# PENDUGAAN PARAMETER REGRESI *ZERO INFLATED GENERALIZED* POISSON DENGAN METODE BAYESIAN
## (Studi Kasus Ketidaklulusan Siswa SMA/SMK Dalam Ujian Nasional di Kota Malang Tahun Ajaran 2015/2016)

**SKRIPSI**


oleh :

**Dewi Kurnia Sari**

**135090501111021**



**JURUSAN MATEMATIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS BRAWIJAYA**

**MALANG**

**2017**

# PENDUGAAN PARAMETER REGRESI *ZERO INFLATED GENERALIZED* POISSON DENGAN METODE BAYESIAN (Studi Kasus Ketidaklulusan Siswa SMA/SMK Dalam Ujian Nasional di Kota Malang Tahun Ajaran 2015/2016)

**SKRIPSI**

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

**oleh :**
**Dewi Kurnia Sari**
**135090501111021**

**JURUSAN MATEMATIKA**
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**
**UNIVERSITAS BRAWIJAYA**
**MALANG**
**2017**

## LEMBAR PENGESAHAN

**PENDUGAAN PARAMETER REGRESI *ZERO INFLATED GENERALIZED* POISSON DENGAN METODE BAYESIAN (Studi Kasus Ketidaklulusan Siswa SMA/SMK Dalam Ujian Nasional di Kota Malang Tahun Ajaran 2015/2016)**

**oleh :**
**Dewi Kurnia Sari**
**135090501111021**

**Setelah dipertahankan di depan Majelis Penguji**
**Pada tanggal 27 April 2017**
**Dan dinyatakan memenuhi syarat untuk memperoleh gelar**
**Sarjana Sains dalam bidang Statistika**

**Pembimbing**

**Dr. Suci Astutik, S.Si., M.Si**
**NIP. 197407221999032001**

**Mengetahui,**
**Ketua Jurusan Matematika**
**Fakultas MIPA Universitas Brawijaya**

**Ratno Bagus Edy Wibowo, S.Si., M.Si., Ph.D**
**NIP. 197509082000031003**

**LEMBAR PERNYATAAN**

**Saya yang bertanda tangan di bawah ini :**

| | | |
|---|---|---|
| **Nama** | **:** | **Dewi Kurnia Sari** |
| **NIM** | **:** | **135090501111021** |
| **Jurusan** | **:** | **Matematika** |
| **Program Studi** | **:** | **Statistika** |
| **Judul Skripsi** | **:** | **Pendugaan Parameter Regresi *Zero Inflated Generalized* Poisson Dengan Metode Bayesian (Studi Kasus Ketidaklulusan Siswa SMA/SMK Dalam Ujian Nasional di Kota Malang Tahun Ajaran 2015/2016)** |

**Dengan ini menyatakan bahwa :**

1. **Isi dari Skripsi yang saya buat adalah benar-benar karya sendiri dan tidak menjiplak karya orang lain, selain nama-nama yang termaktub di isi dan tertulis di daftar pustaka dalam Skripsi ini.**
2. **Apabila di kemudian hari ternyata Skripsi yang saya tulis terbukti hasil jiplakan, maka saya bersedia menanggung segala resiko yang akan saya terima.**

**Demikian pernyataan ini dibuat dengan segala kesadaran.**

**Malang, 27 April 2017**
**Yang menyatakan,**

**<u>Dewi Kurnia Sari</u>**
**135090501111021**

**PENDUGAAN PARAMETER REGRESI *ZERO INFLATED GENERALIZED* POISSON DENGAN METODE BAYESIAN (Studi Kasus Ketidaklulusan Siswa SMA/SMK Dalam Ujian Nasional di Kota Malang Tahun Ajaran 2015/2016)**

## ABSTRAK

Regresi Poisson adalah model regresi yang digunakan untuk data diskrit dengan peubah respon berdistribusi Poisson. Regresi Poisson memiliki asumsi nilai ragam dan rata-rata bernilai sama (*equidispersi*), namun keadaan yang sering terjadi adalah *overdispersi* yang disebabkan karena terlalu banyak nilai nol pada peubah respon. Untuk mengatasi masalah tersebut maka dapat digunakan analisis regresi *Zero Inflated Generalized Poisson* (ZIGP). Pendugaan parameter yang digunakan dalam penelitian ini adalah metode Bayesian dengan simulasi MCMC dan algoritma *gibbs sampling*. Pada metode Bayesian parameter dipandang sebagai peubah acak yang memiliki distribusi sendiri yaitu distribusi prior. Penelitian ini bertujuan untuk mendapatkan model regresi ZIGP dengan metode Bayesian dan mengetahui faktor-faktor yang mempengaruhi jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di kota Malang tahun ajaran 2015/2016. Pada penelitian ini menggunakan prior non informatif dengan distribusi normal dan tiga cara pemeriksaan konvergensi. Dari ketiga cara tersebut dapat disimpulkan bahwa konvergensi pada MCMC sudah terpenuhi. Peubah prediktor yang digunakan dalam penelitian ini yaitu status sekolah, jumlah rombongan belajar, rasio guru-pegawai, rasio guru-siswa, dan rasio rencana penerimaan pendaftaran. Berdasarkan hasil analisis regresi ZIGP dengan metode Bayesian, jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional hanya dipengaruhi oleh jumlah rombongan belajar.

*Kata Kunci : Bayesian, MCMC, Overdispersi, Zero Inflated Generalized Poisson (ZIGP)*

**PARAMETER ESTIMATION FOR ZERO INFLATED GENERALIZED POISSON REGRESSION WITH BAYESIAN METHOD**
**(Case Studies of Unsuccessful SMA/SMK Students In the National Exam In Malang 2015/2016 Year)**

## ABSTRACT

Poisson Regression is regression model used to model discrete data where the response variable follows a Poisson distribution. Poisson regression have assumption mean and variance are equal (equidispersion), however situations often occurs is overdispersion caused by excess number of zero in the response variable. Thus, Zero Inflated Generalized Poisson (ZIGP) regression model can be use to solve this problem. Parameter estimation technique used in this research is Bayesian method with MCMC simulation and gibbs sampling algorithm. In the Bayesian method parameter seen as random variable which have their own distribution called prior distribution. This research have purpose to get ZIGP regression model with Bayesian method and determine the factors that significant effect in case unsuccessful SMA/SMK students in the national exam in Malang 2015/2016 year. In this research used prior non informative with normal distribution and three convergence methods. The result that the convergence methods is convergence of MCMC has been fulfilled. Predictor variables used in this research they are status of the school, learning group, teacher-employee ratio, teacher-student ratio, and the registration plan ratio. Based on the result, variable having a significant effect on the number of unsuccessful SMA/SMK students in the national exam is learning group.

*Keywords : Bayesian, MCMC, Overdispersion, Zero Inflated Generalized Poisson (ZIGP).*

## KATA PENGANTAR

Puji syukur ke hadirat Allah SWT atas segala limpahan rahmat, karunia, dan hidayahNYA sehingga skripsi ini dapat diselesaikan dengan baik. Shalawat dan salam semoga senantiasa tercurah pada Nabi Muhammad SAW, juga segenap keluarga, sahabat, serta umat beliau hingga akhir zaman.

Penulis menyadari bahwa dalam menyusun skripsi ini telah banyak bantuan dan dukungan yang diterima oleh penulis. Oleh karena itu, penulis mengucapkan banyak terima kasih dan rasa hormat kepada :

1. Dr. Suci Astutik, S.Si, M.Si selaku Dosen Pembimbing yang telah memberikan pengarahan dan saran.
2. Achmad Efendi, S.Si, M.Sc., PhD selaku Dosen Penguji I yang telah memberikan pengarahan dan saran.
3. Ir. Heni Kusdarwati, MS selaku Dosen Penguji II yang telah memberikan pengarahan dan saran.
4. Ratno Bagus Edy Wibowo, S.Si., M.Si., Ph.D. selaku Ketua Jurusan Matematika Universitas Brawijaya.
5. Seluruh jajaran Dosen Jurusan Matematika atas ilmu yang diberikan saat perkuliahan.
6. Seluruh staf dan karyawan Jurusan Matematika Universitas Brawijaya.
7. Bapak, Ibu, mas Lili, dan mas Dani beserta seluruh keluarga yang telah memberikan doa, dukungan, dan kasih sayang yang tidak dapat dibandingkan dengan apapun.
8. Keluarga Sukses Soh Soh yang telah menjadi keluarga selama di Malang dan *thanks for everything*.
9. Seluruh sahabat statistika 2013, Pejuang Bayesian, Partner Gupuh Dita RKP, dan Pak Ni yang telah memberikan bantuan dan semangat.
10. Semua pihak yang telah membantu penulis yang tidak bisa disebutkan satu persatu.

Dengan keterbatasan pengetahuan dan kemampuan yang dimiliki, penulis menyadari bahwa dalam menyusun skripsi ini masih jauh dari kesempurnaan. Maka dari itu penulis sangat mengharapkan saran dan kritik yang bersifat membangun demi kesempurnaan skripsi ini. Akhirnya penulis bersama semua pihak yang membantu penyelesaian penulisan skripsi ini berharap semoga skripsi ini dapat memberikan manfaat serta dapat menjadi tambahan pengetahuan bagi pembaca. Amin.

Malang, 27 April 2017

Penulis

## DAFTAR ISI

Halaman

# DAFTAR TABEL

Halaman

## DAFTAR GAMBAR

Halaman

# DAFTAR LAMPIRAN

# BAB I
# PENDAHULUAN

## 1.1 Latar Belakang

Dalam suatu penelitian sering kali ingin diketahui pengaruh dari beberapa peubah dengan menggunakan analisis regresi. Analisis regresi merupakan salah satu ilmu statistika yang digunakan untuk menganalisis hubungan antara satu atau lebih peubah prediktor (X) dengan satu peubah respon (Y). Pada umumnya peubah respon yang digunakan dalam analisis regresi adalah data kontinu, tetapi tidak sedikit penelitian menggunakan data diskrit sebagai peubah respon dalam analisis regresi. Jika data peubah respon adalah diskrit dan berdistribusi Poisson maka analisis yang digunakan adalah regresi Poisson.

Regresi Poisson adalah analisis regresi yang digunakan untuk data diskrit dan termasuk ke dalam *Generalized Linear Model* (GLM). Asumsi yang harus terpenuhi pada regresi Poisson adalah nilai ragam dan rata-rata bernilai sama (*equidispersi*). Dalam suatu penelitian, keadaan yang sering terjadi adalah *overdispersi*. *Overdispersi* merupakan keadaan di mana nilai ragam lebih besar dari nilai rata-ratanya. Salah satu faktor yang menyebabkan *overdispersi* yaitu terlalu banyak nilai nol pada peubah respon. Terdapat beberapa metode yang dapat menyelesaikan masalah *overdispersi*, yaitu *Zero Inflated Poisson* (ZIP), *Zero Inflated Negative Binomial* (ZINB), dan *Zero Inflated Generalized Poisson* (ZIGP).

Penelitian mengenai masalah *overdispersi* sudah pernah dilakukan oleh Famoye dan Singh (2006) dengan menggunakan model regresi *Zero Inflated Generalized Poisson* (ZIGP) pada data kekerasan dalam rumah tangga. Pada penelitian tersebut menjelaskan bahwa model regresi ZIGP merupakan metode yang lebih baik dari pada ZIP dan ZINB. Pada metode ZINB dalam teknik iterasi yang digunakan untuk estimasi parameter sering kali mengalami kegagalan konvergensi. Pada metode ZIP memiliki proporsi data bernilai nol sebesar 63.7% sedangkan pada metode ZIGP sebesar 65.7%, sehingga disarankan metode ZIGP sebagai alternatif yang baik untuk mengatasi masalah *overdispersi*.

