MODEL REGRESI POISSON TERBAIK MENGGUNAKAN *ZERO-INFLATED POISSON* (ZIP) DAN *ZERO-INFLATED NEGATIVE BINOMIAL* (ZINB)

Skripsi

disusun sebagai salah satu syarat

untuk memperoleh gelar Sarjana Sains

Program Studi Matematika

oleh

Ilham Kurniawan
4111413010

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI SEMARANG

2017

PERNYATAAN

Saya menyatakan bahwa skripsi ini bebas plagiat, kecuali yang secara tertulis dirujuk dalam skripsi ini dan disebutkan dalam daftar pustaka. Apabila dikemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan peraturan perundang-undangan.

Semarang, Juli 2017

Ilham Kurniawan

4111413010

UNNES
UNIVERSITAS NEGERI SEMARANG

ii

PENGESAHAN

Skripsi yang berjudul

Model Regresi Poisson Terbaik Menggunakan *Zero-Inflated Poisson* (ZIP) dan *Zero-Inflated Negative Binomial* (ZINB)

disusun oleh

Ilham Kurniawan

4111413010

telah dipertahankan dihadapan sidang Panitia Ujian Skripsi FMIPA UNNES pada tanggal Juli 2017

Panitia,

Ketua	Sekretaris
Prof. Dr. Zaenuri, S.E., M.Si., Akt.	Drs. Arief Agoestanto, M.Si
NIP. 196412231988031001	NIP. 196807221993031005

Ketua Penguji

Dra. Sunarmi, M.Si

NIP. 195506241988032001

Anggota Penguji/ Pembimbing I	Anggota Penguji/ Pembimbing II
Dr. Scolastika Mariani, M.Si	Drs. Arief Agoestanto, M.Si
NIP. 196502101991022001	NIP. 196807221993031005

iii

MOTTO DAN PERSEMBAHAN

MOTTO

Pendidikan merupakan perlengkapan paling baik untuk hari tua (Aristoteles).

Sesungguhnya bersama setiap kesulitan itu ada kemudahan (Q.S.Al Insyirah: 6).

Keberhasilan adalah kemampuan untuk melewati dan mengatasi dari satu kegagalan ke kegagalan berikutnya tanpa kehilangan semangat (Winston Chuchill).

PERSEMBAHAN

1. Untuk kedua orang tua saya Ibu Khunaenah dan Bapak Heru Krisharyanto Alm.
2. Untuk adik-adikku Fahmi Djuniar dan Widya Anggraeni.
3. Untuk keluarga besar tercinta.
4. Untuk teman-teman Matematika angkatan 2013.
5. Untuk teman-teman PKL BPS Kota Pekalongan.
6. Untuk teman-teman KKN Desa Kalimanggis Temanggung.
7. Untuk Almamater tercinta Universitas Negeri Semarang (UNNES).
8. Untuk teman-teman DANUS FMI, JODY, SIGMA, volunteer FIM, dan NEW REGGAB.

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT yang Maha Pengasih dan Penyayang, atas limpahan karunia-Nya dan sholawat serta salam selalu tercurah atas Nabi Muhammad SAW hingga akhir zaman. Pada kesempatan ini, penulis dengan penuh syukur mempersembahkan skripsi yang berjudul "**MODEL REGRESI POISSON TERBAIK MENGGUNAKAN *ZERO-INFLATED POISSON* (ZIP) DAN *ZERO-INFLATED NEGATIVE BINOMIAL* (ZINB)**".

Penulis menyadari dalam menyelesaikan skripsi ini memperoleh banyak bantuan dan dukungan dari berbagai pihak. Untuk itu, dengan rasa hormat, penulis menyampaikan terima kasih kepada:

1. Prof. Dr. Fathur Rokhman, M.Hum., selaku Rektor Universitas Negeri Semarang;
2. Prof. Dr. Zaenuri S.E, M.Si,Akt., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang;
3. Drs. Arief Agoestanto, M.Si., selaku Ketua Jurusan Matematika FMIPA Universitas Negeri Semarang, sekaligus selaku Dosen Pembimbing II yang telah memberikan bimbingan, pengarahan, dan saran-saran selama penyusunan skripsi ini.
4. Drs. Mashuri M.Si., selaku Ketua Program Studi Matematika Jurusan Matematika FMIPA Universitas Negeri Semarang;
5. Dr. Rochmad M.Si., selaku Dosen Wali Program Studi Matematika Jurusan Matematika FMIPA Universitas Negeri Semarang angkatan 2013;

6. Dr. Scolastika Mariani, M.Si., selaku Dosen Pembimbing I yang telah memberikan bimbingan, pengarahan dan saran-saran selama penyusunan skripsi ini;
7. Dra. Sunarmi, M.Si., selaku Dosen Penguji yang telah memberikan penilaian dan saran dalam perbaikan skripsi ini;
8. Bapak dan Ibu Dosen Jurusan Matematika Universitas Negeri Semarang yang telah membekali penulis dengan ilmu selama mengikuti perkuliahan dan penulisan skripsi;
9. Orang tua dan keluarga yang selalu memberikan doa, dukungan, dan semangat;
10. Teman-teman seperjuangan Matematika angkatan 2013 yang selalu menghibur, memberikan motivasi, dorongan semangat dan doa;
11. Semua pihak yang tidak dapat disebutkan satu per satu yang telah memberikan bantuan.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan. Oleh karena itu, penulis mengharapkan saran dan kritik yang membangun dari pembaca.

UNNES
UNIVERSITAS NEGERI SEMARANG

Semarang, Juli 2017

Penulis

ABSTRAK

Kurniawan, Ilham. 2017. *Model Regresi Poisson Terbaik Menggunakan Zero-Inflated Poisson (ZIP) dan Zero-Inflated Negative Binomial (ZINB)*. Skripsi, Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Utama Dr. Scolastika Mariani, M.Si. dan Pembimbing Pendamping Drs. Arief Agoestanto, M.Si.

Kata Kunci: Overdispersi, Regresi Poisson, *Zero-Inflated Poisson* (ZIP), *Zero-Inflated Negative Binomial* (ZINB).

Tujuan dalam penelitian ini adalah mengetahui bentuk model *Zero-Inflated Negative Poisson* (ZIP) dan *Zero-Inflated Negative Binomial* (ZINB) dalam regresi Poisson yang terjadi pelanggaran asumsi equidispersi yaitu berupa overdispersi dengan data berupa kematian balita Puskesmas Tirto Kota Pekalongan tahun 2016 serta mengetahui model terbaik diantara keduanya. Penelitian ini melakukan estimasi dengan metode *maximum likelihood* disertai dengan algoritma *expectation maximization,* dilanjutkan dengan pengujian kesesuaian model dengan menggunakan *likehood ratio test* dan pengujian signifikansi parameter dengan uji Wald dan dilakukan pemilihan model terbaik menggunakan *Akaike Information Criterion* (AIC) dengan mengambil nilai AIC yang terkecil. Software yang digunakan dalam penelitian ini adalah program R 2.10.0.

Kesimpulan yang diperoleh dari penelitian ini model ZIP adalah $ln(\mu) = -1{,}75492 + 0{,}10435X_3$ dan $logit(\omega) = -13{,}49422 + 0{,}66060X_3$ dengan nilai AIC$= 101{,}1387$ dan model ZINB adalah $ln(\mu) = -1{,}75491 + 0{,}10435X_3$ dan $logit(\omega) = -13{,}49431 + 0{,}66060X_3$ dengan nilai AIC$= 103{,}1392$. Model yang terbaik adalah model ZIP dengan nilai AIC$= 101{,}1387$.

DAFTAR ISI

HALAMAN JUDUL i
HALAMAN PERNYATAAN ii
HALAMAN PENGESAHAN iii
MOTTO DAN PERSEMBAHAN iv
KATA PENGANTAR v
ABSTRAK vii
DAFTAR ISI viii
DAFTAR TABEL x
DAFTAR GAMBAR xi
DAFTAR LAMPIRAN xii
DAFTAR SIMBOL xiii

BAB

1. PENDAHULUAN

1.1 Latar Belakang 1
1.2 Rumusan Masalah 5
1.3 Batasan Masalah 5
1.4 Tujuan Penelitian 6
1.5 Manfaat Penelitian 6
1.6 Sistematika Penulisan Skripsi 7

2. TINJAUAN PUSTAKA

2.1 Distribusi Poisson 9
2.2 Distribusi Keluarga Eksponensial 12
2.3 *Generalized Linear Model* (GLM) 12
2.4 Uji Kecocokan Distribusi 14
2.5 Regresi Poisson 14
2.6 Overdispersi 16
2.7 *Excess zeros* 18
2.8 Metode Maksimum Likelihood 18

UNNES
UNIVERSITAS NEGERI SEMARANG

2.9 Regresi *Zero-Inflated Negative Binomial* (ZINB) 19
2.10Estimasi Parameter Regresi *Zero-Inflated Negative Binomial* (ZINB) 25
2.11Regresi *Zero-Inflated Poisson* (ZIP) 31
2.12Estimasi Parameter Model Regresi *Zero-Inflated Poisson* (ZIP) 32
2.13Pengujian Kesesuaian Model 35
2.14Pengujian Signifikansi Parameter 36
2.13.1 Pengujian Signifikansi Parameter β 36
2.13.2 Pengujian Signifikansi Parameter γ 37
2.15Uji Kelayakan Model 38
2.16Program R 38
2.17Kematian Balita 40
2.16.1 Balita 40
2.16.2 Faktor Penyebab Kematian Balita 41
2.18Kerangka Berpikir 42

3. METODE PENELITIAN
3.1 Fokus Penelitian 48
3.2 Pengumpulan Data 49
3.3 Metode Analisis Data 49
3.4 Penarikan Kesimpulan 52

4. HASIL PENELITIAN DAN PEMBAHASAN
4.1 Hasil penelitian 53
4.1.1 Pengujian OLS (*Ordinary Least Square*) 53
4.1.1.1 Uji Normalitas 53
4.1.1.2 Uji Multikolinearitas 55
4.1.1.3 Uji Heteroskedastisitas 56
4.1.1.4 Uji Autokorelasi 56
4.1.1.5 Uji Linearitas 57
4.1.1.6 Persamaan Regresi Terbaik 58
4.1.2 Pengujian Distribusi Poisson pada Variabel Respon Y 59

4.1.3 Pengujian Asumsi Equidispersi 61
4.1.4 Model Awal Regresi *Zero-Inflated Negative Binomial* (ZINB) ...62
4.1.5 Pengujian Kesesuaian Model Regresi *Zero-Inflated Negative Binomial* (ZINB) 64
4.1.6 Pengujian Signifikansi Parameter Regresi ZINB secara individu 66
4.1.1.1. Pengujian Signifikansi Parameter β 66
4.1.1.2. Pengujian Signifikansi Parameter γ 69
4.1.7 Model Akhir Regresi *Zero-Inflated Negative Binomial* (ZINB) ..72
4.1.8 Model Awal Regresi *Zero-Inflated Poisson* (ZIP) 73
4.1.9 Pengujian Kesesuaian Model Regresi *Zero-Inflated Poisson* (ZIP) 76
4.1.10 Pengujian Signifikansi Parameter Regresi ZIP secara individu 78
4.1.1.1. Pengujian Signifikansi Parameter β 78
4.1.1.2. Pengujian Signifikansi Parameter γ 81
4.1.11 Model Akhir Regresi *Zero-Inflated Poisson* (ZIP) 84
4.1.12 Pemilihan Model Terbaik 85
4.2 Pembahasan 85

5. PENUTUP
5.1 Simpulan 90
5.2 Saran 93
DAFTAR PUSTAKA 94
LAMPIRAN 97

UNNES
UNIVERSITAS NEGERI SEMARANG

DAFTAR TABEL

Tabel Halaman

Tabel 4.1 Hasil uji Kolmogorov Smirnov 54

Tabel 4.2 Hasil uji Multikolinearitas 55

Tabel 4.3 Hasil uji Heteroskedastisitas 56

Tabel 4.4 Hasil uji Autokorelasi 57

Tabel 4.5 Hasil uji Linearitas 58

Tabel 4.5 Hasil uji Regresi 59

Tabel 4.6 Hasil uji Chi Kuadrat 60

Tabel 4.7 Hasil uji asumsi equidispersi.....61

Tabel 4.8 Estmasi model *Zero-Inflated Negative Binomial* (ZINB).....62

Tabel 4.9 Likelihood ratio test ZINB.....65

Tabel 4.10 Estmasi model *Zero-Inflated Poisson* (ZIP) 74

Tabel 4.11 Likelihood ratio test ZIP 77

Tabel 4.12 Nilai AIC ZINB dan ZIP.....85

UNNES
UNIVERSITAS NEGERI SEMARANG

DAFTAR GAMBAR

Gambar Halaman

Gambar 2.1. Model Spesifikasi Regresi Count 43

Gambar 2.2. Generalized Count Regression 43

Gambar 2.3. Metode Estimator 44

Gambar 2.4. Pengujian Hipotesis 44

Gambar 2.5. Kerangka Berpikir 47

Gambar 3.1. Diagram Alur Metode Penelitian 51

UNNES
UNIVERSITAS NEGERI SEMARANG

DAFTAR LAMPIRAN

Lampiran Halaman

Lampiran 1. Data Kematian Balita 97

Lampiran 2. Pengujian distribusi Poisson 98

Lampiran 3. Pengujian Equidispersi 100

Lampiran 4. Lanjutan Pengujian Equidispersi 101

Lampiran 5. Model *Zero-Inflated Negative Binomial* (ZINB) 102

Lampiran 6. Lanjutan Model *Zero-Inflated Negative Binomial* (ZINB) 103

Lampiran 7. Model *Zero-Inflated Poisson* 104

Lampiran 8. Lanjutan Model *Zero-Inflated Poisson* 105

Lampiran 9. Surat Permohonan Izin Observasi 106

Lampiran 10. Surat Rekomendasi Research / Survey BAPPEDA 107

Lampiran 11. Surat Izin Penelitian dari Dinas Kesehatan Kota Pekalongan 108

Lampiran 12. Tabel *Chi-square* 109

UNNES
UNIVERSITAS NEGERI SEMARANG

DAFTAR SIMBOL

Y	: Kematian balita
X_1	: Pneumonia
X_2	: Balita Gizi Buruk
X_3	: Diare
$E(Y)$	: Nilai ekspetasi dari variabel random Y
$Var(Y), V(Y)$	: Varian dari variabel random Y
H	: Matriks Hessian
G	: Statistik uji Rasio Likelihood
W	: Statistik uji Wald
W	: Variabel indikator
α	: Taraf signifikansi
η	: Prediktor linier
χ^2	: Statistik Uji Chi-Square
$g(.)$	: fungsi link
Z	: Variabel indikator
$\prod$	: Perkalian dari gugusan data
U	: Vektor skor efisien
θ	: parameter natural
ϕ	: parameter dispersi
μ	: Nilai rata-rata
$\Gamma(.)$	: fungsi gamma

UNNES
UNIVERSITAS NEGERI SEMARANG

n	: Banyaknya percobaan
p	: Banyakya parameter regresi
β_0, γ_0	: Intersep
β_j, γ_j	: Parameter ke-j dari regresi
$\hat{\beta}, \hat{\gamma}$	: Taksiran parameter regresi
k	: Parameter dispersi distribusi Poisson gamma *mixture*
$f(x)$	: Fungsi densitas peluang x
$f(x_2\|x_1)$	: Fungsi densitas probabilitas bersyarat dari X_2 dengan syarat $X_1 = x_1$
$L(\theta)$	: Fungsi likelihood
$L_c(\theta)$	: Fungsi likelihood campuran
$l(\theta)$	: Fungsi ln-likelihood
$\sum$	: Jumlahan
L_0	: likelihood tanpa variabel bebas
L_1	: likelihood dengan variabel bebas

UNNES
UNIVERSITAS NEGERI SEMARANG

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan suatu metode statistika yang digunakan untuk menganalisis hubungan antara peubah respon (*dependent*) dan peubah penjelas (*independent*). Menurut Ruliana (2016), analisis regresi umumnya digunakan untuk menganalisis data variabel respon yang berupa data kontinu. Dewi (2016), variabel respon adalah variabel yang keberadaannya dipengaruhi oleh variabel lain. Namun dalam beberapa penelitian, data variabel respon dapat berupa data diskrit. Salah satu model regresi yang dapat digunakan untuk menganalisis hubungan antara peubah respon yang berupa data diskrit dan peubah penjelas berupa data kontinu, diskrit atau campuran adalah model regresi Poisson.

