dipublikasikan di Jurnal (International Journal of Research Publications) ISSN: 2708-3578 Vol. 110, Issue 1, sebagai artikel dengan judul "Analysis of Diphtheria Disease Mortality in Indonesia Using Zero Inflated Conway Maxwell Poisson Regression Model"

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 31 Januari 2023

Yang Menyatakan,



A. Eka Hermia Fitrianingsy

NIM. H062202007

UCAPAN TERIMA KASIH

Segala puji bagi **Allah SWT Rabb** semesta alam atas segala limpahan berkah, rahmat dan nikmatnya yang tak terhingga serta shalawat serta salam semoga selalu dilimpahkan kepada Nabi yang paling dimuliakan, pemimpin orang-orang bertakwa, **Muhammad bin Abdullah** dan kepada para keluarga serta sahabat-sahabat beliau. Alhamdulillah, berkat pertolongan Allah akhirnya tesis dengan judul “**Pemodelan Regresi Conway Maxwell Poisson Dan Regresi Zero Inflated Conway Maxwell Poisson Pada Kasus Penyakit Difteri Terhadap Mortalitas Pasien Di Indonesia**” yang disusun sebagai salah satu syarat akademik untuk meraih gelar magister pada Program Studi Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin ini dapat dirampungkan. Penulis berharap tesis ini bisa memberikan tambahan pengetahuan bagi pembelajar statistika.

Penulis menyadari bahwa penyelesaian tesis ini tidak luput dari hambatan dan tantangan. Namun, berkat bantuan dan dorongan serta motivasi baik yang terkait secara langsung maupun secara tidak langsung dari berbagai pihak akhirnya kesulitan-kesulitan yang timbul dapat diatasi. Untuk itu suatu kewajiban bagi penulis untuk menyampaikan terima kasih yang tak terhingga kepada kedua orang tua penulis, Ayahanda **Herman** dan Ibunda **Andi Aminah** atas kerja kerasnya mendidik, mendoakan, menjadi inspirasi, membesarkan penulis dengan bertabur cinta, kasih sayang yang tulus serta dengan ikhlas telah mengiringi setiap langkah penulis dengan doa dan restunya. Terima kasih juga kepada keluarga besar penulis atas dukungan baik moral maupun spiritual yang telah ditujukan kepada penulis. Perkenankan pula penulis menyampaikan terima kasih dan penghargaan yang 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 dan sebagai Pembimbing Utama yang telah memberikan petunjuk dan masukannya selama proses pembuatan tesis ini,

4. Ibu **Dr. Dr. Georgina M. Tinungki, M.Si.**, selaku Ketua Program Studi Magister Statistika atas nasehat serta motivasi yang diberikan kepada penulis sejak masih berstatus Mahasiswa Baru sampai penulis menyelesaikan studi di Program Studi Magister Statistika.
5. Ibu **Dr. Erna Tri Herdiani, S.Si., M.Si.**, selaku Pembimbing Pertama yang selalu meluangkan waktu, pemikiran dan energi untuk memberikan arahan, dorongan, dan motivasi kepada penulis mulai dari awal hingga selesainya penulisan tesis ini.
6. Bapak **Dr. Nirwan, M.Si.**, selaku Tim Penguji yang telah memberikan kritikan saran dan pembelajaran yang membangun dalam penyempurnaan tesis ini.
7. Ibu **Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.**, selaku Tim Penguji dan Penasehat Akademik yang bersedia meluangkan waktu untuk memberikan masukan yang membangun dalam penyempurnaan tesis ini.
8. Ibu **Dr. Anna Islamiyati, S.Si., M.Si.** selaku Tim Penguji yang telah memberikan kritikan saran dan pembelajaran yang membangun dalam penyempurnaan tesis 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. Suami penulis, **M. Quraisy** yang telah penuh pengertian dan tulus menanti.
2. Teman seperjuangan **Magister Statistika Angkatan 03** untuk setiap waktu yang tercipta bersama, susah senangnya perkuliahan telah dilewati bersama.

Penulis juga mengucapkan terima kasih untuk semua pihak yang tidak dapat penulis sebutkan satu persatu. Penulis menyadari bahwa masih banyak kekurangan dalam tesis ini. Oleh karena itu, dengan segala kerendahan hati penulis memohon maaf. Akhir kata semoga tulisan ini memberi manfaat bagi para pembaca.

Makassar, 31 Januari 2023

A. Eka Hermia Fitrianingsy

ABSTRAK

A. EKA HERMIA FITRIANINGSY. **Pemodelan Regresi Conway Maxwell Poisson dan Regresi Zero Inflated Conway Maxwell Poisson Pada Kasus Penyakit Difteri Terhadap Mortalitas Pasien Di Indonesia** (dibimbing oleh Dr. Nurtiti Sunusi, M.Si dan Dr. Erna Tri Herdiani, S.Si M.Si)

Difteri adalah penyakit saluran pernapasan bagian atas yang disebabkan oleh bakteri *Corynebacterium Diphtheriae*. Difteri merupakan penyakit menular yang berbahaya, karena terdapat 23 kematian dari 529 kasus pada tahun 2019. Data yang digunakan dalam penelitian tugas akhir ini adalah data sekunder yang diambil dari Dinas Kesehatan tahun 2019. Jumlah kematian pasien difteri adalah data hitung dan asumsi berdistribusi sebagai distribusi Poisson. Jumlah data nol sebesar 73,5 persen menunjukkan adanya overdispersi pada variabel respon. Karena, Regresi Poisson membutuhkan kesetaraan antara mean dan varians, model ini tidak tepat untuk diterapkan pada dispersi di bawah dan di atas. Model regresi Zero Inflated Conway Maxwell Poisson (ZICMP) dapat menjadi alternatif karena fleksibel untuk data under- dan over-dispersi serta dapat menyelesaikan masalah kelebihan nilai nol pada variabel respon. Dalam penelitian ini, ZICMP diterapkan pada kasus kematian pada pasien difteri di Indonesia. Penelitian ini bertujuan untuk mengkaji fungsi kemungkinan dan membentuk algoritma untuk mengestimasi parameter ZICMP serta menerapkan model ZICMP pada kasus kematian pada pasien difteri. Metode Maximum Likelihood Estimation (MLE) digunakan untuk mengestimasi parameter pada ZICMP dan fungsi kemungkinan dimaksimalkan menggunakan algoritma Expectation Maximization (EM).

Kata kunci: Overdispersi, Zero Inflation, Zero Inflated Conway Maxwell Poisson (ZICMP), Difteri.

ABSTRACT

A. EKA HERMIA FITRIANINGSY. **Modeling Conway Maxwell Poisson Regression and Zero Inflated Conway Maxwell Poisson Regression Modeling in Diphtheria Cases on Patient Mortality in Indonesia** (supervised by Dr. Nurtiti Sunusi, S.Si., M.Si and Dr. Erna Tri Herdiani, S.Si M.Si)

Diphtheria is a disease of the upper respiratory tract caused by the bacterium *Corynebacterium Diphtheriae*. Diphtheria is a dangerous infectious disease, because there were 23 deaths from 529 cases in 2019. The data used in this final study is a secondary data taken from the Health Office in 2019. The number of deaths of diphtheria patients is count data and assuming distributed as Poisson distribution. The amount of zero data is 73,5 percent indicates overdispersion on the response variable. Since, Poisson Regression requires an equality between mean and variance, it is not appropriate to apply this model on under- and over-dispersion. The Zero Inflated Conway Maxwell Poisson (ZICMP) regression model can be an alternative because it is flexible for under- and over-dispersion data and can solve the problem of excess zero value on the response variable. In this research, ZICMP is applied on the case of death in diphtheria patients in Indonesia. The purpose of this research is to examine the likelihood function and form an algorithm to estimate the parameter of the ZICMP and also applying ZICMP model in the case of death in diphtheria patients. Maximum Likelihood Estimation (MLE) method is used to estimate the parameter on ZICMP and the likelihood function is maximized using Expectation Maximization (EM) algorithm.