Model regresi ZIGP merupakan salah satu metode yang dapat digunakan untuk mengatasi masalah *overdispersi*. Model ZIGP adalah gabungan antara model regresi *Zero Inflated Poisson* (ZIP) dengan model *Generalized Poisson* (GP). Pendugaan parameter yang digunakan dalam penelitian ini menggunakan metode Bayesian. Perbedaan yang paling utama dari metode Bayesian dan metode klasik adalah pada metode Bayesian parameter dipandang sebagai peubah acak yang memiliki distribusi prior (Ntzoufras, 2009). Kelebihan dari metode Bayesian yaitu dapat digunakan ketika data yang dianalisis memiliki sampel kecil atau tidak memenuhi asumsi. Menurut Pereira (1999) metode Bayesian menghasilkan pendugaan yang lebih baik dari pada metode klasik seperti *Maximum Likelihood Estimation* (MLE). Hal tersebut dikarenakan pada metode Bayesian parameter diperlakukan sebagai peubah acak yang memiliki distribusi sendiri. Distribusi ini sering disebut distribusi prior. Distribusi prior nantinya akan dikombinasikan dengan *likelihood* untuk mendapatkan distribusi posterior.

Pendidikan merupakan salah satu faktor penting dalam suatu kehidupan. Kemajuan di bidang pendidikan akan menciptakan generasi-generasi muda yang cerdas dan sumber daya manusia yang berkualitas. Saat ini pendidikan menjadi fokus utama pembangunan yang dilakukan pemerintah. Mutu pendidikan sangatlah penting untuk meningkatkan perubahan dan perkembangan dalam berbagai aspek kehidupan. Salah satu cara untuk mengukur mutu suatu pendidikan dapat dilihat dari hasil ujian nasional. Ujian nasional adalah ujian yang diadakan oleh pemerintah pusat pada setiap tahun untuk mengevaluasi standar pendidikan secara nasional dan persamaan mutu pendidikan antar daerah. Kualitas pendidikan yang baik diharapkan dapat meminimalkan jumlah ketidaklulusan siswa dalam ujian nasional. Banyak faktor yang mempengaruhi jumlah ketidaklulusan siswa dalam ujian nasional. Dengan diketahuinya faktor-faktor tersebut diharapkan pemerintah dapat meminimalkan jumlah ketidaklulusan siswa dalam ujian nasional.

Model regresi ZIGP dapat diterapkan dalam bidang pendidikan, yaitu untuk memodelkan data pada kasus ketidaklulusan siswa dalam ujian nasional. Ketidaklulusan siswa dalam ujian nasional merupakan kasus yang jarang terjadi, sehingga banyak pengamatan yang bernilai nol. Banyaknya nilai nol pada kasus ketidaklulusan siswa dalam ujian nasional menyebabkan terjadinya *overdispersi*. Data jumlah ketidaklulusan siswa dalam ujian nasional adalah data diskrit sehingga dalam kasus tersebut dapat didekati dengan metode statistika yaitu regresi ZIGP. Berdasarkan penelitian terdahulu, Faridhoh (2015) menerapkan model ZIGP dalam kasus ketidaklulusan siswa SMA di Kota Malang dengan metode MLE. Dari penelitian tersebut tidak terdapat peubah prediktor yang berpengaruh secara signifikan terhadap kasus ketidaklulusan siswa SMA di Kota Malang, sehingga dalam penelitian ini ingin diketahui penerapan metode Bayesian pada model ZIGP dalam kasus ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang, permasalahan yang dapat dirumuskan pada penelitian ini adalah :

1. Bagaimana penerapan model regresi *Zero Inflated Generalized Poisson* (ZIGP) dengan metode Bayesian pada data jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang tahun ajaran 2015/2016?
2. Apa saja peubah prediktor yang mempengaruhi jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang tahun ajaran 2015/2016?

## 1.3 Batasan Masalah

Dalam penelitian ini, batasan masalah yang digunakan adalah pendugaan parameter regresi *Zero Inflated Generalized Poisson* (ZIGP) metode Bayesian dengan teknik simulasi *Markov Chain Monte Carlo* (MCMC) dan algoritma *gibbs sampling* pada kasus ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang tahun ajaran 2015/2016.

## 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, tujuan penelitian ini adalah :

1. Mendapatkan model regresi *Zero Inflated Generalized Poisson* (ZIGP) dengan metode Bayesian pada data jumlah

ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang tahun ajaran 2015/2016.

2. Mengetahui peubah prediktor yang mempengaruhi jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang tahun ajaran 2015/2016.

## 1.5 Manfaat Penelitian

Manfaat penelitian ini adalah :

1. Mengetahui peubah prediktor yang mempunyai pengaruh terhadap jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang tahun ajaran 2015/2016. Informasi tersebut dapat digunakan sebagai pertimbangan bagi pemerintah Kota Malang untuk meminimalkan jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional.
2. Menambah wawasan keilmuwan tentang statistika terutama pada model regresi *Zero Inflated Generalized Poisson* (ZIGP) dengan metode Bayesian serta penerapannya pada data jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang tahun ajaran 2015/2016.

# BAB II
# TINJAUAN PUSTAKA

## 2.1 Tinjauan Statistika

### 2.1.1 Distribusi Poisson

Menurut Walpole (1995), percobaan Poisson adalah percobaan yang menghasilkan nilai-nilai bagi suatu peubah acak Y, yaitu banyaknya hasil percobaan yang terjadi selama suatu selang waktu tertentu atau di suatu daerah tertentu. Percobaan Poisson memiliki ciri-ciri sebagai berikut :

1. Banyaknya hasil percobaan yang terjadi dalam suatu selang waktu atau suatu daerah tertentu, tidak bergantung pada banyaknya hasil percobaan yang terjadi pada selang waktu atau daerah lain yang terpisah.
2. Peluang terjadinya satu hasil percobaan selama suatu selang waktu yang singkat atau daerah yang kecil, sebanding dengan panjang selang waktu atau besarnya daerah tersebut dan tidak bergantung pada banyaknya hasil percobaan yang terjadi di luar selang waktu atau daerah tersebut.
3. Peluang bahwa lebih dari satu hasil percobaan akan terjadi dalam selang waktu yang singkat atau dalam daerah yang kecil tersebut dapat diabaikan.

Sebaran peluang bagi peubah acak Poisson Y, yang menyatakan banyaknya hasil percobaan dalam suatu selang waktu atau daerah tertentu adalah:

$$p(y;\mu) = \frac{e^{-\mu}\mu^{y}}{y!} \tag{2.1}$$

di mana :

$y$ : 0,1,2,3,...

μ : rata-rata banyaknya hasil percobaan

$e$ : 2.71828

Distribusi Poisson memiliki nilai rata-rata dan ragam yang sama yaitu $E(Y) = \mu$ dan $Var(Y) = \mu$.

Pengujian distribusi data dapat menggunakan uji Kolmogorov Smirnov. Menurut Daniel (1989) uji Kolmogorov Smirnov memusatkan perhatian pada dua buah fungsi distribusi kumulatif, yaitu distribusi kumulatif yang dihipotesiskan dan distribusi kumulatif yang teramati. Untuk mengetahui data mengikuti distribusi Poisson dapat digunakan uji Kolmogorov Smirnov dengan hipotesis:

$H_0$ : $F(x) = F_0(x)$ (Data sampel mengikuti distribusi Poisson)
$H_1$ : $F(x) \neq F_0(x)$ (Data sampel tidak mengikuti distribusi Poisson)

dengan statistik yang digunakan adalah:

$$D = \sup|F(x) - F_0(x)| \qquad (2.2)$$

di mana $F(x)$ adalah fungsi kumulatif distribusi sampel dan $F_0(x)$ adalah fungsi kumulatif distribusi Poisson. Kriteria keputusan adalah tolak $H_0$ jika nilai statistik D > tabel statistik Kolmogorov Smirnov.

### 2.1.2 Regresi Poisson

Regresi Poisson merupakan suatu analisis regresi yang digunakan untuk memodelkan data yang menghasilkan peubah respon diskrit. Pada regresi Poisson, peubah respon mengikuti distribusi Poisson. Asumsi yang harus terpenuhi pada regresi Poisson yaitu:

1. Peubah respon berdistribusi Poisson.
2. Tidak terjadi multikolinieritas antar peubah prediktor.
3. Nilai rata-rata dan ragam peubah respon sama (*equidispersi*).

$$E(y) = \mu \qquad (2.3)$$

$$Var(y) = \mu \qquad (2.4)$$

Menurut Agresti (2007), model regresi Poisson merupakan *Generalized Linear Model* (GLM) dengan peubah respon diasumsikan berdistribusi Poisson dan *link function* yang digunakan pada model regresi Poisson adalah *log*. *Link function* merupakan fungsi linier yang menghubungkan rata-rata peubah respon dengan sebuah peubah prediktor. Fungsi $g$ sebagai *link function* dapat ditulis:

$$\begin{aligned} g(\mu_i) &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} \\ g(\mu_i) &= \ln(\mu_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} \\ \mu_i &= \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}) \qquad (2.5) \end{aligned}$$

Model untuk regresi Poisson dapat dilihat pada persamaan (2.6).

$$y_i = \mu_i + \varepsilon_i \qquad (i = 1, 2, 3, \ldots, n) \qquad (2.6)$$

sehingga model regresi Poisson dengan *link function* secara umum dapat ditulis sebagai berikut:

$$y_i = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}) + \varepsilon_i$$
$$y_i = \exp\left(\beta_0 + \sum_{j=1}^{k} \beta_j X_{i,j}\right) + \varepsilon_i \ , \quad i = 1,2,3,\ldots n \qquad (2.7)$$

di mana:

$y_i$ : peubah respon pada pengamatan ke-*i* yang merupakan peubah acak Poisson

$\mu_i$ : rata-rata peubah respon pada pengamatan ke-*i* yang dipengaruhi oleh nilai-nilai peubah prediktor pada pengamatan ke-*i*

$\varepsilon_i$ : galat pada pengamatan ke-*i*

β : nilai parameter yang diduga

$k$ : banyaknya peubah prediktor

$n$ : banyak pengamatan.

### 2.1.3 *Overdispersi*

Salah satu asumsi yang harus terpenuhi dalam regresi Poisson adalah *equidispersi*, yaitu kondisi di mana nilai ragam sama dengan nilai rata-ratanya. Namun dalam regresi Poisson kondisi yang sering terjadi adalah *overdispersi*. Menurut Famoye, dkk (2004), *overdispersi* adalah suatu kondisi data memiliki nilai ragam yang lebih besar dari nilai rata-ratanya. Penyebab terjadinya *overdispersi* pada regresi Poisson diantaranya adalah keheterogenan subjek yang diamati dan terlalu banyaknya nilai nol pada peubah respon.

Pemeriksaan *overdispersi* dapat dilakukan dengan statistik uji *pearson's chi-square*. Statistik uji *pearson's chi-square* yang dapat dipandang sebagai jumlah kuadrat dari *pearson residual*. *Pearson residual* dapat ditulis dengan rumus (Agresti, 2007):

$$\text{Pearson residual} = \frac{(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_{(yi)}}} \qquad (2.8)$$

sehingga statistik uji *pearson's chi-square* dapat ditulis dengan rumus:

$$\sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{V}_{(yi)}} \sim \chi^2_{(n-k-1)} \qquad (2.9)$$

di mana:

$y_i$ : nilai peubah respon pada pengamatan ke-*i*

$\hat{\mu}_i$ : penduga bagi rata-rata peubah respon pada pengamatan ke-*i*
$\hat{V}_{(yi)}$ : penduga bagi ragam peubah respon pada pengamatan ke-*i*
$n$ : banyak pengamatan
$k$ : banyak peubah prediktor.

Terjadi *overdispersi* apabila statistik uji *pearson's chi-square* dibagi derajat bebas bernilai > 1 dengan derajat bebas yang digunakan adalah $n - k - 1$.

### 2.1.4 Multikolinieritas

Pada analisis regresi keadaan yang menunjukkan adanya hubungan atau korelasi antar peubah disebut dengan multikolinieritas. Ukuran hubungan antara peubah dapat diduga dengan koefisien korelasi yang nilainya berkisar antara -1 sampai dengan +1. Salah satu asumsi yang terdapat dalam analisis regresi adalah non-multikolinieritas.

Menurut Kutner, dkk (2004), salah satu cara untuk mendeteksi multikolinieritas adalah dengan menggunakan VIF (*Variance Inflation Factor*). Semakin tinggi nilai VIF menjelaskan tingginya multikolinieritas yang terjadi. VIF juga dapat mengukur seberapa jauh peningkatan ragam penduga koefisien regresi bila terjadi multikolinieritas. Ketetapan multikolinieritas dengan VIF yaitu jika VIF ≥ 10 maka diindikasikan terdapat multikolinieritas, sedangkan jika VIF < 10 maka diindikasikan tidak terdapat multikolinieritas.

