PEMODELAN FREKUENSI DATA KLAIM PADA ASURANSI KENDARAAN DENGAN MENGGUNAKAN DISTRIBUSI POISSON DAN BINOMIAL NEGATIF

SKRIPSI

NUR ANISA

H 121 15 023

PROGRAM STUDI STATISTIKA

DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2020

PEMODELAN FREKUENSI DATA KLAIM PADA ASURANSI KENDARAAN DENGAN MENGGUNAKAN DISTRIBUSI POISSON DAN BINOMIAL NEGATIF

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains

pada Program Studi Statistika Departemen Statistika

Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Hasanuddin

NUR ANISA

H121 15 023

PROGRAM STUDI STATISTIKA

DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2020

LEMBAR PERNYATAAN KEOTENTIKAN

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

PEMODELAN FREKUENSI DATA KLAIM PADA ASURANSI KENDARAAN DENGAN MENGGUNAKAN DISTRIBUSI POISSON DAN BINOMIAL NEGATIF

Adalah benar-benar hasil karya dan pekerjaan saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan sebelumnya.

Makassar, 7 Februari 2020

METERAI TEMPEL
TGL. 20
26EFEAHF738294154
6000
ENAM RIBU RUPIAH

NUR ANISA
NIM : H121 15 023

PEMODELAN FREKUENSI DATA KLAIM PADA ASURANSI KENDARAAN DENGAN MENGGUNAKAN DISTRIBUSI POISSON DAN BINOMIAL NEGATIF

NUR ANISA

H121 15 023

Disetujui oleh :

Pembimbing Utama	**Pembimbing Pertama**
Dr. Amran S.Si.,M.Si	**Andi Kresna Jaya.,S.Si.,M.Si**
NIP. 19701101 199802 1 001	**NIP. 19731228 200003 1 001**

Pada Tanggal : 7 Februari 2020

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : NUR ANISA

NIM : H121 15 023

Departemen : STATISTIKA

Judul Skripsi : Pemodelan Frekuensi Data Klaim Pada Asuransi Kendaraan dengan Menggunakan Distribusi Poisson dan Binomial Negatif

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

DEWAN PENGUJI

Tanda Tangan

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

Ditetapkan di : Makassar

Tanggal : 7 Februari 2020

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIK

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

Nama : Nur Anisa
NIM : H121 15 023
Program Studi : Statistika
Departemen : Statistika
Fakultas :Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas royalti Non-eksklusif** (***Non-exclusive Royalty Free Right***) atas skripsi saya yang berjudul :

"Pemodelan Frekuensi Data Klaim Pada Asuransi Kendaraan dengan Menggunakan Distribusi Poisson dan Binomial Negatif"

beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal 7 Februari 2020

Yang Menyatakan

Nur Anisa

KATA PENGANTAR

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

Alhamdulillahi rabbil ‘alamin, segala puji dan syukur senantiasa penulis haturkan atas segala limpahan rahmat dan karunia yang telah Allah ‘azza Wajalla berikan kepada penulis. Setelah melalui proses yang panjang Alhamdulillah akhirnya penulis dapat menyelesaikan penelitian ini. Semua kemudahan yang penulis dapatkan tidak terlepas atas pertolongan Allah ‘azza Wajalla sebagaimana sabda Rasulullah shallallahu ‘alaihi wasallam bersabda : “Jagalah Allah, niscaya Dia menjagamu. Jagalah Allah, Dia senantiasa bersamamu. Jika kamu memohon sesuatu, mohonlah kepada-Nya. Ketahuilah seandainya semua ummat manusia bersatu untuk memberikan suatu kebaikan kepadamu, mereka tidak akan mampu kecuali sudah ditetapkan Allah untukmu. Dan seandainya semua ummat manusia bersatu untuk mencelakakanmu, mereka tidak akan mampu kecuali keburukan yang telah Allah tetapkan untukmu. Pena sudah diangkat dan tinta sudah kering” (HR.Tirmidzi). Shalawat dan salam penulis haturkan kepada baginda Rasulullah shalallahu ‘alaihi wasallam sebagai sebaik-baik contoh tauladan bagi ummat islam sepanjang masa. Shalawat dan salam juga senantiasa penulis haturkan kepada keluarga Rasulullah, sahabat-sahabatnya beserta pengikut-pengikutnya yang setia. Penulisan skripsi yang berjudul “Pemodelan Frekuensi Data Klaim Pada Asuransi Kendaraan dengan Menggunakan Distribusi Poisson dan Binomial Negatif ” ini adalah sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penyusunan skripsi ini tidak terlepas dari bantuan, pengarahan dan bimbingan dari berbagai pihak. Karena itu pada kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Keluarga penulis terkhusus orang tua tercinta **Ayahanda Syamsul** dan **Ibunda Kurniati** atas semua nasihat, kerja keras, pengorbanan, doa dan kasih sayang yang tulus kepada penulis. Semoga Allah Subhanahu wata’ala senantiasa memberkahi, menjaga dan menyayangi kalian. Kakak-kakak penulis **Rusniati** dan **Rusdi** yang senantiasa membantu penulis serta adik-adik penulis **Muh.Ikhsan** dan **Nur Syahbani** yang juga banyak membantu serta memberikan motifasi kepada penulis.
2. **Ibu Prof. Dr. Dwia Aries Tina Palumbuhu, MA**. Selaku rektor Universitas Hasanuddin dan **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam beserta seluruh jajarannya.
3. **Ibu Dr**. **Nurtiti Sunusi, S. Si.,M.Si** selaku Ketua Departemen Statistika, segenap dosen pengajar, dan staf Departemen Statistika yang senantiasa

memberika ilmu serta pengalaman yang sangat berarti selama penulis menjalani pendidikan di Departemen Statistika.