Keywords: Overdispersion; Zero Inflation; Zero Inflated Conway Maxwell Poisson (ZICMP); Diphtheria.

DAFTAR ISI

TESIS	i
PERNYATAAN KEASLIAN TESIS	ii
UCAPAN TERIMA KASIH.....	v
ABSTRAK.....	vii
ABSTRACT.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA	7
2.1 Generalized Linear Model (Glm).....	7
2.2 Distribusi <i>Poisson</i>	8
2.3 Regresi <i>Poisson</i>	9
2.4 Metode <i>Maximum Likelihood Estimation</i>	12
2.5 Algoritma <i>Expectation Maximization</i>	13
2.6 Regresi <i>Zero-Inflated Poisson</i>	13
2.7 <i>Conway Maxwell Poisson</i>	14
2.8 Regresi <i>Zero-Inflated Conway Maxwell Poisson (Zicmp)</i>	15
2.9 Pengujian kesesuaian model	16
2.10 Pengujian Signifikansi Parameter	17
2.11 Keباikan Model.....	18
2.12 Penyakit Difteri.....	18
BAB III METODOLOGI PENELITIAN	23
3.1 Sumber Data	23
3.2 Variabel Data.....	23

3.3 Metode Penelitian	24
BAB IV HASIL DAN PEMBAHASAN	26
4.1 Estimasi Parameter Regresi Conway Maxwell Poisson.....	26
4.2 Estimasi Parameter Regresi Zero Inflated Conway Maxwell Poisson	29
4.3 Analisis Kasus Mortalitas Pasien Difteri Di Indonesia.....	38
4.4 Kebaikan Model	46
BAB V PENUTUP	47
5.1 Kesimpulan.....	47
5.2 Saran	48
DAFTAR PUSTAKA	49
LAMPIRAN	51

DAFTAR TABEL

1. Jenis <i>Generalized Liner Model</i> Untuk Analisis Statistik	8
2. Variabel Penelitian	23
3. Analisis Deskriptif Variabel	39
4. Hasil Pemeriksaan <i>Zero Inflation</i> pada Variabel Respon	40
5. Hasil Pemeriksaan Multikolinieritas	41
6. Estimasi parameter Model Regresi CMP	42
7. Estimasi parameter Model Regresi ZICMP	43
8. Uji Parsial Parameter β	44
9. Uji Parsial Parameter γ	45
10. Nilai AIC model Regresi CMP dan ZICMP	46

DAFTAR GAMBAR

- | | |
|--|----|
| 1. Sebaran Kasus Difteri Menurut Provinsi Tahun 2020 | 18 |
| 2. Kerangka Konseptual Pemodelan Regresi ZICMP | 21 |

BAB I

PENDAHULUAN

1.1 Latar Belakang

Difteri adalah penyakit menular yang disebabkan oleh *Corynebacterium Diphtheria*. *C.Diphtheria* biasanya berada di saluran pernapasan bagian atas, menyebabkan *ulserasi* pada mukosa, dan menginduksi pembentukan pseudomembran inflamasi. Racun yang sangat kuat diserap ke dalam sirkulasi dan merusak organ-organ lain, berpotensi mengakibatkan kematian. Infeksi Difteri pada manusia dapat berakhir dengan gagal jantung akut atau dapat berakibat fatal beberapa minggu kemudian selama masa pemulihan *C.Diphtheria* biasanya ditularkan melalui kontak langsung atau melalui bersin dan batuk. (Hadfield, 2000).

Sejak tahun 2000-2015, World Health Organization (WHO) mencatat kasus Difteri di *South-East Asia Region* setiap tahunnya menempati urutan pertama kasus Difteri di dunia. Indonesia merupakan negara dengan kasus insiden Difteri terbanyak kedua dibandingkan negara di *South East Asia Region* lainnya, yaitu India. Jumlah kasus Difteri yang dilaporkan di Indonesia dari tahun 2011-2015 sebesar 3.203 kasus, sedangkan Negara India masih menjadi negara dengan kasus Difteri tertinggi dengan jumlah kasus Difteri sebesar 18.350 kasus (WHO, 2017). Kementerian Kesehatan RI mencatat jumlah kasus Difteri pada tahun 2019 kasus Difteri menyebar hampir semua wilayah di Indonesia. Jumlah kasus Difteri pada tahun 2019 sebanyak 529 kasus, jumlah kematian sebanyak 23 kasus, dengan CFR (*Case Fatality Rate*) sebesar 4,35%. Jumlah kasus Difteri tahun 2019 mengalami penurunan yang cukup signifikan jika dibandingkan tahun 2018 (1.386 kasus). Jumlah kematian akibat Difteri juga mengalami penurunan jika dibandingkan tahun sebelumnya (29 kasus), namun CFR pada tahun 2019 lebih tinggi dibandingkan CFR tahun 2018 (Depkes, 2019).

Pada kasus maraknya Difteri tentu saja tidak terlepas dari beberapa faktor yang mempengaruhinya. Faktor resiko penyakit Difteri dapat ditinjau dari berbagai faktor diantaranya faktor ekonomi (penghasilan rendah), faktor gizi rendah, faktor lingkungan, dan pelayanan kesehatan. (Lestari, 2012).

Keterkaitan antara faktor-faktor risiko penyakit Difteri dapat dianalisis menggunakan analisis regresi. Asumsi data regresi adalah variabel respons yang berupa peubah acak kontinu dan mengikuti distribusi normal, namun banyak ditemukan variabel respons yang tidak berdistribusi normal dan tidak linear dalam parameter. Untuk mengatasi hal tersebut dikembangkan *Generalized Linear Model* (GLM). GLM digunakan sebagai perluasan model regresi umum dengan variabel responsnya berdistribusi keluarga eksponensial, meliputi distribusi normal, binomial, *Poisson*, binomial negatif, eksponensial, gamma, dan invers normal (Myers, 2010). Jika variabel respons yang digunakan dalam penelitian ini adalah jumlah mortalitas pasien penderita Difteri yang berupa data diskrit, maka salah satu model regresi yang digunakan adalah model regresi *Poisson*.

Analisis data menggunakan regresi *Poisson* harus memenuhi asumsi seperti nilai varians dan rata-rata dari variabel respons tersebut sama atau *equidispersi* (Myers, 2010). Namun dalam beberapa kasus, tidak sepenuhnya asumsi tersebut terpenuhi, karena data yang bertipe diskrit seringkali mengalami kasus overdispersi (penyimpangan tinggi) maupun underdispersi yaitu nilai *mean* dan *variance* tidak sama, dengan kata lain nilai *variance* lebih besar daripada nilai *mean* ($var(Y) > \mu$) atau nilai *variance* lebih kecil daripada nilai *mean* ($var(Y) < \mu$). Data yang mengandung overdispersi menghasilkan galat baku yang lebih kecil dari nilai sesungguhnya (underestimate). Hal ini menyebabkan kesimpulan yang diperoleh menjadi tidak valid (McCullagh & Nelder, 1989).

Pada umumnya, untuk mengatasi masalah overdispersi digunakan pemodelan dengan regresi *Binomial Negatif*. Akan tetapi pada regresi Binomial Negatif tidak dilengkapi dengan data yang underdispersi (Lord, 2013). Sehingga diperlukan beberapa model yang memungkinkan untuk digunakan pada data overdispersi maupun underdispersi. Sellers dan Shmueli (2010) memperkenalkan metode dalam mengatasi kedua masalah tersebut, yaitu dengan analisis regresi *Conway-Maxwell-Poisson* (COM-*Poisson*). Di mana model tersebut lebih menekankan pada regresi *Poisson* dan regresi logistik. Teknik analisis ini telah berhasil dalam menghitung data cacah dengan berbagai tingkat sebaran (dispersi).