Nilai VIF dapat dihitung dengan rumus:

$$VIF_j = \frac{1}{1-R_j^2} \tag{2.10}$$

di mana $R_j^2$ adalah koefisien determinasi dari model regresi pada peubah prediktor ke-$j$ dan $j$ adalah banyaknya peubah prediktor $(j = 1,2,3, \dots, p)$.

### 2.1.5 Regresi *Generalized Poisson* (GP)

Regresi *Generalized Poisson* (GP) adalah model regresi yang digunakan ketika asumsi *equidispersi* pada regresi Poisson tidak terpenuhi. Menurut Famoye dan Singh (2006), fungsi kepekatan peluang *Generalized Poisson* dapat ditulis seperti persamaan berikut:

$$f(\mu_i, y_i, \omega) = \left(\frac{\mu_i}{1+\omega\mu_i}\right)^{y_i} \frac{(1+\omega y_i)^{y_i-1}}{y_i!} exp\left[\frac{-\mu_i(1+\omega y_i)}{1+\omega\mu_i}\right] \tag{2.11}$$

di mana :

$\omega$ : parameter dispersi
$y_i$ : peubah respon berdistribusi *Generalized Poisson* (GP).

Distribusi *Generalized Poisson* (GP) pertama kali dikenalkan oleh Consul dan Jain (1973) setelah itu dipelajari secara detail oleh Consul (1989). Distribusi *Generalized Poisson* (GP) merupakan perluasan dari distribusi Poisson yang umumnya digunakan untuk mengatasi masalah *overdispersi* atau *underdispersi*. Pada distribusi *Generalized Poisson* (GP) selain parameter μ terdapat $\omega$ sebagai parameter terdispersi. Distribusi tersebut memiliki nilai ragam yang lebih besar atau lebih kecil dari nilai rata-rata. Nilai ragam dan rata-ratanya adalah $\mu_i(1+\omega\mu_i)^2$ dan $\mu_i$. Ketika $\omega = 0$ maka regresi *Generalized Poisson* (GP) akan menjadi regresi Poisson. Jika $\omega > 0$ maka regresi *Generalized Poisson* (GP) terjadi *overdispersi* dan jika $\omega < 0$ maka regresi *Generalized Poisson* (GP) terjadi *underdispersi*. Model regresi *Generalized Poisson* (GP) dapat ditulis sebagai berikut:

$$\ln(\mu_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$$
$$\mu_i = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}) \qquad (2.12)$$

### 2.1.6 *Zero Inflation*

*Zero inflation* adalah nilai nol yang berlebih pada peubah respon. Nilai nol tersebut memiliki arti penting dalam data, maka nilai nol harus dimasukkan dalam analisis. Keadaan *zero inflation* terjadi ketika proporsi nilai nol pada peubah respon mencapai lebih dari 50% (Famoye dkk, 2004). Keadaan tersebut akan menjadi permasalahan ketika menggunakan model regresi Poisson. *Zero inflation* akan menyebabkan terjadinya *overdispersi* dan dapat berakibat pada ketepatan (presisi) dari inferensia.

### 2.1.7 Regresi *Zero Inflated Generalized Poisson* (ZIGP)

Regresi *Zero Inflated Generalized Poisson* (ZIGP) adalah model regresi untuk data diskrit yang mengalami *overdispersi* akibat banyaknya nilai nol pada peubah respon. Menurut Famoye dan Singh (2006), model ZIGP merupakan gabungan dari model *Zero Inflated Poisson* (ZIP) dan *Generalized Poisson* (GP) seperti berikut:

$$P(Y_i = y_i | x_i, z_i) = \begin{cases} \pi_i + (1-\pi_i) f(\mu_i, y_i, \omega), & y_i = 0 \\ (1-\pi_i) f(\mu_i, y_i, \omega), & y_i > 0 \end{cases} \qquad (2.13)$$

di mana $f(\mu_i, y_i, \omega)$ adalah fungsi kepekatan peluang regresi *Generalized Poisson* seperti persamaan (2.11). Dengan $0 < \pi_i < 1$ maka model ZIGP yang diperoleh adalah:

$$P(Y_i = y_i|x_i, z_i) = \begin{cases} \pi_i + (1-\pi_i)exp\left(\frac{-\mu_i}{1+\omega\mu_i}\right), \; y_i = 0 \\ (1-\pi_i)\left(\frac{\mu_i}{1+\omega\mu_i}\right)^{y_i}\frac{(1+\omega y_i)^{y_i-1}}{y_i!}exp\left[\frac{-\mu_i(1+\omega y_i)}{1+\omega\mu_i}\right], \; y_i > 0 \end{cases} \quad (2.14)$$

Rata-rata dan ragam model ZIGP dapat ditulis sebagai berikut secara berturut-turut:

$$E(y_i|x_i) = (1-\pi_i)\mu_i(x_i) \quad (2.15)$$

$$var(y_i|x_i) = (1-\pi_i)[\mu_i{}^2 + \mu_i(1+\omega\pi_i)^2] - (1-\pi_i)^2\mu_i{}^2$$
$$= E(y_i|x_i)[(1+\omega\pi_i)^2 + \pi_i\mu_i] \quad (2.16)$$

dengan

$$\mu_i(x_i) = (1-\omega)\mu_i + \omega y_i \quad (2.17)$$

Dalam fungsi $\mu_i = \mu_i(x_i)$ dan $\pi_i = \pi_i(z_i)$ memenuhi *link function* yang dapat ditulis sebagai berikut:

$$\log(\mu_i) = \sum_{j=1}^{k} x_{ij}\beta_j = \boldsymbol{x^T\beta} \quad (2.18)$$

$$logit(\pi_i) = \log(\pi_i[1-\pi_i])^{-1} = \sum_{j=1}^{m} z_{ij}\delta_j \quad (2.19)$$

di mana:
$x_i$ = $(x_{i1}, x_{i2}, \dots, x_{ik})$ adalah baris ke-*i* dari matriks kovariat **X**
$z_i$ = $(z_{i1}, z_{i2}, \dots, z_{im})$ adalah baris ke-*i* dari matriks kovariat **Z** dengan $z_{i1} = 1$ dan $x_{i1} = 1$
$\boldsymbol{\delta}$ = $(\delta_1, \delta_2, \dots, \delta_m)$ adalah vektor kolom parameter *m*-dimensi
$\boldsymbol{\beta}$ = $(\beta_1, \beta_2, \dots, \beta_k)$ adalah vektor kolom parameter *k*-dimensi.

*Link function* tersebut digunakan untuk menghubungkan $\mu_i$ dengan $x_i$ dan $\pi_i$ dengan $z_i$. Jika matriks kovariat yang sama mempengaruhi $\mu_i$ dan $\pi_i$ maka $\pi_i$ adalah fungsi dari $\mu_i$ yang fungsinya seperti persamaan (2.20). Model regresi ZIGP dengan *logit link* untuk $\mu_i$ dan untuk *logit link* $\pi_i$ dapat dilambangkan dengan ZIGP($\tau$).

$$logit(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = -\tau \sum_{j=1}^{k} x_{ij}\boldsymbol{\beta} = -\tau \boldsymbol{x^T}\boldsymbol{\beta} \qquad (2.20)$$

Model regresi ZIGP akan menjadi model regresi GP ketika $\pi_i = 0$ dan akan menjadi model regresi ZIP ketika $\omega = 0$. Ketika $\tau > 0$ kemungkinan *zero state* terjadi kecil dan ketika $\tau < 0$ kemungkinan *zero state* terjadi lebih besar, sedangkan ketika $\tau = 0$ maka model ZIGP($\tau$) akan berubah menjadi ZIP($\tau$). Model regresi ZIGP terdiri dari model log dan model logit yang disajikan pada persamaan (2.18) dan (2.20).

Fungsi *likelihood* yang digunakan dalam regresi ZIGP adalah sebagai berikut :

$$L(\theta) = \begin{cases} \prod_{i=1}^{n}\left(\frac{1}{1+\mu_i^{-\tau}}\right)\left[\mu_i^{-\tau} + exp\left(\frac{-\mu_i}{1+\omega\mu_i}\right)\right], & y_i = 0 \\ \prod_{i=1}^{n}\left(\frac{1}{1+\mu_i^{-\tau}}\right)\left(\frac{\mu_i}{1+\omega\mu_i}\right)^{y_i}\frac{(1+\omega y_i)^{y_i-1}}{y_i!}exp\left[\frac{-\mu_i(1+\omega y_i)}{1+\omega\mu_i}\right], & y_i > 0 \end{cases} \qquad (2.21)$$

dengan fungsi ln *likelihood* regresi ZIGP adalah :

$$\begin{aligned} ln\,L(\theta) = & -\sum_{i=1}^{n} ln(1+\mu_i^{-\tau}) + \sum_{\substack{i=1 \\ y_i=0}}^{n} ln\left[\mu_i^{-\tau} + exp\left(\frac{-\mu_i}{1+\omega\mu_i}\right)\right] \\ & + \sum_{\substack{i=1 \\ y_i>0}}^{n}\left\{y_i\, ln\left(\frac{\mu_i}{1+\omega\mu_i}\right) + (y_i - 1)\, ln(1+\omega y_i) \right. \\ & \left. - ln(y_i!) - \frac{\mu_i(1+\omega y_i)}{1+\omega\mu_i}\right\} \end{aligned} \qquad (2.22)$$

(Famoye dan Singh, 2006).

### 2.1.8 Metode Bayesian

Dalam analisis data terdapat dua metode untuk pendugaan parameter yaitu metode klasik dan metode Bayesian. Perbedaan yang paling utama dari kedua metode tersebut adalah pada metode Bayesian parameter dipandang sebagai peubah acak yang memiliki distribusi prior. Metode Bayesian menggunakan peluang distribusi untuk menduga parameter. Menurut Walpole (1995), metode

Bayesian adalah suatu metode yang memanfaatkan data sampel dari populasi dengan memperhitungkan suatu distribusi awal. Metode Bayesian merupakan suatu metode yang dapat menggabungkan antara distribusi awal dan informasi dari pengamatan sehingga didapatkan distribusi posterior.

Pada dasarnya distribusi prior tidak mudah ditentukan dan distribusi posterior menjadi sulit diperoleh sehingga metode Bayesian cukup sulit diselesaikan secara analitik. Dalam beberapa penelitian menyebutkan bahwa pendugaan parameter dengan menggunakan metode Bayesian memberikan hasil yang lebih baik jika dibandingkan dengan metode klasik. Hal tersebut karena pada metode Bayesian mempertimbangkan informasi dari distribusi prior, selain itu dibuktikan dengan nilai AIC yang di hasilkan dari metode Bayesian yang sangat kecil.

### 2.1.9 Distribusi Prior dan Posterior

Pada metode Bayesian parameter diperlakukan sebagai peubah acak yang memiliki distribusi, yaitu distribusi prior. Distribusi prior adalah informasi awal yang dibutuhkan untuk membentuk distribusi posterior. Penentuan distribusi prior dalam metode Bayesian sangat penting, pemilihan distribusi prior yang kurang tepat akan berakibat pada distribusi posterior. Perkalian antara distribusi prior dan fungsi *likelihood* akan menghasilkan distribusi posterior. Distribusi posterior ini menyatakan tingkat keyakinan mengenai parameter setelah pengambilan sampel. Menurut Ntzoufras (2009), distribusi posterior dapat ditulis seperti berikut:

$$f(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\boldsymbol{y})} \propto f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}) \tag{2.23}$$

di mana:
$f(\boldsymbol{\theta}|\boldsymbol{y})$ : distribusi posterior
$f(\boldsymbol{y}|\boldsymbol{\theta})$ : fungsi *likelihood*
$f(\boldsymbol{\theta})$ : distribusi prior
$y$ : data
$\theta$ : parameter

Menurut Box dan Tiao (1973) distribusi prior dikelompokkan menjadi dua, yaitu:

1. Berdasarkan bentuk distribusi hasil identifikasi pola data
    a. Distribusi prior sekawan (*conjugate*).

Distribusi prior sekawan atau *conjugate* adalah distribusi prior yang menghasilkan bentuk sama dengan distribusi posterior.

b. Distribusi prior tidak sekawan (*non-conjugate*).

