PEMODELAN REGRESI *HURDLE NEGATIVE BINOMIAL* (HNB) UNTUK MENGATASI OVERDISPERSI DENGAN *EXCESS ZEROS*

(Studi Kasus Banyaknya Siswa SMA yang Gagal Ujian Nasional Tahun Ajaran 2013/2014 di Kota Malang)

SKRIPSI

oleh
AULYA ZHARFANI
115090507111016

PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2015

PEMODELAN REGRESI *HURDLE NEGATIVE BINOMIAL* (HNB) UNTUK MENGATASI OVERDISPERSI DENGAN *EXCESS ZEROS*

(Studi Kasus Banyaknya Siswa SMA yang Gagal Ujian Nasional Tahun Ajaran 2013/2014 di Kota Malang)

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains dalam bidang Statistika

oleh
AULYA ZHARFANI
115090507111016

PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2015

LEMBAR PENGESAHAN SKRIPSI

PEMODELAN REGRESI *HURDLE NEGATIVE BINOMIAL* (HNB) UNTUK MENGATASI OVERDISPERSI DENGAN *EXCESS ZEROS*
(Studi Kasus Banyaknya Siswa SMA yang Gagal Ujian Nasional Tahun Ajaran 2013/2014 di Kota Malang)

Oleh:
AULYA ZHARFANI
115090507111016

Setelah dipertahankan di depan Majelis Penguji pada tanggal 18 Juni 2015 dan dinyatakan memenuhi syarat untuk memperoleh gelar Sarjana Sains dalam bidang Statistika

Pembimbing

Rahma Fitriani, S.Si., M.Sc., Ph.D
NIP. 197603281999032001

Mengetahui,
Ketua Jurusan Matematika
Fakultas MIPA Universitas Brawijaya

Ratno Bagus E.W, S.Si., M.Si., Ph.D
NIP. 197509082000031003

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Aulya Zharfani
NIM : 115090507111016
Jurusan : Matematika
Program Studi : Statistika
Penulis Skripsi berjudul :
PEMODELAN REGRESI *HURDLE NEGATIVE BINOMIAL* (HNB) UNTUK MENGATASI OVERDISPERSI DENGAN *EXCESS ZEROS* (Studi Kasus Banyaknya Siswa SMA yang Gagal Ujian Nasional Tahun Ajaran 2013/2014 di Kota Malang)

Dengan ini menyatakan bahwa :

1. **Skripsi ini adalah benar-benar karya saya sendiri dan bukan hasil plagiat dari hasil karya orang lain. Karya-karya yang tercantum dalam Daftar Pustaka Skripsi ini, semata-mata digunakan sebagai acuan/referensi.**
2. **Apabila kemudian hari diketahui bahwa ini Skripsi saya merupakan hasil plagiat, maka saya bersedia menanggung akibat hukum dari keadaan tersebut.**

Demikian pernyataan ini saya buat dengan segala kesadaran.

Malang, 18 Juni 2015
Yang Menyatakan

Aulya Zharfani
115090507111016

PEMODELAN REGRESI *HURDLE NEGATIVE BINOMIAL* (HNB) UNTUK MENGATASI OVERDISPERSI DENGAN *EXCESS ZEROS*
(Studi Kasus Banyaknya Siswa SMA yang Gagal Ujian Nasional Tahun Ajaran 2013/2014 di Kota Malang)

ABSTRAK

Regresi Poisson adalah salah satu metode regresi yang digunakan untuk menganalisis hubungan antara variabel respon diskrit Y (*count* data) dengan variabel prediktor X. Pada regresi Poisson asumsi yang harus dipenuhi adalah equidispersi di mana ragam dan rata-rata variabel respon bernilai sama. Namun sering kali terjadi pelanggaran asumsi yaitu overdispersi yaitu keadaan di mana ragam suatu variabel lebih besar dari rata-ratanya, sehingga regresi Poisson tidak tepat lagi digunakan. Regresi *Hurdle Negative Binomial* (HNB) dapat digunakan untuk memodelkan data *count* yang mengalami overdispersi dengan banyaknya nilai nol pada respon (*excess zeros*). Metode *Maximum Likelihood Estimation* (MLE) dengan algoritma *Expectation Maximation* (EM) digunakan untuk menduga parameter regresi HNB. Pengujian hipotesis tentang keberartian parameter secara simultan dan parsial dilakukan dengan menggunakan *Likelihood Ratio Test*. Model regresi HNB ini diterapkan pada kasus Ujian Nasional SMA tahun pelajaran 2013/2014 di kota Malang. Variabel respon pada kasus ini adalah banyaknya siswa yang gagal ujian nasional di setiap SMA di Kota Malang. Dari hasil analisis diperoleh peluang SMA yang semua siswanya lulus UN dipengaruhi oleh akreditasi sekolah, rasio rencana penerimaan-pendaftar, rasio guru-murid dan rasio murid-kelas.

***Kata Kunci** : Overdispersi, Excess Zeros, Hurdle Negative Binomial (HNB)*

HURDLE NEGATIVE BINOMIAL(HNB) REGRESSION FOR OVERDISPERSION WITH _EXCESS ZEROS_
(Case Study of The number of National Examination Failure of Senior High School Students in Malang 2013/2014)

ABSTRACT

Poisson regression is one of regression methods used to analyze the relation between the discrete (count data) response variable Y and the predictor variable X. Poisson regression, assumes equidispersion in which the variance equals to the mean the response variable. This assumption will be violated when the variance is greater than the mean, namely overdispersion. In this case Hurdle Negative Binomial (HNB) will be used instead of the Poisson Regression. It will be useful as well when the respons variabel has excess zeros. The parameter of HNB regression will be estimated based on Maximum Likelihood Estimation (MLE) with Maximation Expectation (EM) algorithm. The hypothesis of the parameter significance are tested using Likelihood Ratio Test. This HNB regression model is applied to model the number of failure in the National Examination Senior High School Student in Malang 2013/2014. The number of fail students in each school is the response variable. The analysis indicates that the probability of the Senior High School with zero failure is affected by school accreditation, the ratio of acceptance plan-registrants, the ratio of students-teachers, and the ratio of students-classrooms.

Keyword ***:*** *Overdispersion, Excess Zeros, Hurdle Negative Binomial (HNB)*

KATA PENGANTAR

Syukur Alhamdulillah penulis panjatkan kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya sehingga penulis dapat menyelesaikan Skripsi ini dengan baik.

Dalam penyusunan Skripsi ini tidak sedikit hambatan yang penulis temui, namun berkat bantuan, dukungan dan doa dari semua pihak, akhirnya segala hambatan tersebut dapat teratasi. Untuk itu penulis mengucapkan banyak terima kasih kepada :

1. Rahma Fitriani, S.Si., M.Sc. Ph.D. selaku dosen pembimbing atas bimbingan, saran, waktu dan kesabaran yang telah diberikan.
2. Achmad Efendi, Ph.D. dan Prof. Dr. Ni Wayan Surya Wardhani, MS. Selaku dosen penguji atas ilmu dan saran yang telah diberikan
3. Ratno Bagus E. W., S.Si., M.Si., Ph.D. selaku Ketua Jurusan Matematika FMIPA Universitas Brawijaya.
4. Seluruh dosen, staff dan karyawan Jurusan Matematika FMIPA Universitas Brawijaya.
5. Ayah, Ibu, Afif, Adi, Mbah, Ramah, dan seluruh keluarga atas cinta, kasih sayang, doa, dan dukungannya.
6. Defri Parminta yang selalu menemani, memberikan tawa, semangat, dukungan dan bantuannya.
7. Sahabat-sahabatku tersayang Siti Umrotin, Kisi Puriawati, Rani Supriadi, Lifa Fitria, dan Cicilia P. atas dukungan dan semangat yang diberikan.
8. Teman-temanku yang membantu dalam belajar Olivia, Windy, Risa dan seluruh teman-teman Statistika B 2011.
9. Semua Pihak yang membantu dalam penyelesaian skripsi ini.

Penyusunan Skripsi ini masih jauh dari kesempurnaan. Untuk itu, penulis mengharapkan kritik dan saran yang bersifat membangun demi perbaikan dan penyempurnaan Skripsi. Semoga Skripsi ini bermanfaat bagi mahasiswa secara umum.

Malang, Juni 2015

Penulis

DAFTAR ISI

DAFTAR TABEL

Halaman

DAFTAR GAMBAR

DAFTAR LAMPIRAN

Halaman

BAB I
PENDAHULUAN

1.1 Latar Belakang

Dalam berbagai kasus penelitian, sering ditemukan persoalan yang melibatkan suatu hubungan antara variabel respon atau variabel dependen dan variabel prediktor atau variabel independen. Analisis mengenai hubungan antara variabel respon dengan variabel prediktor dapat dimodelkan dengan analisis regresi. Analisis regresi yang umum digunakan adalah analisis regresi klasik di mana variabel respon bersifat kontinu dan mengikuti sebaran normal. Namun pada kenyataannya sering dijumpai variabel respon bersifat diskrit (*count* data) dan tidak mengikuti sebaran normal.

Model regresi Poisson adalah salah satu model regresi yang dapat digunakan untuk menganalisis hubungan antara variabel respon yang bersifat diskrit (*count* data) dengan variabel prediktor yang bersifat diskrit, kontinu, ataupun campuran. Regresi Poisson termasuk dalam *Generalized Linier Models* (GLM) dengan fungsi logaritma sebagai penghubung (*link function*). Menurut Hinde dan Demetrio (2007) jika variabel acak diskrit Y_i, $i = 1,\ldots,n$ dengan rata-rata μ_i, dan $Y_i \sim$ Poisson (μ_i) maka asumsi yang harus dipenuhi dalam regresi Poisson adalah *equidispersion* di mana ragam harus sama dengan rata-rata dari variabel tersebut Var (Y_i) = μ_i. Namun pada praktiknya sering kali ditemukan masalah overdispersi dan underdispersi. Overdispersi adalah keadaan di mana ragam suatu variabel lebih besar daripada rata-rata, sebaliknya underdispersi adalah keadaan di mana ragam suatu variabel lebih kecil daripada rata-rata. Menurut Hinde dan Demetrio (2007) jika terjadi overdispersi maka regresi Poisson tidak tepat lagi digunakan karena akan menyebabkan nilai *standart error* menjadi *underestimate* dan mengakibatkan kesalahan dalam siginifikansi parameter regresi secara individu.

Masalah overdispersi pada regresi Poisson dapat dimodelkan dengan menggunakan model regresi *Negative Binomial* (NB) dan regresi *Generalized Poisson* (GP). Namun untuk kasus tertentu misalnya pada kasus banyaknya siswa SMA yang gagal ujian nasional, dengan variabel respon diskrit seringkali ditemukan banyak pengamatan yang bernilai nol (*excess zeros*). Banyaknya pengamatan yang bernilai nol dengan proporsi yang besar (lebih dari 50%) dapat menimbulkan kesalahan dalam menarik kesimpulan. Model regresi *Negative Binomial* (NB) dan regresi *Generalized Poisson* (GP) mengesampingkan adanya *excess zeros.* Terdapat beberapa metode yang dapat digunakan untuk mengatasi overdispersi dengan *excess zeros* diantaranya *Zero Inflated Poisson* (ZIP), *Zero Inflated Negative Binomial* (ZINB), *Zero Infalted Generalized Poisson* (ZIGP) dan *Hurdle Negative Binomial* (HNB).

Menurut Saffari, Adnan dan Greene (2012) model yang dapat digunakan untuk mengatasi masalah overdispersi dengan *excess zeros* adalah model regresi *Hurdle Negative Binomial* (HNB). Model *Hurdle Negative Binomial* dapat mengatasi masalah overdispersi dan underdispersi sekaligus dapat mengakomodasi adanya *excess zeros* lebih tepat di antara model lainnya. Oleh karena itu, disarankan untuk menggunakan model regresi *Hurdle Negative Binomial* (HNB) sebagai alternatif dalam mengatasi masalah overdispersi dengan banyaknya nilai nol pada variabel respon (*excess zeros*).

Masalah overdispersi pada data dengan *excess zeros* sering kali terjadi pada kasus-kasus di bidang kesehatan. Contoh penelitian di bidang kesehatan yang dilakukan oleh Lestari dan Wulandari (2014) tentang jumlah kasus tetanus *neonatorum* (TN) di Jawa Timur. Masalah overdispersi dengan *excess zeros* ini juga sering terjadi di bidang pendidikan.

Pendidikan adalah salah satu program utama yang menjadi perhatian bagi pemerintah. Untuk meningkatkan mutu pendidikan diperlukan berbagai upaya yang harus dilakukan di antaranya peningkatan kualitas tenaga pengajar, peningkatan materi dan pelaksanaan kurikulum yang baik serta peningkatan sarana dan prasana. Dengan adanya upaya peningkatan kualitas pendidikan diharapkan dapat mendukung setiap siswa untuk dapat menyelesaikan pendidikannya tepat waktu dan dapat lulus dalam menempuh Ujian Nasional (UN) yang diselenggarakan pemerintah.

Kegagalan Ujian Nasional (UN) merupakan salah satu masalah pendidikan yang tidak hanya menjadi tanggung jawab pemerintah, sekolah, aparatur pendidikan tetapi juga masyarakat. Kegagalan Ujian Nasional adalah suatu kondisi di mana siswa tidak dapat menyelesaikan studinya di kelas terakhir pada jenjang tertentu yang dilaksanakan oleh pemerintah. Kota Malang merupakan salah satu kota yang terletak di Jawa Timur. Malang adalah kota terbesar kedua di Jawa Timur setelah Surabaya, yang setiap tahunnya masih terdapat masalah siswa yang gagal Ujian Nasional (UN). Kota Malang ini memiliki 43 SMA negeri maupun swasta tersebar di 5 kecamatan di Kota Malang.