Pada kenyataannya, data *count* tidak hanya mengalami overdispersi atau underdispersi akan tetapi dapat juga mengalami *excess zero*. *Excess zero* yaitu kondisi ketika proporsi nilai nol pada data lebih besar dari nilai lainnya. Data *count* yang mengandung nilai nol dapat diestimasi menggunakan regresi *Poisson*. Namun untuk data dengan kondisi nilai nol sangat banyak (*excess zero*) memerlukan adanya metode tertentu untuk mengatasinya. Jika regresi *Poisson* tetap digunakan maka estimasi parameternya kurang baik dalam menaksir kelebihan nol tersebut (Nadhiroh,2009). Hal ini menyebabkan adanya pengembangan metode-metode statistik untuk mengatasi masalah tersebut. Overdispersi pada data dapat terjadi karena proporsi nilai nol yang berlebih pada variabel respons (*excess zeros*). Jika pada suatu pemodelan *count data* (data hitung) banyak terdapat amatan yang bernilai nol pada variabel respon (*zero inflation*) dan terjadi Overdispersi maka model yang dapat digunakan adalah model regresi *Zero Inflated Negative Binomial* (Astuti, 2015).

Pada perkembangan terdapat alternatif lain untuk memodelkan kasus dengan banyak data yang bernilai nol dan terjadi *Overdispersi*, selain menggunakan model regresi *Zero Inflated Generalized Poisson* (ZIGP) dan *Zero Inflated Negative Binomial* (ZINB), model tersebut adalah regresi *Zero Inflated Conway Maxwell Poisson*. Baru-baru ini, Sellers and Raim (2016) memperkenalkan model ZICMP, yang pada dasarnya menggantikan distribusi *Poisson* dalam model ZIP dengan distribusi Conway-Maxwell-Poisson (CMP). ZICMP adalah generalisasi dari ZIP karena distribusi *Poisson* adalah kasus khusus dari distribusi CMP. Keuntungan tambahan dari ZICMP adalah dapat menangani tidak hanya Overdispersi tetapi juga underdispersi dalam hitungan. ZICMP sangat fleksibel karena memungkinkan dispersi ke segala arah.

Pemodelan regresi *Zero Inflated Poisson* (ZIP), *Zero Inflated Negative Binomial* (ZINB), dan *Zero Inflated Generalized Poisson* (ZIGP) telah banyak dilakukan, seperti dilakan oleh Lestari (2009), Lestari (2014), dan Astuti (2015). Lestari (2009), menggunakan model regresi *Zero Inflated Poisson* (ZIP) dengan mengaplikasikan model tersebut pada data pekerja seks komersial. Lestari (2014) menggunakan model *Zero Inflated Generalized Poisson* (ZIGP) untuk memodelkan data penderita *Tetanus Neonatorum*. Dan Astuti (2015) menggunakan model *Zero*

Inflated Negative Binomial (ZINB) pada kasus *Tetanus Neonatorum* di provinsi Jawa Timur.

Benyaknya data bernilai nol pada kasus mortalitas pasien Difteri, yaitu sebanyak 73,5 persen mengindikasi adanya Overdispersi dalam variabel respon. Namun, pada pada model regresi *Poisson* terdapat asumsi ragam harus sama dengan rata-rata (*equidispersion*), sehingga model ini tidak tepat digunakan pada data yang mengalami Overdispersi atau *underdispersi*. Model regresi *Zero Inflated Conway Maxwell Poisson* dapat menjadi salah satu alternatif karena dapat mengatasi Overdispersi dan dapat mengatasi masalah nilai nol berlebih pada variabel respon (*zero inflation*).

Untuk mengestimasi parameter, metode *maximum likelihood estimation (MLE)* dapat digunakan pada model regresi yang datanya mengikuti distribusi tertentu. Metode MLE dilakukan dengan memaksimalkan fungsi *likelihood*. Namun pada umumnya, maksimum suatu fungsi *likelihood* tidak bisa diselesaikan secara analitik. Oleh karena itu, jika diperoleh bentuk implisit atau non-linier, maka dapat diselesaikan dengan algoritma Newton-Raphson (NR), Fisher Scoring, atau Expectation Maximization (EM) untuk mendapatkan solusi numeriknya. Banyaknya pengamatan bernilai nol merupakan salah satu penyebab terjadinya overdispersi pada data model regresi ZICMP sehingga algoritma EM cocok digunakan untuk mengestimasi parameter yang menyertakan pengamatan yang bernilai nol (Kusuma dkk., 2013).

Berdasarkan hal tersebut, maka akan dikaji model regresi *Zero Inflated Conway Maxwell Poisson (ZICMP)* pada kasus penyakit Difteri terhadap mortalitas pasien di Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang, dapat dirumuskan masalah sebagai berikut:

1. Bagaimana estimasi parameter model regresi *Conway Maxwell Poisson* dan regresi *Zero Inflated Conway Maxwell Poisson* menggunakan metode *Maximum Likelihood Estimation*?
2. Bagaimana aplikasi model regresi *Conway Maxwell Poisson* dan regresi *Zero Inflated Conway Maxwell Poisson* terhadap data angka mortalitas pasien Difteri di Indonesia?

1.3 Tujuan Penelitian

Tujuan penelitian yang dicapai dalam penelitian ini berdasarkan latar belakang adalah sebagai berikut:

1. Memperoleh estimator parameter model regresi *Conway Maxwell Poisson* dan regresi *Zero Inflated Conway Maxwell Poisson* menggunakan metode *Maximum Likelihood Estimation*
2. Memperoleh variabel yang berpengaruh signifikan terhadap kasus mortalitas pasien Difteri dengan menggunakan model regresi *Zero Inflated Conway Maxwell Poisson* dan *Conway Maxwell Poisson*.

1.4 Batasan Masalah

Dalam penulisan tugas akhir ini, pembatasan masalah sangat diperlukan agar tidak terjadi penyimpangan dari tujuan semula sehingga pemecahan masalah lebih terfokus. Berdasarkan uraian latar belakang, maka pembatasan penelitian ini difokuskan untuk :

1. Estimasi parameter pada model regresi *Zero Inflated Conway Maxwell Poisson* (ZICMP) dilakukan dengan metode *Maximum Likelihood Estimation* (MLE) dan untuk memaksimalkan fungsi likelihood digunakan algoritma EM (*Expectation Maximization*)

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut.

Manfaat untuk pemerintah: Memberikan alternatif model yang dapat dipakai untuk memprediksi banyaknya kasus penyakit Difteri di Indonesia, sehingga diharapkan dapat menjadi tambahan informasi serta bahan rujukan untuk pengambilan kebijakan dan penanggulangan Difteri di masa depan.

Manfaat untuk peneliti : Mengkaji dan menerapkan aplikasi dari ilmu statistik khususnya pemodelan dengan *Zero Inflated Conway Maxwell Poisson (ZICMP)* sehingga dapat diaplikasikan dan diterapkan untuk penyelesaian banyaknya kasus Difteri.