Pada model regresi Poisson dalam analisis terdapat beberapa syarat yang harus dipenuhi, salah satunya keadaan yang equidispersi yaitu nilai *mean* dan varians dari variabel respon sama. Terkadang dalam analisis model regresi Poisson terjadi pelanggaran asumsi tersebut. Ketika nilai varians lebih besar dari nilai *mean* disebut overdispersi, sedangkan ketika nilai varians lebih kecil dari nilai *mean* disebut underdispersi. Jika variabel respon yang digunakan merupakan peubah acak diskret yang berdistribusi Poisson, maka dapat digunakan model regresi Poisson untuk pembentukan model regresi. Pada kenyataannya tidak sepenuhnya asumsi tersebut terpenuhi, seperti nilai varian lebih besar dari nilai rata – ratanya yang disebut overdispersi.

Overdispersi yang disebabkan oleh banyaknya nilai nol yang berlebih pada variabel respon (*excess zeros*) pada dasarnya tetap dapat diestimasi menggunakan regresi Poisson. Namun, untuk data yang banyak mengandung nilai nol memerlukan adanya metode tertentu untuk mengatasinya. Jika regresi poisson tetap digunakan maka estimasi parameternya kurang baik dalam menaksir kelebihan nol tersebut. Hal ini menyebabkan adanya pengembangan model – model statistik untuk mengatasi masalah tersebut diantaranya adalah model regresi *Zero-Inflated Poisson* (ZIP) dan model regresi *Zero-Inflated Negative Binomial* (ZINB).

Menurut Cahyandari (2014), Yulianingsih *et al.*(2012), Analisis regresi untuk data diskrit yaitu dengan menggunakan regresi Poisson. Namun dalam kenyataannya data yang diteliti sering kali tidak memenuhi asumsi equidispersi pada regresi Poisson. Untuk mengatasi masalah overdispersi pada regresi Poisson maka Mouatassim & Ezzahid (2012), Loeys *et al.*(2012), Zamani & Ismail (2013) dan Long *et al.*(2014) mengusulkan untuk melakukan pemodelan Zero-Inflated Poisson. Saffari & Adnan (2012), Sharma & Landge (2013), dan Chipeta *et al.* (2014) melakukan pemodelan *Zero-Inflated Negative Binomial* untuk mengatasi masalah overdispersi pada regresi Poisson.

Adanya masalah overdispersi mengakibatkan kesalahan dalam memodelkan regresi Poisson, model tersebut memiliki beberapa konsekuensi yaitu dalam hal tingkat akurasi model yang rendah dalam mendeteksi besar pengaruh antar variabel bebas dengan variabel terikat. Menurut Mousatassim & Ezzahid (2012) menyatakan bahwa beberapa penyebab terjadinya overdispersi adalah terlalu

banyak nilai nol (*excess zero*) pada variabel respon dan adanya keheterogenan bernilai nol lebih besar dari pada banyaknya harapan amatan nol berdasarkan sebaran poisson.

Nilai varian yang lebih besar dari nilai rata- rata atau sering dikenal overdispersi mengakibatkan perlunya metode dalam penanganan overdispersi. Masalah overdispersi sering kali diabaikan dalam analisis regresi Poisson. Amatan data yang banyak nilai nol memiliki arti penting dalam penelitian khususnya penelitian tentang regresi Poisson.

Beberapa penelitian terkait permasalahan regresi Poisson dan beberapa aplikasinya dari waktu ke waktu selalu mengalami perkembangan. Menurut Mouatassim & Ezzahid (2012) menyatakan bahwa *Zero-Inflated Poisson* digunakan untuk menangani masalah overdispersi. *Zero-Inflated Poisson* dengan menggunakan *Maximum Likelihood Estimator*.

Metode *Maximum Likelihood Estimator* (MLE) adalah metode yang paling sering digunakan dalam mengestimasi parameter. Metode MLE adalah yang baik untuk memperoleh sebuah estimasi tunggal. Metode ini memiliki peran dalam estimasi yang baik dengan melakukan estimasi sampai diperoleh hasil akhir yang konvergen.

UNNES
UNIVERSITAS NEGERI SEMARANG

Menurut Sharma & Landge (2013) menyatakan bahwa kejadian kecelakaan lalu lintas dapat digunakan *Zero-Inflated Negative Binomial*. Uji performa model digunakan nilai *Akaike Information Criterion* (AIC). Menurut Saffari dan Adnan (2012) menyatakan penggunaan *Maximum Likelihood Estimator (MLE)* dan *Akaike Information Criterion (AIC)* dalam *Zero-Inflated Negative Binomial*. AIC

digunakan untuk pemilihan model antara *Zero-Inflated Poisson* dan *Zero-Inflated Negative Binomial* berdasarkan nilai AIC yang terkecil. AIC bertujuan untuk mempermudah menentukan model yang terbaik.

Penelitian ini menerapkan model *Zero-Inflated Poisson* dan *Zero-Inflated Negative Binomial* dalam kasus kematian balita di Puskesmas Tirto Kota Pekalongan tahun 2016. Sehingga variabel respon yang digunakan adalah banyaknya kematian balita di Puskesmas Tirto Kota Pekalongan tahun 2016 dan variabel prediktor yang digunakan yaitu balita yang terkena diare, pneumonia, dan balita gizi buruk.

Balita merupakan masa emas atau "*golden age*" insan manusia yang berusia 0-5 tahun (UU No.20 Tahun 2003). Balita memiliki peranan penting dalam penentuan tingkat kesehatan masyarakat karena dapat menggambarkan kesehatan penduduk secara umum. Salah satu tingkat kesehatan dapat dilihat dari tingkat kematian balita. Menurut Dinas Kesehatan Provinsi Jawa Tengah, Angka Kematian Balita (AKABA) pada tahun 2015 meningkat menjadi 11,64 per 1000 kelahiran hidup sedangkan pada tahun 2014 sebesar 11,54 per 1000 kelahiran hidup. Kematian balita di Indonesia memberikan efek yang sangat signifikan terhadap perubahan akan kelangsungan hidup pada jaman selanjutnya, karena balita adalah generasi penerus bangsa.

Menurut WHO, setiap tahun lebih dari sebelas juta anak meninggal karena menderita sakit dan kurang gizi. Di beberapa negara, satu atau lebih dari lima anak meninggal sebelum mencapai usia lima tahun. Penyebab kematian anak balita di negara berkembang disebabkan oleh pneumonia, diare dan kurang gizi.

Menurut Dinas Kesehaan Provinsi Jawa Tengah di kota Pekalongan jumlah diare pada tahun 2015 sebesar 9.316 meningkat dibandingkan tahun 2014 sebesar 8.084, pneumonia pada tahun 2015 sebesar 1.558 menurun dibandingkan pada tahun 2015 sebesar 1.630 dan gizi buruk pada tahun 2014 sebesar 52 sedangkan pada tahun 2015 mengalami penurunan menjadi 7. Walaupun di Kota Pekalongan terjadi penurun jumlah kematian balita pada tahun 2015 sebesar 74 dibandingkan pada tahun 2014 sebesar 75.

Dengan latar belakang di atas maka judul yang akan dikaji dalam skripsi ini adalah "Model Regresi Poisson Terbaik Menggunakan *Zero-Inflated Poisson* (ZIP) dan *Zero-Inflated Negative Binomial* (ZINB)".

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, maka permasalahan yang yang akan dibahas pada penelitian ini adalah:

1. Bagaimana model *Zero-Inflated Poisson* (ZIP) pada regresi Poisson dalam kasus kematian balita di Puskesmas Tirto Kota Pekalongan?
2. Bagaimana model *Zero-Inflated Negative Binomial* (ZINB) pada regresi Poisson dalam kasus kematian balita di Puskesmas Tirto Kota Pekalongan?
3. Manakah diantara *Zero-Inflated Poisson* (ZIP) dan *Zero-Inflated Negative Binomial* (ZINB) yang menghasilkan model terbaik pada regresi Poisson dalam kasus kematian balita di Puskesmas Tirto Kota Pekalongan?

1.3 Batasan Masalah

Agar pembahasan dalam penelitian ini tidak meluas, maka penelitian ini memberikan batasan – batasan yaitu sebagai berikut:

1. Model yang digunakan adalah *Zero-Inflated Poisson* (ZIP) dan *Zero-Inflated Negative Binomial* (ZINB).
2. Penelitian ini menggunakan bantuan program R.2.10.0.
3. Data yang digunakan dalam penelitian ini adalah data dari Dinas Kesehatan Puskemas Tirto Kota Pekalongan berupa data kematian balita, pneumonia, balita gizi buruk, diare.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini antara lain yaitu

1. Mengetahui model *Zero-Inflated Poisson* (ZIP) pada regresi Poisson dalam kasus kematian balita di Puskesmas Tirto Kota Pekalongan.
2. Mengetahui model *Zero-Inflated Negative Binomial* (ZINB) pada regresi Poisson dalam kasus kematian balita di Puskesmas Tirto Kota Pekalongan.
3. Mengetahui antara *Zero-Inflated Poisson* dan *Zero-Inflated Negative Binomial* yang dihasilkan model terbaik pada regresi Poisson dalam kasus kematian balita di Puskesmas Tirto Kota Pekalongan.

1.5 Manfaat Penelitian

1.5.1 Bagi Mahasiswa Jurusan Matematika UNNES

Menambah wawasan mengenai penerapan matematika dengan model regresi Poisson khususnya mengatasi masalah overdispersi dengan model *Zero-Inflated Poisson* (ZIP) dan *Zero-Inflated Negative Binomial* (ZINB).

1.5.2 Bagi Dinas Instansi

Memberikan masukan kepada Dinas Kesehatan Kota Pekalongan khususnya Puskemas Tirto mengenai hal – hal yang mempengaruhi kematian balita paling dominan sehingga dapat diambil tindakan pencegahan untuk kedepannya.

1.6 Sistematika Penulisan Skripsi

Secara garis besar skripsi ini dibagi menjadi tiga bagian (bab) yaitu bagian awal skripsi, bagian isi skripsi dan bagian akhir skripsi. Berikut ini dijelaskan masing – masing bagian skripsi.

(1) Bagian awal skripsi

Bagian awal skripsi meliputi halaman judul, pernyataan keaslian tulisan, pengesahan, motto, dan persembahan, kata pengantar, abstrak, daftar isi, daftar gambar, daftar tabel, dan daftar lampiran.

(2) Bagian isi skripsi

Bagian isi skripsi secara garis besar terdiri dari lima bab, yaitu :

BAB 1 PENDAHULUAN

Bab ini berisi mengenai latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan skripsi.

BAB 2 TINJAUAN PUSTAKA

Bab ini berisi kajian teori yang mendasari dan berhubungan dengan pemecahan masalah. Teori – teori tersebut digunakan untuk memecahkan masalah yang diangkat dalam skripsi ini.

BAB 3 METODE PENELITIAN

Bab ini mengulas metode yang digunakan dalam penelitian yang berisi langkah-langkah yang dilakukan untuk memecahkan masalah yaitu pengumpulan data, analisis data dan kesimpulan.

BAB 4 HASIL PENELITIAN DAN PEMBAHASAN

Bab ini berisi mengenai penyelesaian dari permasalahan yang diungkapkan

BAB 5 PENUTUP

Bab ini berisi tentang simpulan dari pembahasan dan saran yang berkaitan dengan simpulan

(3) Bagian akhir skripsi

Bagian akhir skripsi meliputi daftar pustaka yang memberikan informasi tentang buku sumber serta literature yang digunakan dan lampiran – lampiran yang mendukung skripsi.

UNNES
UNIVERSITAS NEGERI SEMARANG

BAB 2

TINJAUAN PUSTAKA

2.1 Distribusi Poisson

Menurut Harinaldi (2005:87), dalam eksperimen Poisson yaitu probabilitas memperoleh dengan tepat peristiwa Y sebanyak y kejadian untuk setiap satu satuan unit (waktu dan ruang) tersebut. Ciri – ciri distribusi Poisson antara lain :

1. Banyaknya hasil percobaan yang terjadi dalam suatu interval waktu atau suatu daerah tertentu tidak bergantung pada banyaknya hasil percobaan yang terjadi pada interval waktu atau daerah lain yang terpisah.
2. Probabilitas terjadinya hasil percobaan selama 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 yang terjadi di luar interval waktu atau daerah tersebut

Misalkan y_i, $i = 1, 2, \ldots$ merupakan jumlah kejadian yang muncul dalam selang waktu dengan rata-rata μ_i. Jika Y adalah variabel acak Poisson dengan parameter $\mu > 0$, maka fungsi massa peluang Poisson adalah :

$$f(y, \mu) = \frac{\mu^y e^{-\mu}}{y!} \qquad y = 0,1,2, \ldots \qquad (1)$$

μ menyatakan rata-rata banyaknya sukses yang terjadi dalam selang waktu atau daerah tertentu tersebut dan $e = 2{,}71828$. Distribusi Poisson mempunyai rata-rata dan variansi keduanya sama dengan μ.