Pada penelitian ini dilakukan pendugaan parameter model regresi *Hurdle Negative Binomial* (HNB) untuk memodelkan kasus banyaknya siswa yang gagal ujian nasional. Dalam penelitian ini variabel respon bersifat diskrit yang menyatakan banyaknya siswa yang gagal Ujian Nasional (UN) di setiap SMA baik negeri maupun swasta di Kota Malang. Peluang kegagalan lulus ujian nasional (UN) siswa SMA di Kota Malang sangat kecil sehingga pada variabel respon memiliki banyak pengamatan bernilai nol (*excess zeros*) dan terjadi masalah overdispersi. Oleh sebab itu, peneliti menggunakan regresi *Hurdle Negative Binomial* (HNB) untuk memodelkan kasus banyaknya siswa SMA yang gagal ujian nasional UN di Kota Malang.

1.2 Rumusan Masalah

Rumusan Masalah pada penelitian ini adalah

1. Bagaimana penurunan penduga *Maximum Likelihood Estimation* (MLE) bagi regresi *Hurdle Negative Binomial* (HNB) ?
2. Bagaimana model regresi *Hurdle Negative Binomial* (HNB) pada kasus banyaknya siswa yang gagal ujian nasional pada SMA di Kota Malang yang mengalami overdispersi ?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah

1. Menurunkan penduga *Maximum Likelihood Estimation* (MLE) bagi regresi *Hurdle Negative Binomial* (HNB)

2. Membentuk model regresi *Hurdle Negative Binomial* (HNB) pada kasus banyaknya siswa yang gagal ujian nasional pada SMA di Kota Malang yang mengalami overdispersi

1.4 Batasan Masalah

Batasan masalah dari penelitian ini adalah :

1. Penelitian ini hanya melibatkan Sekolah Menengah Atas (SMA) negeri maupun swasta di Kota Malang pada tahun pelajaran 2013/2014.
2. Pembentukan model regresi *Hurdle Negative Binomial* (HNB) tanpa memperhatikan pencilan

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Memberikan informasi tentang penggunaan regresi *Hurdle Negative Binomial* (HNB) pada kasus yang mengalami overdispersi dengan *excess zeros.*
2. Memberikan informasi mengenai model regresi *Hurdle Negative Binomial* (HNB) pada kasus banyaknya siswa SMA yang gagal ujian nasional di Kota Malang tahun pelajaran 2013/2014.

BAB II
TINJAUAN PUSTAKA

2.1 Distribusi Poisson

Distribusi Poisson ditemukan oleh seorang ahli matematika berkebangsaan Perancis, Simon Denis Poisson pada tahun 1781-1841. Distribusi Poisson merupakan suatu bentuk distribusi untuk peristiwa yang peluang kejadiannya sangat kecil dan bergantung pada interval waktu tertentu dengan hasil pengamatan berupa variabel diskrit. Dalam ekperimen Poisson, peluang memperoleh dengan tepat peristiwa Y sebanyak y kejadian untuk setiap satu satuan unit (waktu dan ruang) yang ditentukan membentuk sebuah distribusi yang fungsi peluangnya adalah :

$$p(y) = \frac{\mu^y e^{-\mu}}{y!}, \quad y = 0,1,2,\ldots \tag{2.1}$$

di mana :

μ : rata-rata banyaknya kejadian dalam satu satuan unit tertentu

e : konstanta dasar (basis) logaritma natural = 2.71828…

Distribusi Poisson memiliki sifat yaitu nilai varians sama dengan nilai rata-ratanya (Gujarati dan Porter, 2009)

$$E(y) = \mu, \tag{2.2}$$

dan

$$V(y) = \mu. \tag{2.3}$$

2.2 Distribusi Binomial Negatif

Apabila dalam sebuah percobaan Binomial Negatif di mana p adalah peluang sukses dan q=1-p adalah peluang gagal dalam setiap percobaan, maka jika variabel acak Y menyatakan banyaknya y gagal sebelum r sukses tercapai pada percobaan tersebut, dapat diperoleh distribusi peluang Binomial Negatif dengan fungsi peluangnya (Harinaldi, 2005):

$$p(y) = \binom{y+r-1}{r-1} p^r (1-p)^y, \quad y = 0,1,2,\ldots.$$

Distribusi Binomial Negatif memiliki banyak cara dalam penurunannya. Menurut McCullagh dan Nelder (1989) distribusi Binomial Negatif merupakan distribusi campuran Poisson Gamma sebagai pendekatan untuk menghadapi data hitung (*count*) yang mengandung overdispersi.

Fungsi peluang distribusi campuran Poisson Gamma dapat diperoleh dengan cara (Cameron dan Trivedi, 1998) :

$$P\,(y|\mu,\delta) = \int \frac{\exp(-\lambda).\lambda^{y}}{y!} \frac{\left(\frac{\delta}{\mu}\right)^{\delta}}{\Gamma(\delta)} \lambda^{(\delta-1)} e^{-\frac{\delta\lambda}{\mu}} d\lambda$$

$$= \frac{\left(\frac{\delta}{\mu}\right)^{\delta}}{\Gamma(\delta)\Gamma(y+1)} \int exp\left(-\lambda\left(1+\frac{\delta}{\mu}\right)\right)\lambda^{(y+\delta-1)}\,d\lambda$$

$$= \frac{(\frac{\delta}{\mu})^{\delta}\;\Gamma(\delta+y)}{\Gamma(\delta)\Gamma(y+1)}\left(\frac{\mu}{\delta+\mu}\right)^{y+\delta}$$

$$= \frac{\Gamma(\alpha^{-1}+y)}{\Gamma(\alpha^{-1})\Gamma(y+1)}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}}\left(\frac{\mu}{\mu+\alpha^{-1}}\right)^{y},\quad y = 0,1,2,\ldots \qquad (2.4)$$

di mana $\frac{1}{\delta} = \alpha$, dan α menunjukkan parameter dispersi.

Distribusi Binomial Negatif memiliki nilai rata-rata $E(Y) = \mu$ dan ragam $Var(Y) = \mu(1+\alpha\mu)$ (Cameron dan Trivedi, 1998).

2.3 Regresi Poisson

Regresi Poisson adalah salah satu metode regresi yang digunakan untuk menganalisis hubungan antara variabel dependen Y berupa data *count* (bersifat diskrit) dengan variabel independen X. Variabel dependen Y yang bersifat diskrit diasumsikan berdistribusi Poisson.

Model regresi Poisson dapat dituliskan sebagai berikut :

$$Y_i = \mu_i + \varepsilon_i \qquad (2.5)$$

di mana Y secara independen berdistribusi sebagai variabel acak Poisson dengan rata-rata μ_i (Gujarati dan Porter, 2009).

Regresi Poisson termasuk dalam *Generalized Linier Models* (GLM), di mana variabel responnya tidak mengharuskan berdistribusi normal. Dalam *Generalized Linier Models* (GLM), fungsi yang menghubungkan mean dari variabel respon μ_i dengan variabel prediktor disebut dengan fungsi penghubung (*link function*) dinyatakan dengan:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} \qquad (2.6)$$

Pada model regresi Poisson, biasanya fungsi penghubung yang digunakan adalah fungsi penghubung log yang berbentuk:

$$g(\mu_i) = \ln \mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta} \qquad (2.7)$$

Hubungan antara mean variabel respond an prediktor linier dengan menggunakan fungsi penghubung log:

$$\ln \mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$$

Kedua ruas diambil fungsi eksponensialnya didapatkan

$$e^{\ln \mu_i} = e^{\boldsymbol{x}_i^T \boldsymbol{\beta}}$$

$$\mu_i = e^{\boldsymbol{x}_i^T \boldsymbol{\beta}} \qquad (2.8)$$

Sehingga fungsi penghubung untuk model regresi Poisson seperti dituliskan pada persamaan di bawah ini:

$$\ln \mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

$$\mu_i = \exp\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right) = exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \qquad (2.9)$$

di mana $\widehat{\boldsymbol{\beta}}$ merupakan penduga kemungkinan maksimum bagi $\boldsymbol{\beta}$.

Berdasarkan persamaan (2.9), maka model umum regresi Poisson pada persamaan 2.5 yaitu $y_i = \mu_i + \varepsilon_i$ dinyatakan sebagai :

$$y_i = \exp\left[\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}\right] + \varepsilon_i \qquad (2.10)$$

yang dapat diduga dengan :

$$\widehat{y}_i = \exp\left[\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip}\right] \qquad (2.11)$$

2.3.1 Pendugaan Parameter Model Regresi Poisson

Pendugaan parameter model regresi Poisson dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE) yaitu dengan cara memaksimumkan fungsi *likelihood* dari regresi Poisson. Pendugaan maksimum Likelihood untuk parameter $\boldsymbol{\beta}$ dinyatakan dengan $\widehat{\boldsymbol{\beta}}$ yang merupakan penyelesaian dari turunan pertama fungsi logaritma natural dari likelihood.

$$\ln L(\boldsymbol{\beta}) = \ln\left(\prod_{i=1}^{n} \frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!}\right)$$

$$= \sum_{i=1}^{n} ln\left(\frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!}\right)$$

$$= \sum_{i=1}^{n} (\ln(e^{-\mu_i}) + \ln(\mu_i^{y_i}) - \ln(y_i!))$$

$$= \sum_{i=1}^{n} \left(-\mu_i + y_i \ln \mu_i - \ln(y_i!)\right)$$

$$= \sum_{i=1}^{n} \left(-e^{\mathbf{x}_i^T \boldsymbol{\beta}} + y_i \ln e^{\mathbf{x}_i^T \boldsymbol{\beta}} - \ln(y_i!)\right)$$

$$= -\sum_{i=1}^{n} e^{x_i^T \boldsymbol{\beta}} + \sum_{i=1}^{n} y_i \, \mathbf{x}_i^T \boldsymbol{\beta} - \sum_{i=1}^{n} \ln(y_i!) \quad (2.12)$$

Fungsi *log-likelihood* di atas, kemudian diturunkan terhadap parameter β yang akan diduga dan menghasilkan persamaan di bawah ini

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\sum_{i=1}^{n} \mathbf{x_i} \exp(\mathbf{x_i^T} \boldsymbol{\beta}) + \sum_{i=1}^{n} y_i \mathbf{x_i} \quad (2.13)$$

Hasil turunan pertama dari fungsi *log-likelihood* pada persamaan (2.13) di atas merupakan fungsi non linier. Oleh karena itu, diperlukan metode numerik untuk memperoleh pendugaan parameternya. Metode numerik yang dapat digunakan adalah metode iterasi *Newton Raphson*.

Langkah-langkah iterasi *Newton Raphson* adalah sebagai berikut (Cameron dan Travedi, 1998)

1. Menentukan nilai dugaan awal parameter ($\hat{\boldsymbol{\beta}}^{(\mathbf{0})}$)
2. Membentuk vektor gradient **g**

$$\mathbf{g}^T_{\mathbf{1}x(\boldsymbol{p}+\mathbf{1})} = \left[\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_0} \frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_1} \frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_2} \cdots \frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_p}\right]_{\beta=\beta^{(m)}}$$

   di mana *p* adalah banyaknya parameter yang akan diduga.

3. Membentuk matriks Hessian **H** :

$$\boldsymbol{H}_{(\boldsymbol{p}+\mathbf{1})x(\boldsymbol{p}+\mathbf{1})} = \begin{bmatrix} \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial {\beta_0}^2} & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_0 \beta_1} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_0 \beta_p} \\ \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_1 \beta_0} & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial {\beta_1}^2} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_1 \beta_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_p \beta_0} & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_p \beta_1} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial {\beta_p}^2} \end{bmatrix}_{\beta=\beta^{(m)}}$$

4. Menggunakan nilai $\hat{\boldsymbol{\beta}}^{(\mathbf{0})}$ ke dalam elemen-elemen vektor **g** dan matriks **H**, sehingga diperoleh vektor $\mathbf{g}(\hat{\boldsymbol{\beta}}^{(\mathbf{0})})$ dan matriks $\mathbf{H}(\hat{\boldsymbol{\beta}}^{(\mathbf{0})})$.

5. Mulai dari m=0 dilakukan iterasi pada persamaan :
$$\widehat{\boldsymbol{\beta}}^{(\boldsymbol{m}+\mathbf{1})} = \widehat{\boldsymbol{\beta}}^{(\boldsymbol{m})} - \left(\boldsymbol{H}^{(\boldsymbol{m})}\right)^{-\mathbf{1}}\mathbf{g}^{(\boldsymbol{m})} \tag{2.14}$$

6. Nilai $\widehat{\boldsymbol{\beta}}^{(\boldsymbol{m})}$ merupakan sekumpulan penaksir parameter yang konvergen pada iterasi ke-*m*.
Jika belum didapatkan penduga parameter yang konvergen $\left|\widehat{\boldsymbol{\beta}}^{(\boldsymbol{m}+\mathbf{1})} - \widehat{\boldsymbol{\beta}}^{(\boldsymbol{m})}\right| \leq \boldsymbol{\varepsilon}$ maka dilanjutkan kembali ke langkah 5 sampai mendapatkan parameter yang konvergen.