Distribusi prior tidak sekawan atau *non-conjugate* adalah distribusi prior yang tidak sama dengan bentuk distribusi posterior.

2. Berdasarkan dengan penentuan setiap parameter pada pola distribusi prior

a. Distribusi prior informatif

Suatu prior disebut informatif jika didasarkan pada informasi dari studi sebelumnya, pengalaman masa lalu maupun opini seorang ahli.

b. Distribusi prior *non-informatif*

Jika tidak ada pengetahuan mengenai distribusi prior sebelumnya, maka akan digunakan prior *non informatif*.

Pada penelitian regresi ZIGP dengan metode Bayesian sering kali prior yang digunakan adalah prior *non-informatif* dengan menggunakan distribusi normal, sehingga distribusi prior dapat ditulis sebagai berikut:

$$\mathrm{f}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}) = \prod_{\mathrm{j}=0}^{\mathrm{m}} \left[ \frac{1}{\sqrt{2\pi\sigma_{\alpha j}}} e^{\left\{-\frac{(\alpha_j - \mu_{\alpha j})^2}{2\sigma_{\alpha j}^2}\right\}} \right] \times \prod_{j=0}^{l} \left[ \frac{1}{\sqrt{2\pi\sigma_{\beta j}}} e^{\left\{-\frac{(\beta_j - \mu_{\beta j})^2}{2\sigma_{\beta j}^2}\right\}} \right] \times \left[ \frac{1}{\sqrt{2\pi\sigma_{\omega}}} e^{\left\{-\frac{(\omega - \mu_{\omega})^2}{2\sigma_{\omega}^2}\right\}} \right] \tag{2.24}$$

Distribusi posterior merupakan hasil perkalian antara distribusi prior dan fungsi *likelihood* seperti berikut ini:

$$f(\boldsymbol{\theta}|\boldsymbol{y}) = -\sum_{i=1}^{n} ln(1+\mu_i^{-\tau}) + \sum_{\substack{i=1\\ y_i=0}}^{n} ln\left[\mu_i^{-\tau} + exp\left(\frac{-\mu_i}{1+\omega\mu_i}\right)\right]$$

$$+\sum_{\substack{i=1\\ y_i>0}}^{n}\left\{y_i\, ln\left(\frac{\mu_i}{1+\omega\mu_i}\right) + (y_i-1)\, ln(1+\omega y_i)\right.$$

$$\left. - ln(y_i!) - \frac{\mu_i(1+\omega y_i)}{1+\omega\mu_i}\right\}$$

$$\times \prod_{\mathrm{j=0}}^{\mathrm{m}}\left[\frac{1}{\sqrt{2\pi\sigma_{\alpha j}}} e^{\left\{-\frac{(\alpha_j-\mu_{\alpha j})^2}{2\sigma_{\alpha j}^2}\right\}}\right]\times$$

$$\prod_{j=0}^{l}\left[\frac{1}{\sqrt{2\pi\sigma_{\beta j}}} e^{\left\{-\frac{(\beta_j-\mu_{\beta j})^2}{2\sigma_{\beta j}^2}\right\}}\right]\times\left[\frac{1}{\sqrt{2\pi\sigma_{\omega}}} e^{\left\{-\frac{(\omega-\mu_{\omega})^2}{2\sigma_{\omega}^2}\right\}}\right] \quad (2.25)$$

### 2.1.10 *Markov Chain Monte Carlo* (MCMC)

Metode Bayesian tidak dapat diselesaikan secara analitik sehingga perlu dikembangkan dengan teknik simulasi.Teknik simulasi yang digunakan dalam metode Bayesian yaitu teknik simulasi *Markov Chain Monte Carlo* (MCMC). Teknik ini mengkombinasikan antara rantai *Markov* dengan simulai *Monte Carlo*. Menurut Ntzoufraz (2009), MCMC merupakan suatu teknik simulasi untuk mendapatkan data sampel dari suatu distribusi tertentu dengan cara kerjanya menggunakan sifat rantai *Markov*. Dengan teknik tersebut perhitungan distribusi posterior yang kompleks dan rumit pada metode Bayesian dapat diselesaikan dan diperoleh hasil yang akurat. Rantai *Markov* merupakan proses stokastik $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}^{(3)}, \dots \boldsymbol{\theta}^{(R)}\}$ maka dapat dinyatakan dalam persamaan sebagai berikut:

$$f\left(\boldsymbol{\theta}^{(r+1)}\middle|\boldsymbol{\theta}^{(r)} \dots \boldsymbol{\theta}^{(1)}\right) = f(\boldsymbol{\theta}^{(r+1)}|\boldsymbol{\theta}^{(r)}) \quad (2.26)$$

Langkah dalam pembangkitan $f(\boldsymbol{\theta}|\boldsymbol{y})$ pada rantai *Markov* adalah sebagai berikut:

1. Menentukan nilai awal $\boldsymbol{\theta}^{(0)}$.
2. Membangkitkan R sampel hingga mencapai distribusi yang stasioner, dengan R merupakan banyaknya iterasi.
3. Memeriksa nilai kekonvergensi nilai simulasi. Jika hasilnya tidak konvergen, maka perlu menambah banyaknya sampel.
4. Membuang B sampel pertama.
5. Menganggap $\{\boldsymbol{\theta}^{(B+1)}, \boldsymbol{\theta}^{(B+2)}, \dots \boldsymbol{\theta}^{(R)}\}$ sebagai sampel untuk analisis posterior di mana B adalah iterasi awal.
6. Membuat plot distribusi posterior.
7. Mendapatkan ringkasan dari distribusi posterior.

(Ntzoufraz, 2009).

Menurut Ntzoufraz (2009), proses dari MCMC menghasilkan sampel acak sebagai berikut:

$$\boldsymbol{\theta} = \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \boldsymbol{\theta}^{(3)}, \dots \boldsymbol{\theta}^{(R)}$$

Dari sampel yang diperoleh, untuk setiap fungsi $G(\boldsymbol{\theta})$ dan parameter $\boldsymbol{\theta}$ dapat diperoleh:

1. Sampel dari parameter yang diinginkan.
2. Ringkasan posterior untuk $G(\boldsymbol{\theta})$ dari sampel dengan menggunakan pendugaan sampel sederhana.
3. Perhitungan dan pemeriksaan korelasi antar parameter.
4. Plot dari distribusi marginal posterior.

### 2.1.11 Algoritma *Gibbs Sampling*

Salah satu pendekatan *Markov Chain Monte Carlo* (MCMC) adalah dengan metode *Gibbs sampling*. Metode tersebut merupakan algoritma yang sering digunakan dalam melakukan simulasi karena kemudahan prosesnya. *Gibbs sampling* menggunakan sampel sebelumnya untuk membangkitkan nilai sampel berikutnya secara acak sehingga membentuk rantai *Markov* dan menggunakan distribusi bersyarat penuh yang kemudian dihubungkan dengan distribusi posterior. Distribusi bersyarat ini memiliki bentuk yang diketahui, sehingga sejumlah nilai acak dapat disimulasikan dengan mudah menggunakan fungsi standar pada *software* statistika dan komputasi. Menurut Ntzoufraz (2009), *Gibbs sampling* selalu bergerak ke nilai-nilai baru tanpa memperhatikan spesifikasi dari distribusi-distribusi yang diajukan dan jika antar parameter memiliki korelasi tinggi maka algoritma ini tidak dapat digunakan.

Menurut Ntzoufraz (2009), proses algoritma *Gibbs Sampling* dapat dilakukan dengan tahapan sebagai berikut:

1. Menentukan nilai awal untuk setiap parameter

$$\boldsymbol{\theta}^{(0)} = \theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)}$$

   di mana $\boldsymbol{\theta}^{(0)}$ merupakan sembarang nilai sesuai batas ketentuan masing-masing distribusi dan $p$ adalah banyaknya parameter.
2. Proses simulasi setelah nilai awal ditentukan adalah
   $\theta_1^{(r)}$ dari $f_1(\theta_1 | \theta_2^{(r-1)}, \theta_3^{(r-1)}, \ldots, \theta_p^{(r-1)}, \boldsymbol{y})$
   $\theta_2^{(r)}$ dari $f_2(\theta_2 | \theta_1^{(r)}, \theta_3^{(r-1)}, \ldots, \theta_p^{(r-1)}, \boldsymbol{y})$
   $\theta_3^{(r)}$ dari $f_3(\theta_3 | \theta_1^{(r)}, \theta_2^{(r)}, \theta_4^{(r-1)}, \ldots, \theta_p^{(r-1)}, \boldsymbol{y})$
   .
   .
   .
   $\theta_p^{(r)}$ dari $f_p(\theta_p | \theta_1^{(r)}, \theta_2^{(r)}, \theta_3^{(r)}, \ldots, \theta_{p-1}^{(r)}, \boldsymbol{y})$
3. Membentuk $\boldsymbol{\theta}^{(r)}$ dan menyimpannya sebagai himpunan nilai-nilai yang dibangkitkan pada iterasi ke $(r+1)$ dari algoritma.
4. Mendapatkan ringkasan hasil dari distribusi posterior.

*Gibbs sampling* menggunakan *full conditional distribution* yang dihubungkan dengan distribusi posterior. *Full conditional distribution* merupakan distribusi parameter setelah parameter lain dan sampel diamati. Densitas $f_1, f_2, f_3, \ldots, f_p$ disebut sebagai *full conditional distribution*.

### 2.1.12 Pemeriksaan Konvergensi

Pemeriksaan konvergensi memiliki tujuan untuk mengetahui pencapaian distribusi target dari algoritma  yang digunakan pada proses simulasi. Jika konvergen tercapai maka sampel yang dibangkitkan dari proses simulasi sesuai dengan distribusi target yaitu distribusi posterior. Pemeriksaan konvergensi MCMC dapat menggunakan *trace plot*, autokorelasi, dan *MC Error*.

1. *Trace plot*
   *Trace plot* adalah plot antara iterasi dengan nilai yang dibangkitkan. Jika hasil *trace plot* tidak membentuk pola

tertentu maka artinya konvergensi MCMC terpenuhi dan tidak perlu menambah banyaknya iterasi.

2. Autokorelasi
   Mendeteksi konvergensi dengan autokorelasi dapat dilihat dengan mengukur tingkat ketergantungan pada rantai *Markov*. Pada plot autokorelasi dapat dikatakan konvergen jika lag pertama mendekati satu dan lag selanjutnya terus menurun menuju nol.
3. *MC Error*
   Pada perhitungan MC *Error*, sampel yang dibangkitkan dibagi menjadi K *batch*. Jika MC *Error* kurang dari 1% simpangan baku, maka konvergensi MCMC terpenuhi.

### 2.1.13 Pengujian Signifikansi Parameter

Pengujian signifikansi parameter bertujuan untuk mengetahui pengaruh peubah prediktor terhadap peubah respon. Menurut Ntzoufraz (2009), pengujian parameter pada metode Bayesian menggunakan *credible interval*. Dengan *credible interval* 95% menunjukkan nilai batas bawah *percentiles* 2,5% dan nilai batas atas *percentiles* 97,5% dari nilai hasil bangkitan distribusi posterior. Hipotesis yang digunakan adalah:

$$H_0: \alpha_j = 0 \text{ vs } H_1: \alpha_j \neq 0$$

dan

$$H_0: \beta_j = 0 \text{ vs } H_1: \beta_j \neq 0$$
$$j = 1,2,3, \dots, q$$

Untuk kriteria keputusan yaitu tolak $H_0$ jika *Credible Interval* tidak mengandung nilai nol yang artinya bahwa peubah prediktor berpengaruh secara signifikan terhadap peubah respon.