BAB II

TINJAUAN PUSTAKA

2.1 Generalized Linear Model (GLM)

Generalized Linear Model merupakan perluasan model regresi umum untuk variabel respon mengikuti sebaran keluarga eksponensial. GLM berhubungan dengan variabel prediktor melalui persamaan dengan bentuk linear Menurut Agresti (2002), terdapat tiga komponen dalam GLM yaitu:

1. *Random component* (komponen acak) yaitu komponen yang ditunjukkan dengan variabel respon Y yang bersifat independen dan mengikuti sebaran keluarga eksponensial. Distribusi dari Y_i adalah anggota dari keluarga eksponensial seperti distribusi Gaussian (normal), binomial, *Poisson*, gamma, atau inverse gaussian.
2. *Systematic component* dari GLM merupakan sebuah variabel prediktor yang merupakan fungsi linear dalam persamaan model pada sisi sebelah kanan. Komponen sistematis yaitu vector $\boldsymbol{\eta}$ yang terdiri dari $\eta_1, \eta_2, \dots, \eta_n]^T$ yang memiliki bentuk umum dari $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ di mana \mathbf{X} merupakan suatu matriks dengan elemen yang terdiri variabel prediktor, sedangkan $\boldsymbol{\beta}$ merupakan bentuk vector dari parameter-parameter model. Masing-masing dari elemen $\boldsymbol{\eta}$ dapat dinyatakan dengan :

$$\eta_i = \sum_{j=1}^p \beta_j x_{ij}, \quad i = 1, 2, \dots, n \text{ dan } j = 1, 2, \dots, p;$$

3. *Link function* (fungsi penghubung) yaitu komponen menghubungkan antara komponen random dan komponen sistematis. Misalkan $\mu_i = E(Y_i)$ dimana $i = 1, 2, \dots, n$. Model untuk menghubungkan μ_i dengan η_i adalah $g(\cdot)$ sehingga $g(\mu_i) = \eta_i$. Fungsi $g(\cdot)$ menghubungkan $E(Y_i)$ dengan variabel prediktor yaitu:

$$g(\mu_i) = \sum_{j=1}^p \beta_j x_{ij}, \quad i = 1, 2, \dots, n \text{ dan } j = 1, 2, \dots, p;$$

Perkembangan teori GLM pada pertengahan 1970-an menyatukan model-model penting untuk variabel respons kontinu dan diskrit. Tabel 2.1 mencantumkan beberapa GLM populer untuk analisis statistik.

Tabel 2.1 Jenis *Generalized Liner Model* Untuk Analisis Statistik

Komponen Acak	Fungsi Link	Model
Normal	Identity	Regresi
Normal	Identity	Analisis Varians
Normal	Identity	Analisis Kovarians
Binomial	Logit	Regresi Logistik
Binomial	Probit And Others	Regresi Biner
Multinomial	Generalized Logit	Multinomial Response
<i>Poisson</i>	Log	Loglinear

2.2 Distribusi *Poisson*

Distribusi *Poisson* merupakan suatu distribusi untuk peristiwa yang probabilitas kejadiannya kecil, dimana kejadian tergantung pada selang waktu tertentu atau di suatu daerah tertentu dengan hasil pengamatan berupa variabel diskrit dan antar variabel prediktor saling independen. Selang waktu tersebut dapat berupa berapa saja panjangnya, misalnya semenitsehari, seminggu, sebulan, bahkan setahun. Daerah tertentu yang dimaksudkan dapat berupa suatu garis, suatu luasan, suatu volume, atau mungkin sepotong bahan (Walpole, 1995).

Distribusi *Poisson* memiliki ciri-ciri sebagai berikut:

1. Banyaknya percobaan yang terjadi dalam suatu selang waktu atau suatu daerah tertentu, tidak tergantung pada banyaknya hasil percobaan yang terjadi pada selang waktu atau daerah lain yang terpisah.
2. Peluang terjadinya satu hasil percobaan selama suatu selang waktu yang singkat sekali atau dalam suatu daerah yang kecil. Sebanding dengan panjang selang waktu tersebut atau besarnya daerah tersebut dan tidak bergantung pada banyak hasil percobaan yang terjadi diluar selang waktu dan daerah tertentu.

3. Peluang bahwa lebih dari satu hasil percobaan akan terjadi dalam selang waktu yang singkat tersebut atau dalam daerah yang kecil tersebut dapat diabaikan.

Fungsi peluang untuk data berdistribusi *Poisson* bergantung pada parameter tunggal, yaitu rata-rata λ . Fungsi peluangnya adalah sebagai berikut:

$$f(y, \lambda) = \Pr(Y = y, \lambda) = \frac{e^{-\lambda} \lambda^y}{y!}, y = 0, 1, 2, \dots \quad (2.1)$$

2.3 Regresi *Poisson*

Regresi *Poisson* secara umum digunakan untuk menganalisis *count data* (data hitung). Menurut Hinde dan Demetrio (2007), pada regresi *Poisson* terdapat asumsi $Y \sim \text{Poisson}(\lambda)$, hal ini berarti variabel respon diasumsikan menyebar *Poisson* dengan parameter λ . Model regresi *Poisson* didapatkan dari sebaran *Poisson* yang mendefinisikan parameter λ sebagai variabel kovariat, dengan y_i adalah pengamatan ke- i dari variabel respon. Jika variabel acak diskrit (Y) berdistribusi *Poisson* dengan parameter $\lambda > 0$, maka fungsi probabilitasnya dinyatakan sebagai berikut.

$$f(y) = \frac{e^{-\lambda} \lambda^y}{y!}, y = 0, 1, 2, \dots$$

Poisson distribution memiliki mean dan varians dari distribusi ini dapat ditunjukkan melalui $E(Y) = \text{Var}(Y) = \lambda$.

Regresi *Poisson* menggunakan generalized linear model (GLM) agar modelnya dapat digunakan dalam data pengamatan, dimana variabel responnya tidak mengharuskan berdistribusi normal. Regresi *Poisson* kemudian digunakan untuk memodelkan suatu peristiwa yang relatif jarang terjadi pada satuan unit tertentu. Secara umum, persamaan regresi *Poisson* dapat dinyatakan sebagaimana persamaan (2.3).

$$\ln \hat{\lambda}_i = \hat{\beta}_0 + \sum_{j=1}^p \hat{\beta}_j x_{ij}, i = 1, 2, \dots, n \text{ dan } j = 1, 2, \dots, p \quad (2.2)$$

$$\lambda_i = \exp(x_i^T \beta)$$

di mana:

p : jumlah variabel prediktor

n : jumlah pengamatan

$\hat{\beta}$: parameter model regresi *Poisson* yang diestimasi

Regresi *Poisson* mengasumsikan bahwa variabel respon (Y) merupakan variabel diskrit dan harus memenuhi asumsi *equidispersi*. *Equidispersi* yaitu nilai mean sama dengan nilai variansi, yaitu $E(Y) = Var(Y) = \lambda$ (Kurniawan, 2017).

2.3.1 Overdispersi

Menurut Hinde dan Demetrio (2007), terdapat beberapa kemungkinan tidak terpenuhi *equidispersion* pada suatu pemodelan, antara lain adalah keragaman hasil pengamatan (keragaman antar individu sebagai komponen yang tidak dijelaskan oleh model), korelasi antar respon individu, terjadi *clustering* (pengelompokan) dalam populasi dan variabel teramati yang dihilangkan. Konsekuensi dari tidak terpenuhi *equidispersion* adalah regresi *Poisson* tidak sesuai untuk memodelkan data karena model yang terbentuk akan menghasilkan estimasi parameter yang bias. Selain itu, Overdispersi juga mengakibatkan nilai *standart error* menjadi lebih kecil (*underestimates*) dari seharusnya, sehingga menghasilkan kesimpulan yang tidak sesuai.