4. **Bapak Dr. Amran S.Si.,M.Si** dan **Bapak Andi Kresna Jaya, S.Si., M.Si** selaku selaku ketua dan sekretaris tim penguji sekaligus sebagai pembimbing yang senantiasa memberikan ilmu serta masukan dalam menyelesaikan tugas akhir ini. Semoga Allah senantiasa menjaga serta memberikan beribu kebaikan kepada beliau.
5. **Bapak Dr. Nirwan Ilyas, M.Si** dan **Bapak Drs. Raupong, M.Si** selaku anggota tim penguji yang senantiasa memberikan masukan dan saran yang sangat membangun dalam menyelesaikan tugas akhir ini. Semoga Allah senantiasa menjaga serta memberikan beribu kebaikan kepada beliau.
6. **Kak Sumarni Abdullah** yang senantiasa berjuang bersama dari awal hingga akhir penyelesaian tugas akhir ini. Semoga Allah senantiasa memberikan kebahagiaan dunia dan akhirat.
7. Sahabat tersayang **Sari Wahyuni** beserta sahabat ‘Friends to Jannah’ **Arni Nengsih, Ainun Fajriah Helmi, Suci Barlian Sari, A. Harismayanti, Lisa Narisda Rahimahullah, Nur Aminah Ahmad, Puji Puspasari, Ina Puspitasari, Sarina dan Nur Hardiyanti Mukhtar** yang senantiasa mendoakan serta memberikan motivasi kepada penulis. Semoga Allah senantiasa menjaga kalian dimanapun berada serta kita semua dapat dikumpulkan lagi di Syurga-Nya kelak.
8. Murobbiyah serta Mudarrisah yang senantiasa memberikan ilmu agama serta mengingatkan penulis dalam jalan kebaikan. Semoga Allah senantiasa menjaga kalian.
9. Teman-teman seperjuangan **Statistika 2015 Universitas Hasanuddin** yang senantiasa membantu serta berbagi ilmu kepada penulis. *Jazakumullahu Khairan Katsiran.*
10. **Akhwat UKM LDK MPM Unhas** yang senantiasa memberikan nasehat serta pengalaman yang berharga selama penulis menempuh pendidikan di kampus merah. *Jazakunnallahu Khairan Katsiran.*
11. Serta semua pihak yang tidak bisa penulis sebut satu persatu yang senantiasa mendoakan serta membantu penulis dalam menyelesaikan tugas akhir dan menyelesaikan studi dikampus merah.

Akhir kata, penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, sehingga saran dan kritik yang bersifat membangun sangat penulis harapkan. Semoga tulisan ini dapat memberikan manfaat terhadap pembaca.

Makassar, 7 Februari 2020

Penulis

ABSTRAK

Penelitian ini, memodelkan frekuensi data klaim asuransi kendaraan dengan menggunakan distribusi poisson dan binomial negatif yang mempertimbangkan karakteristik kendaraan dan profil pengemudi. *Generalized Linear Models* (GLMs) digunakan untuk memodelkan frekuensi data klaim. Karakteristik kendaraan dan profil pengemudi dinyatakan sebagai faktor yang berpengaruh terhadap frekuensi klaim. Estimasi parameter menggunakan metode maksimum *likelihood* dan iterasi numerik Newton Raphson. Pemilihan model berdasarkan nilai *AIC* terkecil. Hasil pemodelan frekuensi klaim pada data *Automobile Claim Datasets* menunjukkan bahwa model distribusi binomial negatif lebih baik dari pada model distribusi poisson. Nilai *AIC* model distribusi poisson dan binomial negatif berturut-turut sebesar 34822 dan 34785.

Kata Kunci : Frekuensi Klaim, *Generalized Linear Model, Distribusi Poisson, Binomial Negatif,* Estimasi Maksimum *Likelihood.*

ABSTRACT

In this research, the frequency of Claims in vehicle insurance were modeled by using a poisson and negative binomial distribution considering the characteristics of the vehicle and driver's profile. Generalized Linear Models (GLMs) are used to model the frequency of claims. The Vehicle characteristics and driver's profiles are stated as factors to influance the frequency of claims. The parameter estimation method was use the Maximum Likelihood Estimation and Newton Raphson Algorithm. The selection of models based on the smallest at AIC value. The results of modeling frequency of claims in the automobile claim datasets showed that a negative binomial distribution model is better than a Poisson distribution model. The AIC value of the Poisson distribution model and negative binomial were 34822 and 34785, respectively.

Keywords: Frequency of Claims, Generalized Linear Model, Poisson distribution, negative binomial, Maximum Likelihood Estimation.

DAFTAR ISI

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Kejadian seperti kecelakaan, sakit, bencana alam dan sebagainya adalah hal yang mungkin terjadi pada seseorang di masa mendatang. Kejadian tersebut dapat menimbulkan kerugian bagi orang atau keluarga yang mengalaminya. Salah satu hal yang dapat dilakukan untuk menyiasati kerugian tersebut adalah dengan mengasuransikan harta atau kendaraan yang dimiliki. Kata asuransi sendiri berasal dari Bahasa Inggris yaitu *insurance* yang artinya pertanggungan. Pengertian asuransi adalah suatu bentuk pengendalian risiko dimana satu pihak mengalihkan risiko ke pihak yang lain dalam hal ini perusahaan asuransi. Salah satu manfaat adanya asuransi yaitu dapat mengalihkan risiko. Asuransi kendaraan merupakan hal menarik untuk dikaji karena merupakan barang investasi untuk kehidupan sehingga banyak masyarakat yang tertarik menggunakan perusahaan asuransi untuk mengatasi risiko pada kendaraan jika terjadi suatu musibah atau kejadian tertentu.

Perusahaan asuransi harus mampu mengatasi risiko yang terjadi dalam suatu periode pertanggungan dan harus mengetahui karakter risikonya. Hal ini bertujuan untuk memprediksi kejadian kerugian dimasa yang akan datang. Risiko adalah ketidakpastian akan terjadinya suatu peristiwa yang dapat menimbulkan kerugian ekonomis. Pada asuransi kendaraan bermotor, penanggung akan memberikan santunan (uang pertanggungan) kepada tertanggung apabila terjadi klaim, sebagai akibat dari kecelakaan kendaraan bermotor yang diasuransikan (Hamda,2018). Dalam asuransi juga dikenal istilah *pure premium* (premi murni). Ohlsson dan Johansson (2010) menyatakan bahwa untuk menghitung harga *pure premium* yaitu hasil kali antara *frekuency claim* dan *severity claim*. Frekuensi klaim adalah jumlah klaim yang terjadi dalam suatu periode waktu tertentu dan severity klaim adalah besarnya biaya yang harus dikeluarkan oleh pihak asuransi dalam suatu periode waktu tertentu. Dimana data frekuensi klaim dan severity klaim merupakan variabel acak dengan distribusi tertentu.