## 2.2 Tinjauan Non Statistika

### 2.2.1 Definisi Pendidikan

Pendidikan merupakan hal yang sangat umum dan tidak asing di dalam suatu kehidupan. Kualitas sumber daya manusia dan tingkat kesejahteraan juga dapat dilihat dari pendidikannya, hal tersebut membuktikan bahwa pendidikan memiliki peranan yang sangat penting dalam kehidupan. Menurut Pramudia (2006), pendidikan adalah proses hominisasi dan proses humanisasi seseorang dalam kehidupan keluarga, masyarakat yang berbudaya kini dan masa depan. Menurut UU No. 20 tahun 2003, pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta ketrampilan yang diperlukan dirinya, masyarakat, bangsa, dan negara.

Pendidikan dapat diperoleh secara formal dan nonformal. Pendidikan formal diperoleh dari program-program yang sudah dirancang secara terstruktur oleh suatu institusi, departemen atau kementrian suatu negara. Pendidikan formal adalah jalur pendidikan yang berjenjang, seperti SD (Sekolah Dasar), SMP (Sekolah Menengah Pertama), dan SMA (Sekolah Menengah Atas). Sedangkan pendidikan nonformal adalah pengetahuan yang didapat manusia (peserta didik) dalam kehidupan sehari-hari (pengalaman) baik yang dapat dirasakan sendiri atau yang dipelajari dari orang lain (mengamati dan mengikuti).

### 2.2.2 Definisi Ujian Nasional

Suatu mutu pendidikan sangatlah penting untuk meningkatkan perubahan dan perkembangan dalam berbagai aspek kehidupan. Dengan mutu pendidikan yang tinggi diharapkan dapat mewujudkan sumber daya manusia Indonesia yang cerdas dan kompetitif. Salah satu cara untuk mengukur mutu suatu pendidikan dapat dilihat dari hasil Ujian Nasional (UN) yang diadakan setiap tahunnya. UN memiliki peranan penting dalam meningkatkan mutu atau kualitas pendidikan. Menurut Rohma (2013), Ujian Nasional digunakan sebagai standarisasi dari pemerintah untuk menguji kelayakan seorang siswa untuk dapat melanjutkan pendidikan ke jenjang yang lebih tinggi dan sebagai pemerataan pendidikan secara nasional. Ujian Nasional juga digunakan sebagai pembanding tingkat pendidikan Indonesia dan negara lain.

Menurut peraturan Menteri Pendidikan Nasional No. 34 tahun 2007 tentang Ujian Nasional tahun pelajaran 2007/2008, pasal 1 ayat 1 menjelaskan bahwa UN adalah kegiatan pengukuran dan penilaian kompetensi peserta didik secara nasional untuk jenjang pendidikan dasar dan menengah. Dalam pasal 2 disebutkan bahwa Ujian Nasional bertujuan untuk menilai pencapaian kompetisi lulusan secara nasional pada mata pelajaran yang ditentukan dari kelompok mata pelajaran ilmu pengetahuan dan teknologi, dalam rangka pencapaian standar nasional pendidikan. Menurut Kementrian Pendidikan dan Kebudayaan Republik Indonesia (2015), hasil Ujian Nasional digunakan sebagai salah satu pertimbangan untuk:

1. Pemetaan mutu program dan/atau satuan pendidikan
2. Dasar seleksi masuk jenjang pendidikan berikutnya
3. Pembinaan dan pemberian bantuan kepada satuan pendidikan dalam upaya untuk meningkatkan mutu pendidikan.

### 2.2.3 Faktor Faktor Yang Mempengaruhi Hasil Ujian Nasional

Hasil dari UN salah satunya digunakan sebagai pertimbangan untuk meningkatkan mutu pendidikan. Dari hasil UN juga dapat menentukan seorang peserta didik dinyatakan lulus atau tidaknya untuk melanjutkan ke jenjang yang lebih tinggi. Kegagalan dalam UN adalah kondisi di mana seorang peserta didik tidak dapat menyelesaikan belajarnya di kelas akhir karena tidak memenuhi kriteria kelulusan yang dibuat oleh pemerintah. Menurut beberapa penelitian yang sudah ada, terdapat beberapa faktor yang mempengaruhi hasil UN. Menurut Suryani (2011), faktor yang menyebabkan ketidaklulusan siswa SMA/SMK dalam UN adalah status akreditasi sekolah, waktu penyelenggaraan, rasio rencana penerimaan pendaftaran, rasio guru-siswa, dan rasio siswa-kelas. Sedangkan menurut Mardiani (2013), faktor yang menyebabkan ketidaklulusan siswa SMA dalam UN adalah peserta ujian nasional, sertifikasi guru, ruang kelas, status sekolah, dan akreditasi sekolah. Dengan mengetahui beberapa faktor yang diduga dapat mempengaruhi hasil UN, diharapkan dapat meminimalkan kegagalan peserta didik dalan UN sehingga tujuan dari UN dapat terpenuhi.

# BAB III
# METODE PENELITIAN

## 3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari Dinas Pendidikan Kota Malang yang beralamat di jalan Veteran No.19, Malang. Data yang digunakan adalah data jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang tahun ajaran 2015/2016. Data dapat dilihat pada Lampiran 1 dan peubah yang digunakan dalam penelitian disajikan pada Tabel 3.1.

Tabel 3.1 Peubah Penelitian

| Peubah | Keterangan |
|---|---|
| Y | Jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang tahun ajaran 2015/2016 |
| $X_1$ | Status sekolah (1=Negeri, 2=Swasta) |
| $X_2$ | Jumlah rombongan belajar (orang) |
| $X_3$ | Rasio Guru-Pegawai |
| $X_4$ | Rasio Guru-Siswa |
| $X_5$ | Rasio rencana penerimaan pendaftaran |

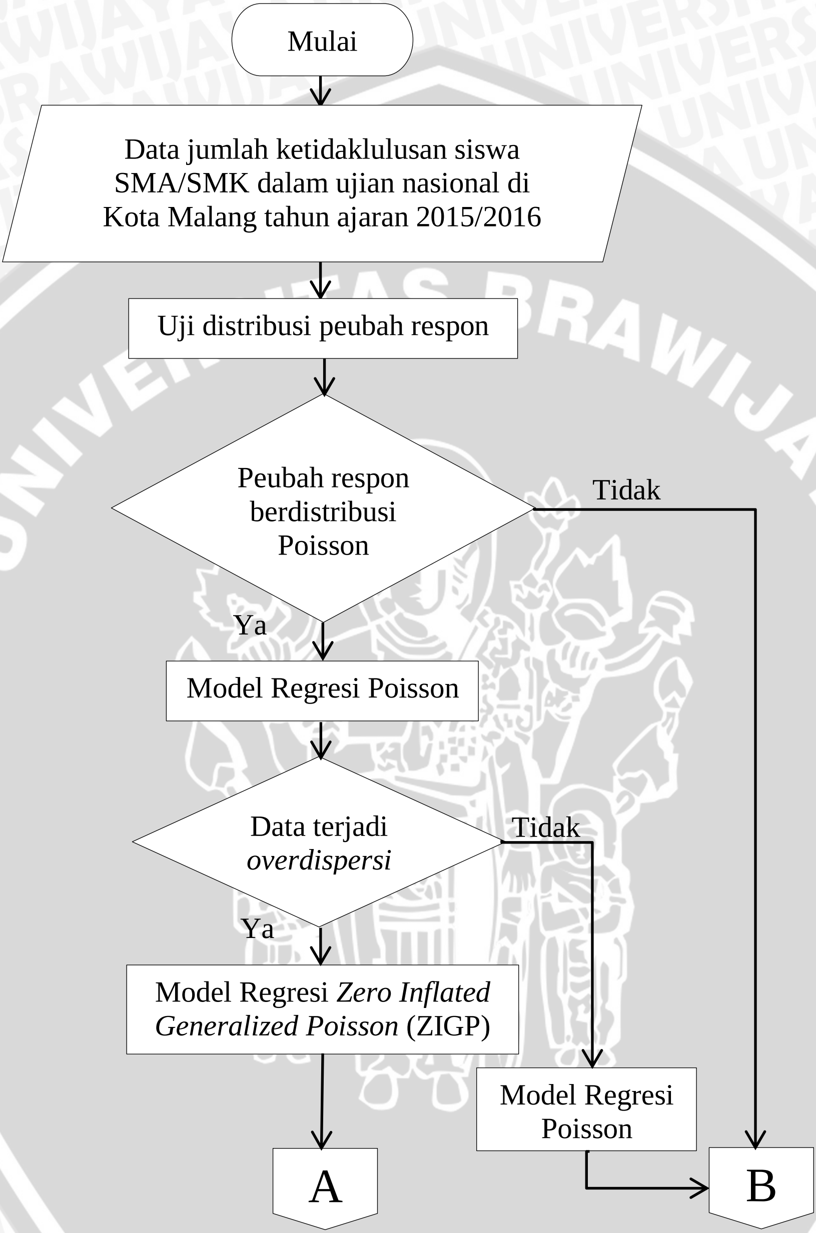
## 3.2 Tahapan Analisis Data

Tahapan untuk pendugaan parameter model regresi *Zero Inflated Generalized Poisson* (ZIGP) dapat dijelaskan sebagai berikut :

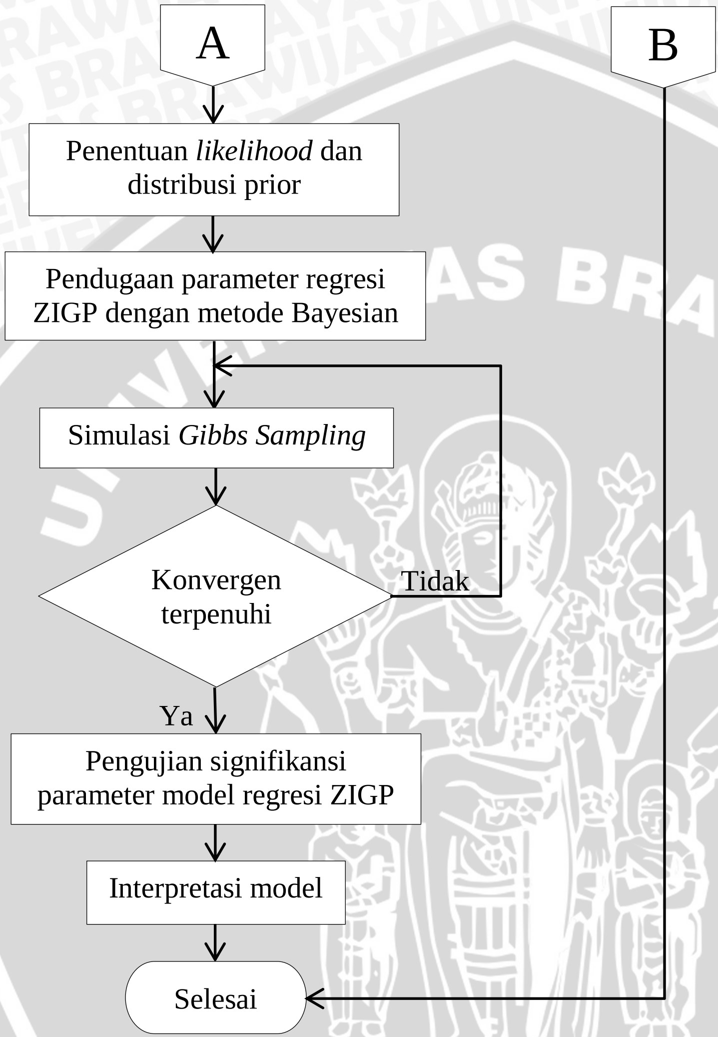
1. Menguji distribusi pada peubah respon (Y) seperti pada persamaan (2.2).
2. Mendeteksi multikolinieritas antar peubah seperti pada persamaan (2.10).
3. Menentukan model regresi Poisson seperti persamaan (2.5).
4. Pemeriksaan *overdispersi* seperti pada persamaan (2.9).
5. Menentukan model umum regresi ZIGP.
6. Menentukan fungsi *likelihood*, distribusi prior, dan distribusi posterior seperti yang dijelaskan pada anak sub bab (2.1.9).
7. Pendugaan parameter regresi ZIGP dengan menggunakan metode Bayesian.
8. Melakukan simulasi *Gibbs Sampling* untuk mendapatkan distribusi posterior.