Pemeriksaan Overdispersi dapat dilakukan menggunakan nilai *Deviance*. Ragam dari sebaran *Poisson* sama dengan rata-rata ($\sigma^2 = \mu$). Overdispersi dideteksi menggunakan nilai *Deviance* dibagi dengan derajat bebas yang mempunyai nilai lebih besar dari 1, sedangkan *underdispersi* dideteksi dengan nilai *Deviance* dibagi dengan derajat bebas yang mempunyai nilai kurang dari satu. Menurut Agresti (2002), nilai *Deviance* dapat dinyatakan sebagaimana persamaan (2.3).

$$D = 2 \sum_{i=1}^n y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) \quad (2.3)$$

di mana :

n : jumlah pengamatan

y_i : variabel respon ke- i dengan $i = 1, 2, \dots, n$

$\hat{\mu}_i$: rata-rata variabel respon y yang dipengaruhi oleh nilai variabel prediktor pada pengamatan ke- i

2.3.2 Zero Inflation

Nilai nol yang berlebihan pada variabel respon (*zero inflation*) sering ditemukan pada analisis regresi *Poisson* baik untuk data diskrit atau *count data*. Apabila nilai nol memiliki arti penting dalam penelitian maka data tersebut tidak dapat dihilangkan namun harus dimasukkan dalam proses analisis. Pada beberapa penelitian dapat dijumpai kondisi terlalu banyak nol pada variabel respon yang lebih dari 50 persen. Besarnya proporsi data yang bernilai nol dapat berakibat pada ketepatan dari inferensia. Selain itu, regresi *Poisson* juga menjadi tidak tepat lagi memodelkan data yang sebenarnya (Famoye, 2006).

2.3.3 Multikolinieritas

Multikolinieritas menunjukkan terdapat hubungan di antara beberapa atau semua variabel yang menjelaskan model regresi. Terdapat dua jenis multikolinieritas yaitu multikolinieritas sempurna dan multikolinieritas tidak sempurna. Pada multikolinieritas sempurna terdapat hubungan linier di antara variabel prediktor di mana satu variabel prediktor adalah fungsi linier dari variabel prediktor yang lain, sedangkan multikolinieritas tidak sempurna terjadi apabila terdapat hubungan linier yang tidak sempurna antar variabel prediktor (Gujarati, 1991).

Menurut Hocking (1996), pendeteksian multikolinieritas dapat dilakukan menggunakan nilai *Variance Inflation Factor* (VIF). Untuk regresi dengan lebih dari dua variabel, persamaan untuk menghitung nilai VIF dapat dinyatakan sebagaimana persamaan (2.5).

$$VIF_j = \frac{1}{1-R_j^2} \quad (2.4)$$

di mana :

R_j^2 : Koefisien determinasi dari *auxiliary regression*

Multikolinieritas sempurna yang terjadi dalam analisis regresi menyebabkan koefisien regresi menjadi *undetermined* (tidak dapat diestimasi), sedangkan pada multikolinieritas tidak sempurna *standart error* cenderung semakin besar seiring dengan meningkatnya tingkat korelasi antar variabel walaupun estimasi parameter masih dapat dilakukan dan bersifat BLUE (*Best Linear Unbiased Estimators*) serta tetap efisien. Besarnya *standart error* pada multikolinieritas tidak sempurna dapat menyebabkan selang kepercayaan menjadi lebih lebar. Selain itu, multikolinieritas

tidak sempurna juga menyebabkan tanda untuk koefisien regresi berkebalikan dengan teori yang ada (Gujarati, 1991).

2.4 Metode Maximum Likelihood Estimation

Metode *maximum likelihood estimation* (MLE) dapat digunakan untuk menaksir parameter suatu model yang diketahui fungsi probabilitasnya. Misalkan Y_1, Y_2, \dots, Y_n adalah sampel acak dari populasi dengan fungsi probabilitas $f(y|\theta)$ yang bergantung pada $\theta = \lambda_1, \beta_1, \beta_2, \gamma$ dengan θ adalah parameter yang tidak diketahui. Karena Y_1, Y_2, \dots, Y_n saling bebas, maka fungsi probabilitas bersama dari Y_1, Y_2, \dots, Y_n dapat dinyatakan sebagai berikut.

$$f(y_1, y_2, \dots, y_n | \theta) = f(y_1 | \theta) \cdot f(y_2 | \theta) \cdot \dots \cdot f(y_n | \theta)$$

Fungsi *likelihood* didefinisikan sebagai fungsi probabilitas bersama dari Y_1, Y_2, \dots, Y_n yang dianggap sebagai fungsi dari θ yang dituliskan sebagai $L(\theta|y)$, yaitu:

$$L(\theta|y) = f(y_1, y_2, \dots, y_n | \theta)$$

$$L(\theta|y) = f(y_1 | \theta) \cdot f(y_2 | \theta) \cdots f(y_n | \theta) = \prod_{i=1}^n f(y_i | \theta) \quad (2.5)$$

(Hogg, McKean, & Craig, 2013)

Estimator maksimum likelihood θ adalah nilai θ yang memaksimumkan fungsi likelihood $L(\theta|y)$. Namun, lebih mudah bekerja dengan logaritma natural dari fungsi likelihood, yaitu $l(\theta|y) = \ln L(\theta|y)$. Karena fungsi logaritma natural adalah fungsi yang monoton naik, maka nilai yang memaksimumkan fungsi $l(\theta|y)$ sama dengan memaksimumkan fungsi $L(\theta|y)$ sehingga dapat dituliskan sebagai berikut.

$$l(\theta|y) = \ln L(\theta|y)$$

$$l(\theta|y) = \ln \left\{ \prod_{i=1}^n f(y_i | \theta) \right\} = \sum_{i=1}^n \ln f(y_i | \theta)$$

Untuk memperoleh nilai θ yang memaksimumkan fungsi $l(\theta|y)$, maka $l(\theta|y)$ diturunkan terhadap θ dan kemudian menyamakannya dengan nol seperti pada Persamaan (2.7) berikut.

$$l'(\theta|y) = \frac{\partial l(\theta|y)}{\partial \theta} = 0 \quad (2.7)$$

(Hogg, McKean, & Craig, 2013)

2.5 Algoritma Expectation Maximization

Algoritma Expectation Maximization (EM) pertama kali diperkenalkan oleh Dempster, Laird, dan Rubin pada tahun 1977. Algoritma EM digunakan untuk mengestimasi suatu parameter dengan memaksimumkan fungsi likelihood yang mengandung data hilang atau tidak lengkap (incomplete data). Algoritma EM terdiri dari dua tahap, yaitu tahap ekspektasi dan maksimalisasi. Misalnya diasumsikan terdapat data pengamatan y berdistribusi tertentu yang mengandung data hilang z . Oleh karena itu, dibentuk distribusi gabungan antara y dan z , yaitu:

$$f(y, z | \theta) = f(z) \cdot f(y|z) \quad (2.8)$$

Tahap ekspektasi (*E-Step*) pada algoritma EM dilakukan dengan menghitung ekspektasi dari fungsi *ln-likelihood* dari data hilang berdasarkan data pengamatan yang ada (tidak hilang), yang digunakan untuk mengganti keberadaan data yang dianggap hilang. Oleh karena itu, fungsi Q didefinisikan sebagai berikut:

$$Q(\theta | \theta(0); y) = E[\ln L(\theta(0) | y, z)] \quad (2.9)$$

Pada tahap maksimalisasi (*M-Step*) dilakukan dengan mencari nilai estimator yang dapat memaksimumkan fungsi Q yang telah didefinisikan pada tahap ekspektasi. Metode yang dapat digunakan untuk memaksimumkan fungsi Q adalah metode *Newton-Raphson* yang dilakukan secara numerik. Langkah E dan M dilakukan secara iteratif sampai didapatkan estimator θ yang konvergen, yaitu saat $\|\theta(r+1) - \theta(r)\| \leq \varepsilon$, dengan ε adalah nilai galat yang sangat kecil (Pratama, 2017).