Untuk memodelkan frekuensi klaim dan severity klaim dapat dilakukan salah satunya dengan pendekatan *Generalized linier model* (GLM). Pendekatan GLM merupakan perluasan dari model regresi linier dengan asumsi prediktor memiliki efek linier akan tetapi tidak mengasumsikan distribusi tertentu dari variabel respon dan digunakan ketika variabel respon merupakan anggota dari keluarga eksponensial. GLM bertujuan untuk mengetahui hubungan sebab-akibat, pengaruh dari variable independent terhadap variable dependent. Keunggulan GLM dibandingkan dengan regresi linier biasa terletak pada distribusi (bentuk kurva) variabel dependent. Variable dependent pada GLM tidak hanya berasal dari distribusi normal (kurva lonceng simetris), akan tetapi distribusi-distribusi yang termasuk keluarga eksponensial, yaitu; Binomial, Poisson, Binomial Negative, Normal, Gamma, Invers Gaussian (Zahro,2018).

Beberapa penelitian sebelumnya yang pernah membahas tentang asuransi kendaraan diantaranya yaitu Silvie Kafkova dan Lenka Křivankova (2014) dalam penelitiannya yang berjudul Generelized Linear Model in Vehicle Insurance. Dalam penelitian ini menggunakan distribusi Poisson dan fungsi log-link dengan mempertimbangkan faktor-faktor resiko seperti usia kendaraan, type kendaraan, usia pengemudi, area tempat tinggal dan jenis kelamin pengemudi. Adapun hasil dari penelitian ini yaitu pendekatan GLM lebih tepat digunakan untuk memodelkan asuransi kendaraan dengan faktor-faktor yang signifikan berpengaruh terhadap model yaitu umur pengemudi, usia kendaraan dan area tempat tinggal. Penelitian lain juga dilakukan oleh Tina Diningrum (2012) dalam penelitiannya yang berjudul Model Asuransi Kendaraan Bermotor Menggunakan Distribusi Mixed Poisson. Adapun hasil dari penelitiannya bahwa distribusi binomial negatif lebih tepat digunakan untuk data klaim asuransi kendaraan bermotor sedangkan distribusi geometrik tidak bisa digunakan untuk menggantikan distribusi poisson. Adapun pada penelitian kali ini akan membandingkan model poisson dan binomial negatif untuk diterapkan pada frekuensi data klaim asuransi kendaraan.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang maka dirumuskan permasalahan yaitu Bagaimana menentukan model frekuensi data klaim dengan menggunakan distribusi poisson dan binomial negatif pada data asuransi kendaraan?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini yaitu mengkaji model frekuensi klaim dengan menggunakan distribusi poisson dan binomial negatif untuk data asuransi kendaraan.

1.4 Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah untuk menentukan model terbaik frekuensi data klaim antara distribusi poisson dengan binomial negatif pada data asuransi kendaraan.

1.5 Manfaat

Adapun manfaat dari penelitian ini adalah dapat mengetahui model terbaik frekuensi data klaim antara distribusi poisson dengan binomial negatif pada data asuransi kendaraan. Serta mengetahui faktor-faktor apasaja yang berpengaruh terhadap kenaikan ataupun penurunan dari data frekuensi klaim.

BAB II
TINJAUAN PUSTAKA

2.1 Pengertian Asuransi

Menurut Kamus Besar Bahasa Indonesia (KBBI) asuransi atau pertanggungan diartikan sebagai perjanjian antara dua pihak, pihak yang satu berkewajiban membayar iuran dan pihak yang lain berkewajiban memberikan jaminan sepenuhnya kepada pembayar iuran apabila terjadi sesuatu yang menimpa pihak pertama atau barang miliknya sesuai dengan perjanjian yang dibuat.

Definisi asuransi menurut Kitab Undang-Undang Hukum Dasar (KUHD) tentang asuransi atau pertanggungan seumurnya, Bab 9, Pasal 246 : Asuransi atau pertanggungan adalah suatu perjanjian dengan mana seorang penanggung mengikatkan diri kepada seorang tertanggung, dengan menerima suatu premi, untuk memberikan penggantian kepadanya karena suatu kerugian, kerusakan atau kehilangan keuntungan yang diharapkan, yang mungkin akan dideritanya karena suatu peristiwa yang tak tertentu. Jadi asuransi secara umum adalah menyerahkan pertanggungan risiko kepada penanggung yaitu perusahaan asuransi untuk jangka waktu dan perjanjian-perjanjian yang telah disepakati (Diningrum, 2012).

2.2 Model Resiko Asuransi

Menurut (L.J Gitman, 2012) perusahaan dikatakan memiliki risiko yang kecil jika presentase kerugiannya $(0-5\%)$, risiko sedang $(5-10\%)$, dan risiko tinggi jika $(>10\%)$ terdapat dua pendekatan untuk membentuk model distribusi klaim selama periode asuransi yaitu model risiko kolektif dan model risiko individual.

2.2.1 Model Risiko Individual

Dalam risiko individual, klaim-klaim dari organisasi pertanggungan dimodelkan sebagai jumlah klaim dari beberapa tertanggung individual. Pada model risiko individu, misalkan kerugian *(loss)* untuk setiap polis, X_i untuk i= 1,2,3,..n, terjadi pada suatu blok. Asumsi kerugian-kerugian tersebut saling bebas dan berdistribusi identik; $X_i, \dots X_n$ sampel acak dari X (Supriyadi, 2005)

2.2.2 Model Risiko Kolektif

Untuk model risiko yang bersifat kolektif, dapat digunakan suatu asumsi proses acak yang klaimnya berkelanjutan pada portofolio polis asuransi. Secara umum dapat diformulasikan dengan:

Misalkan N sebagai jumlah klaim yang dihasilkan dari portofolio polis yang diberikan dalam periode waktu tertentu dan X_1 dinotasikan sebagai besar klaim ke 1, X_2 sebagai besar klaim ke 2 dan seterusnya, sehingga total besar klaim adalah penjumlahan dari besar klaim individual, sehingga dapat dituliskan sebagai berikut :

$$S = X_1 + X_2 + \cdots + X_N \qquad (2.1)$$

Persamaan (2.1) menyatakan *aggregate* klaim yang dihasilkan oleh portofolio dari periode tersebut. Jumlah individual klaim $X_1, X_2, \ldots X_N$ adalah variabel acak dan disebut *measure the severity* of claims (Mentari, 2017).