Teorema

Misalkan Y variabel random berdistribusi Poisson dengan parameter μ maka rata-rata dan variansi Y adalah μ

Bukti

Rata – rata dari Y yang berdistribusi Poisson dengan parameter μ

$$E(Y) = \sum_{y=0}^{\infty} y.f(y,\mu)$$

$$= \sum_{y=0}^{\infty} y.\frac{\mu^y e^{-\mu}}{y!}$$

$$= \sum_{y=0}^{\infty} y.\frac{\mu^y e^{-\mu}}{y(y-1)!}$$

$$= \sum_{y=0}^{\infty} \frac{\mu^y e^{-\mu}}{(y-1)!}$$

$$= \sum_{y=1}^{\infty} \frac{\mu\, \mu^{(y-1)} e^{-\mu}}{(y-1)!}$$

$$= \mu \sum_{y=1}^{\infty} \frac{\mu^{(y-1)} e^{-\mu}}{(y-1)!}$$

$$= \mu \sum_{z=0}^{\infty} \frac{\mu^z e^{-\mu}}{z!} \text{ (misal } z = y-1, y = 1 \text{ maka} z = 0)$$

$$= \mu \sum_{z=0}^{\infty} f(z;\mu)$$

$$= \mu.1$$

$$= \mu$$

UNNES
UNIVERSITAS NEGERI SEMARANG

Sedangkan variansi dari Y yang berdistribusi Poisson dengan parameter μ adalah

$$V(Y) = E(Y^2) - [E(Y)]^2$$

$$= E(Y^2 - Y + Y) - [E(Y)]^2$$

$$= E[Y(Y-1) + Y] - [E(Y)]^2$$

$$= E[Y(Y-1)] + E(Y) - [E(Y)]^2$$

$$= \sum_{y=0}^{\infty} y(y-1).\frac{\mu^y e^{-\mu}}{y!} + E(Y) - [E(Y)]^2$$

$$= \sum_{y=0}^{\infty} y(y-1).\frac{\mu^y e^{-\mu}}{y(y-1)(y-2)!} + E(Y) - [E(Y)]^2$$

$$= \sum_{y=0}^{\infty} \frac{\mu^y e^{-\mu}}{(y-2)!} + E(Y) - [E(Y)]^2$$

$$= \sum_{y=2}^{\infty} \frac{\mu^2 \mu^{y-2} e^{-\mu}}{(y-2)!} + E(Y) - [E(Y)]^2$$

$$= \mu^2 \sum_{y=2}^{\infty} \frac{\mu^{y-2} e^{-\mu}}{(y-2)!} + E(Y) - [E(Y)]^2$$

$$= \mu^2 \sum_{z=0}^{\infty} \frac{\mu^z e^{-\mu}}{z!} + E(Y) - [E(Y)]^2 \text{ (misal } z = y-2, y = 2 \text{ maka } z = 0)$$

$$= \mu^2 + \mu - \mu^2 = \mu$$

Beberapa data cacah yang memungkinkan model distribusi poisson dapat digunakan adalah sebagai berikut :

1. Jumlah kasus penyakit X pada suatu negara dalam interval tahun tertentu.
2. Jumlah kasus pembunuhan pada suatu daerah dalam setahun.
3. Jumlah kecelakaan lalu lintas dalam interval bulan tertentu.

4. Jumlah bencana alam yang terjadi pada daerah tertentu dalam sepuluh tahun terakhir.
5. Jumlah panggilan telepon selama jam sibuk.

2.2 Distribusi Keluarga Eksponensial

Menurut Mc.Cullagh & Nelder (1989:28), sebuah variabel random Y mempunyai distribusi yang termasuk dalam keluarga eksponensial apabila fungsi densitas probabilitasnya dalam bentuk :

$$f(y;\theta,\phi) = exp\left\{\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y;\phi)\right]\right\} \quad (2)$$

untuk beberapa fungsi yang diketahui $a(\theta), b(\theta)$, dan $c(y;\phi)$. θ adalah parameter natural. ϕ adalah parameter skala atau dispersi, seperti σ^2 pada distribusi normal. Beberapa distribusi yang termasuk dalam keluarga eksponensial antara lain distribusi Normal, Poisson, Gamma, Inverse Gaussian, dan sebagainya.

2.3 *Generalized Linear Model* (GLM)

Analisis regresi yang responnya termasuk salah satu keluarga eksponensial disebut *Generalized Linear Models* (GLM). *Generalized Linear Model* (GLM) merupakan perluasan dari proses pemodelan linier untuk pemodelan data yang mengikuti distribusi probabilitas selain distribusi normal, seperti Poisson, Binomial, Multinomial, dan lain – lain.

Menurut Agresti (2002:116) menyatakan *Generalized Linear Model* didefinisikan menjadi tiga kompenen utama yaitu:

1. Komponen random

Variabel respon $Y = (y_1, y_2, \dots, y_n)$ dalam suatu n observasi diasumsikan saling bebas dan memiliki distribusi yang termasuk dalam keluarga eksponensial, dengan fungsi kepadatan peluang sebagai berikut :

$$f(y; \theta_i, \phi) = exp\left\{\left[\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i; \phi)\right]\right\} \quad (3)$$

Parameter θ_i disebut dengan parameter natural dan nilainya dapat berbeda untuk $i = 1,2, \dots, n$. Parameter ϕ disebut parameter dispersi. Jika ϕ konstan yang diketahui, maka persamaan (3) dinyatakan dalam bentuk :

$$f(y_i; \theta_i) = exp\{y_i\theta_i - b(\theta_i) + c(y_i)\} \quad (4)$$

2. Komponen Sistematis

Kontribusi variabel prediktor dalam model dinyatakan dalam bentuk kombinasi linier antara parameter (η) dengan parameter regresi yang akan diestimasi

$$\eta_i = \beta_o + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

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

3. Fungsi *Link*

Fungsi *Link* adalah suatu fungsi yang menghubungkan komponen acak dengan komponen sistematis. Diketahui $\mu_i = E(y_i)$. Model yang menghubungkan μ_i dengan prediktor linier η_i dinyatakan dengan

$$g(\mu_i) = \eta_i = \sum_j \beta_j x_{ij} \quad (5)$$

dengan fungsi g menunjukkan fungsi *link*. Suatu fungsi *link* disebut fungsi *link* kanonik jika $g(\mu_i) = \theta_i$

$$\theta_i = \sum_j \beta_j x_{ij} \quad (6)$$

Memilih *link* kanonik g yang sesuai dengan suatu distribusi variabel respon Y sangat mempermudah estimasi, meskipun dengan komputasi modern hal ini tidak lagi menjadi permasalahan utama.

2.4 Uji Kecocokan Distribusi

Menurut Sudjana (2005: 291) uji kecocokan menggunakan chi kuadrat dapat digunakan untuk mendeteksi distribusi Poisson. Uji chi kuadrat untuk menguji hipotesis dengan rumus

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

dengan O_i = banyak pengamatan data asli

E_i = banyak pengamatan yang diharapkan (jumlah data keseluruhan yang dikalikan dengan peluang untuk masing – masing pengamatan)

Adapun prosedur pengujiannya adalah sebagai berikut :

Hipotesis

H_0 : populasi mengikuti distribusi Poisson

H_1 : populasi tidak mengikuti distribusi Poisson

Kriteria uji H_0 ditolak jika $\chi^2_{\text{hitung}} > \chi^2_{\alpha;dk}$ dengan $dk = k - 2$

2.5 Regresi Poisson

Menurut Rini Cahyandari (2012), regresi Poisson merupakan salah satu dari model regresi yang berasal dari distribusi Poisson yang biasanya digunakan untuk menganalisis data dengan respon berupa variabel diskrit yang nilainya berupa integer tidak negatif. Regresi Poisson merupakan penerapan dari *Generalized Linear Model* (GLM). *Generalized Linear Model* (GLM) merupakan perluasan

dari model regresi yang menggambarkan hubungan antara variabel dependen dengan variabel independen dengan variabel dependen yang memiliki sebaran eksponensial. Regresi Poisson digunakan untuk menganalisis data *count* (berjenis diskrit).

Misalkan $y_i,\ i = 1, 2, \ldots$ merupakan jumlah kejadian yang muncul dalam selang waktu dengan rata-rata μ_i. Jika Y adalah variabel acak Poisson dengan parameter $\mu > 0$, fungsi massa peluangnya adalah:

$$f(y, \mu) = \frac{\mu^y e^{-\mu}}{y!} \qquad y = 0,1,2, \ldots \tag{7}$$

dengan asumsi $E(Y) = V(Y) = \mu$ (Cameron & Trivedi, 1998 : 10).

Persamaan (7) dapat ditulis dalam bentuk

$$f(y_i; \mu_i) = exp[y_i \ln(\mu_i) - \mu_i - \ln(y_i!)]$$

$$= exp\left[\frac{y_i \ln(\mu_i) - \mu_i}{1} - \ln(y_i!)\right] \tag{8}$$

dengan membandingkan persamaan (3) dengan persamaan (8) maka diperoleh :

$\theta_i = \ln \mu_i,\ b(\theta_i) = \mu_i = exp(\theta_i), a(\phi) = 1, c(y_i, \phi) = -\ln(y_i!)$

Distribusi poisson mempunyai rata- rata dan varian keduanya sama dengan μ. Diketahui dalam GLM, kontribusi variabel prediktor X dalam model dinyatakan dalam bentuk kombinasi linier antara parameter (η) dengan parameter regresi yang akan diestimasi,yakni

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} \tag{9}$$

Dalam distribusi Poisson diketahui bahwa $\mu_i = E(y_i)$. Model yang menghubungkan μ_i dengan prediktor linier η_i, dinyatakan dengan $g(\mu_i) = \eta_i$. Berdasarkan konsep GLM untuk distribusi Poisson bahwa pada saat $g(\mu_i)$ sama

dengan parameter natural $\theta_i(g(\mu_i) = \theta_i = ln(\mu_i))$, sehingga kanonikal link (fungsi yang mentranformasikan nilai *mean* ke parameter natural) adalah log natural link : $g(\mu_i) = ln(\mu_i)$. Sehingga hubungan μ_i dengan prediktor linier η_i, dinyatakan dengan $ln(\mu_i) = \eta_i$. Dengan menggunakan fungsi link log natural tersebut diperoleh model regresi Poisson dalam bentuk :

$$ln\,\mu_i = \eta_i$$

$$ln\,\mu_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

$$\mu_i = exp\,(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}) \qquad (10)$$

dimana μ_i nilai ekspektasi y_i berdistribusi Poisson. Dalam model regresi Poisson koefisien regresi β_j menyatakan perubahan yang diharapkan terhadap logaritma natural *mean* per unit perubahan pada prediktor X_j. Dengan demikian regresi Poisson memenuhi 3 komponen GLM sehingga terbukti bahwa regresi Poisson merupakan salah satu penerapan GLM.

Pada regresi Poisson diasumsikan variabel respon (Y) berdistribusi Poisson dan tidak terjadi multikolinearitas diantara masing – masing variabel prediktor (X). Dalam regresi Poisson terdapat asumsi yang harus dipenuhi yaitu variabel respon (Y) diskrit dan asumsi equidispersi. Equidispersi yaitu nilai rata – rata sama dengan nilai varian atau $V(Y) = E(Y) = \mu$.

2.6 Overdispersi

Menurut Cameron & Trivedi (1998: 4), suatu ciri dari distribusi Poisson adalah adanya equidispersi, yakni keadaan dimana nilai *mean* dan varian dari variabel respon bernilai sama. Namun kadang-kadang ditemukan keadaan yang disebut overdispersi yaitu nilai variannya lebih besar dari nilai rata-ratanya.

Menurut Hilbe (2011: 141), overdispersi pada regresi Poisson terjadi ketika varian dari variabel respon lebih besar dari rata-ratanya. Jika pada data diskrit terjadi overdispersi tetapi tetap digunakan model regresi Poisson, maka estimasi parameter koefisien regresinya tetap konsisten tetapi tidak efisien karena berdampak pada nilai standar *error* yang tinggi. Hal–hal penyebab overdispersi antara lain :

a. terjadi korelasi antara pengamatan
b. asumsi equidispersi regresi Poisson tidak terpenuhi
c. terdapat *excess zero*

Kasus data yang memiliki nilai variansinya lebih besar atau lebih kecil dari nilai rata–ratanya sering ditemukan dalam analisis data *count*. Dalam menganalisis data *count* yang variabel random Y diskrit biasanya digunakan regresi Poisson. Regresi Poisson memiliki syarat asumsi equidispersi atau nilai variannya sama dengan nilai rata – ratanya.

Overdispersi ataupun underdispersi akan menghasilkan nilai *devians* model yang sangat besar sehingga model yang dihasilkan kurang tepat. Untuk menguji asumsi equidispersi pada regresi Poissson digunakan dengan melihat nilai *Pearson's chi-square* yang dibagi dengan derajat bebasnya $(n - p - 1)$ atau dengan melihat nilai *deviance residual* dibagi dengan derajat kebebasan. Salah satu model yang dapat digunakan untuk mengatasi masalah overdispersi adalah dengan model *Zero-Inflated Poisson* dan *Zero-Inflated Negative Binomial*.

UNNES
UNIVERSITAS NEGERI SEMARANG

2.7 *Excess zeros*

Menurut Rainer Winkelmann (2008: 174), salah satu permasalahan pada regresi Poisson yaitu nilai nol yang berlebih (*Excess Zeros*). Pada variabel respon pada data diskrit mungkin ditemukan data bernilai kosong/nol. Akan tetapi, dalam banyak kasus, kosong memiliki arti penting pada penelitian yang bersangkutan. Jika nilai nol memiliki arti penting dalam data diskrit maka data tersebut harus dimasukkan dalam analisis.

Dalam penelitian dapat dijumpai kondisi dimana terlalu banyak nol. *Excess zeros* dapat dilihat pada proporsi variabel respon yang bernilai nol lebih besar dari data diskrit lainnya. Selain itu regresi Poisson juga menjadi tidak tepat lagi menggambarkan data yang sebenarnya. *Excess zeros* merupakan salah satu penyebab terjadinya overdispersi.