2.6 Regresi Zero-Inflated Poisson

Model regresi ZIP adalah model campuran yang sederhana untuk data diskrit dengan banyak peristiwa nol (Lambert, 1992). Jika merupakan variabel acak bebas yang berdistribusi ZIP, maka nilai nol yang terdapat pada observasi diduga telah terjadi dengan dua cara yang sesuai dengan keadaan (*state*) yang terpisah. Keadaan pertama disebut dengan *zero state* dengan probabilitas dan keadaan kedua disebut dengan *Poisson state* dengan probabilitas. Kedua keadaan memberikan distribusi campuran dua komponen. Fungsi peluang dari model regresi ZIP adalah:

Untuk setiap pengamatan Y_i yang saling bebas $i = 1, \dots, n$ dan

$$Y_i \sim \begin{cases} 0, & \text{dengan peluang } \pi_i \\ \text{Poisson } (\lambda_i), & \text{dengan peluang } (1 - \pi_i) \end{cases}$$

maka :

$$P(Y_i = y_i) = \begin{cases} \pi_i + (1 - \pi_i)e^{-\lambda}, & \text{untuk } y_i = 0 \\ \frac{(1-\pi_i)e^{-\lambda}\lambda^{y_i}}{y_i!}, & \text{untuk } y_i > 0 \end{cases} \quad (2.11)$$

Lambert (1992) menyarankan model gabungan λ_i dan π_i , yaitu

$$\ln(\lambda_i) = x_i^T \beta \text{ dan } \text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = x_i^T \gamma \quad (2.12)$$

Hal ini didasarkan karena komponen berdistribusi poisson, sehingga fungsi linear logaritma.

X adalah matriks variabel yang memuat himpunan-himpunan yang berbeda dari faktor eksperimen yang berhubungan dengan peluang pada *zero state* dan *mean Poisson* pada *Poisson state*, sedangkan β dan γ adalah parameter regresi yang akan ditaksir.

2.7 CONWAY MAXWELL POISSON

Distribusi Conway-Maxwell-Poisson (CMP) pertama kali diperkenalkan oleh Conway dan Maxwell pada tahun 1962 dalam konteks sistem linier, yang kemudian dikenal sebagai generalisasi dari distribusi *Poisson* termasuk Bernoulli dan Geometris sebagai kasus khusus dan fleksibel untuk mengontrol underdispersed dan *Overdispersi*. Distribusi COM-Poisson yang dikembangkan oleh Conway dan Maxwell dengan parameter λ yang sama dengan parameter *Poisson* dan parameter disperse (ϕ). variabel acak $Y \sim \text{COM-Poisson}(\lambda, \phi)$ dan parameter $\lambda > 0$ dan $\phi \geq 0$, bentuk fungsi probabilitas COM-Poisson didefinisikan sebagai berikut ;

$$P(Y = y|\lambda, \phi) = \frac{\lambda^y}{(y!)^\phi Z(\lambda, \phi)}, \text{ untuk } y = 0, 1, 2, \dots \text{ n } \lambda > 0, \phi \geq 0 \quad (2.13)$$

Dimana

$$Z(\lambda, \phi) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\phi}$$

λ : location parameter

ϕ : Parameter Dispersi

$Z(\lambda, \phi)$: Normalisasi Konstan

dimana $\phi > 1$ *underdispersi* dan $\phi < 1$ *Overdispersi*

COM-Poisson untuk setiap generalisasi dari distribusi diskrit yang dapat diamati seperti:

- *Poisson* : dimana ($\phi = 1$), hasil dari

$$Z(\lambda, \phi) = \exp(\lambda) \rightarrow Y \sim Poi(\lambda)$$

$$P(Y = y | \lambda, 1) = \frac{e^{-\lambda} \lambda^y}{y!} = \frac{\lambda^y}{y! e^\lambda}$$

- Geometric : dimana ($\lambda < 1, \phi = 0$), hasil dari

$$Z(\lambda, \phi) = \sum_{j=0}^{\infty} \lambda^j = \frac{1}{1-\lambda} \rightarrow Y \sim Geo(1-\lambda)$$

$$P(Y = y | \lambda < 1, \phi = 0) = \lambda^y (1-\lambda)$$

- Bernoulli : dimana ($\phi \rightarrow \infty$), hasil dari

$$Z(\lambda, \phi) = 1 + \lambda \rightarrow Y \sim Ber\left(\frac{\lambda}{1+\lambda}\right)$$

$$P(Y = y | \lambda, \phi \rightarrow \infty) = \frac{\lambda}{1+\lambda}$$

Selanjutnya, dengan mensubstitusikan persamaan (2.12) kedalam persamaan (2.14) diperoleh seperti persamaan sebagai berikut

$$P(Y = y | \lambda, \phi) = \frac{[\exp(x_i^T \beta)]^{y_i}}{(y_i!)^\phi Z(\lambda_i, \phi)} \quad (2.14)$$

2.8 Regresi zero-inflated Conway Maxwell Poisson (ZICMP)

Sellers and Raim (2016) memperkenalkan model ZICMP, yang pada dasarnya menggantikan distribusi *Poisson* dalam model ZIP dengan distribusi *Conway-Maxwell-Poisson* (CMP) (Conway dan Maxwell, 1962; Shmueli et al., 2005). Keuntungan dari ZICMP adalah dapat menangani tidak hanya Overdispersi tetapi juga *underdispersi* dalam hitungan. ZICMP sangat fleksibel karena memungkinkan dispersi ke segala arah.

Distribusi ZICMP merupakan hasil pencampuran distribusi CMP dan distribusi degenerasi pada nol. Untuk setiap pengamatan Y_i yang saling bebas $i = 1, \dots, n$ dan

$$Y_i \sim \begin{cases} 0, & \text{dengan peluang } \pi_i \\ CMP(\lambda_i \phi_i), & \text{dengan peluang } (1 - \pi_i) \end{cases}$$

Untuk *excess zeros*, parameter π_i ditambahkan dalam model *zero inflated*, dimana π_i adalah probabilitas *excess zeros* dan $(1 - \pi_i)$ adalah probabilitas perhitungan yang dihasilkan dari distribusi CMP. Fungsi peluang model regresi *zero-inflated Conway Maxwell Poisson* (ZICMP) dapat dinyatakan sebagaimana persamaan (2.15).

$$P(Y_i = y_i) = \begin{cases} \pi_i + \frac{(1-\pi_i)}{Z(\lambda, \phi)}, & \text{untuk } y_i = 0 \\ (1 - \pi_i) \frac{\lambda^{y_i}}{(y_i!)^\phi Z(\lambda, \phi)}, & \text{untuk } y_i > 0 \end{cases} \quad (2.15)$$

dimana $0 \leq \pi_i \leq 1, \phi_i \geq 0, \phi$ adalah parameter dispersi dan $Z(\lambda_i, \phi) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\phi}$ menormalkan distribusi. Diasumsikan bahwa λ_i dan π_i bergantung pada vektor dari variabel prediktor x_i , yang dapat didefinisikan sebagai berikut.

a. Model data *count* untuk λ_i , adalah

$$\ln(\lambda_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \lambda_i \geq 0, i = 1, 2, \dots, n. \quad (2.16)$$

b. Model *zero-inflation* untuk π_i , adalah

$$\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i^T \boldsymbol{\gamma} \text{ atau } \pi_i = \frac{e^{\mathbf{x}_i^T \boldsymbol{\gamma}}}{1+e^{\mathbf{x}_i^T \boldsymbol{\gamma}}} \quad (2.17)$$

di mana :

p : jumlah variabel prediktor

n : jumlah pengamatan

$\boldsymbol{\beta}$: parameter model regresi ZICMP yang di estimasi

$\boldsymbol{\gamma}$: parameter model regresi ZICMP yang di estimasi

2.9 Pengujian kesesuaian model

Pengujian kesesuaian model regresi *Poisson* adalah dengan menggunakan *Likelihood Ratio* (LR), dengan prosedur pengujian,