9. Pemeriksaan konvergensi *Markov Chain Monte Carlo* (MCMC).
10. Melakukan pengujian signifikansi parameter model regresi ZIGP.
11. Menentukan peubah prediktor yang mempengaruhi jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang tahun ajaran 2015/2016.
12. Menyatakan model akhir regresi ZIGP.
13. Interpretasi model regresi ZIGP.

Pada penelitian ini *software* yang digunakan adalah Rstudio, Minitab 17, dan Excel. Diagram alir disajikan pada Gambar 3.1.



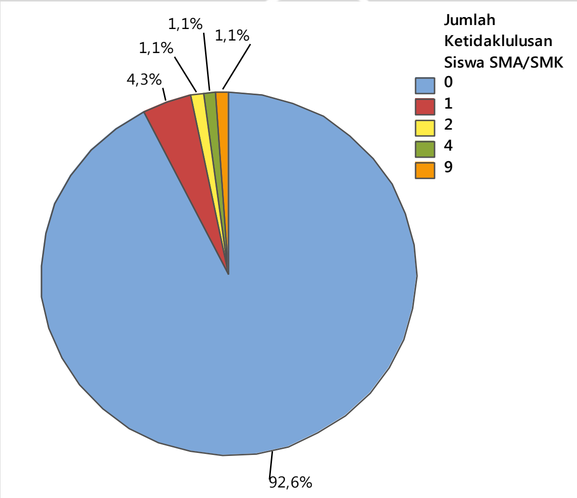

Gambar 3.1. Diagram Alir Pendugaan Parameter Regresi ZIGP Dengan Metode Bayesian


A
B
Penentuan *likelihood* dan distribusi prior
Pendugaan parameter regresi ZIGP dengan metode Bayesian
Simulasi *Gibbs Sampling*
Konvergen terpenuhi
Tidak
Ya
Pengujian signifikansi parameter model regresi ZIGP
Interpretasi model
Selesai

# BAB IV
# HASIL DAN PEMBAHASAN

## 4.1 Pengujian Distribusi Peubah Respon

Kejadian ketidaklulusan siswa SMA/SMK dalam ujian nasional pada setiap sekolah di Kota Malang merupakan kejadian yang jarang terjadi, sehingga data observasi banyak mengandung nilai nol. Pada tahun ajaran 2015/2016 kejadian ketidaklulusan siswa SMA/SMK dalam ujian nasional pada setiap sekolah di Kota Malang dapat dikatakan rendah, hal tersebut juga dapat dilihat dari diagram lingkaran pada Gambar 4.1. Data jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di kota Malang tahun ajaran 2015/2016 dapat dilihat pada Lampiran 1.




Gambar 4.1. Diagram Lingkaran Jumlah Ketidaklulusan Siswa SMA/SMK Dalam Ujian Nasional di Kota Malang

Pada regresi Poisson mengasumsikan peubah respon berdistribusi Poisson. Pengujian distribusi untuk peubah respon dapat menggunakan uji Kolmogorov Smirnov. Hipotesis yang digunakan dalam pengujian distribusi Poisson adalah sebagai berikut :

$H_0$ : $F(x) = F_0(x)$ (Peubah respon mengikuti distribusi Poisson)
$H_1$ : $F(x) \neq F_0(x)$ (Peubah respon tidak mengikuti distribusi Poisson)

Hasil uji Kolmogorov Smirnov pada Lampiran 2 dengan $n = 94$ dan $\alpha = 0.05$ menghasilkan nilai statistik $D = 0.1085$ dan $D_{tabel} = 0.137$. Karena nilai statistik $D < D_{tabel}$ maka keputusan terima $H_0$, sehingga dapat disimpulkan bahwa peubah respon (jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang tahun ajaran 2015/2016) berdistribusi Poisson.

## 4.2 Pemeriksaan *Zero Inflation*

Pemeriksaan *zero inflation* pada peubah respon dapat dilakukan dengan melihat persentase jumlah observasi yang bernilai nol. Jika persentase jumlah observasi yang bernilai nol lebih dari 50% maka dapat dikatakan *zero inflation*. Pada data jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang memiliki data observasi sebanyak 94. Data observasi yang bernilai nol berjumlah 87 dan persentasenya sebesar 92.6%. Karena persentase jumlah observasi yang bernilai nol lebih dari 50% maka dapat disimpulkan bahwa terdapat *zero inflation* pada peubah respon.

## 4.3 Pengujian Multikolinieritas

Pengujian multikolinieritas dapat menggunakan nilai VIF (*Variance Inflation Factor*). Jika nilai VIF ≥ 10 maka diindikasikan terdapat multikolinieritas, sedangkan jika nilai VIF < 10 maka diindikasikan tidak terdapat multikolinieritas. Nilai VIF untuk setiap peubah prediktor dapat dilihat pada Tabel 4.1.

Tabel 4.1. Nilai VIF

| Peubah | VIF |
|---|---|
| X1 | 1.767 |
| X2 | 1.920 |
| X3 | 1.038 |
| X4 | 1.276 |
| X5 | 1.245 |

Dari Tabel 4.1 menunjukkan semua peubah prediktor memiliki nilai VIF < 10, sehingga dapat disimpulkan bahwa tidak terdapat multikolinieritas. Hasil pengujian multikolinieritas secara lengkap dapat dilihat pada Lampiran 3.

## 4.4 Regresi Poisson dan Pengujian *Overdispersi*

Analisis regresi Poisson antara jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang dengan lima peubah prediktor dapat dilihat pada Lampiran 4. Hasil dari analisis tersebut menghasilkan model regresi Poisson seperti persamaan (2.5) berikut ini :

$$\ln(\mu_i) = \widehat{\beta_0} + \widehat{\beta_1}X_1 + \widehat{\beta_2}X_2 + \widehat{\beta_3}X_3 + \widehat{\beta_4}X_4 + \widehat{\beta_5}X_5$$

$$\ln(\mu_i) = -5.923 + 1.628X_1 + 0.043X_2 + 0.093X_3$$
$$-2.959X_4 + 0.129X_5$$

Pada regresi Poisson memiliki asumsi *equidispersi*, tetapi keadaan yang sering terjadi adalah *overdispersi*. Pengujian *overdispersi* dapat dilakukan dengan menggunakan *Pearson's Chi-Square*. Keadaan *overdispersi* terjadi ketika nilai statistik uji *Pearson's Chi-Square* dibagi derajat bebas bernilai > 1. Pengujian *overdispersi* pada Lampiran 5 menghasilkan nilai statistik uji *Pearson's Chi-Square* sebesar 262.66, sehingga didapatkan nilai statistik uji *Pearson's Chi-Square* dibagi derajat bebas sebesar 2.985. Karena nilai statistik uji *Pearson's Chi-Square* dibagi derajat bebas bernilai > 1, maka dapat disimpulkan bahwa terjadi *overdispersi* dan regresi Poisson tidak dapat digunakan. Untuk mengatasi masalah *overdispersi* karena adanya *zero inflation* pada peubah respon maka dapat digunakan regresi *Zero Inflated Generalized Poisson* (ZIGP).

## 4.5 Regresi *Zero Inflated Generalized Poisson* (ZIGP) dengan Metode Bayesian

Pada penelitian ini pendugaan parameter model regresi ZIGP dilakukan menggunakan metode Bayesian dengan simulasi MCMC dan algoritma *gibbs sampling*. Pada metode Bayesian parameter diperlakukan sebagai peubah acak yang memiliki distribusi sendiri, sehingga menghasilkan pendugaan yang lebih baik dari pada metode klasik. Tujuan utama metode Bayesian adalah mendapatkan distribusi posterior yang merupakan hasil perkalian antara distribusi prior dan fungsi *likelihood*. Proses awal pendugaan parameter dengan metode Bayesian yaitu menentukan distribusi prior. Penentuan distribusi prior dalam metode Bayesian cukup sulit ditentukan, sehingga distribusi posterior juga sulit didapatkan. Karena kurangnya informasi mengenai distribusi prior, maka beberapa penelitian sebelumnya sering kali menggunakan prior *non informatif* untuk model regresi ZIGP dengan metode Bayesian. Oleh karena itu, pada penelitian ini prior yang dipilih adalah prior *non informatif*.

Pada penelitian ini menggunakan prior distribusi normal karena distribusi tersebut yang biasa digunakan untuk parameter $\beta$. Parameter regresi ZIGP ditetapkan mengikuti distribusi normal yang memiliki dua parameter, yaitu $\mu$ dan $\sigma^2$. Nilai awal rata-rata dan ragam yang digunakan pada penelitian ini secara berurutan adalah 0 dan 1000. *Source code* untuk regresi ZIGP dengan metode Bayesian secara lengkap dapat dilihat pada Lampiran 6. Pada simulai MCMC algoritma yang digunakan adalah *gibbs sampling*. Simulasi dilakukan dengan iterasi sebanyak 30,000 + 5000 *burn in* + 50 thin. Proses *burn in* bertujuan untuk menghilangkan pengaruh nilai awal. Dari hasil analisis ringkasan distribusi posterior dapat dilihat pada Lampiran 7.

### 4.6 Pemeriksaan Konvergensi

Pada penelitian ini pemeriksaan konvergensi dilihat dari hasil MC *error, trace plot,* dan autokorelasi. Tujuan dari pemeriksaan konvergensi ini adalah untuk mengetahui apakah nilai sampel yang dibangkitkan sesuai dengan distribusi stasioner yaitu distribusi posterior. Pemeriksaan konvergensi menggunakan MC *error* dapat dilihat seperti pada Tabel 4.2.

Tabel 4.2. Konvergensi Simulasi MCMC

| Parameter | sd | sd*1% | MC *error* | Keputusan |
|---|---|---|---|---|
| a0 | 0.0316 | 0.000316 | 0.000188 | Konvergen |
| a1 | 0.0314 | 0.000314 | 0.000196 | Konvergen |
| a2 | 0.0134 | 0.000134 | 0.000076 | Konvergen |
| a3 | 0.0304 | 0.000304 | 0.000193 | Konvergen |
| a4 | 0.0316 | 0.000316 | 0.000207 | Konvergen |
| a5 | 0.0308 | 0.000308 | 0.000197 | Konvergen |
| b0 | 0.0315 | 0.000315 | 0.00019 | Konvergen |
| b1 | 0.0317 | 0.000317 | 0.000228 | Konvergen |
| b2 | 0.0049 | 0.000049 | 0.000032 | Konvergen |

Tabel 4.2. Konvergensi Simulasi MCMC (Lanjutan)

| Parameter | Sd | sd*1% | MC *error* | Keputusan |
|---|---|---|---|---|
| b3 | 0.0296 | 0.000296 | 0.000195 | Konvergen |
| b4 | 0.0317 | 0.000317 | 0.000234 | Konvergen |
| b5 | 0.0295 | 0.000295 | 0.00019 | Konvergen |
| W | 0.0314 | 0.000314 | 0.000195 | Konvergen |

Dari Tabel 4.2 menunjukkan bahwa nilai MC *error* untuk setiap parameter kurang dari 1% simpangan baku, artinya setiap parameter sudah konvergen. Jika dilihat dari *trace plot* pada Lampiran 8 menunjukkan bahwa *trace plot* tidak membentuk pola yang artinya konvergensi sudah terpenuhi. Sedangkan jika dilihat dari plot *Autocorrelation Function* (ACF) pada Lampiran 9, autokorelasi antar parameter rendah sehingga sampel yang dibangkitkan saling bebas dan konvergensi terpenuhi. Dari ketiga cara tersebut maka dapat disimpulkan bahwa konvergensi simulasi MCMC untuk setiap parameter dalam penelitian ini sudah terpenuhi. Ketika konvergensi sudah terpenuhi maka sampel yang dibangkitkan dari proses simulasi berasal dari distribusi posterior yang diinginkan. Pada Lampiran 10 menunjukkan *density plot* yang gambarnya sudah sesuai dengan bentuk distribusi prior yang dipilih, yaitu distribusi normal.