2.8 Metode Maksimum Likelihood

Menurut Casella dan Berger (1990: 289) metode maksimum likelihood merupakan metode yang paling sering digunakan untuk memperoleh taksiran. Misalkan $X_1, X_2, \dots, X_n$ adalah sampel random dari populasi dengan densitas $f(x|\theta_1, \dots, \theta_k)$ maka fungsi likelihood didefinisikan dengan

$$L(\theta|x) = L(\theta_1, \dots, \theta_k|x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i|\theta_1, \dots, \theta_k).$$

Estimator maksimum likelihood $\hat{\theta}$ adalah nilai θ yang memaksimalkan fungsi likelihood $L(\theta)$. Menurut Bayu Ariawan, Suparti, dan Sudarno (2012), untuk memperoleh nilai $\hat{\theta}$ yang memaksimumkan $L(\theta)$ harus diderivatifkan dengan langkah-langkah sebagai berikut:

1. Nilai $\hat{\theta}$ diperoleh dari derivative pertama

   $\frac{\partial}{\partial \theta_j} L(\theta) = 0$ dengan $j = 1,2, \dots, p$

2. Nilai $\hat{\theta}$ dikatakan memaksimumkan $L(\theta)$ jika

   $\frac{\partial^2}{\partial \theta_j^2} L(\theta)|_{\theta=\hat{\theta}} < 0$ dengan $j = 1,2, \dots, p$

Selain memaksimumkan fungsi likelihood, nilai $\hat{\theta}$ juga dapat diperoleh dengan memaksimumkan log natural-likelihood $(ln\, L(\theta))$. Dalam banyak kasus dengan diferensiasi digunakan, akan lebih mudah bekerja pada logaritma natural yang dinotasikan dengan $l(\theta) = ln\, L(\theta)$. Untuk memperoleh nilai $\hat{\theta}$ yang memaksimumkan $lnL(\theta)$ dapat dilakukan dengan langkah- langkah yang sama seperti dalam memperoleh nilai $\hat{\theta}$ yang memaksimumkan $L(\theta)$.

2.9 Regresi *Zero-Inflated Negative Binomial* (ZINB)

Menurut Hilbe (2011:186), regresi *Zero-Inflated Negative Binomial* (ZINB) merupakan model yang dibentuk dari distribusi campuran Poisson gamma. Distirubsi campuran Poisson gamma terbentuk jika suatu distribusi Poisson (μ) dengan μ merupakan nilai variabel random yang berdistribusi gamma maka akan dihasilkan distribusi campuran gamma Poisson yang dinamakan distribusi binomial negatif. Menurut Hilbe (2011: 189) fungsi kepadatan peluangnya adalah

$$f(y|\alpha,\beta) = \frac{\Gamma(y+\alpha)}{y!\Gamma(\alpha)}\left(\frac{1}{1+\beta}\right)^{\alpha}\left(1-\frac{1}{1+\beta}\right)^{y} \qquad y = 0,1,2, \dots$$

dengan rata-rata dan variansi distribusi binomial negatif yaitu:

$E[Y] = \alpha\beta$ dan $V[Y] = \alpha\beta + \alpha\beta^2$ (Hilbe, 2011: 197)

Bukti

Rata – rata dari Y yang berdistribusi binomial negatif

$$E(Y) = \sum_{y=0}^{\infty} y \binom{\alpha + y - 1}{\alpha - 1} \left(\frac{1}{1+\beta}\right)^{\alpha} \left(1 - \frac{1}{1+\beta}\right)^{y}$$

$$= \sum_{y=0}^{\infty} y \binom{\alpha + y - 1}{\alpha - 1} \left(\frac{1}{1+\beta}\right)^{\alpha} \left(\frac{\beta}{1+\beta}\right)^{y}$$

$$= \sum_{y=0}^{\infty} y \frac{(\alpha + y - 1)!}{\big((\alpha + y - 1) - (\alpha - 1)\big)!\,(\alpha - 1)!} \left(\frac{1}{1+\beta}\right)^{\alpha} \left(\frac{\beta}{1+\beta}\right)^{y}$$

$$= \sum_{y=0}^{\infty} y \frac{(\alpha + y - 1)!}{y(y-1)!\,(\alpha - 1)!} \left(\frac{1}{1+\beta}\right)^{\alpha} \left(\frac{\beta}{1+\beta}\right)^{y}$$

$$= \sum_{y=1}^{\infty} \frac{(\alpha + y - 1)!}{(y-1)!\,(\alpha - 1)!} \left(\frac{1}{1+\beta}\right)^{\alpha} \left(\frac{\beta}{1+\beta}\right)^{y}$$

$$= \sum_{y=1}^{\infty} \alpha \binom{\alpha + y - 1}{y - 1} \left(\frac{1}{1+\beta}\right)^{\alpha} \left(\frac{\beta}{1+\beta}\right)^{y}$$

Misalkan $y - 1 = z$

$$E(Y) = \sum_{z=0}^{\infty} \alpha \binom{\alpha + z}{z} \left(\frac{1}{1+\beta}\right)^{\alpha} \left(\frac{\beta}{1+\beta}\right)^{z-1}$$

$$= \alpha \frac{\left(\frac{\beta}{1+\beta}\right)}{\left(\frac{1}{1+\beta}\right)} \sum_{z=0}^{\infty} \binom{(\alpha + 1) + (z - 1)}{z} \left(\frac{1}{1+\beta}\right)^{\alpha+1} \left(\frac{\beta}{1+\beta}\right)^{z}$$

$$= \alpha\beta$$

Varian dari Y yang berdistribusi binomial negatif

$$E\big(Y(Y-1)\big) = \sum_{y=0}^{\infty} y(y-1) \binom{\alpha + y - 1}{\alpha - 1} \left(\frac{1}{1+\beta}\right)^{\alpha} \left(\frac{\beta}{1+\beta}\right)^{y}$$

$$= \sum_{y=0}^{\infty} y(y-1)\frac{(\alpha+y-1)!}{y!\,(\alpha-1)!}\left(\frac{1}{1+\beta}\right)^{\alpha}\left(\frac{\beta}{1+\beta}\right)^{y}$$

$$= \sum_{y=2}^{\infty} \frac{(\alpha+y-1)!}{(y-2)!\,(\alpha-1)!}\left(\frac{1}{1+\beta}\right)^{\alpha}\left(\frac{\beta}{1+\beta}\right)^{y}$$

$$= \sum_{y=2}^{\infty} \alpha(\alpha+1)\binom{\alpha+y-1}{y-2}\left(\frac{1}{1+\beta}\right)^{\alpha}\left(\frac{\beta}{1+\beta}\right)^{y}$$

Misalkan $z = y - 2$

$$E\big(Y(Y-1)\big) = \sum_{z=0}^{\infty} \alpha(\alpha+1)\binom{\alpha+z+1}{z}\left(\frac{1}{1+\beta}\right)^{\alpha}\left(\frac{\beta}{1+\beta}\right)^{z+2}$$

$$= \alpha(\alpha+1)\sum_{z=0}^{\infty}\binom{(\alpha+2)+(z-1)}{z}\left(\frac{1}{1+\beta}\right)^{\alpha+2}\left(\frac{1}{1+\beta}\right)^{-2}\left(\frac{\beta}{1+\beta}\right)^{z}\left(\frac{\beta}{1+\beta}\right)^{2}$$

$$= \alpha(\alpha+1)\frac{\left(\frac{\beta}{1+\beta}\right)^{2}}{\left(\frac{1}{1+\beta}\right)^{2}}\sum_{z=0}^{\infty}\binom{(\alpha+2)+(z-1)}{z}\left(\frac{1}{1+\beta}\right)^{\alpha+2}\left(\frac{\beta}{1+\beta}\right)^{z}$$

$$= \alpha(\alpha+1)\beta^2$$

Sehingga variannya

$$V(Y) = E(Y^2) - \big(E(Y)\big)^2$$

$$= E\big(Y(Y-1)\big) + E(Y) - \big(E(Y)\big)^2$$

$$= \alpha(\alpha+1)\beta^2 + \alpha\beta - (\alpha\beta)^2$$

$$= (\alpha\beta)^2 + \alpha\beta^2 + \alpha\beta - (\alpha\beta)^2$$

$$= \alpha\beta + \alpha\beta^2$$

Untuk membentuk suatu model regresi pada distribusi binomial negatif, maka nilai parameter dari distribusi campuran Poisson gamma dinyatakan dalam

bentuk $\mu = \alpha\beta$ dan $k = 1/\alpha$ sehingga diperoleh rata-rata dan variansi dalam bentuk:

$E(Y) = \mu$ dan $V(Y) = \mu + k\mu^2$

Bukti

$E(Y) = \alpha\beta = \mu$

$V(Y) = \alpha\beta + \alpha\beta^2$

$= \mu + \mu\beta$

$= \mu + \mu\dfrac{\mu}{\alpha}$

$= \mu + \mu^2 k$

Kemudian fungsi massa peluangnya menjadi:

$$f(y|\mu, k) = \frac{\Gamma\left(y+\frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right)y!}\left(\frac{1}{1+k\mu}\right)^{\frac{1}{k}}\left(\frac{k\mu}{1+k\mu}\right)^{y} \qquad y = 0,1,2,\ldots \qquad (11)$$

Fungsi distribusi pada persamaan (11) dinamakan fungsi kepadatan peluang binomial negatif dengan *mean* $E[Y] = \mu$ dan variansi $V[Y] = \mu + k\mu^2$, k dinamakan parameter dispersi. Jika parameter k konstan (tetap) maka dapat ditunjukkan bahwa fungsi distribusi binomial negatif pada persamaan (11) termasuk dalam keluarga eksponensial seperti berikut:

$$f(y|\mu, k) = exp\left\{y\, ln\left(\frac{k\mu}{1+k\mu}\right) + \frac{1}{k}ln\left(\frac{1}{1+k\mu}\right) + ln\left(\frac{\Gamma(y+1/k)}{\Gamma(1/k)y!}\right)\right\} \qquad (12)$$

Dari persamaan (12), sesuai dengan persamaan (4) dapat disimpulkan bahwa distribusi binomial negatif merupakan suatu keluarga eksponensial dengan

$\theta = ln\left(\frac{k\mu}{1+k\mu}\right)$ $b(\theta) = -\frac{1}{k}ln\left(\frac{1}{1+k\mu}\right)$ dan $c(y) = ln\left(\frac{\Gamma(y+1/k)}{\Gamma(1/k)y!}\right)$

saat $k \to 0$ maka distribusi binomial negatif memiliki varians $V[Y] \to \mu$. Keadaan tersebut memungkinkan distribusi binomial negatif akan mendekati suatu distribusi Poisson yang mengasumsikan *mean* dan variansi yang sama yaitu $E[Y] = V[Y] = \mu$ (Agresti, 2002 :7)

$1/k$ adalah bentuk parameter yang mengukur jumlah overdispersi. Jika Y_i adalah variabel random independen yang diskrit dengan $i = 1,2,3, \dots, n$ nilai nol pada observasi diduga muncul dalam dua cara yang sesuai untuk keadaan (*state*) yang terpisah. Menurut Garay & Hashimoto (2011), regresi ZINB dengan keadaan pertama disebut *zero state* terjadi dengan probabilitas p_i dan menghasilkan hanya observasi bernilai nol, sementara keadaan kedua disebut *Negative Binomial State* terjadi dengan probabilitas $(1 - p_i)$ dan berdistribusi Binomial Negatif dengan *mean* μ dengan $0 \leq p_i \leq 1$. Proses dua keadaan ini dengan variabel Y_i memberikan distribusi campuran dua komponen dan didapat fungsi probabilitas sebagai berikut:

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)\left(\frac{1}{1+k\mu_i}\right)^{1/k} & \text{, untuk } y_i = 0 \\ (1 - p_i)\frac{\Gamma(y_i+1/k)}{\Gamma(1/k)\Gamma(y_i+1)}\left(\frac{1}{1+k\mu_i}\right)^{1/k}\left(\frac{k\mu_i}{1+k\mu_i}\right)^{y_i} & \text{, untuk } y_i = 1,2, \dots \end{cases} \tag{13}$$

dengan $i = 1,2,3 \dots, n; 0 \leq p_i \leq 1,\ \mu_i \geq 0, k$ adalah parameter dispersi dengan $1/k > 0$ dan $\Gamma(.)$ adalah fungsi gamma. Diasumsikan bahwa parameter μ_i dan p_i masing – masing bergantung pada variabel x_i dan z_i, sehingga menurut Garay & Hashimoto (2011) menyatakan model dari regresi ZINB dibagi menjadi dua komponen model yaitu:

1. Model data diskrit untuk μ_i adalah

$$ln(\mu_i) = x_i^T\beta, \quad \mu_i \geq 0, i = 1, \dots, n. \tag{14}$$

x_i adalah matriks variabel yang memuat himpunan – himpunan yang berbeda dari faktor eksperimen yang berhubungan dengan peluang pada *mean Negative Binomial* pada *Negative Binomial state.*

2. Model *zero-Inflation* untuk p_i adalah

$$logit(p_i) = ln\left(\frac{p_i}{1-p_i}\right) = z_i^T\gamma, \quad 0 \leq p_i \leq 1, i = 1, \dots, n. \tag{15}$$

x_i adalah matriks variabel yang memuat himpunan – himpunan yang berbeda dari faktor eksperimen yang berhubungan dengan peluang pada *zero state.*

Pengaruh dari masing – masing matriks kovariat x_i dan z_i terhadap μ_i dan p_i bias sama atau tidak sama, jika masing-masing matriks kovariat memberikan pengaruh yang sama terhadap μ_i dan p_i maka matriks $x_i = z_i$, sehingga model (14) dan (15) menjadi:

1. Model data diskrit untuk μ_i adalah

$$ln(\mu_i) = x_i^T\beta, \text{ atau } \mu_i = e^{x_i^T\beta} \quad \mu_i \geq 0, \mathrm{i} = 1, \dots, \mathrm{n}. \tag{16}$$

2. Model *zero-inflation* untuk p_i adalah

$$logit(p_i) = ln\left(\frac{p_i}{1-p_i}\right) = x_i^T\gamma, \text{ atau } p_i = \frac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}} \quad 0 \leq p_i \leq 1, i = 1, \dots, n \tag{17}$$

x_i adalah matriks variabel yang memuat himpunan – himpunan yang berbeda dari faktor eksperimen yang berhubungan dengan peluang pada *mean Negative Binomial* pada *Negative Binomial state*, sedangkan β dan γ adalah parameter regresi yang akan ditaksir.

UNNES
UNIVERSITAS NEGERI SEMARANG

2.10 Estimasi Parameter Regresi *Zero-Inflated Negative Binomial* (ZINB)

Menurut Bayu Ariawan, Suparti, dan Sudarno (2012) menyatakan estimasi parameter model regresi ZINB menggunakan metode *Maximum Likelihood Estimation* (MLE) dengan prosedur Algoritma EM (*Expectation Maximization*) dan Newton Rhapson. Metode ini biasanya digunakan untuk menduga parameter suatu model yang diketahui fungsi densitasnya.