N merupakan peubah acak yang menyatakan jumlah klaim bernilai bulat non negatif. N memuat N_1, N_2,... yaitu kejadian dengan jumlah klaim yang mungkin terjadi selama satu periode asuransi. Jumlah klaim merupakan barisan kejadian acak yang dapat diprediksi melalui distribusi dengan peubah acak yang bebas. Sehingga klaim pada tahun kedua tidak dipengaruhi oleh klaim pada tahun pertama dan seterusnya. Jumlah klaim yang terjadi selama periode tunggal tertentu menjadi data yang diperlukan dalam penentuan distribusi jumlah klaim (Pramesti, 2011).

2.3 Generelized Linear Models (GLMs)

GLMs merupakan bentuk umum atau general dari model linear. Diketahui vektor **y** memiliki *n* komponen, yang merupakan realisasi dari sebuah matrik respon $\boldsymbol{Y}$. Setiap komponennya independen dan berdistribusi dengan mean atau E(**Y**)= $\boldsymbol{\mu}$. Jika model yang terbentuk memiliki prediktor $\boldsymbol{X}$, dengan beberapa parameter yang tidak diketahui $\beta_1, \ldots, \beta_n$, maka modelnya adalah berupa kombinasi linear $\boldsymbol{\mu} = \sum_{j=1}^{p} x_j \beta_j$, atau jika dituliskan dalam bentuk matriks menjadi $\boldsymbol{\mu} = \boldsymbol{X\beta}$.

Sebagai transisi dari model linear ke Generalized Linear model, maka dijabarkan bentuk melalui tiga buah komponen, yaitu:

1. *Random Component,* yaitu nilai-nilai pengamatan respon **Y** yang saling bebas dari berdistribusi tertentu.

2. *Systematic Component*,yaitu kombinasi linear dari variabel **X** dengan Parameter β yang dilambangkan dengan $\boldsymbol{\mu}= \boldsymbol{X\beta}$

3. *Link between random and Systematic/ link function,* yaitu suatu fungsi yang menjelaskan nilai ekspektasi dari variabel respon (**Y**) yang menghubungkan dengan variabel-variabel penjelas melalui persamaan linier. Dituliskan dengan η_i = g (μ) fungsi g (μ)inilah yang disebut dengan fungsi penghubung atau *link function*.

Ketiga komponen tersebut, *link function* akan menentukan model yang akan digunakan dalam GLMs. *Link function* paling sederhana adalah $g(\mu) = \eta$ disebut sebagai penghubung identitas *(identity link).* Pada GLMs, variable respon **Y** dapat berdistribusi selain Normal, yang masuk dalam *Exponential Family* (Zahro, 2018).

Menurut Mc Cullagh dan Nelder (1989), fungsi hubung adalah suatu fungsi yang menghubungkan prediktor linier η dengan nilai harapan respon *y* yaitu μ. Berikut ini penghubung kanonik (η) untuk beberapa distribusi (Jong dan Heller, 2008). Fungsi *link* dapat menjelaskan hubung linier antara fungsi transformasi dari mean, g (μ) dengan variable predictor yang menghubungkan komponen sistematis η terhadap nilai *mean* μ. Fungsi hubung atau *link function* merupakan nilai harapan dari komponen acak. Fungsi link dimodelkan sebagai berikut :

$$\eta = g(\mu\ ) \text{ atau } \mu = g^{-1}(\eta)$$

Berikut ini adalah tabel yang menjelaskan tiga komponen utama *GLM* pada model regresi poisson dan binomial negative

Tabel 2.1 *Komponen GLM*

Model Regresi	Komponen Acak	Komponen Sistematik	Fungsi Hubung
Poisson	$Y_i \sim Poisson\ (\mu_i)$	$x_i^T \beta$	Log
Binomial Negatif	$Y_i \sim Poisson\ (\mu_i, K)$	$x_i^T \beta$	Log

Sebuah variabel random Y_i diasumsikan sebagai *exponential dispersion model (EDM)* yang mempunyai fungsi densitas yang berasal dari (Landriault,2019) :

$$f_{Y_i}(\, y_i, \theta_i, \phi) = \exp\left\{\frac{y_i\theta_i - b\,(\theta_i)}{\frac{\phi}{w_i}} + c\,(y_i, \phi, w_i)\right\}$$

keterangan :

θ_i : merupakan parameter kanonik

$\phi > 0$: merupakan parameter dispersi

w_i : merupakan besaran pasti (*exposure).*

Nilai mean dan variansi dari EDM ini yaitu :

- Mean

  $E\ (Y_i) \equiv\ \mu_i = b'(\theta_i)$

- Variansi

  Var $(Y_i) = \frac{\phi b''(\theta_i)}{w_i} = \frac{\phi\, v(\mu_i)}{w_i}$

  dimana :

  $v(\mu_i) = b''((b')^{-1}\mu_i)$

Pembuktian:

Fungsi kepadatan untuk semua distribusi keluarga eksponensial mempunyai bentuk umum sebagai berikut:

$$f(y_i; \theta_i, \phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi/w_i} + c(y_i, \phi, w_i)\right\}$$

Untuk menentukan mean, mula-mula ditentukan turunan pertama dari fungsi $f(\cdot)$ sebagai berikut

$$\frac{\partial f(y_i;\theta_i,\phi)}{\partial\theta_i}=\frac{y_i-b'(\theta_i)}{\phi/w_i}f(y_i;\theta_i,\phi)$$