Hipotesis :

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = \gamma_1 = \gamma_2 = \dots = \gamma_p = 0$$

H_1 : paling sedikit ada satu $\beta_j \neq 0$ atau $\gamma_j \neq 0$, dengan $j = 1, 2, 3, \dots, p$

Statistika uji :

$$G = -2 \ln \left[\frac{L_0}{L_1} \right] = -2(\ln L_0 - \ln L_1) \quad (2.18)$$

dengan L_0 adalah likelihood dengan tanpa variabel bebas

L_1 adalah likelihood dengan variabel bebas

Kriteria uji : Tolak H_0 pada taraf signifikansi α jika $G_{hitung} > \chi_{\alpha; 2p}^2$

2.10 Pengujian Signifikansi Parameter

Menurut Myers (2010; 218) pengujian signifikansi dibagi menjadi 2 model yaitu:

2.10.1 Pengujian Signifikansi Parameter β

Pengujian signifikansi parameter model $\ln(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$ dengan $i = 1, \dots, n$ untuk model regresi ZICM

Hipotesis :

$$H_0: \boldsymbol{\beta}_j = 0$$

$$H_1: \boldsymbol{\beta}_j \neq 0$$

Statistika uji :

$$W_j = \left(\frac{\hat{\beta}}{SE(\hat{\beta})} \right)^2 \quad (2.19)$$

Kriteria uji :

Tolak H_0 pada taraf signifikansi α jika $W_j > X_{\alpha;1}^2$ pada taraf alpha.

Penolakan H_0 mengindikasikan bahwa penjelas X_{ij} memiliki pengaruh terhadap peubah respon Y pada taraf signifikansi

2.10.2 Pengujian Signifikansi Parameter γ

Pengujian signifikansi parameter model $\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i^T \boldsymbol{\gamma}$ dengan $i = 1, \dots, n$

Hipotesis :

$H_0: \boldsymbol{\gamma}_j = 0$ (tidak ada pengaruh yang signifikan antara variabel bebas terhadap variabel terikat)

$H_1: \boldsymbol{\gamma}_j \neq 0$ (ada pengaruh yang signifikan antara variabel bebas terhadap variabel terikat)

Statistika uji :

$$W_j = \left(\frac{\hat{\gamma}}{SE(\hat{\gamma})} \right)^2 \quad (2.20)$$

Kriteria uji :

Tolak H_0 pada taraf signifikansi α jika $W_j > X_{\alpha;1}^2$ pada taraf alpha.

Penolakan H_0 mengindikasikan bahwa penjelas X_{ij} memiliki pengaruh terhadap peubah respon Y pada taraf signifikansi

2.11 Kebaikan Model

Menurut Akaike (1978), nilai AIC (Akaike Information Criterion) dapat digunakan untuk pemilihan model terbaik. Nilai AIC dihitung berdasarkan nilai maximum likelihood dan jumlah parameter pada model regresi yang terbentuk. Persamaan untuk menghitung nilai AIC dinyatakan sebagaimana persamaan (2.21).

$$AIC = -2\ln(\text{maximum likelihood}) + 2(\text{number of parameter}) \quad (2.21)$$

Model regresi terbaik adalah model regresi yang mempunyai nilai AIC terkecil.

2.12 Penyakit Difteri

Difteri adalah salah satu penyakit infeksi akut yang sangat menular. Penyakit Difteri disebabkan oleh bakteri *Corynebacterium diphtheriae*. Bakteri ini menginfeksi tenggorokan, saluran udara bagian atas. Karakteristik utama penyakit ini adalah sakit tenggorokan, demam, dan kelenjar bengkak di leher serta bisa menyebabkan miokarditis atau neuropati perifer (WHO, 2018). Menurut Burkovski (2014), Difteri adalah infeksi klasik yang biasanya menyebar dari orang ke orang lain melalui droplet pernafasan oleh orang yang terinfeksi berupa batuk atau bersin. Semua golongan umur bias tertular, namun anak usia kurang dari 5 tahun dan orang tua diatas 60 tahun sangat beresiko tertular penyakit Difteri.

Di Indonesia wabah Difteri muncul kembali sejak tahun 2001 di Cianjur, Semarang, Tasikmalaya, Garut, dan Jawa Timur dengan *CFR* 11,7-31,9%. Pada tahun 2012 telah terjadi 1192 kasus dengan 76 kematian. Dari 18 provinsi yang melaporkan adanya kasus Difteri, kasus tertinggi terjadi di Jawa Timur sebanyak 954 kasus. Di Kabupaten Bojonegoro kasus dan kematian akibat Difteri sangat fluktuatif. Penyebaran kasus Difteri di Kabupaten Bojonegoro berawal pada tahun 2009 ditemukan 4 kasus. Pada tahun 2010-2012 kasus Difteri mengalami peningkatan dengan di temukan 1 kasus kematian di tahun 2012, akan tetapi pada tahun 2013-2014 kasus Difteri mengalami penurunan dan terjadi 1 kasus kematian pada periode 2 tahun berturut-turun. Di tahun 2015 kasus Difteri mengalami peningkatan kembali dari tahun sebelumnya, lalu mengalami penurunan hingga 1 kasus di tahun 2017 dan meninggal (Kiki, 2017).

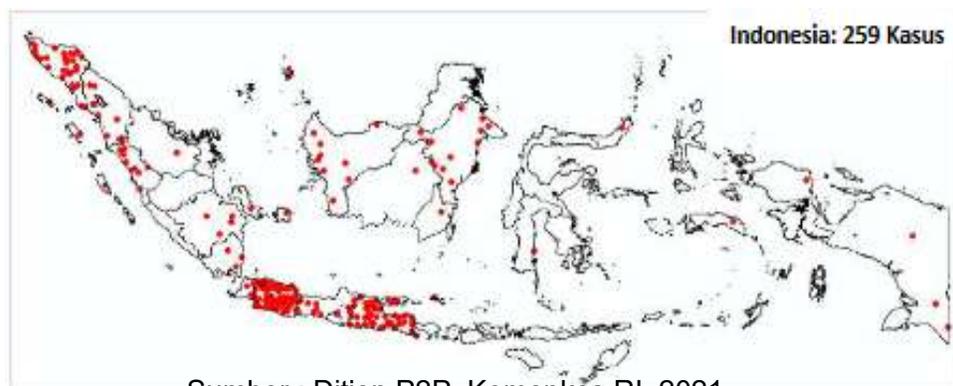
2.15.1 Definisi Penyakit Difteri

Menurut kementerian kesehatan (2020) penyakit Difteri adalah salah satu penyakit yang sangat menular yang dapat dicegah dengan imunisasi, disebabkan oleh

bakteri *Corynebacterium diphtheriae strain toksigenik*. Manusia adalah satu-satunya *reservoir Corynebacterium diphtheriae*. Penularan terjadi secara droplet (percikan ludah) dari batuk, bersin, muntah, melalui alat makan, atau kontak erat langsung dari lesi di kulit. Apabila tidak diobati dan kasus tidak mempunyai kekebalan, angka kematian sekitar 50%, sedangkan dengan terapi angka kematiannya sekitar 10% (CDC Manual for the Surveilan of Vaccine Preventable Diseases, 2017). Angka kematian Difteri rata rata 5 – 10% pada anak usia kurang 5 tahun dan 20% pada dewasa diatas 40 tahun (CDC Atlanta, 2016).