Dari persamaan (16) dan (17) didapat :

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

$$\mu_i = e^{x_i^T \beta} \tag{18}$$

$$ln\left(\frac{p_i}{1-p_i}\right) = x_i^\gamma \gamma$$

$$p_i = \frac{e^{x_i^\gamma \gamma}}{1+e^{x_i^\gamma \gamma}} \tag{19}$$

$$(1-p_i) = \frac{1}{1+e^{x_i^\gamma \gamma}} \tag{20}$$

Dari persamaan (18), (19) dan (20) disubstitusikan ke persamaan (13) diperoleh :

$$P(Y_i = y_i) = \begin{cases} \dfrac{e^{x_i^\gamma \gamma}}{1+e^{x_i^\gamma \gamma}} + \dfrac{1}{1+e^{x_i^\gamma \gamma}}\left(\dfrac{1}{1+ke^{x_i^T \beta}}\right)^{1/k} & \text{, untuk } y_i = 0 \\ \dfrac{1}{1+e^{x_i^\gamma \gamma}} \dfrac{\Gamma\left(y_i + \frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right)\Gamma(y_i+1)}\left(\dfrac{1}{1+ke^{x_i^T \beta}}\right)^{\frac{1}{k}}\left(\dfrac{ke^{x_i^T \beta}}{1+ke^{x_i^T \beta}}\right)^{y_i} & \text{, untuk } y_i = 1,2,\ldots \end{cases} \tag{21}$$

Misalkan diambil sampel $(y_1, x_1, z_1), \dots, (y_n, x_n, z_n)$ dari n percobaan yang saling bebas. Fungsi likelihood dari persamaan (21) untuk parameter $\theta = \left(\frac{1}{k}, \beta^T, \gamma^T\right)^T$ adalah

$$L(\theta|y_i) = \begin{cases} \prod \dfrac{e^{x_i^T\gamma} + \left(\dfrac{1}{1+ke^{x_i^T\beta}}\right)^{1/k}}{1+e^{x_i^T\gamma}}, \text{untuk } y_i = 0 \\ \prod \dfrac{\Gamma(y_i + 1/k)\left(\dfrac{1}{1+ke^{x_i^T\beta}}\right)^{1/k}\left(\dfrac{ke^{x_i^T\beta}}{1+ke^{x_i^T\beta}}\right)^{1/k}}{1+e^{x_i^T\gamma}\,\Gamma(1/k)\Gamma(y_i+1)}, \text{untuk } y_i = 1,2,.. \end{cases} \tag{22}$$

Sehingga fungsi log-likelihood dari persamaan (22) adalah

$$lnL(\theta|y_i) = \begin{cases} \sum_{i=1}^{n} ln\left\{e^{x_i^T\gamma} + \left(\dfrac{1}{1+ke^{x_i^T\beta}}\right)^{1/k}\right\} - \sum_{i=1}^{n} ln\left[1+e^{z_i^T\gamma}\right], \text{untuk } y_i = 0 \\ -\sum_{i=1}^{n} ln\left[1+e^{z_i^T\gamma}\right] + \sum_{i=1}^{n} ln[\Gamma(1/k+y_i)] - \sum_{i=1}^{n} ln[\Gamma(y_i+1)] - \sum_{i=1}^{n} ln[\Gamma(1/k)] \\ +y_i \sum_{i=1}^{n} ln\left(\dfrac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}}\right)^{y_i} + 1/k \sum_{i=1}^{n} ln\left(\dfrac{1}{1+ke^{x_i^T\beta}}\right)^{1/k}, \text{untuk } y_i = 1,2,\dots \end{cases} \tag{23}$$

dengan $i = 1,2,\dots,n$. Estimasi dengan maksimum likelihood rasio dihitung dengan memaksimalkan log-likelihoodnya pada persamaan (23). Menurut Bayu Ariawan, Suparti, dan Sudarno (2012), penjumlahan fungsi log-likelihood pada persamaan (23) tidak linear, sehingga fungsi likelihood ini tidak dapat diselesaikan dengan metode numerik biasa. Sehingga digunakanlah algoritma EM (*Expectation Maximization*) yang merupakan salah satu metode yang digunakan untuk menemukan estimasi suatu parameter melalui kerangka metode *Maximum*

Likelihood Estimation (MLE) dari suatu fungsi distribusi yang dengan informasi data yang tidak lengkap atau data hilang (*missing).*

Misalkan variabel $y_i(i = 1,2, \dots, n)$ berkaitan dengan vektor variabel indikator $\mathrm{W} = (w_1, \dots, w_n)^T$ yaitu:

$$w_i = \begin{cases} 1, \text{jika } \mathrm{y_i} \text{ berasal dari } zero\ state \\ 0, \text{jika } \mathrm{y_i} \text{ berasal dari } Negative\ Binomial\ state \end{cases}$$

dengan $i = 1,2, \dots, n$ jika nilai variabel respon $y_i = 1,2, \dots.$ maka nilai $w_i = 0$. Sedangkan jika nilai variabel respon $y_i = 0$ maka nilai w_i mungkin 0 mungkin 1. Oleh karena itu, nilai w_i dianggap hilang. Peluang dari w_i dapat dinyatakan :

$$P(w_i = 1) = p_i$$

$$P(w_i = 0) = 1 - p_i$$

dengan $i = 1,2, \dots, n$. Sehingga distribusi dari variabel W adalah $w_i \sim Binomial(n, p_i)$ mempunyai rataan dan variansi $E(w_i) = np_i$ dan $V(w_i) = np_i(1 - p_i)$.

Bukti

Rataan dari distribusi binomial adalah

$$M_x = E(e^{tx})$$

$$= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{1-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{1-x}$$

$$= (pe^t + q)^n$$

$$M'_x(t) = npe^t(pe^t + q)^{n-1}$$

$$M'_x(0) = npe^0(pe^0 + q)^{n-1}$$

$$M'_x(0) = np(p+q)^{n-1}$$

$$M'_x(0) = np(p+1-p)^{n-1}$$

$$= np(1)^{n-1}$$

$$E(X) = np$$

$$M''_x(t) = npe^t(pe^t+q)^{n-1} + (n-1).pe^t.npe^t(pe^t+q)^{n-2}$$

$$M''_x(0) = np + (n-1)np^2$$

$$E(X^2) = np + n^2p^2 - np^2$$

$$V(x) = E(X^2) - \left(E(X)\right)^2$$

$$= np + n^2p^2 - np^2 - np^2$$

$$= np - np^2$$

$$= np(1-p)$$

Distribusi gabungan antara y_i dan w_i yang terbentuk yaitu

$$f(w_i, y_i|p_i, \mu_i) = (p_i)^{w_i}(1-p_i)^{(1-w_i)}\left[\frac{\Gamma(y_i+1/k)}{\Gamma(1/k)\Gamma(y_i+1)}\left(\frac{1}{1+k\mu_i}\right)^{1/k}\left(\frac{k\mu_i}{1+k\mu_i}\right)^{y_i}\right]^{(1-w_i)} \tag{24}$$

Substitusikan persamaan (18), (19) dan (20) ke persamaan (24) didapat persamaan log-likelihoodnya:

$$lnL(\beta, \gamma|y_i, w_i) = \sum_{i=1}^{n}\{w_i x_i^T\gamma - ln[1+exp(x_i^T\gamma)] + (1-w_i)\, ln[g(y_i;\beta, 1/k)]\} \tag{25}$$

dimana $g(y_i;\beta,1/k) = \frac{\Gamma(y_i+1/k)}{\Gamma(1/k)\Gamma(y_i+1)}\left(\frac{1}{1+k\mu_i}\right)^{1/k}\left(\frac{k\mu_i}{1+k\mu_i}\right)^{y_i}$ dan $\mu_i = e^{x_i^T\beta}$ dengan $i = 1,2,\dots,n$. Persamaan (25) yang akan dimaksimumkan menggunakan algoritma EM, dengan parameter β dan γ dapat diestimasi secara terpisah menjadi:

$$ln\, L(\beta,\gamma|y_i,w_i) = ln\, L(\beta|y_i,w_i) + ln\, L(\gamma|y_i,w_i)$$

Dengan

$$ln\, L(\gamma|y_i,w_i) = \sum_{i=1}^{n}[w_i x_i^T\gamma - ln[1+exp(x_i^T\gamma)]] \tag{26}$$

dan

$$ln\, L(\beta|y_i,w_i) = \sum_{i=1}^{n}(1-w_i)\left\{\frac{\Gamma(y_i+1/k)}{\Gamma(1/k)\Gamma(y_i+1)}\left(\frac{1}{1+ke^{x_i^T\beta}}\right)^{1/k}\left(\frac{ke^{x_i^T\beta}}{1+ke^{x_i^T\beta}}\right)^{y_i}\right\} \tag{27}$$

Algoritma EM dibagi menjadi dua langkah yaitu

2.1 Tahap ekspektasi (E-Step)

Tahap E-Step dengan cara mengganti variabel w_i dengan $w_i^{(m)}$ yang merupakan ekspetasi dari w_i

$$w_i^{(m)} = E\left(w_i|y_i,\gamma^{(m)},\beta^{(m)}\right)$$

$$= \begin{cases} \left(1+\left(e^{x_i^T\gamma^{(m)}}\right)\left[\frac{1}{1+k^{(m)}+e^{x_i^T\beta^{(m)}}}\right]^{1/k^{(m)}}\right)^{-1} & \text{, jika } y_i = 0 \\ 0, \text{jika } y_i = 1,2,\ldots \end{cases}$$

Sehingga

$$Q\left(\beta,\gamma;\beta^{(m)},\gamma^{(m)}\right) = E_{\theta^{(k)}}\{ln\, L(\beta,\gamma|y_i,w_i)|y_i,\gamma^{(m)},\beta^{(m)}\}$$

$$= \sum_i^n ln\, L\left(\gamma^{(m)}|y_i,w_i^{(m)}\right) + \sum_i^n ln\, L\left(\beta^{(m)}|y_i,w_i^{(m)}\right)$$

dimana persamaan (26) dan (27) menjadi

$$ln\, L\left(\gamma^{(m)}|y_i,w_i^{(m)}\right) = \sum_{i=1}^{n}\left[w_i^{(m)}x_i^T\gamma - ln\left(1+e^{x_i^T\gamma}\right)\right] \tag{28}$$

$$ln\, L\left(\beta^{(m)}|y_i,w_i^{(m)}\right) = \sum_{i=1}^{n}\left(1-w_i^{(m)}\right)\left\{\frac{\Gamma(y_i+1/k)}{\Gamma(1/k)\Gamma(y_i+1)}\left(\frac{1}{1+ke^{x_i^T\gamma}}\right)^{1/k}\left(\frac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}}\right)^{y_i}\right\} \tag{29}$$

2.2 Tahap maksimalisasi (M-step)

Memaksimalkan β dan γ pada persamaan (28) dan (29) dengan menghitung $\beta^{(m+1)}$ dan $\gamma^{(m+1)}$ dengan metode Newton-Raphson. Muhammad Taufan, Suparti dan Agus Rusgiyono (2012), adapun dengan langkah – langkah sebagai berikut :

a. Misalkan $\beta^{(m)}$ dan $\gamma^{(m)}$ adalah aproksimasi metode maksimum likelihood untuk mengestimasi $\hat{\beta}$ dan $\hat{\gamma}$

b. Menghitung $\beta^{(m+1)}$ dan $\gamma^{(m+1)}$ dengan cara :

$$\beta^{(m+1)} = \beta^{(m)} - \left(H^{(m)}\right)^{-1} U^{(m)}$$

dan

$$\gamma^{(m+1)} = \gamma^{(m)} - \left(H^{(m)}\right)^{-1} U^{(m)}$$

Dengan H adalah turunan kedua dari $\ln L\left(\beta^{(m)} | y_i, w_i^{(m)}\right)$ dan $\ln L\left(\gamma^{(m)} | y_i, w_i^{(m)}\right)$, U adalah turunan pertama dari $\ln L\left(\beta^{(m)} | y_i, w_i^{(m)}\right)$ dan $\ln L\left(\gamma^{(m)} | y_i, w_i^{(m)}\right)$

c. Mengganti $\beta^{(m)}$ dan $\gamma^{(m)}$ dengan $\beta^{(m+1)}$ dan $\gamma^{(m+1)}$ pada iterasi selanjutnya, kemudian kembali lakukan tahap ekspektasi (E-Step)

Tahap *E-step dan M-step* ini dilakukan berulang-ulang sampai diperoleh penaksir parameter yang konvergen $\left|\beta^{(m)} - \beta^{(m+1)}\right| \leq \varepsilon$ dan $\left|\gamma^{(m)} - \gamma^{(m+1)}\right| \leq \varepsilon$ biasanya ε merupakan nilai bilangan positif yang sangat kecil, misalnya $\varepsilon = 10^{-5}$.

2.11 Regresi *Zero-Inflated Poisson* (ZIP)

Salah satu penyebab terjadinya overdispersi adalah lebih banyak observasi bernilai nol daripada yang ditaksir untuk model regresi Poisson. Lambert (1992) mengusulkan model regresi *Zero-Inflated Poisson* yang digunakan untuk menganalisis lebih banyak observasi bernilai nol daripada yang ditaksir. Menurut Jansakul & Hinde (2002) menyatakan jika y_i adalah variabel random independen yang mempunyai distribusi ZIP, nilai nol pada observasi diduga muncul dalam dua cara yang sesuai untuk keadaan (*state*) yang terpisah. Keadaan pertama disebut *zero state* terjadi dengan probabilitas ω_i dan menghasilkan hanya observasi bernilai nol, sementara keadaan kedua disebut *Poisson State* terjadi dengan probabilitas $(1-\omega_i)$ dan berdistribusi Poisson dengan *mean* μ_i. Proses dua keadaan ini memberikan distribusi campuran dua komponen dengan fungsi probabilitas sebagai berikut :

$$P(Y=y_i)=\begin{cases}\omega_i+(1-\omega_i)e^{-\mu_i}, \text{untuk } y_i=0\\ \frac{(1-\omega_i)e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \text{untuk } y_i>0, 0\leq\omega_i\leq 1\end{cases} \tag{30}$$

yang dinotasikan dengan $Y_i \sim ZIP(\mu,\omega)$

Menurut Lambert (1992) menyatakan hubungan model untuk μ dan ω adalah sebagai berikut :

$$log(\mu)=X\beta \; dan \; logit(\omega)=log\left(\frac{\omega}{1-\omega}\right)=X\gamma \tag{31}$$

X adalah matriks variabel yang memuat himpuan-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.