Kemudian masing-masing ruas di integralkan, sehingga diperoleh :

$$\int_0^\infty \frac{\partial f(y_i;\theta_i,\phi)}{\partial\theta_i}\,dy_i=\int_0^\infty \frac{y_i-b'(\theta_i)}{\phi/w_i}f(y_i;\theta_i,\phi)\,dy_i$$

$$\frac{\partial}{\partial\theta_i}\int_0^\infty f(y_i;\theta_i,\phi)\,dy_i=\frac{1}{\phi/w_i}\int_0^\infty f\{y_i-b'(\theta_i)\}f(y_i;\theta_i,\phi)\,dy_i$$

$$\frac{\partial}{\partial\theta_i}1=\frac{1}{\phi/w_i}\int_0^\infty y_i f(y_i;\theta_i,\phi)dy_i-b'(\theta_i)\int_0^\infty f(y_i;\theta_i,\phi)\,dy_i$$

$$0=\frac{1}{\phi/w_i}\{E[Y_i]-b'(\theta_i)\}$$

$$E[Y_i]=b'(\theta_i)$$

Untuk variansi, dapat diperoleh dari turunan kedua fungsi $f(\cdot)$, yaitu

$$\frac{\partial^2 f(y_i;\theta_i,\phi)}{\partial\theta_i^2}=-\frac{b''(\theta_i)}{\phi/w_i}f(y_i;\theta_i,\phi)+\left\{\frac{y_i-b'(\theta_i)}{\phi/w_i}\right\}^2 f(y_i;\theta_i,\phi)$$

Kemudian masing-masing ruas di integralkan, sehingga diperoleh

$$\int_0^\infty \frac{\partial^2 f(y_i;\theta_i,\phi)}{\partial\theta_i^2}\,dy_i=\int_0^\infty \left\{-\frac{b''(\theta_i)}{\phi/w_i}f(y_i;\theta_i,\phi)+\left(\frac{y_i-b'(\theta_i)}{\phi/w_i}\right)^2 f(y_i;\theta_i,\phi)\right\}dy_i$$

$$\frac{\partial^2}{\partial\theta_i^2}\int_0^\infty f(y_i;\theta_i,\phi)dy_i=-\frac{b''(\theta_i)}{\phi/w_i}\int_0^\infty f(y_i;\theta_i,\phi)dy_i+\int_0^\infty\left(\frac{y_i-b'(\theta_i)}{\phi/w_i}\right)^2 f(y_i;\theta_i,\phi)dy_i$$

$$\frac{\partial^2}{\partial\theta_i^2}1=-\frac{b''(\theta_i)}{\phi/w_i}+\frac{1}{(\phi/w_i)^2}\int_0^\infty\{y_i-E[Y_i]\}^2 f(y_i;\theta_i,\phi)dy_i$$

$$0=-\frac{b''(\theta_i)}{\phi/w_i}+\frac{1}{(\phi/w_i)^2}\,E\{(y_i-E[Y_i])^2\}$$

$$0=-\frac{b''(\theta_i)}{\phi/w_i}+\frac{Var\,[Y_i]}{(\phi/w_i)^2}$$

$$Var\,[Y_i]=\frac{\phi b''(\theta_i)}{w_i}$$

2.4 Distribusi Poisson

Peubah acak Y_i dikatakan berdistribusi poisson jika memiliki fungsi peluang sebagai berikut (Hamda, 2018):

$$Pr\ (Y_i = y_i) = \frac{e^{-\lambda}\,\lambda^{y_i}}{y_i!}\ ;\ y_i = 0,1,2,\ldots$$

dimana y_i adalah banyaknya klaim yang terjadi selama suatu selang waktu atau daerah tertentu dan λ adalah rata-rata banyaknya kejadian dalam selang waktu atau daerah tertentu. Distribusi poisson memiliki suatu karakteristik khusus yaitu nilai rataan dan variansi nya sama yang disebut equidispersi. Nilai rataan dan variansi dari distribusi poisson adalah sebagai berikut :

$$E\ (Y) = \lambda$$

$$Var\ (Y) = \lambda$$

Beberapa karakteristik dari percobaan yang mengikuti sebaran distribusi Poisson (Eko dan Hasbi, 2017) antara lain :

1. Kejadian yang terjadi pada populasi yang besar dengan probabilitas yang kecil
2. Bergantung pada panjang interval waktu yang sama
3. Kejadian yang termasuk dalam *counting process* atau termasuk kedalam lingkupan proses stokastik.

Distribusi Poisson memberikan suatu model yang realistis untuk berbagai macam fenomena random selama nilai dari variabel random Poisson adalah bilangan integer non negatif, banyak fenomena random untuk suatu *count* dari beberapa respon (variabel yang diteliti) merupakan suatu calon untuk pemodelan yang mengasumsikan distribusi Poisson. Misalkan suatu *count* mungkin berupa jumlah kecelakaan lalu lintas tiap minggu, jumlah panggilan telepon per jam dalam suatu perusahaan yang masuk lewat operator, banyaknya kerusakan per unit dari beberapa material, jumlah aliran listrik tiap satuan panjang kabel, dan lain-lain.

Pada model regresi Poisson diasumsikan bahwa variabel dependen Y_i yang menyatakan jumlah kejadian berdistribusi Poisson, diberikan sejumlah variabel independen x_1,..., x_k. Y_i mengikuti distribusi Poisson dengan fungsi kepadatan peluang yaitu :