2.4 Overdispersi dan Underdispersi

Salah satu asumsi yang harus dipenuhi dalam regresi Poisson adalah nilai rata-rata dan ragam sama atau disebut dengan *equidispersi* (Ismail dan Jemain, 2005). Namun pada kenyataannya sering dijumpai kasus di mana ragam lebih besar atau lebih kecil dari rata-rata. Keadaan di mana nilai ragam suatu variabel lebih besar dari nilai rata-rata disebut overdispersi, sedangkan underdispersi adalah keadaan di mana nilai ragam lebih kecil dari nilai rata-rata (Famoye *et al*, 2004).

Menurut McCullagh dan Nelder (1989), overdispersi dapat terjadi karena adanya pengelompokan (*clustering*) dalam populasi. Cameron dan Trivedi (1998) menjelaskan bahwa fenomena overdispersi dapat terjadi karena adanya sumber keragaman yang tidak teramati (*unobserved heterogeneity*). Overdispersi dapat pula terjadi karena adanya pengamatan hilang pada variabel penjelas, adanya pencilan pada data dan banyaknya pengamatan yang bernilai nol pada variabel respon (*excess zeros*). Konsekuensi terjadinya overdispersi adalah nilai penduga bagi kesalahan baku (*standard error)* yang diperoleh dari model menjadi lebih kecil (*underestimate*) dan menimbulkan kesalahan dalam signifikansi parameter. Akhirnya interpretasi model menjadi tidak benar dan regresi Poisson tidak tepat lagi untuk memodelkan data (Hinde dan Demetrio, 2007).

Menurut Rodriguez (2013) cara untuk mendeteksi overdispersi adalah dengan menggunakan statistik *Pearson chi-squre* yang dibagi dengan derajat bebas sebagai berikut
$$\hat{\phi} = \frac{\chi^2{}_{pearson}}{n-p} \tag{2.15}$$

di mana *n* adalah banyaknya observasi dan *p* adalah banyaknya parameter dan

$$\chi^2_{pearson} = \sum \frac{(y_i - \mu_i)^2}{var\ (y_i)} = \sum \frac{(y_i - \mu_i)^2}{\phi \mu_i} \quad (2.16)$$

di mana,

$\chi^2_{pearson}$: statistik *Pearson chi-squre*

μ_i : nilai harapan y_i

y_i : nilai variabel respon Y

Jika k bernilai 1 maka dapat dikatakan terjadi equidispersi. Jika k >1, maka terjadi keadaan overdispersi sedangkan pada saat k < 1 terjadi underdispersi.

2.5 Model Regresi *Hurdle Negative Binomial*

Salah satu penyebab terjadinya overdispersi adalah banyaknya pengamatan yang bernilai nol. Model yang dapat digunakan untuk mengatasi overdispersi dengan adanya *excess zeros* adalah model *Hurdle Negative Binomial* (HNB).

Dimisalkan Y_i adalah variabel random yang diskrit dengan *i* adalah bilangan bulat non negatif (*i* = 1,2,..., *n*) dan Y_i merupakan variabel respon dari model regresi *Hurdle Negative Binomial*. Model regresi HNB ini memiliki dua keadaan yaitu keadaan pertama terjadi dengan peluang p_i dan menghasilkan hanya observasi bernilai nol *zero state*, sementara keadaan kedua disebut *Negative Binomial state* terjadi dengan peluang (1-p_i) dan berdistribusi Binomial Negatif dengan mean μ dan 0< p_i <1.

$$\mathrm{P(Y_i = y_i)} = \begin{cases} 0 & \text{, dengan peluang } p_i \\ negative\ binomial & \text{, dengan peluang } (1 - p_i) \end{cases} \quad (2.17)$$

Proses kedua keadaan ini dengan variabel Y_i menghasilkan distribusi campuran dua komponen dan di dapatkan fungsi peluang dari Y_i adalah (Saffari *et al*, 2012):

$$Pr\ (Y_i = y_i) = \begin{cases} p_{i,} & y_i = 0 \\ (1 - p_i) \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \frac{(1 + \alpha\mu_i)^{-\alpha^{-1} - y_i} \alpha^{y_i} \mu_i^{y_i}}{1 - (1 + \alpha\mu_i)^{-\alpha^{-1}}}, & y_i > 0 \end{cases} \quad (2.18)$$

di mana *i* = 1,2,3,.. ,n ; 0 < p_i <1 ; $\mu_i \geq 0$ dan α ≥ 0 merupakan parameter dispersi dan memenuhi persamaan :

$$\text{logit}(pi) = \log\left(\frac{\mathrm{p_i}}{1-\mathrm{p_i}}\right) = \mathbf{x_i^T}\boldsymbol{\delta} \tag{2.19}$$

dan

$$\log(\mu_i) = \mathbf{x_i^T}\boldsymbol{\beta} \tag{2.20}$$

di mana δ dan β adalah parameter regresi HNB yang akan diduga, sedangkan $\mathbf{x_i}$ adalah matrik variabel yang memuat himpunan-himpunan yang berbeda dari faktor eksperimen yang berhubungan dengan mean *Negative Binomial* pada *Negative Binomial state* dan peluang pada *zero state*.

2.6 Pendugaan Parameter Regresi *Hurdle Negative Binomial*

Pendugaan parameter regresi *Hurdle Negative Binomial* (HNB) dapat menggunakan metode *Maximum Likelihood Estimation* (MLE) dengan prosedur algoritma *Expectation Maximization* (EM). Algoritma EM pertama kali diperkenalkan oleh Dempster, Laird dan Rubin (1977), merupakan salah satu metode optimasi yang banyak digunakan sebagai alternatif dalam memaksimumkan fungsi *likelihood* yang mengandung data missing. Dua tahap dilakukan dalam algoritma EM, yaitu Ekspektasi dan Maksimalisasi.

Misal diasumsikan terdapat data observasi *x* berdistribusi tertentu yang mengandung data *missing y* yang diasumsikan mengikuti mekanisme *Missing completely at random*. Data *x* selanjutnya disebut *incomplete data set* dan data lengkap yang terangkum dalam $\hat{Z}(x, y)$, selanjutnya disebut *complete data set*. Fungsi likelihood yang umum digunakan adalah berdasarkan *x* yang lengkap, yaitu :

$$L(\hat{\theta}|x) = \prod_{i=1}^{n} f(x_i|\hat{\theta}) \tag{2.21}$$

Namun karena x adalah *incomplete* maka fungsi likelihood (2.21) tidak dapat dimaksimumkan, dan umum disebut *incomplete data likelihood*. Untuk mengatasinya terlebih dahulu dibentuk distribusi gabungan antara *x* dan *y*, dan fungsi likelihood (2.21) kembali disusun berdasarkan distribusi gabungan antara *x* dan *y*, yaitu :

$$\begin{aligned} L(\theta|x, y) &= \prod_{i=1}^{n} f(x_i, y_i|\theta) \\ &= \prod_{i=1}^{n} f(x_i|y_i|\theta) f(y_i|\theta) \end{aligned} \tag{2.22}$$

Tahap ekspektasi pada algoritma EM adalah menghitung nilai ekspektasi dari fungsi likelihood $L(\theta|x, y)$ sedangkan tahap maksimalisasi adalah memaksimalkan nilai ekspektasi dari *likelihood* yang dihasilkan pada tahap ekspektasi. Berikut ini adalah langkah dalam algoritma EM menurut Lestari (2008)

1. Menentukan inisialisasi parameter $\hat{\theta}_k\ k = 0$
2. Langkah Ekspektasi :
   Menghitung *complete data likelihood* dengan cara subtitusi $\hat{\theta}_k$ pada fungsi
   $$Q(\hat{\theta}_k) = E\left[L(\hat{\theta}_k|x,y)|x\right]$$
3. Langkah Maksimalisasi
   Dilakukan dengan mengacu pada kondisi berikut:
   $$\frac{\partial Q(\hat{\theta}_k)}{\partial \hat{\theta}_k} = 0$$
   Untuk mendapatkan nilai inisialisasi yang baru $\hat{\theta}_{k+1}$. Langkah Ekspektasi dan Maksimalisasi dilakukan secara iteratif sampai didapatkan perbedaan antara ($\hat{\theta}_{k+1}$-$\hat{\theta}_k$) lebih kecil dari 10^{-5}.

2.7 Pengujian Parameter Model Regresi HNB

2.7.1 Pengujian Parameter Model secara Simultan

Pengujian secara simultan dilakukan menggunakan *Likelihood Ratio Test* dengan statistik uji G. Pengujian ini dilakukan untuk mengetahui pengaruh variabel prediktor di dalam model secara bersama-sama. Hipotesis yang digunakan adalah :

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = \ \delta_1 = \delta_2 = \cdots = \delta_p = 0$

H_1 : paling sedikit ada satu $\beta_j \neq 0$ atau $\delta_j \neq 0$, dengan j=1,..p

Statistik uji *G* didefinisikan sebagai berikut :

$$G = -2\,[L_0 - L_p] \text{ dan } G \sim \chi^2_{\alpha,p} \qquad (2.23)$$

di mana :

L_0 : *lnlikelihood* model tanpa variabel prediktor (model intersep)

L_p : *lnlikelihood* model dengan p variabel prediktor (model penuh)

Hipotesis nol (H_0) akan ditolak jika statistik uji $G > \chi^2_{\alpha,p}$ atau nilai-*p* < α= 0.1 sebaliknya hipotesis nol (H_0) akan diterima jika statistik uji $G < \chi^2_{\alpha,p}$ atau nilai-*p* > α= 0.1

2.7.2 Pengujian Parameter Model secara Parsial

Pengujian secara parsial digunakan untuk mengetahui pengaruh parameter regresi dari masing-masing variabel prediktor pada model. Pengujian secara parsial dilakukan dengan menggunakan uji Wald.

a. Uji signifikansi parameter δ_j

$H_0 : \delta_j = 0$

$H_1 : \delta_j \neq 0$ dengan j = 1,2,..p

Statistik uji Wald adalah sebagai berikut :

$$W^2 = \left(\frac{\hat{\delta}_j}{SE\ (\hat{\delta}_j)}\right)^2 \quad (2.24)$$

di mana $SE\ (\hat{\delta}_j)$ adalah nilai *standard error* dari parameter δ_j. $SE\ (\hat{\delta}_j) = \sqrt{var(\hat{\delta}_j)}$, di mana $var(\hat{\delta}_j)$ merupakan elemen diagonal dari matriks varian kovarian $(\boldsymbol{H}(\boldsymbol{\theta}))^{-1}$ dengan $\boldsymbol{H}(\boldsymbol{\theta}) = \boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}$, di mana **X** adalah matrik berukuran nx(p+1)

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}$$

b. Uji signifikansi parameter β_j

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$ dengan j = 1,2,..p

Statistik uji Wald adalah sebagai berikut :

$$W^2 = \left(\frac{\hat{\beta}_j}{SE\ (\hat{\beta}_j)}\right)^2 \quad (2.25)$$

$SE\ (\hat{\beta}_j)$ adalah nilai *standard error* dari parameter β_j.

$SE\ (\hat{\beta}_j) = \sqrt{var(\hat{\beta}_j)}$, di mana $var(\hat{\delta}_j)$ merupakan elemen diagonal dari matriks varian kovarian $(\boldsymbol{H}(\boldsymbol{\theta}))^{-1}$ dengan $\boldsymbol{H}(\boldsymbol{\theta}) = \boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}$, di mana **X** adalah matrik berukuran nx(p+1)

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}$$

Hipotesis nol (H_0) akan ditolak jika statistik uji $W^2 > \chi^2_{1,\alpha}$ atau nilai-*p* < α= 0.1 sebalikanya Hipotesis nol (H_0) akan diterima jika statistik uji $W^2 < \chi^2_{1,\alpha}$ atau nilai-*p* > α= 0.1 (Agresti, 2007).

2.8 Landasan Kebijakan dan Fungsi Ujian Nasional (UN)

Hasil belajar siswa adalah tingkat penguasaan pengetahuan yang dicapai oleh siswa dalam mengikuti program pembelajaran sesuai dengan tujuan pendidikan yang diterapkan. Faktor-faktor yang mempengaruhi hasil belajar siswa dapat digolongkan menjadi dua yaitu faktor internal dan faktor eksternal. Faktor internal terkait dengan disiplin, respon dan motivasi siswa sedangkan faktor eksternal adalah lingkungan belajar, tujuan pembelajaran, kreatifitas media belajar oleh pendidik serta metode pembelajaran. Faktor tersebut mempengaruhi satu sama lain dan merupakan satu kesatuan yang mendasari hasil belajar siswa (Maisaroh dan Rostrieningsih, 2010).

Penilaian hasil belajar oleh pemerintah bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu dalam kelompok mata pelajaran ilmu pengetahuan teknologi dan dilakukan dalam bentuk Ujian Nasional (UN). Pada PP No 19 tahun 2005 tentang SNP pasal 68 menyatakan bahwa hasil Ujian Nasional digunakan sebagai :

1. Pemetaan mutu program dan satuan pendidikan
2. Dasar seleksi masuk jenjang pendidikan berikutnya
3. Penentuan kelulusan peserta didik dari program dan satuan pendidikan
4. Pembinaan dan pemberian bantuan kepada satuan pendidikan dalam upayanya untuk meningkatkan mutu pendidikan

Sementara itu mengenai kelulusan peserta didik mengacu pada PP No. 19 tahun 2005 pasal 72 ayat (1) yang menyatakan peserta didik dinyatakan lulus dari satuan pendidikan pada pendidikan dasar dan menengah setelah:

a. Menyelesaikan seluruh program pembelajaran
b. Memperoleh nilai minimal baik pada penilaian akhir untuk seluruh mata pelajaran yang terdiri atas kelompok mata pelajaran agama dan akhlak mulia, kelompok mata pelajaran kewarganegaraan dan kepribadian, kelompok mata pelajaran estetika, dan kelompok mata pelajaran jasmani, olah raga, dan kesehatan
c. Lulus ujian sekolah untuk kelompok mata pelajaran ilmu pengetahuan dan teknologi
d. Lulus Ujian Nasional.