2.12 Estimasi Parameter Model Regresi *Zero-Inflated Poisson* (ZIP)

Menurut Muhammad Taufan, Suparti dan Agus Rusgiyono (2012) menyatakan estimasi parameter regresi *Zero-Inflated Poisson* menggunakan metode *Maksimum Likelihood Estimation* (MLE). Metode ini biasanya digunakan untuk menaksir parameter suatu model yang diketahui fungsi densitasnya.

Dari persamaan (31) diperoleh :

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

$$\mu_i = e^{x_i^T \beta} \tag{32}$$

$$\ln\left(\frac{\omega_i}{1-\omega_i}\right) = x_i^T \gamma$$

$$\omega_i = \frac{e^{x_i^T \gamma}}{1+e^{x_i^T \gamma}} \tag{33}$$

$$(1-\omega_i) = \frac{1}{1+e^{x_i^T \gamma}} \tag{34}$$

Persamaan (32), (33) dan (34) disubstitusikan ke persamaan (30) didapat

$$P(Y=y_i) = \begin{cases} \frac{e^{x_i^T \gamma}}{1+e^{x_i^T \gamma}} + \frac{1}{1+e^{x_i^T \gamma}} e^{-e^{x_i^T \beta}}, \text{untuk } y_i = 0 \\ \frac{\frac{1}{1+e^{x_i^T \gamma}} e^{-e^{x_i^T \beta}} e^{x_i^T \beta^{y_i}}}{y_i!}, \text{untuk } y_i > 0 \end{cases} \tag{35}$$

Sehingga fungsi likelihood dari persamaan (35) adalah

$$L(\beta,\gamma|y_i) = \begin{cases} \prod_{i=1}^{n} \frac{e^{x_i^T \gamma} + e^{-e^{x_i^T \beta}}}{1+e^{x_i^T \gamma}}, \text{untuk } y_i = 0 \\ \prod_{i=1}^{n} \frac{\frac{1}{1+e^{x_i^T \gamma}} e^{-e^{x_i^T \beta}} e^{x_i^T \beta^{y_i}}}{y_i!}, \text{untuk } y_i > 0 \end{cases} \tag{36}$$

Sehingga fungsi ln-likelihood dari persamaan (36) adalah :

$$ln\, L(\beta,\gamma|y_i) = \begin{cases} \sum_{i=1}^{n} ln\left(e^{x_i^T\gamma} + exp\left(-e^{x_i^T\beta}\right)\right) - \sum_{i=1}^{n} ln\left(1 + e^{x_i^T\gamma}\right), untuk\ y_i = 0 \\ \sum_{i=1}^{n}(x_i^T\beta)y_i - e^{x_i^T\beta} - \sum_{i=1}^{n} ln\left(1 + e^{x_i^T\gamma}\right) - \sum_{i=1}^{n} ln\, y_i!\,, untuk\ y_i > 0 \end{cases} \tag{37}$$

Penjumlahan fungsi ln-likelihood pada persamaan (37) akan membuat sulit perhitungan karena tidak diketahui nilai nol mana yang berasal dari *zero state* dan mana yang berasal dari *Poisson state*, sehigga fungsi likelihood ini tidak dapat diselesaikan dengan metode numerik biasa.

Memaksimalkan fungsi ln-likelihood digunakan algoritma EM (*Expectation Maximization*) yang merupakan salah satu metode yang digunakan untuk menemukan estimasi suatu parameter melalui kerangka metode *Maximum Likelihood Estimation (MLE)* dari suatu fungsi distribusi yang dengan informasi data yang tidak lengkap atau data hilang (*missing).*

Misalkan variabel *Y* berkaitan dengan variabel indikator *Z* yaitu :

$$Z = \begin{cases} 1, \text{jika } y_i \text{berasal dari } zero\ state \\ 0, \text{jika } y_i \text{berasal dari } Poisson\ state \end{cases}$$

Permasalahannya adalah jika nilai variabel respon $y_i = 1,2,3,\ldots,$ maka nilai $z_i = 0$ sedangkan jika nilai variabel respon $y_i = 0$, maka nilai z_i mungkin 0 mungkin 1. Oleh karena itu, nilai z_i dianggap hilang. Untuk mengatasi hal ini dilakukan estimasi parameter dengan algoritma EM. Bahwa langkah – langkahnya adalah sebagai berikut :

1. Menentukan distribusi dari variabel Z

   $P(z_i = 1) = \omega_i$

   $P(z_i = 0) = 1 - \omega_i$

   Sehingga $z_i \sim$Binomial $(1, \omega_i)$, $E(z_i) = \omega_i$dan $\text{var}(z_i) = \omega_i(1 - \omega_i)$

2. Membentuk distribusi gabungan antara y_i dan z_i yaitu

$$f(y_i, z_i|\omega_i, \mu_i) = f(z_i)f(y_i|z_i) = (1-\omega_i)^{(1-z_i)}(\omega_i)^{z_i}\left(\frac{exp(-\mu_i)\mu_i^{y_i}}{y_i!}\right)^{(1-z_i)} \quad (38)$$

dengan mensubstitusikan nilai $\mu_i = exp(x_i^T\beta)$, $\omega_i = \frac{e^{x_i^T\gamma}}{1+e^{x_i^T\gamma}}$ dan

$(1-\omega_i) = \frac{1}{1+e^{x_i^T\gamma}}$ ke persamaan (38) diperoleh

$$ln\, L(\beta, \gamma|y, z) = \sum_{i=1}^{n}[z_i x_i^T\gamma - ln(1 + exp(x_i^T\gamma))] + \sum_{i=1}^{n}(1-z_i)(y_i x_i^T\beta - exp(x_i^T\beta)) - \sum_{i=1}^{n}(1-z_i)\, ln\, y_i! \quad (39)$$

Persamaan (39) yang akan dimaksimumkan menggunakan algoritma EM, dimana parameter β dan γ dapat diestimasi secara terpisah, dengan menuliskan persamaan (39) menjadi :

$$ln\, L(\beta, \gamma, y, z) = ln\, L(\beta, y, z) + ln\, L(\gamma, y, z) - \sum_{i=1}^{n}(1-z_i)\, ln(y_i!)$$

dengan

$$ln\, L(\gamma, y, z) = \sum_{i=1}^{n}[z_i x_i^T\gamma - ln(1 + exp(x_i^T\gamma))] \quad (40)$$

$$ln\, L(\beta, y, z) = \sum_{i=1}^{n}(1-z_i)(y_i x_i^T\beta - exp(x_i^T\beta)) \quad (41)$$

3. Tahap ekspektasi

Ganti variabel z_i dengan $z_i^{(k)}$ yang merupakan ekspektasi dari z_i

$$z_i^{(k)} = E\left(z_i|y_i, \gamma^{(k)}, \beta^{(k)}\right)$$

$$= \begin{cases} \frac{1}{1+exp\left(-x_i^T\gamma^{(k)} - exp\left(x_i^T\beta^{(k)}\right)\right)}, \text{jika } y_i = 0 \\ 0, \text{jika } y_i > 0 \end{cases}$$

Sehingga persamaan (40) dan (41) menjadi

$$ln\, L\left(\gamma^{(k)}, y, z^{(k)}\right) = \sum_{i=1}^{n}\left[z_i^{(k)} x_i^T\gamma^{(k)} - ln\left(1 + exp\left(x_i^T\gamma^{(k)}\right)\right)\right] \quad (42)$$

$$ln\, L\left(\beta^{(k)}, y, z^{(k)}\right) = \sum_{i=1}^{n}\left(1 - z_i^{(k)}\right)\left(y_i x_i^T\beta^{(k)} - exp\left(x_i^T\beta^{(k)}\right)\right) \quad (43)$$

4. Tahap maksimalisasi

   Memaksimalkan β dan γ pada persamaan (42) dan (43) dengan menghitung $\beta^{(k+1)}$ dan $\gamma^{(k+1)}$ dengan metode Newton-Raphson.

   Misalkan $\beta^{(k)}$ dan $\gamma^{(k)}$ adalah aproksimasi metode maksimum likelihood untuk mengestimasi $\hat{\beta}$ dan $\hat{\gamma}$. Dengan menggunakan metode Newton – Raphson maka :

   $\beta^{(k+1)} = \beta^{k} - \left(H^{(k)}\right)^{-1} U^{(k)}$, dan $\gamma^{(k+1)} = \gamma^{k} - \left(H^{(k)}\right)^{-1} U^{(k)}$

   Dimana H adalah turunan kedua dari $ln\, L\left(\beta^{(k)}, y, z^{(k)}\right)$ dan $\ln L\left(\gamma^{(k)}, y, z^{(k)}\right)$, U adalah turunan pertama dari $ln\, L\left(\beta^{(k)}, y, z^{(k)}\right)$ dan $\ln L\left(y^{(k)}, y, z^{(k)}\right)$.

5. Ganti $\beta^{(k)}$ dan $\gamma^{(k)}$ dengan $\beta^{(k+1)}$ damn $\gamma^{(k+1)}$ pada iterasi selanjutnya, kemudian kembali lakukan tahap ekspektasi.

6. Tahap ke-3 dan ke-4 ini dilakukan berulang-ulang sampai diperoleh penaksir parameter yang konvergen $\left(\left|\beta^{(k)} - \beta^{(k+1)}\right|\right) \leq \varepsilon$ dan $\left|\gamma^{(k)} - \gamma^{(k+1)}\right| \leq \varepsilon$, biasanya $\varepsilon = 10^{-5}$.

2.13 Pengujian Kesesuaian Model

Menurut Zamzani & Ismail (2013) pengujian kesuaian model regresi ZINB dan ZIP dengan menggunakan *Likelihood Ratio* (LR) *Test* dengan prosedur pengujian :

Hipotesis :

H_0 : $\beta_1 = \beta_2 = \cdots = \beta_p = \gamma_1 = \gamma_2 = \cdots = \gamma_p = 0$ (Variabel bebas secara bersama – sama tidak mempunyai pengaruh terhadap variabel terikat)

H_1 : paling sedikit ada satu $\beta_j \neq 0$ atau $\gamma_j \neq 0$, dengan j $= 1,2, \dots$, p (Variabel bebas secara bersama – sama mempunyai pengaruh terhadap variabel terikat)

dengan model ZINB β_j adalah parameter ke-j dari model $\ln(\mu_i) = x_i^T \beta$ dengan i $= 1, \dots$, n, γ_j adalah parameter ke-j dari model $\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = x_i^T \gamma$ dengan i $= 1, \dots$, n.

dengan model ZIP β_j adalah parameter ke-j dari model $\ln(\mu) = X\beta$, γ_j adalah parameter ke-j dari model $\text{logit}(\omega) = X\gamma$

Statistika uji :

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

dengan L_0 adalah likelihood tanpa variabel bebas

L_1 adalah likelihood dengan variabel bebas

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

2.14 Pengujian Signifikansi Parameter

Menurut Myers *et al.* (2010: 218) pengujian signifikansi dibagi menjadi 2 model yaitu :

2.14.1 Pengujian Signifikansi Parameter $\boldsymbol{\beta}$

Pengujian signifikansi parameter model $\ln(\mu_i) = x_i^T\beta$ dengan $i = 1, \dots, n$ untuk model regresi ZINB dan parameter model $\ln(\mu) = X\beta$ untuk model regresi ZIP

Hipotesis :

$H_0: \beta_j = 0$ (tidak ada pengaruh yang signifikan antara variabel bebas terhadap variabel terikat)

$H_1: \beta_j \neq 0$ (ada pengaruh yang signifikan antara variabel bebas terhadap variabel terikat)

Untuk setiap $j = 1, 2, \dots, p$

Statistika uji :

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

Kriteria uji :

Tolak H_0 pada taraf signifikansi α jika $W_j > X^2_{a:1}$ pada taraf alpha. Penolakan H_0 mengindikasikan bahwa penjelas X_j memiliki pengaruh terhadap peubah respon Y pada taraf signifikansi.

2.14.2 Pengujian Signifikansi Parameter γ

Pengujian signifikansi parameter model $\ln\left(\frac{p_i}{1-p_i}\right) = X_i^T\gamma$ dengan $i = 1, \dots, n$ untuk ZINB dan model logit $(\omega) = X\gamma$ untuk ZIP

Hipotesis :

$H_0: \gamma_j = 0$ (tidak ada pengaruh yang signifikan antara variabel bebas terhadap variabel terikat)

$H_1: \gamma_j \neq 0$ (ada pengaruh yang signifikan antara variabel bebas terhadap variabel terikat)

Untuk setiap $j = 1,2, \dots, p$

Statistika uji :

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

Kriteria uji :

Tolak H_0 pada taraf signifikansi α jika $W_j > X^2_{\alpha:1}$ pada taraf alpha. Penolakan H_0 mengindikasikan bahwa penjelas X_j memiliki pengaruh terhadap peubah respon Y pada taraf signifikansi

2.15 Uji Kelayakan Model

Menurut Sharma dan Landge (2013) *Akaike Information Criterion* (AIC) digunakan untuk menilai kinerja model. AIC didefinisikan oleh

$$AIC = -2 \ln L(\hat{\theta}) + 2k$$

dimana $L(\hat{\theta})$ adalah nilai likelihood, dan k adalah jumlah parameter. Model yang terbaik yaitu dengan memilih model yang mempunyai nilai AIC terkecil.

2.16 Program R

R adalah suatu kesatuan software yang terintegrasi dengan beberapa fasilitas untuk manipulasi, perhitungan dan penampilan grafik yang handal. R berbasis pada bahasa pemrograman S, yang dikembangkan oleh AT&T Bell Laboratories (sekarang Lucent Technologies) pada akhir tahun ’70 an. R merupakan versi gratis dari bahasa S dari software (berbayar) yang sejenis yakni

S-PLUS yang banyak digunakan para peneliti dan akademisi dalam melakukan kegiatan ilmiahnya.

Pada awalnya, versi pertama R dibuat oleh Ross Ihaka dan Robert Gentleman dari Universitas Auckland, namun selanjutnya R dikembangkan oleh tim yang disebut tim inti. Tim inti (core team) terdisi dari ahli statistik, ahli komputer dan pemrograman, geografi, ekonomi dari institusi yang berbeda dari seluruh dunia yang mencoba membangun sebuah sistem (software) yang handal namun dengan biaya yang sangat murah. Menurut kutipan dari penghargaan Association for Computing Machinery Software bagi John Chamber 1998, menyatakan bahwa (bahasa pemrograman) S telah merubah orang dalam memanipulasi, visualisasi dan menganalisis data untuk selamanya. R dibuat searah dengan ide yang ada pada bahasa pemrograman S. Banyak projek lainnya yang berkaitan / berbasis / perluasan dari R, seperti geoR, Rattle, R Commander, SciViews R GUI dan lainnya.