## 4.7 Pengujian Signifikansi Parameter

Pengujian signifikansi parameter pada metode Bayesian menggunakan *credible interval* dengan hipotesis yang digunakan yaitu :

$$H_0\colon \alpha_j = 0 \text{ vs } H_1\colon \alpha_j \neq 0$$

dan

$$H_0\colon \beta_j = 0 \text{ vs } H_1\colon \beta_j \neq 0$$
$$j = 1,2,3,\dots,q$$

Nilai *credible interval* dilihat dari nilai percentil 2.5% dan percentil 97.5%. Jika dalam rentang batas tersebut tidak mengandung nilai nol, maka keputusan tolak $H_0$ yang artinya parameter berpengaruh secara signifikan. Nilai *credible interval* untuk setiap parameter dapat dilihat pada Tabel 4.3.

Tabel 4.3. *Credible Interval*

| Parameter | Percentil 2.5% | Percentil 97.5% | Keputusan |
|---|---|---|---|
| a0 | -0.0505 | 0.0726 | Terima $H_0$ |
| a1 | -0.0522 | 0.0701 | Terima $H_0$ |
| a2 | 0.0117 | 0.0643 | Tolak $H_0$ |
| a3 | -0.0299 | 0.0887 | Terima $H_0$ |
| a4 | -0.0603 | 0.0635 | Terima $H_0$ |
| a5 | -0.0425 | 0.0779 | Terima $H_0$ |
| b0 | -0.0691 | 0.0539 | Terima $H_0$ |
| b1 | -0.0677 | 0.0567 | Terima $H_0$ |
| b2 | 0.0059 | 0.0254 | Tolak $H_0$ |
| b3 | -0.0782 | 0.0381 | Terima $H_0$ |
| b4 | -0.0632 | 0.0605 | Terima $H_0$ |
| b5 | -0.0781 | 0.0372 | Terima $H_0$ |
| W | -0.0609 | 0.0613 | Terima $H_0$ |

Dari Tabel 4.3 menunjukkan bahwa parameter a2 dan b2 yang berpengaruh secara signifikan terhadap model. Jadi peubah prediktor yang mempengaruhi jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang tahun ajaran 2015/2016 adalah jumlah rombongan belajar (X2).

## 4.8 Interpretasi Model Regresi *Zero Inflated Generalized Poisson* (ZIGP)

Model regresi ZIGP terdiri dari model log dan model logit. Model tersebut didapatkan setelah proses simulasi mencapai konvergen dan signifikan. Nilai duga parameter model regresi ZIGP dapat dilihat pada Tabel 4.4.

Tabel 4.4. Nilai Hasil Pendugaan Parameter

| Parameter | Koefisien Parameter |
|---|---|
| a0 | 0.0112 |
| a1 | 0.0092 |
| a2 | 0.0364 |

Tabel 4.4. Nilai Hasil Pendugaan Parameter (Lanjutan)

| Parameter | Koefisien Parameter |
|---|---|
| a3 | 0.0295 |
| a4 | 0.0019 |
| a5 | 0.0172 |
| b0 | -0.0074 |
| b1 | -0.0057 |
| b2 | 0.0165 |
| b3 | -0.0198 |
| b4 | -0.0010 |
| b5 | -0.0198 |
| W | 0.0005 |

Model regresi ZIGP yang terbentuk dari metode Bayesian seperti persamaan (2.18) dan (2.20) yaitu :

$$\log(\mu) = -0.0074 - 0.0057X_1 + 0.0165X_2 - 0.0198X_3 \\ -0.0010X_4 - 0.0198X_5$$

$$\text{logit}(\pi) = 0.0112 + 0.0092X_1 + 0.0364X_2 + 0.0295X_3 \\ +0.0019X_4 + 0.0127X_5$$

Pada model log dan model logit, peubah prediktor yang berpengaruh secara signifikan adalah jumlah rombongan belajar (X2). Interpretasi dari regresi ZIGP adalah:

1. Model log menjelaskan bahwa untuk setiap peningkatan 1 orang siswa dalam rombongan belajar maka akan meningkatkan banyak siswa SMA/SMK yang tidak lulus dalam ujian nasional di Kota Malang sebesar $(1 - \exp(0.0165)) \times 100\% = 1.7\%$ dibandingkan sebelumnya dengan asumsi nilai peubah prediktor yang lain dianggap tetap.
2. Model logit menjelaskan bahwa untuk setiap peningkatan 1 orang siswa dalam rombongan belajar maka akan meningkatkan kemungkinan terjadinya ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang sebesar

exp(0.0364)=1.037 kali dengan asumsi nilai peubah prediktor yang lain dianggap tetap.

# BAB V
# PENUTUP

## 5.1 Kesimpulan

Berdasarkan hasil analisis regresi ZIGP yang sudah dilakukan maka kesimpulan yang didapatkan adalah:

1. Model regresi ZIGP dengan metode Bayesian:

$$\log(\mu) = -0.0074 - 0.0057X_1 + 0.0165X_2 - 0.0198X_3 -0.0010X_4 - 0.0198X_5$$

$$\text{logit}(\pi) = 0.0112 + 0.0092X_1 + 0.0364X_2 + 0.0295X_3 +0.0019X_4 + 0.0127X_5$$

2. Jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional di Kota Malang tahun ajaran 2015/2016 dipengaruhi secara signifikan oleh jumlah siswa dalam rombongan belajar.

## 5.2 Saran

1. Bagi peneliti selanjutnya, disarankan dapat menggunakan regresi ZIGP metode Bayesian dengan membandingkan berbagai macam distribusi prior untuk mendapatkan hasil posterior yang terbaik.
2. Dalam rangka meminimumkan jumlah ketidaklulusan siswa SMA/SMK dalam ujian nasional maka sebaiknya setiap sekolah menerapkan jumlah rombongan belajar yang ideal, yaitu 32 siswa dalam satu kelas.

# DAFTAR PUSTAKA

Agresti, A. 2007. *An Introduction to Categorical Data Analysis Second Edition*. John Wiley & Sons. New York.

Box, G.E.P. dan Tiao, G.C. 1973. *Bayesian Inference In Statistical Analysis*. Addision-Wesley Publishing Company, Inc. Madison.

Consul, P. 1989. *Generalized Poisson Distributions Properties and Ap-plications*. Marcel Dekker, Inc. New York.

Consul, P. dan Jain, G. 1973. *A generalization of the Poisson distribu-tion*. *Technometrics*.

Daniel, W.W. 1989. *Statistika NonParametrik Terapan*. PT Gramedia. Jakarta.

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).

Famoye, F., dan Singh, K. P. 2006. *Zero Inflated Generalized Poisson Regression Model With an Application to Domestic Violence Data*. Journal of Data Science 4 (2006).

Faridhoh, B.U. 2015. *Pemodelan Ketidaklulusan Siswa SMA Dalam Ujian Nasional di Kota Malang Menggunakan Regresi Zero Inflated Poisson dan Regresi Zero Inflated Generalized Poisson*. Universitas Brawijaya. Malang.

Kementerian Pendidikan dan Kebudayaan Republik Indonesia. 2015. *Kebijakan Perubahan Ujian Nasional* [online]. Tersedia: https://mustafatope.files.wordpress.com/2015/01/ujian-nasional-2015-v0-4.pdf. [Diakses pada tanggal 16 November 2016].

Kutner, M.H., Nachtsheim, C.J. dan Neter, J. 2004. *Applied Liniar Regression Models*,. Mc. Graw-Hill Companies, Inc. New York.

Mardiani, LK., Sukarsa, I.K.G dan Srinadi, I.G.A.M. 2013. *Penerapan Regresi Zero Inflated Poisson Untuk Mengatasi Overdispersi Pada Regresi Poisson*. E-Jurnal Matematika.

Ntzoufras, I. 2009. *Bayesian Modelling Using WinBUGS*. John Wiley & Sons, Inc. New Jersey.

Pereira, F. 1999. *Practical "Modern" Bayesian Statistics In Actuarial Science*. General Insurance Convention. City University. London.

Pramudia, J.R. 2006. *Orientasi Baru Pendidikan*. Jurnal Pendidikan Luar Sekolah, Universitas Pendidikan Indonesia.

Rohma, F. 2013. *Kecurangan Dalam Ujian Nasional Di Sekolah Menengah Atas Cheating On Nasional Exam In Senior High School*. Artikel Ilmiah, Universitas Jember (UNEJ). Jember.

Suryani, N. 2011. *Penerapan Model Regresi Zero Inflated Generalized Poisson (ZIGP) Pada Banyaknya Siswa Gagal Ujian Nasional di SMA/SMK Kabupaten Malang*. Universitas Brawijaya. Malang.

Walpole, R.E. 1995. *Pengantar Statistika Edisi 3*. PT Gramedia Pustaka Utama. Jakarta.

# LAMPIRAN

Lampiran 1. Data Jumlah Ketidaklulusan Siswa SMA/SMK Dalam Ujian Nasional di Kota Malang Tahun Ajaran 2015/2016

| No | Nama Sekolah | Y | X1 | X2 |
|---|---|---|---|---|
| 1 | SMA NEGERI 1 MALANG | 0 | 1 | 33 |
| 2 | SMA NEGERI 2 MALANG | 0 | 1 | 22 |
| 3 | SMA NEGERI 3 MALANG | 0 | 1 | 28 |
| 4 | SMA NEGERI 4 MALANG | 0 | 1 | 30 |
| 5 | SMA NEGERI 5 MALANG | 0 | 1 | 32 |
| 6 | SMA NEGERI 6 MALANG | 0 | 1 | 28 |
| 7 | SMA NEGERI 7 MALANG | 0 | 1 | 53 |
| 8 | SMA NEGERI 8 MALANG | 0 | 1 | 35 |
| 9 | SMA NEGERI 9 MALANG | 0 | 1 | 31 |
| 10 | SMA NEGERI 10 MALANG | 0 | 1 | 31 |
| 11 | SMA ADVENT DWI ABDI MALANG | 0 | 2 | 5 |
| 12 | SMA ARDJUNA MALANG | 0 | 2 | 3 |
| 13 | SMA BAHRUL MAGHFIROH | 0 | 2 | 5 |
| 14 | SMA BRAWIJAYA SMART SCHOOL (BSS) | 0 | 2 | 18 |
| 15 | SMA CHARIS | 0 | 2 | 21 |
| 16 | SMA DARUL ULUM AGUNG | 0 | 2 | 4 |
| 17 | SMA ISLAM BAITURROHMAH | 0 | 2 | 3 |
| 18 | SMA 'ISLAM' TERAKREDITASI A | 0 | 2 | 28 |
| 19 | SMA KATOLIK COR JESU | 1 | 2 | 15 |
| 20 | SMA KATOLIK FRATERAN MALANG | 1 | 2 | 13 |
| 21 | SMA KATOLIK KOLESE SANTO YUSUP MALANG | 0 | 2 | 35 |
| 22 | SMA KATOLIK SANTA MARIA MALANG | 0 | 2 | 18 |

Lampiran 1. (Lanjutan)

| No | Nama Sekolah | Y | X1 | X2 |
|---|---|---|---|---|
| 23 | SMA KATOLIK ST. ALBERTUS MALANG | 9 | 2 | 34 |
| 24 | SMA KERTANEGARA MALANG | 0 | 2 | 4 |
| 25 | SMA KR. KALAM KUDUS MALANG | 0 | 2 | 9 |
| 26 | SMA KRISTEN 2 YPK JATIM MALANG | 0 | 2 | 3 |
| 27 | SMA KRISTEN PETRA MALANG | 0 | 2 | 6 |
| 28 | SMA KRISTEN SETIA BUDI MALANG | 0 | 2 | 5 |
| 29 | SMA LABORATORIUM UM | 0 | 2 | 29 |
| 30 | SMA MUHAMMADIYAH 1 MALANG | 0 | 2 | 7 |
| 31 | SMA NASIONAL MALANG | 0 | 2 | 10 |
| 32 | SMA NURUL MUTTAQIN ALBAROKHAH | 0 | 2 | 6 |
| 33 | SMA PANJURA | 0 | 2 | 18 |
| 34 | SMA PGRI 6 MALANG | 0 | 2 | 6 |
| 35 | SMA SHALAHUDDIN TERAKREDITASI A MALANG | 0 | 2 | 6 |
| 36 | SMA SURYA BUANA | 0 | 2 | 3 |
| 37 | SMA TAMAN HARAPAN MALANG | 0 | 2 | 6 |
| 38 | SMA TAMAN MADYA MALANG | 0 | 2 | 7 |
| 39 | SMA WAHID HASYIM MALANG | 0 | 2 | 4 |
| 40 | SMA WASKITA DHARMA | 0 | 2 | 3 |
| 41 | SMA WIDYA GAMA | 0 | 2 | 6 |
| 42 | SMA WISNUWARDHANA MALANG | 0 | 2 | 3 |
| 43 | SMK NEGERI 1 MALANG | 0 | 1 | 53 |
| 44 | SMK NEGERI 2 MALANG | 0 | 1 | 50 |
| 45 | SMK NEGERI 3 MALANG | 0 | 1 | 37 |
| 46 | SMK NEGERI 4 MALANG | 0 | 1 | 83 |
| 47 | SMK NEGERI 5 MALANG | 0 | 1 | 60 |
| 48 | SMK NEGERI 6 MALANG | 0 | 1 | 78 |