BAB III
METODE PENELITIAN

3.1 Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder yaitu data banyaknya siswa SMA yang gagal Ujian Nasional Tahun Pelajaran 2013/2014 di Kota Malang yang disajikan pada Lampiran 1. Data didapatkan dari Dinas Pendidikan Kota Malang Bagian SMP,SMA dan SMK, Sub. Bagian Penyusunan Program dan Bidang Fungsional Pendidikan.

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah :

1. Variabel respon Y yaitu banyaknya siswa yang gagal Ujian Nasional (UN) di setiap SMA di Kota Malang pada tahun pelajaran 2013/2014
2. Variabel-variabel yang mempengaruhi banyaknya siswa yang gagal Ujian Nasional (UN) adalah :

   X1: Status Akreditasi Sekolah

   Status sekolah yang telah diakreditasi oleh Badan Akreditasi Nasional Sekolah/Madrasah (BAN-S/M), yang mempunyai peringkat A, B,C dan belum atau tidak terakreditasi. Kategori A ditulis 1, Kategori B ditulis 2, Kategori C ditulis 3, dan sekolah yang belum terakreditasi ditulis dengan kode 4.

   X2: Rasio rencana penerimaan-pendaftaran

   Rasio ini diperoleh dengan menghitung perbandingan antara banyaknya siswa yang akan diterima dengan banyaknya pendaftar pada sekolah yang bersangkutan. Angka yang diperoleh merupakan indikator seleksi yang terjadi di sekolah yang bersangkutan. Rumus yang digunakan adalah: $\frac{\text{banyaknya siswa yang akan diterima}}{\text{banyaknya siswa pendaftar pada PSB}}$

   X3: Rasio Guru-Murid

   Rasio ini diperoleh dengan menghitung perbandingan antara banyaknya murid pada suatu sekolah dengan banyaknya guru yang ada pada sekolah bersangkutan. Indikator ini digunakan untuk menggambarkan beban guru dalam mengajar. Indikator ini juga digunakan untuk melihat mutu pengajaran di kelas karena semakin tinggi nilai rasio ini berarti semakin berkurang tingkat pengawasan atau perhatian guru terhadap murid sehingga mutu pengajaran cenderung semakin rendah. Rumus yang digunakan adalah $\frac{\text{banyaknya Guru}}{\text{banyaknya Murid}}$. Standar ideal rasio guru dan murid adalah 1:40.

X4: Rasio murid-kelas

Rasio ini diperoleh dengan menghitung perbandingan antara banyaknya murid pada suatu sekolah dengan banyak kelas yang ada pada sekolah yang bersangkutan. Angka yang ada merupakan indikator kepadatan kelas yang bersangkutan. Rumus yang digunakan adalah $\frac{\text{banyaknya murid}}{\text{banyaknya kelas}}$. Standar ideal rasio murid dan kelas adalah 36:1

(Cahyosetiono, 2009).

3.3 Metode Analisis Data

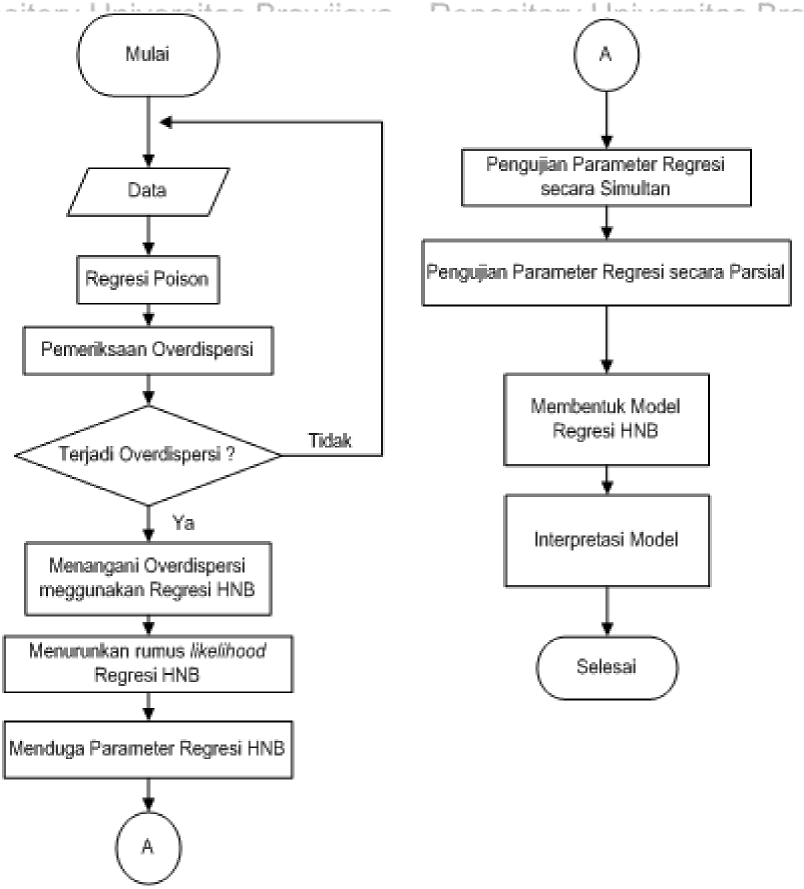
Tahapan-tahapan yang dilakukan dalam penelitian ini untuk mencapai tujuan penelitian adalah sebagai berikut :

1. Pemeriksaan terjadinya overdispersi pada variabel respon *Y* menggunakan statistik *Pearson chi- square* yang dibagi dengan derajat bebasnya (*db*= *n-p*) sesuai dengan persamaan (2.15). Jika terjadi overdispersi maka $\phi > 1$ sedangkan jika terjadi underdispersi maka $\phi < 1$. Jika terjadi overdispersi atau underdispersi maka dilakukan pembentukan model regresi *Hurdle Negative Binomial* (HNB).
2. Menurunkan fungsi *likelihood* dari regresi *Hurdle Negative Binomial* (HNB).
3. Menduga parameter model *Hurdle Negative Binomial* (HNB) menggunakan *maximum likelihood* dengan algoritma *Expectation Maximization* (EM).
4. Pengujian signifikansi parameter regresi

1. Pengujian secara simultan dengan menggunakan statistik uji *G* dengan rumus sesuai dengan persamaan (2.23).
2. Pengujian secara parsial dengan menggunakan statistik uji Wald sesuai dengan persamaan (2.24) dan (2.25).

5. Membentuk Model *Hurdle Negative Binomial* (HNB)
6. Interpretasi model

Diagram alir dari metode penelitian disajikan pada Gambar 3.1. Analisis data dilakukan dengan bantuan *software* statistika yaitu SAS 9.3



Gambar 3.1 Diagram Alir Metode Analisis

BAB IV
HASIL DAN PEMBAHASAN

4.1 Pemodelan Regresi *Poisson*

Regresi Poisson merupakan salah satu metode regresi yang digunakan untuk menganalisis hubungan antara variabel dependen Y yang bersifat diskrit (*count* data) dengan variabel independen X. Pendugaan Parameter model regresi *Poisson* dilakukan dengan menggunakan metode *maximum likelihood* dengan iterasi Newton Raphson. Proses pendugaan parameter regresi *Poisson* untuk data SMA di Kota Malang tahun pelajaran 2013/2014 dilakukan dengan bantuan software SAS 9.3 dan diperoleh hasil berdasarkan Lampiran 2 tersaji pada Tabel 4.1 di bawah ini

Tabel 4.1 Nilai Duga Parameter Regresi *Poisson*

Parameter	Penduga Parameter	Salah Baku	Wald Chi-Square	Nilai-*P*
Intersep	4.1528	2.1639	3.68	0.055
X1	-1.1391	0.7191	2.51	0.1132
X2	-0.5226	0.4777	1.2	0.2739
X3	1.8444	9.9036	0.03	0.8523
X4	-0.1664	0.0639	6.78	0.0092

Berdasarkan Tabel 4.1 di atas, diperoleh model regresi Poisson adalah sebagai berikut :

$y = \exp(4.1528 - 1.1391\ X_1 - 0.5226\ X_2 + 1.8444\ X_3 - 0.1664\ X_4)$

4.2 Pemeriksaan Overdispersi pada Regresi *Poisson*

Pemeriksaan overdispersi dilakukan dengan menggunakan nilai statistik *Pearson chi-square* yang dibagi dengan derajat bebasnya. Jika hasilnya sama dengan satu (ϕ=1) maka asumsi equidispersi terpenuh. Jika $\phi > 1$ maka terjadi overdispersi sedangkan jika $\phi < 1$ maka terjadi underdispersi.

Hasil pengujian overdispersi berdasarkan Lampiran 2 dapat diketahui bahwa nilai *Pearson chi-square* sebesar 40.0786 dan derajat bebas sebesar 36 sehingga nilai ϕ adalah 1.0832. Nilai ϕ

tersebut lebih besar dari 1 sehingga dapat disimpulkan bahwa terjadi overdispersi dan regresi *Poisson* tidak tepat digunakan untuk memodelkan kasus banyaknya siswa SMA yang tidak lulus Ujian Nasional.

4.3 Penurunan Fungsi *Likelihood* Regresi *Hurdle Negative Binomial* (HNB)

Untuk menduga parameter regresi HNB, digunakan metode *Maximum Likelihood Estimation* (MLE). Metode ini biasanya digunakan untuk menduga parameter suatu model yang diketahui fungsi kepekatan peluangnya. Terlebih dahulu akan diturunkan *likelihood* dari fungsi HNB. Dengan mensubtitusi *link function* dari model regresi HNB ke fungsi kepekatan peluangnya.

Dari persamaan (2.19) dan (2.20) diperoleh :

$$\log\left(\frac{\mathrm{p_i}}{1-\mathrm{p_i}}\right) = \mathbf{x_i^T\delta} \text{ yang ekuivalen dengan } \left(\frac{\mathrm{p_i}}{1-\mathrm{p_i}}\right) = e^{\boldsymbol{x_i^T\delta}}$$

$$\mathrm{p_i} = e^{\boldsymbol{x_i^T\delta}} - \mathrm{p_i}e^{\boldsymbol{x_i^T\delta}}$$

$$\mathrm{p_i}\,(1 + e^{\boldsymbol{x_i^T\delta}}) = e^{\boldsymbol{x_i^T\delta}}$$

Sehingga $\mathrm{p_i} = \dfrac{e^{\boldsymbol{x_i^T\delta}}}{1+e^{\boldsymbol{x_i^T\delta}}}$ (4.1)

dan $(1-\mathrm{p_i}) = \dfrac{\mathrm{p_i}}{e^{\boldsymbol{x_i^T\delta}}}$

$$= \frac{e^{\boldsymbol{x_i^T\delta}}}{1+e^{\boldsymbol{x_i^T\delta}}}\,\frac{1}{e^{\boldsymbol{x_i^T\delta}}}$$

Sehingga $(1-\mathrm{p_i}) = \dfrac{1}{(1+e^{\boldsymbol{x_i^T\delta}})}$ (4.2)

Dan $\log(\mu_i) = \mathbf{x_i^T\beta}$, maka $\mu_i = \mathrm{e}^{\boldsymbol{x_i^T\beta}}$ (4.3)

Dengan mensubtitusi persamaan (4.1), (4.2) dan (4.3) ke persamaan (2.18) maka diperoleh

$$P\,(Y=y_i)\begin{cases} \dfrac{e^{x_i^T\delta}}{1+e^{x_i^T\delta}} & y_i = 0 \\ \dfrac{1}{(1+e^{x_i^T\delta})}\dfrac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})}\dfrac{(1+\alpha\mathrm{e}^{\boldsymbol{x_i^T\beta}})^{-\alpha^{-1}-y_i}\alpha^{y_i}(\mathrm{e}^{\boldsymbol{x_i^T\beta}})^{y_i}}{1-(1+\alpha\mathrm{e}^{\boldsymbol{x_i^T\beta}})^{-\alpha^{-1}}}, & y_i > 0 \end{cases}$$

Sehingga fungsi *likelihood* bagi HNB adalah :

$$L(\delta,\beta) = \begin{cases} \prod_{i=1}^{n} \dfrac{e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}} & y_i = 0 \\ \prod_{i=1}^{n} \dfrac{1}{(1+e^{\boldsymbol{x}_i^T\boldsymbol{\delta}})} \dfrac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})} \dfrac{(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{-\alpha^{-1}-y_i}\alpha^{y_i}(e^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{y_i}}{1-(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{-\alpha^{-1}}}, & y_i > 0 \end{cases}$$

dengan fungsi ln*likelihood* sebagai berikut :

$$\ln L(\delta,\beta) = \begin{cases} \sum_{i=1}^{n} \ln e^{\boldsymbol{x}_i^T\boldsymbol{\delta}} - \sum_{i=1}^{n} \ln\left(1+e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right) & y_i = 0 \\ -\sum_{i=1}^{n} \ln\left(1+e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right) + \sum_{i=1}^{n} \ln\left(\Gamma\left(y_i + {}^{1}/_{\alpha}\right)\right) - \sum_{i=1}^{n} \ln(\Gamma(y_i+1)) & \\ - \sum_{i=1}^{n} \ln\left(\Gamma\left({}^{1}/_{\alpha}\right)\right) - \left({}^{1}/_{\alpha} + y_i\right) \sum_{i=1}^{n} \ln\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right) & \\ + y_i \sum_{i=1}^{n} \left(ln\,\alpha + ln\, e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right) - \sum_{i=1}^{n} \ln\left(1 - \left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-1/\alpha}\right) & , y_i > 0 \end{cases} \quad (4.4)$$

4.4 Pendugaan Parameter Regresi *Hurdle Negative Binomial* (HNB)

Fungsi ln*likelihood* dari HNB pada persamaan (4.4) akan menyulitkan perhitungan pendugaan parameter karena tidak diketahui nilai nol mana yang berasal dari *zero state* dan mana yang berasal dari *negative binomial state*, sehingga fungsi likelihood ini tidak dapat diselesaikan dengan metode numerik biasa.