2.16.1 Kelebihan Program R

R mempunyai karakteristik tersendiri, dimana selalu dimulai dengan prompt “>” pada console-nya. R mempunyai beberapa kelebihan, diantaranya :

1. Efektif dalam pengelolaan data dan fasilitas penyimpanan. Ukuran file yang disimpan jauh lebih kecil dibanding software lainnya.
2. Lengkap dalam operator perhitungan array
3. Lengkap dan terdiri dari koleksi tools statistik yang terintegrasi untuk analisis data, diantaranya mulai statistik deskriptif, fungsi probabilitas, berbagai macam uji statistik hingga time series.

UNNES
UNIVERSITAS NEGERI SEMARANG

4. Tampilan grafik yang menarik dan fleksibel ataupun costumized
5. Dapat dikembangkan sesuai keperluan dan kebutuhan dan sifatnya yang terbuka, setiap orang dapat menambahkan fitur – fitur tambahan dalam bentuk paket ke dalam software R.

2.16.2 Fungsi Penting dalam R

a) Fungsi dasar matematika
b) Operasi vektor matriks
c) Fungsi dasar statistika
d) Fungsi pembangkit data peubah acak
e) Fungsi untuk menangani grafik.

2.17 Kematian Balita

2.17.1 Balita

Anak Balita adalah sebagai masa emas atau "golden age" yaitu insan manusia yang berusia 0-5 tahun (UU No.20 Tahun 2003). Saat usia batita, anak masih tergantung penuh kepada orang tua untuk melakukan kegiatan penting, seperti mandi, buang air dan makan. Perkembangan berbicara dan berjalan sudah bertambah baik. Namun kemampuan lain masih terbatas.

Balita berperan penting dalam penentuan tingkat kesehatan masyarakat karena dapat menggambarkan kesehatan penduduk secara umum. Salah satu tingkat kesehatan di suatu negara dapat dilihat dari tingkat kematian balita. Menurut Kemenkes RI. (2015: 125) kematian balita di Indonesia sebesar 26,29 per 1000 kelahiran hidup.

2.17.2 Faktor Penyebab Kematian Balita

Menurut Kemenkes RI (2015), beberapa faktor penyebab kematian balita yaitu :

1. Pneumonia

Menurut Kemenkes RI (2016), pneumonia adalah infeksi akut yang mengenai jaringan paru – paru (alveoli). Pneumonia balita ditandai adanya gejala batuk dan atau kesukara bernapas seperti napas cepat, tarikan dinding dada bagian bawah ke dalam (TDDK) atau gambaran radiologi foto thorax/dada menunjukkan infiltrate paru akut. Demam bukan merupakan gejala yang spesifik pada balita.

Menurut Kemenkes RI (2015: 172), pneumonia merupakan penyebab dari 15% kematian balita, yaitu diperkirakan sebanyak 922 balita di tahun 2015. Pneumonia menyerang semua umur di semua wilayah terbanyak terjadi di Asia Selatan dan Afrika sub-sahara. Populasi yang rentan terserang pneumonia adalah anak – anak usia kurang dari 2 tahun usia lanjut lebih dari 65 tahun dan orang yang memiliki masalah kesehatan.

2. Diare

Menurut Depkes RI (2011: 2), diare adalah suatu kondisi dimana seseorang buang air besar dengan konsistensi lembek atau cair, bahkan dapat berupa air saja dan frekuensinya lebih sering (biasanya tiga kali atau lebih) dalam satu hari. Diare disebabkan oleh bakteri, virus, alergi, keracunan, imunodefisiensi.

Menurut Kemenkes RI (2015: 179), penyakit diare merupakan penyakit endemis di Indonesia dan juga merupakan penyakit potensial yang sering disertai

UNNES
UNIVERSITAS NEGERI SEMARANG

dengan kematian. Pada tahun 2015 terjadi kematian sebanyak 30 orang dengan jumlah penderita diare 1.213 orang.

3. Balita Gizi Buruk

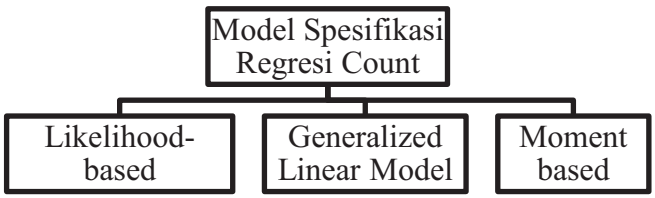
Menurut Depkes RI (2008), gizi buruk adalah suatu keadaan kurang gizi tingkat berat pada anak berdasarkan indeks berat badan menurut tinggi badan (BB/TB) <-3 standar deviasi WHO-NCHS dan atau ditemukan tanda – tanda klinis marasmus, kwashiorkor dan marasmus-kwashiorkor.

Menurut Kemenkes RI (2015: 147), gizi buruk dapat terjadi pada semua kelompok umur, tetapi yang perlu lebih diperhatikan yaitu pada kelompok bayi dan balita. Pada usia 0-2 tahun merupakan masa tumbuh kembang yang optimal (*golden period*) terutama untuk pertumbuhan janin sehingga bila terjadi gangguan pada masa ini tidak dapat dicukupi pada masa berikutnya dan akan berpengaruh negative pada kualitas generasi penerus.

2.18 Kerangka Berpikir

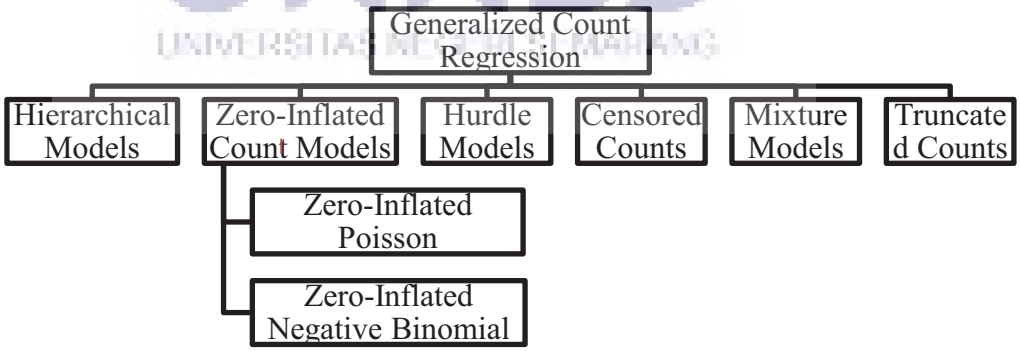
Menurut Cameron & Trivedi (1998: 19) statistik inferensi untuk model – model regresi nonlinear didasarkan pada teori asimtotik. Model dan hasil regresi bervariasi sesuai dengan kekuatan asumsi distribusi yang dibuat. Hasil pengujian hipotesis juga tergantung pada kekuatan asumsi distribusi. Dalam menghitung analisis data cacahan memiliki beberapa pendekatan pemodelan umum yang sering digunakan yaitu likelihood-based, generalized linear models, dan moment based. Secara umum model regresi data cacahan dapat dilihat pada Gambar 2.1

.



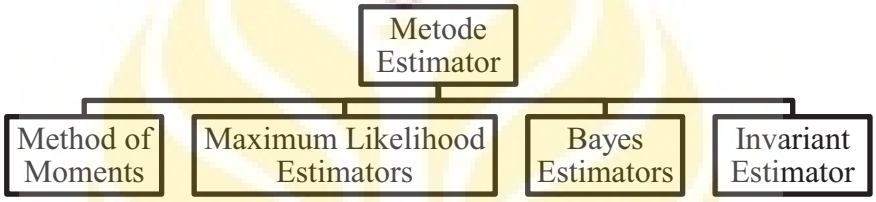
Gambar 2.1. Model Spesifikasi Regresi Count

Salah satu alasan atas kegagalan regresi Poisson adalah bahwa proses Poisson memiliki heterogenitas yang tidak teramati yang memberikan kontribusi keacakan tambahan. Beberapa masalah standar regresi Poisson yang sering terjadi antara lain overdispersi, kelebihan nilai nol atau inflasi nol, pengamatan berkorelasi, tren ketergantungan waktu. Generalized dari suatu model dasar mampu mengatasi masalah dari regresi Poisson. Menurut Cameron & Trivedi (1998: 96) beberapa generalisasi regresi cacahan antara lain *Mixture Models, Truncated Count, Censored Counts, Hurdle Models, Zero-Inflated Count Models,* dan *Hierarchical Models*. Penggunaan *Hierarchical* dan *Censored Models* banyak digunakan dalam kasus underdispersi. Macam – macam *Zero-Inflated Count Models* yaitu *Zero-Inflated Poisson* dan *Zero-Inflated Negative Binomial*. Secara umum model *Generalized Count Regression* dapat dilihat pada Gambar 2.2.



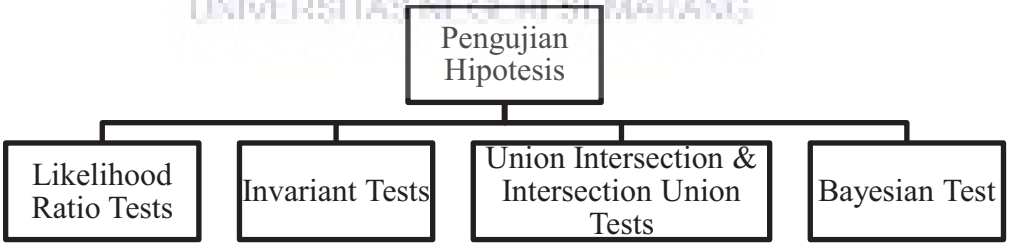
Gambar 2.2. Generalized Count Regression

Menurut Casella dan Berger (1990: 285) dalam beberapa kasus tugas yang mudah untuk memutuskan bagaimana memperkirakan parameter adalah dengan menggunakan metode estimator. Estimator adalah fungsi dari sampel. Seringkali metode mengevaluasi estimator akan menyarankan yang baru. Beberapa metode untuk menemukan estimasi yaitu *Method of Moments, Maximum Likelihood Estimators, Bayes Estimators,* dan *Invariant Estimators*. Secara umum metode estimator dapat dilihat pada Gambar 2.3.



Gambar 2.3. Metode Estimator

Menurut Casella dan Berger (1990: 346) dalam pengujian hipotesis memiliki berbagai macam pengujian dengan berbagai situasi dan keuntungan yang berbeda-beda dari tiap pengujian. Ada 4 metode prosedur pengujian hipotesis yaitu *Likelihood Ratio Tests, Invariant Tests, Bayesian Tests, Union Intersection* dan *Intersection Union Tests*. Secara umum pengujian hipotesis dapat dilihat pada Gambar 2.4.



Gambar 2.4. Pengujian Hipotesis

Analisis regresi ada beberapa macam yaitu analisis regresi linier, regresi berganda, regresi Spasial, regresi Panel dan regresi Poisson. Penelitian ini hanya

akan membahas tentang regresi Poisson. Model regresi yang digunakan dalam menganalisis variabel respon berupa data diskrit dan variabel penjelas berupa kontinu, diskrit atau campuran adalah model regresi Poisson. Namun terkadang dalam analisis regresi Poisson sering terjadi permasalahan dalam asumsinya berupa keadaan equidispersi yaitu nilai *mean* dan varians dari variabel respon sama.

Permasalahan equidispersi yang sering terjadi yaitu berupa nilai varians yang lebih besar dari nilai *mean*nya (overdispersi). Overdispersi disebabkan oleh banyaknya nilai nol yang berlebih pada variabel respon (*excess zero*). Jika regresi Poisson tetap digunakan maka estimasi parameternya kurang baik. Hal ini menyebabkan pengembangan model untuk mengatasi overdispersi yaitu *Zero-Inflated Poisson* dan *Zero-Inflated Negative Binomial.* Kedua model tersebut menggunakan *Maximum Likelihood Estimation* (MLE).

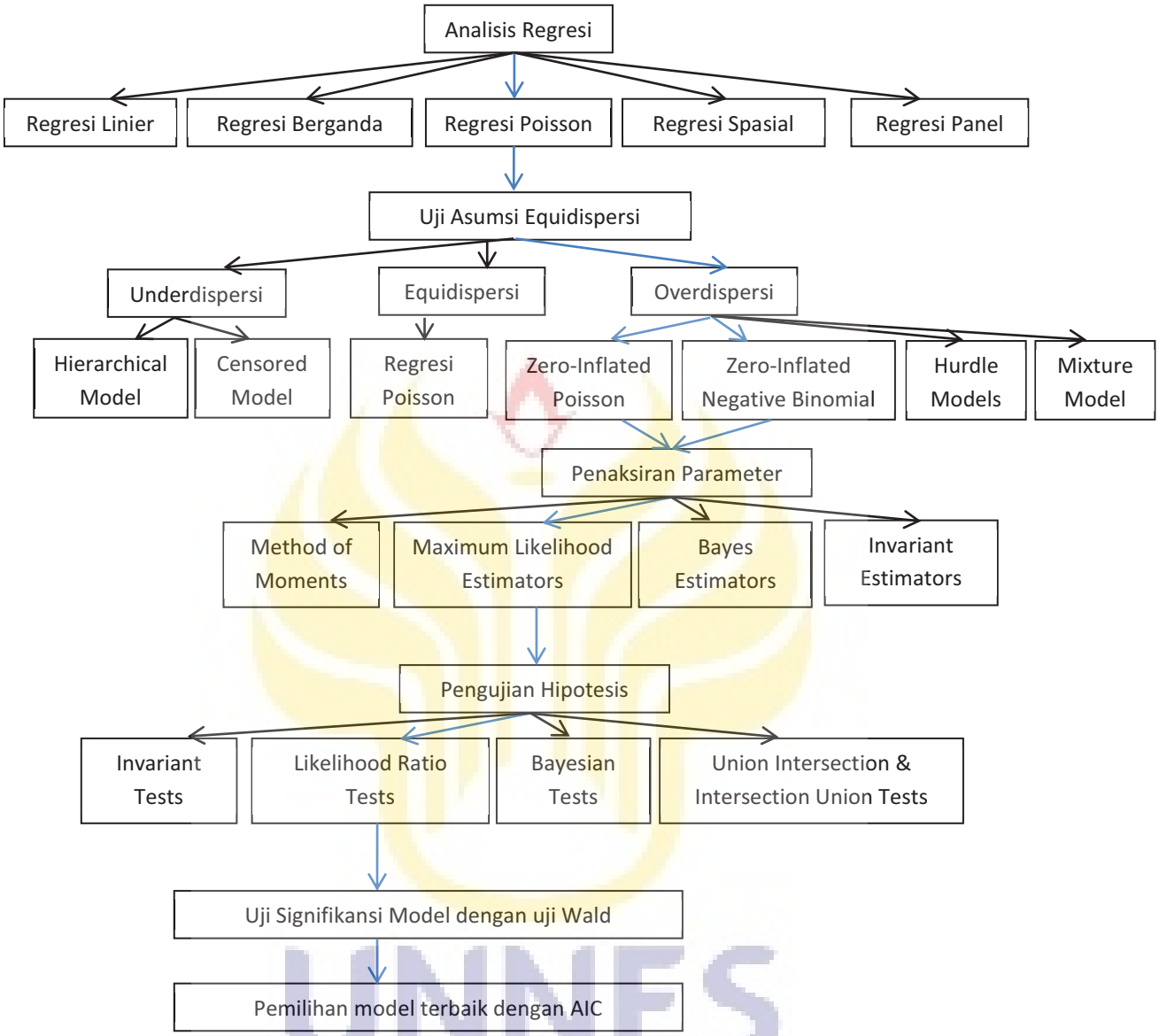
Menurut Casella dan Berger (1990 : 289) menyatakan bahwa MLE digunakan untuk estimasi nilai realisasi. MLE adalah pilihan yang wajar untuk estimasi. MLE digunakan ketika titik parameter yang sampel diamati kemungkinan besar. Menurut Casella dan Berger (1990 : 346) menyatakan bahwa pengujian hipotesis dari estimasi MLE adalah dengan likelihood ratio dengan likelihood berlaku secara luas untuk MLE. Menurut Loeys *et al.* (2012) dalam model poisson atau binomial negatif dapat dilakukan uji menggunakan Likelihood ratio. Likelihood rasio test memungkinkan membantu dalam memilih dari dua model yang diestimasi dengan metode MLE.