$Pr\,(Y_i|x_1,..., x_k\,) = \frac{e^{-\lambda}\,\lambda^{y_i}}{y_i!}$; $\lambda > 0$ dan nol untuk lainnya ; $y_i = 0,1,2,\ldots$

atau

$Y_i \sim Poiss(\lambda); i = 1,2,\ldots,n$

Dalam regresi Poisson hubungan antara variabel respon dan variabel penjelas dapat dituliskan dalam bentuk :

$$E[Y_i|x_i] = \lambda = \beta_0 + x_1\beta_1 + \cdots + x_k\beta_k$$

atau dalam bentuk vektor ditulis :

$$E[Y_i|x_i] = \lambda = \boldsymbol{x_i^T\beta}$$

Jika nilai $\lambda > 0$, maka digunakan fungsi link η_i= exp($\boldsymbol{x_i^T\beta}$) atau η_i= log λ = $\boldsymbol{x_i^T\beta}$ untuk menghubungkan $\lambda = E[Y_i|x_i]$ dengan fungsi linear $\boldsymbol{x_i^T\beta}$, sehingga hubungan antara $\lambda = E[Y_i|x_i]$ dan $\boldsymbol{x_i^T\beta}$ menjadi tepat. Dengan demikian model regresi dapat ditulis dalam bentuk :

$$E[Y_i|x_i] = \lambda = \exp\left(\boldsymbol{x_i^T\beta}\right); i = 1,2,\ldots,n \tag{2.3}$$

dengan $\boldsymbol{\beta}$ merupakan paremeter yang tidak diketahui dalam model dan harus diestimasi, dan $\boldsymbol{x_i}$ merupakan vektor $px1$ dari variabel penjelas serta $\boldsymbol{\beta}$ merupakan vektor $px1$ dari parameter regresi.

2.5 Distribusi Binomial Negatif

Berdasarkan Poisson, rataan λ diasumsikan konstan atau homogen dalam kelas. Dengan asumsi λ berdistribusi gamma dimana rataan $E(\lambda)$ =μ_i dan variansi $var(\lambda) = \mu_i{}^2 v_i{}^{-1}$, dan $Y_i|\lambda$ menjadi Poisson dengan rataan bersyarat $E(Y_i|\lambda) = \lambda$ dapat ditunjukkan bahwa distribusi marjinal Y_i mengikuti distribusi binomial negatif dengan fungsi kepadatan peluang sebagai berikut (Irwan & Devni, 2013) :

$$\Pr(Y_i = y_i) = \int_0^\infty \Pr(Y_i = y_i|\lambda_i) f(\lambda_i) d\lambda_i$$

$$= \frac{\Gamma\,(y_i + v_i)}{\Gamma\,(y_i + 1)\Gamma(v_i)} \left(\frac{v_i}{v_i + \mu_i}\right)^{v_i} \left(\frac{\mu_i}{v_i + \mu_i}\right)^{y_i} \tag{2.4}$$

dimana rataan $E(Y_i) = \mu_i$ dan variansi $Var(Y_i) = \mu_i + {\mu_i}^2 {v_i}^{-1}$

Misalkan $Y_i \sim Poisson(\lambda)$, terhadap λ itu sendiri adalah variabel random dengan distribusi Gamma. Misalkan $Y_i|\lambda \sim Poisson(\lambda)$ dan $\lambda \sim Gamma(v_i, \frac{\mu_i}{v_i})$ dengan $Gamma(v_i, \frac{\mu_i}{v_i})$ adalah distribusi gamma dengan rataan $E(\lambda) = \mu_i$ dan $Var(\lambda) = {\mu_i}^2 {v_i}^{-1}$, dengan fungsi kepadatan peluang :

$$\Pr(\lambda) = \frac{1}{\left(\frac{\mu_i}{v_i}\right)^{v_i} \Gamma(v_i)} \lambda^{v_i - 1} exp\left(\frac{-\lambda}{\frac{\mu_i}{v_i}}\right); \lambda > 0 \text{ dan nol untuk lainnya.}$$

Dapat ditunjukkan bahwa distribusi tak bersyarat dari y_i adalah binomial negatif, sebagai berikut :

Fungsi kepadatan peluang bersyarat dari y_i adalah

$$Pr(Y_i = y_i \,|\lambda) = \frac{e^{-\lambda} \lambda^{y_i}}{y_i!}$$

Fungsi kepadatan peluang dari λ adalah

$$\Pr(\lambda) = \frac{1}{\left(\frac{\mu_i}{v_i}\right)^{v_i} \Gamma(v_i)} \lambda^{v_i - 1} exp\left(\frac{-\lambda v_i}{\mu_i}\right)$$

Memanfaatkan definisi dari fungsi kepadatan peluang bersyarat, maka didapat fungsi peluang bersama dari y_i dan λ adalah

$$Pr(y_i, \lambda) = \Pr(\lambda)\, Pr(Y_i = y_i|\lambda)$$

$$= \frac{1}{\left(\frac{\mu_i}{v_i}\right)^{v_i} \Gamma(v_i)} \lambda^{v_i - 1} e^{\left(\frac{-\lambda v_i}{\mu_i}\right)} \frac{e^{-\lambda} \lambda^{y_i}}{y_i!}$$

$$= \frac{1}{\left(\frac{\mu_i}{v_i}\right)^{v_i} \Gamma(v_i) y_i!} \lambda^{v_i + y_i - 1} e^{-\left(1 + \frac{v_i}{\mu_i}\right)\lambda}$$

Dengan diperolehnya fungsi peluang bersama dari y_i dan λ maka fungsi peluang tak bersyarat dari y_i adalah :

$$Pr(Y_i = y_i) = \int_0^\infty \frac{1}{\left(\frac{\mu_i}{v_i}\right)^{v_i} \Gamma(v_i) y_i!} \lambda^{v_i + y_i - 1} e^{-\left(1 + \frac{v_i}{\mu_i}\right)\lambda} d\lambda$$

$$= \frac{1}{\left(\frac{\mu_i}{v_i}\right)^{v_i} \Gamma(v_i) y_i!} \int_0^\infty \lambda^{v_i + y_i - 1} e^{-\left(1 + \frac{v_i}{\mu_i}\right)\lambda} d\lambda$$

Misalkan $z_i = \left(1 + \frac{v_i}{\mu_i}\right)\lambda$ maka $dz_i = \left(1 + \frac{v_i}{\mu_i}\right) d\lambda$ dan $\lambda = \left(\frac{\mu_i}{v_i + \mu_i}\right) z_i$ maka dengan demikian didapatkan

$$Pr(Y_i = y_i) = \frac{1}{\left(\frac{\mu_i}{v_i}\right)^{v_i} \Gamma(v_i) y_i!} \int_0^\infty \left[\left(\frac{\mu_i}{v_i + \mu_i}\right) z_i\right]^{v_i + y_i - 1} \left(\frac{\mu_i}{v_i + \mu_i}\right) e^{-z_i} dz_i$$

$$= \frac{\Gamma(y_i + v_i)}{\Gamma(y_i + 1)\Gamma(v_i)} \left(\frac{v_i}{v_i + \mu_i}\right)^{v_i} \left(\frac{\mu_i}{v_i + \mu_i}\right)^{y_i}$$

untuk $y_i = 0,1,2 \ldots$ dan $Pr(Y_i = y_i)$ adalah fungsi kepadatan peluang binomial negatif.