Misalkan variabel y_i berkaitan dengan vektor variabel indikator **W** = $(w_1,\ldots,w_n)^T$ yaitu :

$$w_i = \begin{cases} 1 & \text{jika } y_i \text{ berasal dari } \textit{zero state} \quad (y_i=0) \\ 0 & \text{jika } y_i \text{ berasal dari } \textit{negative binomial state} \quad (y_i=0,1,2..) \end{cases} \quad (4.5)$$

Berdasarkan persamaan (2.17) dan persamaan (4.5), maka

$P(w_i=1) = p_i$

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

Sehingga distribusi dari variabel W adalah $w_i \sim$ Binomial $(1,p_i)$.

Jika nilai variabel respon y_i=1,2,.. maka w_i= 0. Namun jika variabel respon y_i = 0 maka nilai w_i dapat bernilai 0 atau 1. Oleh karena itu w_i dianggap hilang.

Untuk mengatasi hal ini maka pendugaan parameter akan dilakukan dengan algoritma *Expectation Maximization* (EM). Langkah-langkahnya adalah

1. Membentuk distribusi gabungan antara w_i dan y_i yaitu

$$f(y,w|p,\mu) = f(w_i)\, f(y_i|w_i)$$
$$= f(w_i|1,\mathrm{p_i})\, f(y_i|w_i,\mu_i)$$

Dengan $f(w_i|1,\mathrm{p_i}) = (p_{i,})^{w_i}(1-p_{i,})^{(1-w_i)}$, $\quad w_i = 0{,}1$

Dan $f(y_i|w_i,\mu_i) = \left[\frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})}\frac{(1+\alpha\mu_i)^{-\alpha^{-1}-y_i}\alpha^{y_i}\mu_i^{y_i}}{1-(1+\alpha\mu_i)^{-\alpha^{-1}}}\right]^{(1-w_i)}$

Sehingga diperoleh

$$f(y,w|p,\mu) = (p_{i,})^{w_i}(1-p_{i,})^{(1-w_i)}\left[\frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})}\frac{(1+\alpha\mu_i)^{-\alpha^{-1}-y_i}\alpha^{y_i}\mu_i^{y_i}}{1-(1+\alpha\mu_i)^{-\alpha^{-1}}}\right]^{(1-w_i)} \tag{4.6}$$

Subtitusikan persamaan (4.1), (4.2) dan (4.3) ke persamaan (4.6) dan diperoleh :

$f(y,w|p,\mu)$

$$= \left(\frac{e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}}\right)^{w_i}\left(\frac{1}{(1+e^{\boldsymbol{x}_i^T\boldsymbol{\delta}})}\right)^{(1-w_i)}\left[\frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})}\frac{(1+\alpha\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{-\alpha^{-1}-y_i}\alpha^{y_i}(\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{y_i}}{1-(1+\alpha\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{-\alpha^{-1}}}\right]^{(1-w_i)}$$

$$= \left(\frac{e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}}\right)^{w_i}\left(1+e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)^{w_i}\left(\frac{1}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}}\right)\left[\frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})}\frac{\left(1+\alpha\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\alpha^{-1}-y_i}\alpha^{y_i}(\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{y_i}}{1-\left(1+\alpha\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\alpha^{-1}}}\right]^{(1-w_i)}$$

$$= \left(e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)^{w_i}\left(\frac{1}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}}\right)\left[\frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})}\frac{(1+\alpha\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{-\alpha^{-1}-y_i}\alpha^{y_i}(\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{y_i}}{1-(1+\alpha\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{-\alpha^{-1}}}\right]^{(1-w_i)}$$

Sehingga fungsi likelihoodnya menjadi :

$L(\delta,\beta|y,w)$

$$= \prod_{i=1}^{n}\left[\left(e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)^{w_i}\left(\frac{1}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}}\right)\left[\frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})}\frac{(1+\alpha\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{-\alpha^{-1}-y_i}\alpha^{y_i}(\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{y_i}}{1-(1+\alpha\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{-\alpha^{-1}}}\right]^{(1-w_i)}\right]$$

Dan fungsi ln*likelihood*nya adalah :

$$\ln L(\delta,\beta|y,w) = \sum_{i=1}^{n} w_i\,\boldsymbol{x}_i^T\boldsymbol{\delta} - \sum_{i=1}^{n}\ln\left(1+\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right) + \sum_{i=1}^{n}(1-wi\ln g(yi;\beta;\alpha) \tag{4.7}$$

di mana

$$\ln g(y_i;\beta;\alpha) = \ln\left[\frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})}\frac{\left(1+\alpha\,\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\alpha^{-1}-y_i}\alpha^{y_i}(\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{y_i}}{1-\left(1+\alpha\,\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\alpha^{-1}}}\right]$$

$$= \sum_{i=1}^{n}\ln\left(\Gamma\left(y_i+{}^{1}\!/_{\alpha}\right)\right) - \sum_{i=1}^{n}\ln(\Gamma(y_i+1)) -$$

*i=1n*ln (Γ*(1α))− 1α+ yii=1n*ln *1+αe**xiTβ**+yii=1nln α+ln e**xiTβ**−i=1n*ln *1−1+αe**xiTβ**−1α*

vektor parameter $\boldsymbol{\delta}$ dan $\boldsymbol{\beta}$ dapat di duga secara terpisah, dengan menuliskan persamaan (4.7) menjadi

$\ln \mathrm{L}\ (\delta,\beta,y,w) = \ln L\ (\delta,y,w) + \ln L\ (\beta,y,w)$

dengan

$$\ln\ L\ (\delta,y,w) = \sum_{i=1}^{n} w_i\ \boldsymbol{x}_i^T\boldsymbol{\delta} - \sum_{i=1}^{n}\ln\left(1+\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right) \qquad (4.8)$$

dan

$$\ln L\ (\beta,y,w) = \sum_{i=1}^{n}(1-w_i)\left[-\left({}^{1}\!/_{\alpha}+y_i\right)\sum_{i=1}^{n}\ln\left(1+\right.\right.$$

***exiTβ**+yii=1nln α+ln **exiTβ** −i=1n*ln*1−1+αe**xiTβ**−1α*

(4.9)

Sedangkan $\sum_{i=1}^{n}\ln\left(\Gamma\left(y_i+{}^{1}\!/_{\alpha}\right)\right)$; $\sum_{i=1}^{n}\ln\left(\Gamma(y_i+1)\right)$ dan $\sum_{i=1}^{n}\ln\ (\Gamma(\ {}^{1}\!/_{\alpha}))$dapat di abaikan karena tidak mengandung vektor parameter $\boldsymbol{\delta}$ dan $\boldsymbol{\beta}$

2. Selanjutnya dilakukan tahap Ekspektasi dan Maksimalisasi :
   a. Tahap Ekspektasi (*E-Step*)
      Mengganti variabel w_i dengan $w_i{}^{(m)}$ yang merupakan ekspetasi dari w_i sehingga persamaan (4.7) dan persamaan (4.8) menjadi :

$$E\left[lnL\left(\boldsymbol{\delta}|\boldsymbol{y},\boldsymbol{w}^{(m)}\right)\right] = \sum_{i=1}^{n}\left[w_i{}^{(m)}\boldsymbol{x}_i^T\boldsymbol{\delta} - ln\left(1+\exp(\boldsymbol{x}_i^T\boldsymbol{\delta})\right)\right] \quad (4.10)$$

$$E\left[lnL(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{w}^{(m)})\right] = \sum_{i=1}^{n}\left(1 - w_i^{(m)}\right)\left[\left(1/_\alpha + y_i\right)\sum_{i=1}^{n}\ln\left(1 + \right.\right.$$

exiTβ+yii=1nln *α+lnexiTβ*

*−i=1n*ln*1−1+αexiTβ−1α* (4.11)

b. Tahap Maksimalisasi (*M-Step*)

(i) Maksimalisasi untuk $\boldsymbol{\beta}$

Maksimalisasi untuk $\boldsymbol{\beta}$ dengan menghitung $\boldsymbol{\beta}^{(\boldsymbol{m+1})}$ yang diperoleh dari memaksimumkan (4.11) dengan metode Newton Raphson.

Langkah-langkah dalam Newton Raphson :

1) Menentukan nilai taksiran awal parameter ($\widehat{\boldsymbol{\beta}}^{(\boldsymbol{0})}$)

2) Menentukan vektor gradien **g**

$$\mathbf{g}^T = \left[\frac{lnL(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\beta_0}\frac{lnL(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\beta_1}\cdots\frac{lnL(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\beta_p}\right]$$

Dengan

$$\frac{lnL(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}(1 - w_i^{(m)})\left[-\left(\frac{1}{\alpha} + y_i\right)\sum_{i}^{n}\frac{\alpha\boldsymbol{x}_i^T e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}{1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}} + \right.$$

yii=1nxiT−i=1n1α 1+αexiTβ−(α+1)α

αxiTexiTβ1−1+αexiTβ−1α

proses penurunan $\frac{lnL(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\beta}}$ tersaji pada Lampiran 3

3) Membentuk Matriks Hessian H :

$$\boldsymbol{H}_{(\boldsymbol{p}+1)\boldsymbol{x}(\boldsymbol{p}+1)} = \begin{bmatrix} \frac{\partial^2\mathrm{lnL}(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\beta_0}^2} & \frac{\partial^2\mathrm{lnL}(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\beta_0}\boldsymbol{\beta_1}} & \cdots & \frac{\partial^2\mathrm{lnL}(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\beta_0}\boldsymbol{\beta_p}} \\ \frac{\partial^2\mathrm{lnL}(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\beta_1}\boldsymbol{\beta_0}} & \frac{\partial^2\mathrm{lnL}(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\beta_1^2} & \cdots & \frac{\partial^2\mathrm{lnL}(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\beta_1}\boldsymbol{\beta_p}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2\mathrm{lnL}(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\beta_1}\boldsymbol{\beta_p}} & \frac{\partial^2\mathrm{lnL}(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\beta_2}\boldsymbol{\beta_p}} & \cdots & \frac{\partial^2\mathrm{lnL}(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\beta_p}^2} \end{bmatrix}$$

Dengan

$$\frac{\partial^2 \ln L\left(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)}\right)}{\partial \boldsymbol{\beta}^2} = \sum_{i=1}^{n}\left(1-\mathrm{w_i}^{(\mathrm{m})}\right)$$

$$\left[\frac{\left(\begin{array}{c}\left(\frac{1}{\alpha}+y_i\right)\sum_{i=1}^{n}\left(\frac{\alpha(\boldsymbol{x}_i^T)^2 e^{2\boldsymbol{x}_i^T\boldsymbol{\beta}}}{\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^2}\right)-\sum_i^n\left(\alpha(\boldsymbol{x}_i^T)^2 e^{2\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)-(\alpha+1) \\ (1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}})^{-\frac{(2\alpha+1)}{\alpha}}+\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\frac{(\alpha+1)}{\alpha}}-\alpha\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\frac{2(\alpha+1)}{\alpha}}\end{array}\right)}{\alpha^2\left(1-\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\frac{1}{\alpha}}\right)^2}\right]$$

Proses penurunan $\frac{\partial^2 \ln L\left(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)}\right)}{\partial \boldsymbol{\beta}^2}$ tersaji pada Lampiran 3

4) Mensubtitusikan nilai $\widehat{\boldsymbol{\beta}}^{(\mathbf{0})}$ ke dalam elemen-elemen vektor **g** dan matriks **H**, sehingga diperoleh vektor $\mathbf{g}(\widehat{\boldsymbol{\beta}}^{(\mathbf{0})})$ dan matriks $\mathbf{H}(\widehat{\boldsymbol{\beta}}^{(\mathbf{0})})$.
5) Mulai dari m=0 dilakukan iterasi sesuai persamaan :

$$\widehat{\boldsymbol{\beta}}^{(\boldsymbol{m}+\mathbf{1})} = \widehat{\boldsymbol{\beta}}^{(\boldsymbol{m})} - \left(\boldsymbol{H}^{(\boldsymbol{m})}\right)^{-\mathbf{1}}\mathbf{g}^{(\boldsymbol{m})}$$

Nilai $\widehat{\boldsymbol{\beta}}^{(\boldsymbol{m})}$ merupakan sekumpulan penduga parameter yang konvergen pada iterasi ke-*m*.
6) Jika belum didapatkan penduga parameter yang konvergen $\left|\widehat{\boldsymbol{\beta}}^{(\boldsymbol{m}+\mathbf{1})} - \widehat{\boldsymbol{\beta}}^{(\boldsymbol{m})}\right| \leq \boldsymbol{\varepsilon}$ maka dilanjutkan kembali ke langkah 5 sampai mendapatkan parameter yang konvergen.