Kasus Difteri pada tahun 2020 menyebar di hampir semu wilayah di Indonesia. Jumlah kasus Difteri pada tahun 2020 sebanyak 259 kasus, jumlah kematian sebanyak 13 kasus, dengan CFR sebesar 5,02%. Jumlah kasus Difteri tahun 2020 mengalami penurunan yang cukup signifikan jika dibandingkan tahun 2019 (529 kasus). Jumlah kematian akibat Difteri juga mengalami penurunan jika dibandingkan tahun sebelumnya (23 kasus). Tetapi, CFR pada tahun 2020 meningkat jika dibandingkan dengan tahun 2019 (4,5%). Berdasarkan provinsi, jumlah kasus terbanyak terdapat di Jawa Barat yakni sebanyak 75 kasus. Sementara itu, pada tahun 2020 terdapat 12 provinsi tidak ditemukan kasus Difteri.

Gambar 2.1 : Sebaran Kasus Difteri Menurut Provinsi Tahun 2020



Difteri merupakan jenis penyakit menular yang dapat menimbulkan KLB (Kejadian Luar Biasa)/Wabah seperti tercantum dalam Permenkes 1501 tahun 2010. Kegiatan penanggulangan KLB Difteri dilakukan dengan melibatkan program-program terkait yaitu *surveilans epidemiologi*, program imunisasi, klinisi, laboratorium dan program kesehatan lainnya serta lintas sektor terkait. KLB Difteri terjadi jika suatu

wilayah kab/kota dinyatakan KLB Difteri jika ditemukan satu suspek Difteri dengan konfirmasi laboratorium kultur positif atau jika ditemukan suspek Difteri yang mempunyai hubungan epidemiologi dengan kasus kultur positif.

Kebijakan dalam penanggulangan KLB Difteri antara lain:

1. Setiap Kejadian Luar Biasa (KLB) harus dilakukan penyelidikan dan penanggulangan sesegera mungkin untuk menghentikan penularan dan mencegah komplikasi dan kematian
2. Dilakukan tatalaksana kasus di rumah sakit dengan menerapkan prinsip kewaspadaan seperti menjaga kebersihan tangan, penempatan kasus di ruang tersendiri /isolasi, dan mengurangi kontak erat kasus dengan orang lain
3. Setiap suspek Difteri dilakukan pemeriksaan laboratorium dengan kultur
4. Setiap kontak erat diberi kemoprofilaksis
5. Kontak erat diberikan imunisasi pada saat penyelidikan epidemiologi
6. Pengambilan spesimen pada kontak erat dapat dilakukan jika diperlukan sesuai dengan kajian epidemiologi
7. Setiap suspek Difteri dilakukan ORI (respon pemberian imunisasi pada KLB) sesegera mungkin, sebaiknya luas wilayah ORI dilakukan untuk satu (1) kab/kota tetapi jika tidak memungkinkan karena sesuatu hal maka ORI minimal dilakukan satu (1) kecamatan dengan sasaran sesuai kajian epidemiologi dan interval ORI 0-1-6 bulan
8. ORI dilanjutkan sampai selesai walaupun status KLB Difteri di suatu wilayah kabupaten/kota dinyatakan telah berakhir

2.15.2 Faktor Risiko Penyakit Difteri

Difteri disebabkan oleh bakteri *Corynebacterium diphtheriae*, yang dapat menular melalui partikel di udara, benda pribadi, peralatan rumah tangga yang terkontaminasi, serta menyentuh luka yang terinfeksi kuman Difteri. Selain penularan Difteri juga bisa terjadi melalui air liur seseorang. Contoh penularan tersebut adalah berbagi makanan atau minuman kepada orang yang terjangkit Difteri atau melakukan kontak fisik yang melibatkan air liur, dengan pengidap penyakit Difteri (WHO, 2019).

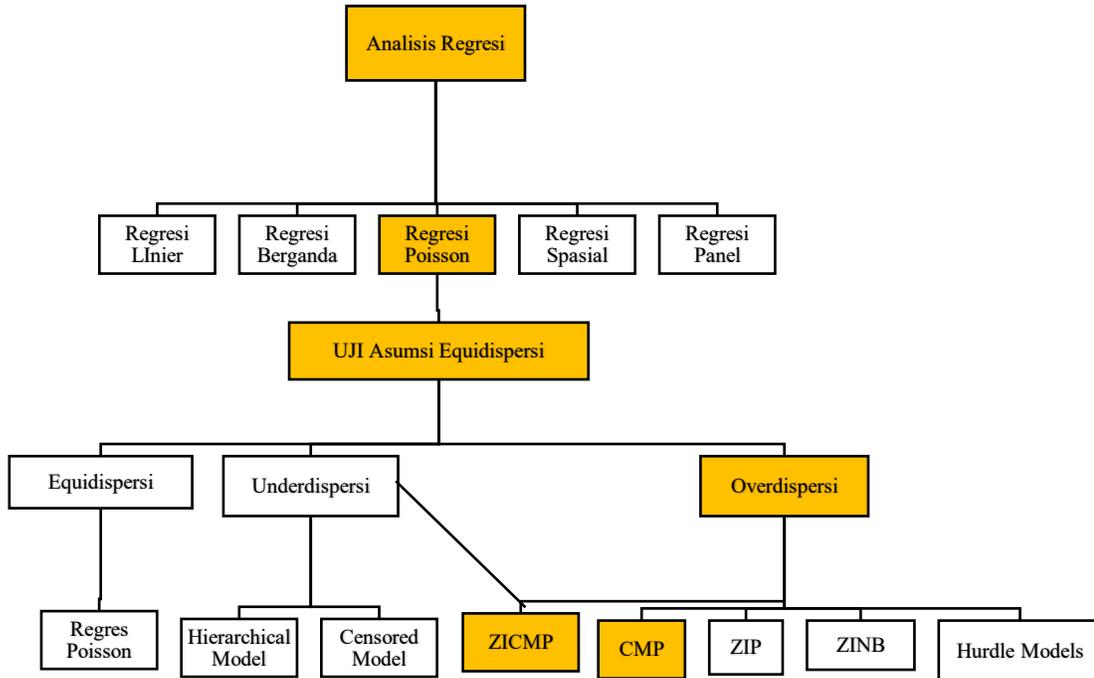
Menurut Markus (2018) faktor risiko penyakit Difteri dapat ditinjau dari beberapa faktor antara lain:

1. Anak-anak di bawah usia 5 tahun dan orang tua di atas usia 60 tahun.
2. Belum mendapatkan vaksinasi Difteri
3. Berkunjung ke daerah dengan cakupan imunisasi Difteri yang rendah
4. System kekebalan tubuh yang lemah, seperti penderita HIV/AIDS
5. Gaya hidup yang tidak sehat
6. Lingkungan dengan kebersihan dan sanitasi yang buruk

Faktor risiko penyakit Difteri di Provinsi Jawa Timur tahun 2010 dan 2011 berhubungan dengan cakupan imunisasi DPT dan cakupan imunisasi DT (Izza, 2015). Faktor risiko penyakit Difteri di Kabupaten Blitar pada Tahun 2015 adalah kelembaban ruangan dan pencahayaan, status imunisasi, ventilasi dan keberadaan sarana pelayanan kesehatan (Nanang, 2016). Di kota Padang ditemukan faktor risiko penyakit Difteri disebabkan oleh status imunisasi DPT dan lingkungan rumah (Ridha, 2017).

Pada penelitian yang telah dikemukakan diatas, disimpulkan bahwa ada banyak faktor yang menyebabkan risiko penyakit Difteria diantaranya dari partikel udara, peralatan rumah yang terkontaminasi, persentase kemiskinan dan masih banyak lagi.

2.13 Kerangka Konseptual



Gambar 2.2. Kerangka Konseptual Pemodelan Regresi *Zero inflated Conway Maxwell Poisson*