Parameter berbeda dapat menghasilkan berbagai jenis distribusi binomial negatif. Misalnya dengan mengambil $v_i = \alpha^{-1}$, Y_i mengikuti sebuah distribusi binomial negatif dengan rataan $E(Y_i) = \mu_i$ dan $Var(Y_i) = \mu_i(1 + \alpha\mu_i)$, dimana α menunjukkan parameter dispersi (Lawless, 1987) sehingga persamaan (2.4) menjadi :

$$Pr(Y_i = y_i) = \int Pr(Y_i = y_i | \lambda) f(\lambda) d\lambda$$

$$= \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i}\right)^{\alpha^{-1}} \left(\frac{\mu_i}{\alpha^{-1} + \mu_i}\right)^{y_i}$$

$$Pr(Y_i = y_i) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i!\ \Gamma(\alpha^{-1})} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i} (1 + \alpha\mu_i)^{-\alpha^{-1}}$$

2.6 Pengujian Kecocokan Model Distribusi dengan Kolmogorov-Smirnov

Pengujian *Kolmogorov-Smirnov* satu sampel adalah suatu alat untuk menguji kesesuaian antara distribusi pengamatan dengan suatu distribusi teoritis tertentu (Siegel dan Castellan, 1992). Jika terdapat $x_1, x_{2,\ldots} x_N$ sebagai nilai-nilai dari data sample yang telah disusun secara meningkat, *N* adalah besarnya sampel dan *m* merupakan banyak observasi yang sama atau kurang dari *x*, maka dari data sampel yang terurut dapat dibentuk suatu fungsi distribusi komulatif observasi, $S_N(x)$ sebagai berikut :

$$S_N(x) = \begin{cases} 0 & , x < x_1 \\ \frac{m}{N}, & x_m \le x \le x_{m+1} \\ 1 & , x \ge x_N \end{cases}$$

Kemudian dari nilai $S_N(x)$ dibandingkan dengan $F^*(x)$ sebagai fungsi distribusi kumulatif teoritik untuk melihat apakah ada alasan yang kuat untuk menyatakan bahwa $F^*(x)$ adalah fungsi distribusi yang sebenarnya.

Terdapat rumusan hipotesis 2 arah yang dinyatakan dalam :

$$H_0 : F(x) = F^*(x), \qquad -\infty < x < \infty$$

$$H_1 : F(x) \neq F^*(x), \qquad \text{untuk semua x}$$

dengan $F(x)$ merupakan fungsi distribusi yang tidak diketahui.

Pada uji *Kolmogorov-Smirnov*,D_N (D_{hitung}) merupakan selisih maksimum antara $F^*(x)$ dan $S_N(x)$ untuk seluruh x yang merupakan pengukur perbedaan antara model teoritik dan data observasi, sehingga statistik uji dua arah *Kolmogorov-Smirnov* dapat dinyatakan dalam bentuk :

$$D_N = sup\ |F^*(x) - S_N(x)| \qquad (2.5)$$

Untuk suatu taraf nyata α tertentu, pengujian *Kolmogorov-Smirnov* Membandingkan selisih maksimum pengamatan dalam persamaan (2.5) dengan nilai kritis D_N^α (D_{tabel}), yang didefinisikan dengan

$$P(D_N < D_N^\alpha) = 1 - \alpha$$

Nilai-nilai kritis D_N^α untuk taraf nyata α yang berbeda-beda disajikan dalam tabel uji *Kolmogorof-Smirnov* untuk berbagai nilai N. Aturan pengambilan keputusan yaitu terima H_0 jika D_N yang diamati kurang dari nilai kritis D_N^α, maka distribusi yang ditentukan dapat diterima pada taraf nyata α yang ditentukan, jika sebaliknya maka H_0 ditolak atau distribusi yang ditentukan ditolak.

2.7 Jenis Variabel

2.7.1 Variabel Dummy

Variabel dummy atau biasa juga disebut sebagai variabel kategorik adalah variabel yang digunakan untuk membuat kategori data yang bersifat kualitatif (data kualitatif tidak memiliki satuan ukur) agar data kualitatif dapat digunakan dalam analisa regresi maka data terlebih dahulu ditransformasikan ke dalam bentuk kuantitatif. Contoh data kuanlitatif yaitu misalnya jenis kelamin yang terdiri atas laki-laki dan perempuan. Jika ditransformasikan menjadi data kuantitatif menjadi : laki-laki$=0$; perempuan$=1$, atau misalnya tingkat pendidikan yang terdiri atas SMA dan Sarjana, jika ditransformasikan menjadi data kuantitatif menjadi : SMA =0; Sarjana=1.

Variabel dummy dalam regresi sedikit berbeda dengan variabel lainnya baik dalam pengolahan data ataupun saat membaca hasil regresi. Regresi linear atau regresi berganda merupakan suatu fungsi yang menjelaskan hubungan varaibel independen dengan variabel dependen. Satu variabel dependen (Y) biasanya dipengaruhi oleh beberapa variabel independen (X). misalnya variabel produksi dipengaruhi oleh luas lahan, pupuk, jumlah tenaga kerja, modal.