(ii) Maksimalisasi untuk δ

Langkah-langkah memaksimalkan δ identik dengan tahap maksimalisasi untuk β di atas. Maksimalisasi δ dilakukan dengan menghitung $\delta^{(m+1)}$ yang diperoleh dari memaksimumkan persamaan (4.10) dengan metode Newton Raphson. Langkah-langkahnya adalah sebagai berikut :

1) Menentukan nilai taksiran awal parameter ($\widehat{\boldsymbol{\delta}}^{(\mathbf{0})}$)
2) Menentukan vektor gradien **g**

$$\mathbf{g}^T = \left[\frac{lnL\left(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)}\right)}{\partial\delta_0}\ \frac{lnL\left(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)}\right)}{\partial\delta_1} \cdots \frac{lnL\left(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)}\right)}{\partial\delta_p}\right]$$

dengan $\frac{lnL\left(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)}\right)}{\partial\boldsymbol{\delta}} = \sum_{i=1}^{n}\left[w_i^{(m)}x_i^T - \frac{\boldsymbol{x}_i^T e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}}\right]$

proses penurunan $\frac{lnL\,(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\delta}}$ tersaji pada Lampiran 3

3) Membentuk Matriks Hessian H :

$$\boldsymbol{H}_{(\boldsymbol{p}+1)\boldsymbol{x}(\boldsymbol{p}+1)} = \begin{bmatrix} \frac{\partial^2 \ln L(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\delta_0}^2} & \frac{\partial^2 \ln L(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\delta_0}\boldsymbol{\delta_1}} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\delta_0}\boldsymbol{\delta_p}} \\ \frac{\partial^2 \ln L(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\delta_1}\boldsymbol{\delta_0}} & \frac{\partial^2 \ln L(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\delta_1}^2} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\delta_1}\boldsymbol{\delta_p}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 \ln L(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\delta_1}\boldsymbol{\delta_p}} & \frac{\partial^2 \ln L(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\delta_2}\boldsymbol{\delta_p}} & \ddots \cdots & \frac{\partial^2 \ln L(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\boldsymbol{\delta_p}^2} \end{bmatrix}$$

Dengan $\frac{\partial^2 \ln L\,(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)})}{\partial\,\boldsymbol{\delta}^2} = -\sum_{i=1}^{n}\left[\frac{(\boldsymbol{x}_i^T)^2 e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}}{\left(1+e^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)^2}\right]$

4) Mensubtitusikan nilai $\hat{\boldsymbol{\delta}}^{(\mathbf{0})}$ ke dalam elemen-elemen vektor **g** dan matriks **H**, sehingga diperoleh vektor $\mathbf{g}(\hat{\boldsymbol{\delta}}^{(\mathbf{0})})$ dan matriks $\mathbf{H}(\hat{\boldsymbol{\delta}}^{(\mathbf{0})})$.

5) Mulai dari m=0 dilakukan iterasi sesuai persamaan :

$$\hat{\boldsymbol{\delta}}^{(\boldsymbol{m}+\mathbf{1})} = \hat{\boldsymbol{\delta}}^{(\boldsymbol{m})} - \left(\boldsymbol{H}^{(\boldsymbol{m})}\right)^{-\mathbf{1}}\mathbf{g}^{(\boldsymbol{m})}$$

Nilai $\hat{\boldsymbol{\delta}}^{(\boldsymbol{m})}$ merupakan sekumpulan penduga parameter yang konvergen pada iterasi ke-*m*.

6) Jika belum didapatkan penduga parameter yang konvergen $\left|\hat{\boldsymbol{\delta}}^{(\boldsymbol{m}+\mathbf{1})} - \hat{\boldsymbol{\delta}}^{(\boldsymbol{m})}\right| \leq \varepsilon$ maka dilanjutkan kembali ke langkah 5 sampai mendapatkan parameter yang konvergen.

3. Selanjutnya mengganti $\boldsymbol{\beta}^{(m)}$ dan $\boldsymbol{\delta}^{(m)}$ dengan $\boldsymbol{\beta}^{(m+1)}$ dan $\boldsymbol{\delta}^{(m+1)}$ pada iterasi selanjutnya, kemudian kembali melakukan tahap Ekspektasi.
4. Tahap *E-step* dan *M-step* dilakukan berulang-ulang sampai diperoleh penduga parameter yang konvergen. Dikatakan konvergen jika $\left|\boldsymbol{\delta}^{(m)} - \boldsymbol{\delta}^{(m+1)}\right| \leq \varepsilon$ dan $\left|\boldsymbol{\beta}^{(m)} - \boldsymbol{\beta}^{(m+1)}\right| \leq \varepsilon$ di mana $\varepsilon = 10^{-5}$.

4.5 Pemodelan Regresi *Hurdle Negative Binomial* (HNB)

Sebelum melakukan pendugaan parameter pada model regresi *Hurdle Negative Binomial* (HNB) dilakukan penurunan rumus *likelihood* regresi HNB menggunakan metode *Maximum Likelihood Estimation* (MLE) sesuai pada sub bab 4.3. Proses pendugaan dilakukan dengan bantuan software SAS 9.3 sesuai *code* pada Lampiran 4 dan diperoleh model penuh regresi *Hurdle Negative Binomial* (HNB) sesuai pada Lampiran 5 sebagai berikut:

Tabel 4.2 Nilai Pendugaan Parameter Model Regresi *Hurdle Negative Binomial* (HNB)

Parameter	Penduga Parameter	Salah Baku	Statistik uji-t	Nilai-*P*
δ_0	-4.4108	3.3746	-1.31	0.1983
δ_1	1.2021	1.0054	1.2	0.2385
δ_2	0.5388	0.6118	0.88	0.3835
δ_3	-1.3833	13.9846	-0.1	0.9217
δ_4	0.1741	0.09222	1.89	0.066
β_0	-0.7447	4.908	-0.15	0.8801
β_1	-1.3547	4.7388	-0.29	0.7764
β_2	11.6739	5.2596	2.22	0.0319
β_3	0.6174	160.28	0.00	0.9969
β_4	-1.6898	0.7853	-2.15	0.0372

Dari hasil pendugaan parameter model regresi *Hurdle Negative Binomial* (HNB) pada Tabel 4.2 diperoleh model regresi *Hurdle Negative Binomial* (HNB) adalah :

$$\log(\mu_i) = -\ 0.7447 - 1.3547\ X_1 + 11.6739X_2 + 0.6174\ X_3 - 1.6898\ X_4$$

$$\text{logit}(pi) = -\ 4.4108 + 1.2021\ X_1 + 0.5388\ X_2 - 1.3833X_3 + 0.1741X_4$$

Model di atas menjelaskan bahwa rata-rata SMA yang siswanya gagal ujian nasional (UN) dan peluang SMA yang semua siswanya lulus dipengaruhi oleh keempat variabel prediktor yaitu akreditasi sekolah, rasio rencana penerimaan-pendaftar, rasio guru-murid, dan rasio murid-kelas. Namun pengaruh signifikansi keempat variabel prediktor tersebut belum diketahui. Untuk mengetahui sejauh mana keempat variabel prediktor tersebut berpengaruh terhadap model maka dilakukan uji signifikansi parameter regresi dari model HNB tersebut.

4.6 Pengujian Signifikansi Parameter Model Regresi HNB

4.6.1 Uji Simultan

Pengujian signifikansi parameter secara simultan didasarkan pada *Likelihood Rasio Test* dengan statistik uji G. Statistik uji G ini juga dapat digunakan untuk menguji kelayakan atau kesesuaian model. Hipotesis yang digunakan pada uji simultan ini adalah

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = \delta_1 = \delta_2 = \cdots = \delta_p = 0$

H_1 : paling sedikit ada satu $\beta_j \neq 0$ atau $\delta_j \neq 0$, dengan j=1,..p

Berdasarkan hasil perhitungan dengan bantuan SAS pada Lampiran 5 diperoleh nilai G sebesar 34.1 dengan $\chi^2_{0.1,4} = 7.7794$. Nilai satatistik uji G > $\chi^2_{0.1,4}$ maka keputusannya H_0 ditolak sehingga dapat dikatakan bahwa minimal terdapat satu variabel prediktor yang memberikan pengaruh yang berbeda terhadap variabel respon dan hal ini juga berarti bahwa model telah sesuai pada taraf nyata 90%.

4.6.2 Uji Parsial

Pengujian signifikansi parameter secara parsial digunakan untuk mengetahui pengaruh masing-masing variabel prediktor terhadap variabel respon. Pengujian secara parsial berdasarkan statistik uji Wald sesuai output pada Lampiran 5 dapat dilihat pada Tabel 4.3 berikut:

Tabel 4.3 Hasil Uji secara Parsial Koefisien Regresi HNB

Parameter	Penduga Parameter	Statistik Wald	Nilai-*p*
δ_0	-4.4108	1.7084	0.1983
δ_1	1.2021	1.4296	0.2385
δ_2	0.5388	0.7756	0.3835
δ_3	-1.3833	0.0098	0.9217
δ_4	0.1741	3.5641	0.066
β_0	-0.7447	0.0230	0.8801
β_1	-1.3547	0.0817	0.7764
β_2	11.6739	4.9264	0.0319
β_3	0.6174	0.0000	0.9969
β_4	-1.6898	4.6302	0.0372

Pengujian parameter secara parsial dengan hipotesis :

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

Dari hasil uji secara parsial koefisien regresi HNB pada Tabel 4.4 menunjukkan bahwa nilai statistik uji Wald parameter β_1 dan β_3 berturut-turut 0.0817 dan 0.000 dengan nilai-p berturut-turut 0.7764 dan 0.9969 yang keduanya lebih besar dari α=0.1. Hal ini berarti bahwa parameter β_1 dan β_3 tidak berpengaruh secara signifikan dalam model. Namun untuk parameter β_2 dan β_4 nilai statistik uji Wald berturut-turut 4.9264 dan 4.6302 serta nilai-p berturut-turut 0.0319 dan 0.0372 yang keduanya lebih kecil dari α=0.1. Hal ini berarti bahwa parameter β_2 dan β_4 berpengaruh secara signifikan dalam model.

Selain itu, pengujian parameter secara parsial dengan hipotesis :

$H_0 : \delta_j = 0$

$H_1 : \delta_j \neq 0$

menghasilkan nilai statistik uji Wald parameter δ_1, δ_2, dan δ_3 berturut-turut 1.4296, 0.7756 dan 0.0098 serta nilai-p berturut-turut 0.2385, 0.3835 dan 0.9217 yang semuanya lebih besar dari α=0.1. Hal ini berarti bahwa parameter δ_1, δ_2, dan δ_3 tidak berpengaruh secara signifikan dalam model. Sebaliknya nilai statistik uji Wald parameter δ_4 sebesar 3.5641 dan nilai-p sebesar 0.066 yang lebih kecil dari α=0.1. Hal ini berarti bahwa parameter δ_4 berpengaruh secara signifikan dalam model. Tetapi karena nilai pendugaan parameter berubah ketika parameter yang tidak signifikan dikeluarkan dari model maka semua parameter dimasukkan dalam model sehingga model regresi HNB untuk kasus ini adalah :

$\log(\mu_i) = -0.7447 - 1.3547\ X_1 + 11.6739\ X_2 + 0.6174\ X_3 - 1.6898\ X_4$

dan

$\text{logit}(pi) = -4.410 + 1.2021\ X_1 + 0.5388\ X_2 - 1.3833\ X_3 + 0.1741\ X_4$

di mana X_1 adalah akreditasi sekolah, X_2 adalah rasio rencana penerimaan-pendaftar, X_3 adalah rasio guru-murid dan X_4 adalah rasio murid-kelas.

Dari model di atas dapat dijelaskan bahwa model logit pada kasus ini menjelaskan bahwa peluang SMA yang semua siswanya lulus dipengaruhi oleh rasio murid-kelas, sedangkan akreditasi sekolah, rasio rencana penerimaan-pendaftar, dan rasio guru-murid meskipun dimasukkan dalam model, pengaruhnya tidak signifikan pada kasus ini.