UNNES
UNIVERSITAS NEGERI SEMARANG

Mouatassim & Ezzahid (2012), Loeys *et al.* (2012), dan Sharma (2013) model *zero inflated* digunakan dalam mengatasi permasalahan overdispersion khususnya pada data yang banyak amatan nol pada regresi Poisson. Model *zero inflated* terbagi menjadi bermacam – macam yaitu model *Zero-Inflated Poisson* dan *Zero-Inflated Negative Binomial.* Model *Zero-Inflated Poisson* dan *Zero-Inflated Negative Binomial* dalam penelitian ini menggunakan estimasi yaitu *Maximum Likelihood Estimation* dengan pengujian hipotesis menggunakan *Likelihood Ratio Tests.* Penelitian ini menggunakan *Akaike Information Criterion* (AIC) untuk mendapatkan model terbaik.

Secara sistematik alur kerangka berpikir dijelaskan melalu gambar 2.5.

UNNES
UNIVERSITAS NEGERI SEMARANG



Gambar 2.5. Kerangka Berpikir

UNNES
UNIVERSITAS NEGERI SEMARANG

BAB 5

PENUTUP

5.1 Simpulan

Penelitian ini memberikan model Poisson yang terbaik antara model *Zero-Inflated Poisson* dan *Zero-Inflated Negative* Binomial pada data kematian balita di Puskesmas Tirto Kota Pekalongan pada tahun 2016. Dari rumusan masalah dan hasil pembahasan pada BAB 4 maka dapat disimpulkan bahwa

1. Model *Zero-Inflated Poisson* (ZIP) adalah

   Berdasarkan uji kesesuaian model dengan uji *Likelihood Ratio* (LR) *Test* dengan kriteria tolak H_0 pada taraf signifikansi α jika $G_{\text{hitung}} > \chi^2_{\alpha;2p}$ atau p-value $< \alpha$ bahwa model telah sesuai digunakan untuk data kematian balita dan berdasarkan pengujian signifikansi parameter masing – masing individu dengan kriteria tolak H_0 jika $W_i > X^2_{\alpha:1}$ atau p-value $< \alpha$ maka diperoleh model yaitu

   1. Model data diskrit untuk μ adalah

   $$\ln(\mu) = -1{,}75492 + 0{,}10435X_3$$

   Interpretasi model tersebut sebagai berikut :

   a. Konstanta sebesar $-1{,}75492$, artinya jika variabel pneumonia (X_1), balita gizi buruk (X_2), dan diare (X_3), bernilai nol maka banyaknya kematian balita bernilai $exp(-1{,}75492) = 0{,}172921$. Hal ini karena kematian balita dipengaruhi oleh faktor variabel bebas selain pada model.

b. Koefisien X_3 sebesar 0,10435, artinya setiap terjadi diare menyebabkan peningkatan nilai harapan dari banyaknya kematian balita sebesar $exp(0.10435) = 1{,}10998$

2. Model *zero-inflation* untuk ω adalah

$$logit\ (\omega) = -13{,}49422 + 0{,}66060X_3$$

Interpretasi model tersebut sebagai berikut :

a. Konstanta sebesar $-13{,}49422$, artinya jika variabel pneumonia (X_1), balita gizi buruk (X_2), dan diare (X_3), bernilai nol maka banyaknya kematian balita bernilai $exp(-13{,}49422) = 1{,}378906$. Hal ini karena kematian balita dipengaruhi oleh faktor variabel bebas selain pada model

b. Koefisien X_3 sebesar 0,66060, artinya setiap terjadi diare menyebabkan peningkatan resiko tidak terjadi kematian balita sebesar $exp(0{,}66060) = 1{,}93595$.

2. Model *Zero-Inflated Negative Binomial* (ZINB) adalah

Berdasarkan uji kesesuaian model dengan uji *Likelihood Ratio* (LR) *Test* dengan kriteria tolak H_0 pada taraf signifikansi α jika $G_{\text{hitung}} > \chi^2_{\alpha;2p}$ atau p-value $< \alpha$ bahwa model telah sesuai digunakan untuk data kematian balita dan berdasarkan pengujian signifikansi parameter masing – masing individu dengan kriteria tolak H_0 jika $W_i > X^2_{\alpha:1}$ atau p-value $< \alpha$ maka diperoleh model yaitu

1. Model data diskrit untuk μ_i adalah

$$\ln(\mu_i) = -\ 1{,}75491 + 0{,}10435X_3$$

Interpretasi model tersebut sebagai berikut :

a. Konstanta sebesar $-1{,}75491$, artinya jika variabel pneumonia (X_1), balita gizi buruk (X_2), dan diare (X_3), bernilai nol maka banyaknya kematian balita bernilai $exp(-1{,}75491) = 0{,}172922$. Hal ini karena kematian balita dipengaruhi oleh faktor variabel bebas selain pada model.

b. Koefisien X_3 sebesar $0{,}10435$, artinya setiap terjadi diare menyebabkan peningkatan nilai harapan dari banyaknya kematian balita sebesar $exp(0{,}10435) = 1{,}10998$

2. Model *zero-inflation* untuk p_i adalah

$$logit\ (p_i) = -13{,}49431 + 0{,}66060X_3$$

Interpretasi model tersebut sebagai berikut :

a. Konstanta sebesar $-13{,}49431$, artinya jika variabel pneumonia (X_1), balita gizi buruk (X_2), dan diare (X_3), bernilai nol maka banyaknya kematian balita bernilai $exp(-13{,}49431) = 1{,}37878$. Hal ini karena kematian balita dipengaruhi oleh faktor variabel bebas selain pada model

b. Koefisien X_3 sebesar $0{,}66060$, artinya setiap terjadi diare menyebabkan peningkatan resiko tidak terjadi kematian balita sebesar $exp(0{,}66060) = 1{,}93595$

3. Model Poisson terbaik antara *Zero-Inflated Poisson* (ZIP) dan *Zero-Inflated Negative Binomial* (ZINB) adalah model *Zero-Inflated Poisson* (ZIP) dengan nilai AIC (101,1387) yang lebih kecil daripada *Zero-Inflated Negative Binomial* (ZINB) dengan nilai AIC (103,1392).

UNNES
UNIVERSITAS NEGERI SEMARANG

5.2 Saran

Berikut adalah beberapa saran yang dapat diberikan kepada pembaca.

1. Perhitungan estimasi parameter dalam penelitian ini hanya menggunakan program R 2.10.0, penelitian selanjutnya dapat menggunakan perbandingan manual menggunakan *Microsoft Excel.*
2. Diharapkan Puskesmas Tirto Kota Pekalongan hendaknya meningkatkan pencengahan dan penanganan terhadap penyakit diare karena berdasarkan hasil diare memiliki pengaruh penyebab kematian balita.
3. Penelitian selanjutnya untuk menyelesaikan kasus overdispersi dikarenakan *excess zero* pada regresi poisson dapat menggunakan model *Zero-Inflated Generalized Poisson (ZIGP)* atau *Zero-altered Poisson (ZAP)*.

UNNES
UNIVERSITAS NEGERI SEMARANG

DAFTAR PUSTAKA

Agresti, A. 2002. *Categorical Data Analysis* (2th ed.). New York : John Wiley & Sons, Inc.

Ariawan, B., Suparti, & Sudarno. 2012. Pemodelan Regresi Zero-Inflated Negative Binomial (ZINB) Untuk Data Respon Diskrit dengan Excess Zeros. *Jurnal GAUSSIAN*, 1(1) : 55-64.

Cahyandari, R. 2014. Pengujian Overdispersi Pada Model Regresi Poisson. *Statistika,* 14(2) : 69-76.

Cameron, A.C, & P.K. Trivedi. 1998. *Regression Analysis of Count Data*. New York : Cambridge University Press.

Casella, G. & R.L. Berger. 1990. *Statistical Inference*. California: Wadsworth, Inc.

Chipeta, M.G., B.M. Ngwira, C. Simoonga & L.N. Kazembe. 2014. Zero Adjusted Models with Applications to Analysing Helminths Count Data. *BMC*, 7 : 856.

Depkes R.I. 2008. Sistem Kewaspadaan Dini (SKD) KLB-Gizi buruk. Jakarta : Departemen Kesehatan RI.

Depkes, R.I. 2011. *Buku Saku Petugas Kesehatan Lintas Diare Lima Langkah Tuntaskan Diare.* Jakarta: Departemen Kesehatan RI.

Dewi, E.T.K., A. Agoestanto, & Sunarmi. 2016. Metode Least Trimmed Square (LTS) dan MM-Estimation Untuk Mengestimasi Parameter Regresi Ketika Terdapat Outlier. *UNNES Journal of Mathematics*, 5(1) : 47-54

Dinkes Jateng. 2014. *Profil Kesehatan Provinsi Jawa Tengah Tahun 2014.* Semarang: Dinas Kesehatan.

Dinkes Jateng. 2015. *Profil Kesehatan Provinsi Jawa Tengah Tahun 2015.* Semarang: Dinas Kesehatan.

Garay, A.M., E.M. Hashimoto, E.M.M. Ortega, & V.C. Lachos. 2011. On Estimation And Influence Diagnostics for Zero-Inflated Negative Binomial Regression Models. *Computational Statistics and Data Analysis,* 55 : 1304-1318.

Hall, D.B. 2000. Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study. *Biometrics*, 56 : 1030-1039

Harinaldi. 2005. *Prinsip – Prinsip Statistik Untuk Teknik dan Sains*. Jakarta : Erlangga.

UNNES
UNIVERSITAS NEGERI SEMARANG

Hilbe, J.M. 2011. Negative Binomial Regression (2th ed.). New York : Cambridge University Press.

Jansakul, N. & J.P. Hinde. 2002. Score Tests for Zero-Inflated Poisson Models. *Computational Statistics & Data Analysis*, 40 : 75-96.

Kemenkes, R.I. 2015. *Profil Kesehatan Indonesia*. Jakarta : Kementrian Kesehatan Republik Indonesia.

Kemenkes, R.I. 2016. *Pedoman Pencegahan dan Pengendalian Infeksi Saluran Pernapasan Akut*. Jakarta : Kementerian Kesehatan Republik Indonesia.

Lambert, D. 1992. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*, 34(1) : 1-14.

Loeys, T., B. Moerkerke, O.D. Smet, & A. Buysse. 2012. The Analysis of Zero-Inflated Count Data : Beyond Zero-Inflated Poisson Regression. *British Journal of Mathematical and Statistical Psychology*, 65 : 163-180.

Long, D.L., J.S. Preisser, A.H. Herring, & C.E. Golin. 2014. A Marginalized Zero-Inflated Poisson Regression Model with Overall *Exp*osure Effects. *Statistics in Medicine*, 33 : 5151-5165.

Mc.Cullagh, P. & J.A. Nelder. 1989. Generalized Linear Models (2th ed.). London : Chapman and Hall.

Mouatassim, Y. & E.H. Ezzahid. 2012. Poisson Regression and Zero-Inflated Poisson Regression : Application to Private Health Insurance Data. *Eur.Actuar. J.*, 2 : 187-204.

Myers, R.H., D.C. Montgomery, G.G. Vining, & T.J. Robinson. 2010. *Generalized Linear Models with Application in Engineering and The Sciences* (2th ed.). New Jersey : John Wiley and Sons.

Ruliana, P. Hendikawati, & A. Agoestanto. 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

Saffari, S.E., & R. Adnan. 2012. Parameter Estimation on Zero-Inflated Negative Binomial Regression with Right Truncated Data. *Sains Malaysiana*, 41(11) : 1483 – 1487.

Sharma, A.K. & V.S. Landge. 2013. Zero Inflated Negative Binomial for Modelling Heavy Vehicle Crash Rate on Indian Rural Highway. *International Journal of Advances in Engineering & Technology,* 5(2) : 292-301.

Sudjana. 2005. *Metoda Statistika*. Bandung : Tarsito.

Taufan, M., Suparti, & A. Rusgiyono. 2012. Analisis Faktor-Faktor Yang Mempengaruhi Banyaknya Klaim Asuransi Kendaraan Bermotor Menggunakan Model Regresi *Zero-Inflated Poisson* (Studi Kasus di PT. Asuransi Sinar Mas Cabang Semarang Tahun 2010). *Media Statistika,* 5(1): 49-61.

Winkelman, R. 2008. *Econometric Analysis of Count Data* (5th ed.). Berlin : Springer.

Yulianingsih, K.A., K.G. Sukarsa, & L.P. Suciptawati. 2012. Penerapan Regresi Poisson Untuk Mengetahui Faktor-Faktor Yang Memengaruhi Jumlah Siswa SMA/SMK Yang Tidak Lulus UN di Bali. *e-Jurnal Matematika*, 1(1) : 59-63.

Zamani, H. & N. Ismail. 2013. Score Test for Testing Zero-Inflated Poisson Regression against Zero-Inflated Generalized Poisson Alternatives. *Journal of Applied Statistic*, 40(9) : 2056-2068.

UNNES
UNIVERSITAS NEGERI SEMARANG