Regresi memiliki beberapa persyaratan yang harus dipenuhi. Karena regresi masuk dalam statistik parametrik, tentunya variabel-variabel didalamnya memiliki skala interval atau rasio. Selain itu data-data yang akan digunakan juga harus memenuhi kaidah asumsi klasik. Tetapi, dari beberapa variable yang kita gunakan, bisa saja satu atau dua variabel tersebut berupa variabel dalam skala nominal atau ordinal. Variabel skala nominal atau ordinal di dalam regresi tersebut biasa dikenal sebagai variabel dummy. Jika data kualitatif tersebut memiliki *m* kategori maka jumlah variabel dummy yang dicantumkan didalam model tersebut adalah (*m-1*)

a. Variabel skala nominal

Variabel dengan skala nominal adalah variabel kategorik yang nilainya tidak memperhatikan urutan. Misalnya jenis kelamin, dimana 0 merepresentasikan jenis kelamin laki-laki dan 1 merepresentasikan jenis kelamin perempuan.

b. Variabel skala ordinal

Variabel dengan skala ordinal adalah variabel kategorik yang nilainya memperhatikan urutan. Misalnya tingkat luka yang dialami dalam kecelakaan (1. Ringan 2. Sedang 3.Berat)

2.7.2 Variabel Kontinue

Variabel kontinue merupakan yang memiliki nilai-nilai yang terdapat dalam suatu selang/interval bilangan real. Contoh variabel kontinue adalah harga kendaraan.

2.8 Estimasi Generelized Linear Model

Metode yang digunakan untuk mengestimasi parameter regresi logistic adalah *Maximum Likelihood Estimation* (MLE). MLE adalah teknik yang digunakan untuk mencari titik tertentu untuk memaksimumkan sebuah fungsi. Langkah yang perlu dilakukan dengan metode MLE adalah dengan membuat fungsi *likelihood* distribusi, membuat transformasi fungsi tersebut dalam bentuk *ln*, menurunan secara parsial terhadap parameter dan menyamakannya dengan nol (Zahro,2018). Fungsi distribusi bersama $P\ (y_1, y_2, \dots, y_n|\beta)$ dari variabel acak Y dinamakan fungsi *likelihood.* Misal terdapat n sampel $y_1, y_2, \dots, y_n$ yang masing-masing memiliki fungsi kepadatan peluang $P\ (y_i|\beta)$, dan β adalah vektor parameter, fungsi *likelihood* dari β yang dinotasikan dengan L(β) adalah (Ntzoufras,2009) :

$$L\ (\boldsymbol{\beta}) = \prod_{i=1}^{n} P\ (y_i, \boldsymbol{\beta}) \qquad (2.6)$$

2.9 Metode Newton Raphson

Estimasi parameter dengan metode Newton Raphson merupakan metode untuk menyelesaikan persamaan non-linear secara iteratif seperti persamaan *likelihood* yang mencari lokasi yang memaksimalkan suatu fungsi. Keunggulan metode ini adalah memiliki laju konvergensi kuadratik, sehingga metode ini lebih cepat untuk kovergen menuju akar pendekatan daripada metode numerik lainnya. Format iteratif dari metode ini adalah :

$$\hat{\beta}^{(i+1)} = \hat{\beta}^{(i)} - \left(D^i\right)^{-1} G^{(i)} \qquad (2.7)$$

dengan :

$\hat{\beta}^{(i)}$: Parameter taksiran pada iterasi ke-i

$\hat{\beta}^{(i+1)}$: Parameter taksiran pada iterasi ke-(i+1)

$G^{(i)}$: turunan pertama fungsi *log-likelihood*, sehingga entri dari $G^{(i)}$ adalah $\frac{\partial l}{\partial \theta\prime}$

$\left(D^i\right)$: turunan kedua fungsi *log-likelihood*, sehingga entri dari $\left(D^i\right)$ adalah $\frac{\partial^2 l}{\partial (\theta)^{2\prime}}$

Iterasi dilakukan sampai $\left|\left|\hat{\theta}^{(i+1)} - \hat{\theta}^{(i)}\right|\right| < \varepsilon,$ dengan $\varepsilon = 10^{-5}$ (Rochmad, 2013). Apabila iterasi berhenti, maka akan diperoleh nilai estimasi parameter.

2.10 Uji Kecocokan (*Goodness Of Fit*) Model

Beberapa uji kecocokan model yang sering digunakan pada GLM adalah Pearson's Chi-Square dan deviance (Ismail dan Jemain, 2007).

2.10.1 Uji Kecocokan (*Goodness Of Fit*) Model Regresi Poisson dengan menggunakan *pearson's Chi-Square*

Menurut McCullagh & Nelder (1989) uji *Goodness Of Fit* dengan menggunakan *pearson's Chi-Square* didefinisikan sebagai berikut:

$$\mathcal{X}^2 = \sum_i \frac{(y_i - \mu_i)^2}{Var\ (\mu_i)}$$

dimana $Var\ (\mu_i)$ yaitu fungsi variansi dari distribusi peluang datanya dan $Var\ (\mu_i) = \frac{Var\ (Y_i)}{\phi}$ karena data yang diobservasi itu berdistribusi Poisson, maka $Var\ (\mu_i) = \mu_i$ sehingga statistik *pearson's Chi-square* untuk regresi Poisson adalah sebagai berikut :

$$\mathcal{X}^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\mu_i} \qquad (2.8)$$

2.10.2 Uji Kecocokan (*Goodness Of Fit*) Model Regresi Binomial Negatif dengan menggunakan *pearson's Chi-Square*

Karena data yang diobservasi berdistribusi Binomial Negatif, maka $Var\ (Y_i) = \mu_i(1 + \alpha\mu_i)$ sehingga statistik *Pearson chi-square* untuk model Binomial Negatif adalah :

$$\mathcal{X}^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\mu_i(1 + \alpha\mu_i)} \qquad (2.9)$$

2.11 Pemilihan Model Terbaik

Pemilihan model terbaik dari beberapa model yang disajikan, dipilih berdasarkan kriteria yang digunakan. Adapun pilihan kriteria sebagai perbandingan model salah satunya dapat menggunakan *Akaike's Information Criterian* (AIC). AIC adalah suatu kriteria yang menyeimbangkan *goodness of fit model* berdasarkan nilai *likelihood* dengan banyaknya parameter dari model. Sehingga, model yang terbaik adalah model dengan nilai *Akaike's Information Criterian* (AIC) terkecil. Nilai AIC diperoleh dari persamaan sebagai berikut :

$$AIC = -2 \log l_{fit} + 2\,k \tag{2.10}$$