4.7 Interpretasi Model

Berdasarkan uji signifikansi parameter regresi *Hurdle Negative Binomial* (HNB) diperoleh model :

$$\log(\mu_i) = -0.7447 - 1.3547\ X_1 + 11.6739\ X_2 + 0.6174\ X_3 - 1.6898\ X_4 \quad (4.13)$$

dan

$$\text{logit}(p_i) = -4.4108 + 1.2021 X_1 + 0.5388\ X_2 - 1.3833\ X_3 + 0.1741\ X_4 \quad (4.14)$$

Untuk interpretasi model (4.13) memperhatikan setiap parameter dalam model dengan bentuk $ln\, y = \beta x$ kemudian diturunkan terhadap setiap variabel prediktor seperti berikut :

$$ln\, y = \beta x$$
$$\frac{\partial ln y}{\partial x} = \beta$$
$$\frac{\partial y}{\partial x}\frac{\partial ln y}{\partial y} = \beta$$
$$\frac{\partial y}{\partial x}\frac{1}{y} = \beta$$
$$\frac{\partial y\,/y}{\partial x} = \beta \quad (4.15)$$

Bentuk $\partial y/y$ merupakan bentuk perubahan y secara persentase, maka untuk mempermudah interpretasi ruas kiri dan ruas kanan dikalikan 100% seperti persamaan:

$$\frac{\partial y\,/y}{\partial x}100\% = \beta\, 100\% \quad (4.16)$$

Parameter yang signifikan pengaruhnya terhadap model (4.13) hanya parameter β_2 dan β_4 maka interpretasi yang diperlukan adalah interpretasi untuk parameter β_2 dan β_4. Parameter β_2 adalah koefisien regresi untuk variabel prediktor X_2 yaitu rasio penerimaan dan pendaftar dan β_4 adalah koefisien regresi untuk variabel prediktor X_4 yaitu rasio murid dan kelas. Sehingga interpretasi untuk model (4.13) adalah :

1. Setiap kenaikan rasio penerimaan dan pendaftar sebesar 0.01 maka akan meningkatkan rata-rata SMA yang siswanya gagal UN sebesar 11.67%
2. Setiap kenaikan rasio murid dan kelas sebesar 0.01 maka akan menurunkan rata-rata SMA yang siswanya gagal UN sebesar 1.68%

Untuk model (4.14) dapat dituliskan sebagai berikut

$$\frac{p_i}{1-p_i} = \exp(-\ 4.4108 + 1.2021X_1 + 0.5388X_2\ -1.3833X_3 + 0.1741X_4)$$

dengan parameter yang berpengaruh signifikan hanya δ_4 yang merupakan koefisien regresi untuk variabel prediktor X_4 yaitu rasio murid dan kelas. Sehingga interpretasi untuk model logit di atas adalah setiap peningkatan rasio murid dan kelas sebesar 0.01 menyebabkan peningkatan peluang SMA yang semua siswanya lulus sebesar exp(0.1741) = 1.19 kali peluang SMA yang tidak semua siswanya lulus.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Salah satu metode alternatif untuk mengatasi overdispersi dengan adanya *Excess Zeros* adalah regresi *Hurdle Negative Binomial* (HNB), setelah dilakukan analisis tentang metode ini, diperoleh hasil sebagai berikut :

1. a. Fungsi ln*likelihood* bagi regresi HNB adalah

$$\ln L\,(\delta,\beta) = \begin{cases} \sum_{i=1}^{n} \ln e^{x_i^T\delta} - \sum_{i=1}^{n} \ln\left(1+e^{x_i^T\delta}\right) & y_i = 0 \\ -\sum_{i=1}^{n} \ln\left(1+e^{x_i^T\delta}\right) + \sum_{i=1}^{n} \ln\left(\Gamma\left(y_i + {}^1\!/_\alpha\right)\right) - \sum_{i=1}^{n} \ln(\Gamma(y_i+1)) & \\ -\sum_{i=1}^{n} \ln\ (\Gamma({}^1\!/_\alpha)) - \left({}^1\!/_\alpha + y_i\right)\sum_{i=1}^{n} \ln\left(1+\alpha e^{x_i^T\beta}\right) & \\ +y_i\sum_{i=1}^{n}\left(ln\,\alpha + ln\, e^{x_i^T\beta}\right) - \sum_{i=1}^{n} \ln\left(1-\left(1+\alpha e^{x_i^T\beta}\right)^{-1/\alpha}\right) & ,\ y_i > 0 \end{cases}$$

   b. Pendugaan parameter model Regresi HNB menghasilkan penduga yang implisit, sehingga diperlukan prosedur iteratif untuk memperoleh pendugaan parameternya. Metode yang digunakan adalah Algoritma EM (*Expectation Maximization*) yang terdiri dari dua tahap yaitu tahap ekspektasi dan tahap maksimaliasasi. Pada tahap maksimalisasi digunakan metode *Newton Rahpson* untuk memaksimalkan fungsi *likelihood* yang diperoleh pada tahap ekspektasi.

2. Model regresi *Hurdle Negative Binomial* (HNB) untuk kasus banyaknya siswa SMA yang gagal Ujian Nasional (UN) tahun pelajaran 2013/2014 di Kota Malang adalah

   $\log(\mu_i)$ = -0.744 - 1.354 X_1 + 11.673 X_2 + 0.617 X_3 - 1.689 X_4

   dan

   logit(pi) = - 4.410 + 1.202 X_1 + 0.538 X_2 - 1.383 X_3 + 0.174 X_4

5.2 Saran

Saran yang dapat diberikan untuk penelitian selanjutnya adalah :

1. Sebaiknya menambahkan lebih banyak variabel prediktor seperti jarak lokasi sekolah terhadap pusat kota kecamatan, rata-rata nilai NUM siswa pendaftar serta metode pembelajaran yang digunakan.
2. Mengaplikasikan regresi HNB pada kasus yang mengalami underdispersi.
3. Mengaplikasikan regresi HNB pada kasus yang berbeda misalnya pada kasus di bidang kesehatan dan ekonomi.

DAFTAR PUSTAKA

Agresti, A. 2007. An Introduction to Categorical Data Analysis. Edisi Kedua. Canada: John Wiley and Sons.

Cahyosetiyono, D. 2009. *Pemodelan banyaknya Siswa Gagal Ujian Nasional dengan Regresi Zero Inflated Generalized Poisson*. Thesis Program Magister. ITS

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

Dempster, A., Laird, N. M., dan Rubin, D.B. 1977. *Maximum Likelihood from Incomplete Data via EM Algorithm*. Jurnal The Royal Statistical Society Series B (Methodological), Vol 39 Page 1-38. http://web.mit.edu/6.435/www/Dempster77.pdf. Tanggal Akses: 7 Januari 2015

Famoye, F., Wulu, J. T. dan Singh, K.P. 2004. *On the Generalized Poisson Regression Model with an Application to Accident Data*. Journal of Data Science 2 (2004) 287-295. http://www.jds-online.com/file_download/53/. Tanggal Akses: 6 Januari 2015

Gujarati, D. dan Porter, D.C. 2009. *Dasar-Dasar Ekonometrika Edisi 5*. Jakarta: Salemba Empat.

Harinaldi. Dr.,Ir. 2005. *Prinsip-Prinsip Statistika untuk Teknik dan Sains*. Jakarta: Gelora Aksara Pratama.

Hinde, J. dan Demetrio, C.G.B. 2007. *Overdispersion: Models and Estimation*.http://ce.esalq.usp.br/arquivous/aulas/2012/LCE5868/overdispersionBook.pdf. Tanggal Akses: 24 September 2014.

Ismail, N. dan Jemain, A.A. 2005. *Generalized Poisson Regression: An Alternative for Risk Classification*. Jurnal Teknologi Universiti Teknologi Malaysia, Vol 43 Page 39-54.

http://www.penerbit.utm.my/onlinejournal/43/C/JTDIS43C4.pdf. Tanggal Akses: 24 September 2014.

Lestari, A. 2008. *Pemodelan Regresi Zero Inflated Poisson (Aplikasi Pada Data Pekerja Seks Komersial di Klinik Reproduksi Putat Jaya Surabaya)*. Thesis Program Magister. ITS

Lestari, S.P. dan Wulandari, S.P. 2014. *Pemodelan Faktor-Faktor yang Mempengaruhi Jumlah Kasus Tetanus Neonatorum (TN) di Jawa Timur dengan Metode Regresi Zero Inflated Generalized Poisson (ZIGP)*. Jurnal Sains dan Seni Pomits. Vol. 3 Nomor 2 Hal. 2337-3520. http://ejurnal.its.ac.id/index.php/sains_seni/article/download/8079/2011. Tanggal Akses: 4 Mei 2015.

Maisaroh, dan Roestrieningsih. 2010. *Peningkatan Hasil Belajar Siswa Dengan menggunakan Metode Pembelajaran Active Learning Tipe Quiz Team Pada Mata Pelajaran Keterampilan Dasar Komunikasi di SMK Negeri 1 Bogor*. Jurnal Ekonomi dan Pendidikan, Vol. 8 Nomor 2 Hal 157-172. http://journal.uny.ac.id/index.php/jep/article/viewFile/571/427. Tanggal Akses: 11 Januari 2015.

McCullagh, P., dan Nelder, J.A. 1989. *Generalized Linear Models,Second Edition*. Chapman and Hall, London.

Rodriguez, G. 2013. *Model for Count Data With Overdispersi*. http://data.princeton.edu/wws509/notes/c4a.pdf. Tanggal Akses: 21 Juni 2015

Saffari, S. E, Adnan, R., dan Greene, W. 2012. *Hurdle Negative binomial Regression Model with Right Cencored Count Data*. Journal of Statistics and Operations Research Transactions, Vol 36 iss.2 Page 181-194. http://www.idescat.cat/sort/sort362/36._2.4.saffari-etal.pdf. Tanggal Akses: 9 September 2014.

LAMPIRAN

Lampiran 1. Data Banyaknya Siswa SMA yang Gagal Ujian Nasional (UN) Tahun Pelajaran 2013/2014 di Kota Malang

NO	NAMA SEKOLAH	Y	X1	X2	X3	X4
1	SMA N 01 Kota Malang	0	1	0.71	0.07	29.13
2	SMA N 02 Kota Malang	0	1	0.97	0.07	26.85
3	SMA N 03 Kota Malang	0	1	0.37	0.07	29.70
4	SMA N 04 Kota Malang	0	1	0.59	0.07	27.43
5	SMA N 05 Kota Malang	0	1	1.00	0.06	30.47
6	SMA N 06 Kota Malang	0	1	0.83	0.06	30.14
7	SMA N 07 Kota Malang	0	1	1.02	0.06	33.44
8	SMA N 08 Kota Malang	0	1	0.82	0.06	31.45
9	SMA N 09 Kota Malang	0	1	1.09	0.06	32.50
10	SMA N 10 Kota Malang	0	1	1.00	0.06	24.58
11	SMA Islam Baiturrohmah	0	3	2.86	0.11	22.00
12	SMA Islam	0	1	0.73	0.04	32.72
13	SMA Katolik Cor Jesu	0	1	0.86	0.08	21.47
14	SMA Islam Maarif	0	2	2.27	0.08	20.00
15	SMA Katolik st. Albertus	0	1	0.68	0.06	32.33
16	SMA PGRI 6 Malang	0	2	2.43	0.25	10.00
17	SMAK Santa Maria	0	1	0.76	0.04	39.29
18	SMA Kristen Charis	0	2	1.28	0.09	15.00
19	SMA Laboratorium UM	0	1	0.43	0.05	32.48
20	SMA Taman Madya	0	1	5.66	0.19	24.50
21	SMA Shalahuddin	0	1	4.09	0.19	12.33
22	SMA 2 YPK JATIM	0	1	3.33	0.22	13.75
23	SMA Nasional	1	2	1.10	0.16	23.50
24	SMA Katolik Frateran	0	1	1.11	0.05	32.93
25	SMA Darul 'Ulum Agung	0	2	0.70	0.06	12.00
26	SMA Kristen Petra	0	1	1.71	0.14	18.50
27	SMA Kristen Kalam Kudus	1	1	1.07	0.09	13.78
28	SMA Taman Harapan	1	1	2.40	0.10	24.17
29	SMA Panjura	0	1	0.92	0.07	25.71
30	SMA Kertanegara	0	2	1.00	0.14	26.67
31	SMA Dharma Raya Bakti	0	2	1.67	0.11	26.67
32	SMAK Kolese Santo Yusup	0	1	0.81	0.05	39.91
33	SMA Surya Buana	0	4	1.33	0.12	11.00
34	SMA Advent Dwi Abdi	0	2	2.14	0.06	16.20

Lampiran 1 (Lanjutan)

35	SMA Brawijaya Smart School	0	1	1.01	0.08	21.59
36	SMA Widya Gama	3	1	1.92	0.12	17.50
37	SMA Kristen Setia Budi	2	2	1.15	0.11	11.75
38	SMA Muhammadiyah 1	1	1	1.96	0.08	23.71
39	SMA Cokroaminoto	0	3	10.00	0.03	17.00
40	SMA Wisnuwardhana	0	2	7.50	0.22	9.00
41	SMA Waskita Dharma	0	3	6.40	0.30	7.67
42	SMA Wahid Hasyim	0	2	9.38	0.40	5.00

Sumber : Dinas Pendidikan Kota Malang (2013-2014)

Lampiran 2. Output Program SAS untuk Regresi Poisson

The GENMOD Procedure

Model Information

Data Set	WORK.SMA
Distribution	Poisson
Link Function	Log
Dependent Variable	Y

Number of Observations Read	42
Number of Observations Used	42

Class Level Information

Class	Levels	Values
Y	4	0 1 2 3

Parameter Information

Parameter	Effect
Prm1	Intercept
Prm2	X1
Prm3	X2
Prm4	X3
Prm5	X4

Lampiran 2. (Lanjutan)

Iteration History For Parameter Estimates

Iter	**Ridge**	**LogLikelihood**	**Prm1**	**Prm2**	**Prm3**	**Prm4**	**Prm5**
0	**0**	**-30.93114**	1.65498	-0.3235	-0.07342	-1.30367	-0.05506
1	**0**	**-20.81869**	2.03180	-0.5107	-0.13339	-1.71996	-0.08596
2	**0**	**-18.38049**	3.09151	-0.7998	-0.25374	-1.41966	-0.12944
3	**0**	**-17.90928**	3.90054	-1.0405	-0.41675	0.46916	-0.15870
4	**0**	**-17.86793**	4.12788	-1.1280	-0.50766	1.67239	-0.16587
5	**0**	**-17.86725**	4.15233	-1.1389	-0.52225	1.84092	-0.16638
6	**0**	**-17.86725**	4.15278	-1.1391	-0.52258	1.84440	-0.16638
7	**0**	**-17.86725**	4.15278	-1.1391	-0.52258	1.84440	-0.16638

Criteria For Assessing Goodness Of Fit

Criterion	**DF**	**Value**	**Value/DF**
Deviance	37	27.0988	0.7324
Scaled Deviance	37	27.0988	0.7324
Pearson Chi-Square	37	40.0786	1.0832
Scaled Pearson X2	37	40.0786	1.0832
Log Likelihood		-17.8673	
Full Log Likelihood		-20.3522	
AIC (smaller is better)		50.7043	
AICC (smaller is better)		52.3710	
BIC (smaller is better)		59.3927	

Lampiran 2. (Lanjutan)

Last Evaluation Of The Negative Of The Gradient and Hessian

	Prm1	Prm2	Prm3	Prm4	Prm5
Gradient	3.3485E-7	6.0112E-7	1.8374E-6	7.2646E-8	4.1095E-6
Prm1	9.0000003	12.000001	14.595797	1.0133678	161.15873
Prm2	12.000001	18.503171	19.866613	1.4045329	200.15113
Prm3	14.595797	19.866613	36.807572	2.1909885	231.79372
Prm4	1.0133678	1.4045329	2.1909885	0.1486739	16.186923
Prm5	161.15873	200.15113	231.79372	16.186923	3298.4852

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	4.1528	2.1639	-0.0884	8.3939	3.68	0.055
X1	1	-1.1391	0.7191	-2.5485	0.2703	2.51	0.1132
X2	1	-0.5226	0.4777	-1.4588	0.4136	1.2	0.2739
X3	1	1.8444	9.9036	-17.5663	21.2551	0.03	0.8523
X4	1	-0.1664	0.0639	-0.2916	-0.0412	6.78	0.0092
Scale	0	1	0	1	1		

Lampiran 3. Pembuktian Penurunan Rumus *Log-Likelihood* Model HNB

- **Turunan terhadap parameter β**

$$lnL\left(lnL\left(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)}\right)\right)=\sum_{i=1}^{n}(1-\boldsymbol{w_i}^{(m)})\left[-\left({}^{1}/_{\alpha}+y_i\right)\sum_{i=1}^{n}\ln\left(1+\alpha e^{\boldsymbol{x_i^T\beta}}\right)+y_i\sum_{i=1}^{n}\left(ln\,\alpha+ln\,e^{\boldsymbol{x_i^T\beta}}\right)-\right.$$

*i=1n*ln*1−1+αe**xiTβ**−1α*

$$\frac{\partial}{\partial\boldsymbol{\beta}}\,lnL\left(lnL\left(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)}\right)\right)=\sum_{i=1}^{n}(1-w_i^{(m)})\left[-\left({}^{1}/_{\alpha}+y_i\right)\sum_{i=1}^{n}\frac{1}{1+\alpha e^{\boldsymbol{x_i^T\beta}}}\left(\alpha\boldsymbol{x_i^T}e^{\boldsymbol{x_i^T\beta}}\right)+y_i\sum_{i=1}^{n}\left(\frac{1}{e^{\boldsymbol{x_i^T\beta}}}\left(\boldsymbol{x_i^T}e^{\boldsymbol{x_i^T\beta}}\right)\right)\right.$$

$$\left.-\sum_{i=1}^{n}\left(\frac{1}{\left(1-\left(1+\alpha e^{\boldsymbol{x_i^T\beta}}\right)^{-1/\alpha}\right)}\left(\frac{1}{\alpha}\left(1+\alpha e^{\boldsymbol{x_i^T\beta}}\right)^{-\frac{(\alpha+1)}{\alpha}}\right)\left(\alpha\boldsymbol{x_i^T}e^{\boldsymbol{x_i^T\beta}}\right)\right)\right]$$

$$=\sum_{i=1}^{n}(1-w_i^{(m)})\left[-\left({}^{1}/_{\alpha}+y_i\right)\sum_{i=1}^{n}\frac{\alpha\boldsymbol{x_i^T}e^{\boldsymbol{x_i^T\beta}}}{1+\alpha e^{\boldsymbol{x_i^T\beta}}}+y_i\sum_{i=1}^{n}\boldsymbol{x_i^T}-\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}\frac{1}{\alpha}\frac{\left(1+\alpha e^{\boldsymbol{x_i^T\beta}}\right)^{-\frac{(\alpha+1)}{\alpha}}\left(\alpha\boldsymbol{x_i^T}e^{\boldsymbol{x_i^T\beta}}\right)}{\left(1-\left(1+\alpha e^{\boldsymbol{x_i^T\beta}}\right)^{-1/\alpha}\right)}\right]$$

Lampiran 3. (Lanjutan)

$$\frac{\partial^2}{\partial\boldsymbol{\beta}^2} lnL\left(lnL\left(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)}\right)\right)=\sum_{i=1}^{n}(1-w_i{}^{(m)})\left[-\left({}^{1}/_{\alpha}+\ y_i\right)\sum_{i=1}^{n}\left(\frac{\alpha\left(\boldsymbol{x}_i^T\right)^2 e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)-\left(\alpha\boldsymbol{x}_i^T e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)\left(\alpha\boldsymbol{x}_i^T e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)}{\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^2}\right)-\right.$$

$$\left.\sum_{i=1}^{n}\left(\frac{-\frac{(\alpha+1)}{\alpha}\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\frac{(2\alpha+1)}{\alpha}}\left(\alpha\boldsymbol{x}_i^T e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)\left(\alpha\boldsymbol{x}_i^T e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)+\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\frac{(\alpha+1)}{\alpha}}\left(\alpha(\boldsymbol{x}_i^T)^2 e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)\left(\alpha\left(1-\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\frac{1}{\alpha}}\right)\right)-\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\frac{(\alpha+1)}{\alpha}}\left(\alpha\boldsymbol{x}_i^T e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)\alpha\left(\frac{1}{\alpha}\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\frac{(\alpha+1)}{\alpha}}\left(\alpha\boldsymbol{x}_i^T e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)\right)}{\alpha^2\left(1-\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-1/\alpha}\right)^2}\right)\right]$$

$$\frac{\partial^2}{\partial\boldsymbol{\beta}^2} lnL\left(lnL\left(\boldsymbol{\beta};\boldsymbol{y};\boldsymbol{w}^{(m)}\right)\right)=\sum_{i=1}^{n}(1-w_i{}^{(m)})\left[-\left({}^{1}/_{\alpha}+\ y_i\right)\sum_{i=1}^{n}\left(\frac{\alpha\left(\boldsymbol{x}_i^T\right)^2 e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}+\left(\alpha\boldsymbol{x}_i^T\right)^2 e^{2\boldsymbol{x}_i^T\boldsymbol{\beta}}-\left(\alpha\boldsymbol{x}_i^T\right)^2 e^{2\boldsymbol{x}_i^T\boldsymbol{\beta}}}{\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^2}\right)-\right.$$

$$\left.\sum_{i=1}^{n}\left(\frac{\left(\alpha\boldsymbol{x}_i^T\right)^2 e^{2\boldsymbol{x}_i^T\boldsymbol{\beta}}-\frac{(\alpha+1)}{\alpha}\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\frac{(2\alpha+1)}{\alpha}}+\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\frac{(\alpha+1)}{\alpha}}\left(\alpha\left(\boldsymbol{x}_i^T\right)^2 e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)\left(\alpha\left(1-\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-1/\alpha}\right)\right)-\left(\alpha\left(\boldsymbol{x}_i^T\right)^2 e^{2\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)\left(\frac{1}{\alpha}\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\frac{(\alpha+1)}{\alpha}}\right)\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-\frac{(\alpha+1)}{\alpha}}}{\alpha^2\left(1-\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^{-1/\alpha}\right)^2}\right)\right]$$

$$=\sum_{i=1}^{n}(1-w_i{}^{(m)})\left[\left(-\left({}^{1}/_{\alpha}+\ y_i\right)\sum_{i=1}^{n}\left(\frac{\alpha\left(\boldsymbol{x}_i^T\right)^2 e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}{\left(1+\alpha e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}\right)^2}\right)\right)-\right.$$

*i=1nα**xiT2**e2**xiTβ**−(α+1)α1+αe**xiTβ**−(2α+1)α+1+αe**xiTβ**−(α+1)αα**xiT2**e**xiTβ**α1−1+αe**xiTβ**−1α−α**xiT2**e2**xi** **Tβ**1α1+αe**xiTβ**−2(α+1)αα21−1+αe**xiTβ**−1α2*

Lampiran 3. (lanjutan)

- **Turunan terhadap parameter $\boldsymbol{\delta}$**

$$lnL\left(lnL\left(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)}\right)\right)=\sum_{i=1}^{n}\left[w_i^{(m)}\boldsymbol{x}_i^T\boldsymbol{\delta}-ln\left(1+\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)\right]$$

$$\frac{\partial}{\partial\boldsymbol{\delta}}\,lnL\left(lnL\left(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)}\right)\right)=\sum_{i=1}^{n}\left[w_i^{(m)}\boldsymbol{x}_i^T-\frac{\mathbf{1}}{1+\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}}\left(\boldsymbol{x}_i^T\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)\right]$$

$$=\sum_{i=1}^{n}\left[w_i^{(m)}\boldsymbol{x}_i^T-\frac{\boldsymbol{x}_i^T\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}}{1+\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}}\right]$$

$$\frac{\partial^2}{\partial\boldsymbol{\delta}^2}\,lnL\left(lnL\left(\boldsymbol{\delta};\boldsymbol{y};\boldsymbol{w}^{(m)}\right)\right)=-\sum_{i=1}^{n}\left[\frac{\left(\boldsymbol{x}_i^T\boldsymbol{x}_i^T\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)\left(1+\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)-\left(\boldsymbol{x}_i^T\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)\left(\boldsymbol{x}_i^T\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)}{\left(1+\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)^2}\right]$$

$$=-\sum_{i=1}^{n}\left[\frac{\left(\boldsymbol{x}_i^T\right)^2\left(\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)+\left(\boldsymbol{x}_i^T\right)^2\left(\mathrm{e}^{2\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)-\left(\boldsymbol{x}_i^T\right)^2\left(\mathrm{e}^{2\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)}{\left(1+\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)^2}\right]$$

$$=-\sum_{i=1}^{n}\left[\frac{\left(\boldsymbol{x}_i^T\right)^2\left(\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)}{\left(1+\mathrm{e}^{\boldsymbol{x}_i^T\boldsymbol{\delta}}\right)^2}\right]$$

Lampiran 4. Program SAS untuk Menduga Parameter Regresi HNB dan Pengujian Hipotesis

```
/* Poisson Regression Model* /

proc genmod data = SMA;
  class Y  /param=glm;
  model y = x1 x2 x3 x4 / dist = poisson link = log itprint;

run;


/* HNB models */

proc nlmixed data = sma ;
  parms a0 = 0 a1 = 0 a2 = 0 a3 = 0 a4 = 0
       b0 = 0 b1 = 0 b2 = 0 b3 = 0 b4 = 0;
  eta0 = a0 + a1 * X1 + a2 * X2 + a3 * X3 + a4 * X4;
  exp_eta0 = exp(eta0);
  p0 = exp_eta0 / (1 + exp_eta0);
  etap = b0 + b1 * X1 + b2 * X2 + b3 * X3 + b4 * X4;
  exp_etap = exp(etap);
  phi = 1/alpha;

  if Y = 0 then ll = log(p0);
  else ll = (-log(1 + exp_eta0)) + (log(gamma(Y + phi))) –
           (log(gamma(Y +   1))) - (log(gamma(phi))) + (-(phi + Y) *
            log(1 + (phi * exp_etap))) + (Y * (log(phi) + log(exp_etap)))
           - (log(1-(1+(phi * exp_etap)) ** (-phi)));

  model Y ~ general(ll);
  ods output ParameterEstimates=pe;

run;
```

Lampiran 5. Output Program SAS untuk Pendugaan Parameter Regresi HNB dan Pengujian Hipotesis

The NLMIXED Procedure

Specifications

Data Set	WORK.SMA
Dependent Variable	Y
Distribution for Dependent Variable	General
Optimization Technique	Dual Quasi-Newton
Integration Method	None

Dimensions

Observations Used	42
Observations Not Used	0
Total Observations	42
Parameters	11

Parameters

a0	**a1**	**a2**	**a3**	**a4**	**b0**	**b1**	**b2**	**b3**	**b4**	**alpha**	**NegLogLike**
0	0	0	0	0	0	0	0	0	0	1	35.3505062

Iteration History

Iter	**Calls**	**NegLogLike**	**Diff**	**MaxGrad**	**Slope**
1	4	23.4734	11.8771	51.2346	-1860.7
2	6	22.2773	1.19616	22.1527	-0.8501
3	7	22.0504	0.22691	8.07403	-0.396
…	…	…	…	…	…
166	632	17.0280104	7.07E-13	18.84553	-0.054

Lampiran 5. (Lanjutan)

Fit Statistics

-2 Log Likelihood	34.1
AIC (smaller is better)	56.1
AICC (smaller is better)	64.9
BIC (smaller is better)	75.2

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|
a0	-4.4108	3.3746	42	-1.31	0.1983
a1	1.2021	1.0054	42	1.20	0.2385
a2	0.5388	0.6118	42	0.88	0.3835
a3	-1.3833	13.9846	42	-0.10	0.9217
a4	0.1741	0.09222	42	1.89	0.0660
b0	-0.7447	4.9080	42	-0.15	0.8801
b1	-1.3547	4.7388	42	-0.29	0.7764
b2	11.6739	5.2596	42	2.22	0.0319
b3	0.6174	160.28	42	0.00	0.9969
b4	-1.6898	0.7853	42	-2.15	0.0372
alpha	0.006013	0.01450	42		