Lampiran 1. (Lanjutan)

| No | Nama Sekolah | Y | X1 | X2 |
|---|---|---|---|---|
| 49 | SMK NEGERI 7 MALANG | 0 | 1 | 24 |
| 50 | SMK NEGERI 8 MALANG | 0 | 1 | 29 |
| 51 | SMK NEGERI 9 MALANG | 1 | 1 | 28 |
| 52 | SMK NEGERI 10 MALANG | 0 | 1 | 41 |
| 53 | SMK NEGERI 11 MALANG | 0 | 1 | 36 |
| 54 | SMK NEGERI 12 MALANG | 2 | 1 | 40 |
| 55 | SMK NEGERI 13 MALANG | 0 | 1 | 14 |
| 56 | SMK ARJUNA 1 MALANG | 0 | 2 | 10 |
| 57 | SMK ARJUNA 2 MALANG | 0 | 2 | 9 |
| 58 | SMK BAITUL MAKMUR | 0 | 2 | 8 |
| 59 | SMK BHAKTI LUHUR MALANG | 0 | 2 | 5 |
| 60 | SMK BINA BANGSA MALANG | 0 | 2 | 6 |
| 61 | SMK BINA CENDIKA MALANG | 0 | 2 | 5 |
| 62 | SMK BINA MANDIRI | 0 | 2 | 9 |
| 63 | SMK COR JESU MALANG | 0 | 2 | 9 |
| 64 | SMK EL-HAYAT | 0 | 2 | 3 |
| 65 | SMK FARMASI MAHARANI | 0 | 2 | 5 |
| 66 | SMK GRAF. KARYA NASIONAL | 0 | 2 | 6 |
| 67 | SMK KARTIKA IV-1 MALANG | 0 | 2 | 13 |
| 68 | SMK KERTAWIISATA MALANG | 0 | 2 | 3 |
| 69 | SMK KESEHATAN ADI HUSADA | 0 | 2 | 13 |
| 70 | SMK MUHAMMADIYAH 2 MALANG | 0 | 2 | 17 |
| 71 | SMK MUHAMMADYAH 1 KOTA MALANG | 0 | 2 | 6 |
| 72 | SMK MUHAMMADYAH 3 KOTA MALANG | 1 | 2 | 3 |
| 73 | SMK NASIONAL MALANG | 0 | 2 | 33 |
| 74 | SMK P U MALANG | 0 | 2 | 15 |
| 75 | SMK PETRA YPK | 0 | 2 | 6 |

Lampiran 1. (Lanjutan)

| No | Nama Sekolah | Y | X1 | X2 |
|---|---|---|---|---|
| 76 | SMK PGRI 2 MALANG | 0 | 2 | 17 |
| 77 | SMK PGRI 3 MALANG | 4 | 2 | 95 |
| 78 | SMK PGRI 6 MALANG | 0 | 2 | 14 |
| 79 | SMK PGRI 7 SINGHASARI | 0 | 2 | 15 |
| 80 | SMK PRAJNAPARAMITA MALANG | 0 | 2 | 12 |
| 81 | SMK PUTRA INDONESIA KOTA MALANG | 0 | 2 | 15 |
| 82 | SMK SHALAHUDDIN 2 | 0 | 2 | 5 |
| 83 | SMK SHALAHUDDIN MALANG | 0 | 2 | 4 |
| 84 | SMK SRI WEDARI MALANG | 0 | 2 | 5 |
| 85 | SMK TARUNA BHAKTI | 0 | 2 | 3 |
| 86 | SMK TELKOM SANDHY PUTRA | 0 | 2 | 28 |
| 87 | SMK TUMAPEL MALANG | 0 | 2 | 5 |
| 88 | SMK TUNAS BANGSA KOTA MALANG | 0 | 2 | 10 |
| 89 | SMK WASKITA DHARMA | 0 | 2 | 1 |
| 90 | SMK WIDYAGAMA MALANG | 0 | 2 | 9 |
| 91 | SMK WISNUWARDHANA MALANG | 0 | 2 | 9 |
| 92 | SMK YP 17-1 MALANG | 0 | 2 | 12 |
| 93 | SMK YP 17-2 MALANG | 0 | 2 | 6 |
| 94 | SMKS KESEHATAN KENDEDES | 0 | 2 | 5 |

Lampiran 1. (Lanjutan)

| No | X3 | X4 | X5 |
|---|---|---|---|
| 1 | 2.739 | 0.068 | 0.290 |
| 2 | 3.278 | 0.063 | 1.000 |
| 3 | 2.148 | 0.069 | 0.335 |
| 4 | 3.238 | 0.077 | 0.211 |
| 5 | 2.429 | 0.069 | 1.000 |
| 6 | 3.050 | 0.068 | 1.033 |
| 7 | 2.880 | 0.066 | 1.017 |
| 8 | 2.680 | 0.070 | 1.190 |
| 9 | 4.833 | 0.065 | 1.047 |
| 10 | 1.786 | 0.058 | 0.775 |
| 11 | 8.000 | 0.133 | 2.000 |
| 12 | 5.000 | 0.200 | 5.556 |
| 13 | 0.000 | 0.039 | 1.400 |
| 14 | 3.250 | 0.085 | 1.007 |
| 15 | 5.000 | 0.117 | 0.833 |
| 16 | 0.000 | 0.071 | 1.250 |
| 17 | 4.500 | 0.184 | 2.857 |
| 18 | 3.273 | 0.036 | 1.000 |
| 19 | 5.500 | 0.069 | 1.037 |
| 20 | 2.364 | 0.070 | 1.111 |
| 21 | 3.750 | 0.053 | 0.809 |
| 22 | 2.583 | 0.049 | 0.766 |
| 23 | 3.273 | 0.068 | 0.923 |
| 24 | 3.000 | 0.160 | 1.905 |
| 25 | 0.000 | 0.062 | 1.190 |
| 26 | 2.200 | 0.478 | 6.000 |
| 27 | 6.500 | 0.140 | 1.875 |
| 28 | 0.000 | 0.156 | 1.154 |

Lampiran 1. (Lanjutan)

| No | X3 | X4 | X5 |
|---|---|---|---|
| 29 | 2.882 | 0.055 | 0.627 |
| 30 | 2.667 | 0.118 | 2.069 |
| 31 | 6.000 | 0.068 | 0.571 |
| 32 | 0.000 | 0.063 | 1.351 |
| 33 | 5.000 | 0.063 | 1.000 |
| 34 | 8.500 | 0.202 | 2.308 |
| 35 | 1.889 | 0.179 | 2.308 |
| 36 | 0.000 | 0.155 | 1.667 |
| 37 | 1.417 | 0.126 | 3.158 |
| 38 | 9.000 | 0.235 | 6.122 |
| 39 | 0.000 | 0.571 | 5.357 |
| 40 | 7.000 | 0.304 | 5.143 |
| 41 | 1.000 | 0.081 | 1.000 |
| 42 | 0.000 | 0.214 | 1.333 |
| 43 | 3.121 | 0.058 | 0.761 |
| 44 | 2.610 | 0.062 | 1.000 |
| 45 | 3.630 | 0.089 | 1.094 |
| 46 | 3.111 | 0.050 | 0.806 |
| 47 | 4.231 | 0.056 | 0.972 |
| 48 | 3.326 | 0.058 | 0.681 |
| 49 | 3.385 | 0.059 | 1.071 |
| 50 | 2.500 | 0.064 | 1.115 |
| 51 | 3.647 | 0.068 | 0.945 |
| 52 | 2.800 | 0.043 | 1.000 |
| 53 | 2.909 | 0.052 | 0.907 |
| 54 | 8.714 | 0.045 | 0.927 |
| 55 | 2.667 | 0.086 | 1.111 |
| 56 | 3.800 | 0.186 | 1.000 |

Lampiran 1. (Lanjutan)

| No | X3 | X4 | X5 |
|---|---|---|---|
| 57 | 2.333 | 0.072 | 1.639 |
| 58 | 0.000 | 0.194 | 1.250 |
| 59 | 5.500 | 0.063 | 0.889 |
| 60 | 0.000 | 0.188 | 4.286 |
| 61 | 3.500 | 0.139 | 2.162 |
| 62 | 2.667 | 0.115 | 0.752 |
| 63 | 2.833 | 0.068 | 0.972 |
| 64 | 3.000 | 0.085 | 1.176 |
| 65 | 1.333 | 0.099 | 1.132 |
| 66 | 9.000 | 0.118 | 3.810 |
| 67 | 5.500 | 0.079 | 1.630 |
| 68 | 0.000 | 0.091 | 1.000 |
| 69 | 9.000 | 0.094 | 1.000 |
| 70 | 5.400 | 0.189 | 1.882 |
| 71 | 2.857 | 0.109 | 1.961 |
| 72 | 0.000 | 0.455 | 12.500 |
| 73 | 3.563 | 0.057 | 1.124 |
| 74 | 6.000 | 0.070 | 1.417 |
| 75 | 0.000 | 0.143 | 1.000 |
| 76 | 3.000 | 0.042 | 1.798 |
| 77 | 2.861 | 0.036 | 0.794 |
| 78 | 1.000 | 0.021 | 1.220 |
| 79 | 6.500 | 0.029 | 1.149 |
| 80 | 0.000 | 0.109 | 3.571 |
| 81 | 5.333 | 0.088 | 1.579 |
| 82 | 0.000 | 0.067 | 5.625 |
| 83 | 4.500 | 0.196 | 3.000 |
| 84 | 0.000 | 0.167 | 1.212 |

Lampiran 1. (Lanjutan)

| No | X3 | X4 | X5 |
|---|---|---|---|
| 85 | 0.000 | 0.128 | 12.308 |
| 86 | 6.286 | 0.041 | 0.331 |
| 87 | 3.000 | 0.074 | 1.063 |
| 88 | 8.500 | 0.148 | 1.471 |
| 89 | 0.000 | 1.000 | 0.000 |
| 90 | 7.500 | 0.104 | 1.000 |
| 91 | 8.500 | 0.126 | 1.087 |
| 92 | 3.800 | 0.133 | 1.724 |
| 93 | 0.000 | 0.161 | 3.077 |
| 94 | 7.500 | 0.113 | 1.250 |

Keterangan :
Y = jumlah ketidaklulusan siswa dalam Ujian Nasional (orang)
X1 = status sekolah (1=Negeri, 2=Swasta)
X2 = jumlah rombongan belajar (orang)
X3 = rasio Guru-Pegawai
X4 = rasio Guru-Siswa
X5 = rasio rencana penerimaan pendaftaran

Lampiran 2. Uji Kolmogorov Smirnov

```
> #input data
> data=read.table("D:\\DATA KETIDAKLULUSAN TAHUN 2016.
txt",header=TRUE)
> Y=data[,1]
> X1=data[,2]
> X2=data[,3]
> X3=data[,4]
> X4=data[,5]
> X5=data[,6]
> a=cbind(Y,X1,X2,X3,X4,X5)
> b=data.frame(a)

> #uji kolmogorov smirnov
> K=function(Y){
+   q=sort(unique(Y),decreasing=FALSE)
+   F1=ppois(q,mean(Y),lower.tail = TRUE)
+   frek=as.matrix(table(Y))
+   F0=cumsum(frek)/length(Y)
+   D=max(abs(F0-F1))
+ }
> c=K(Y)
> c
[1] 0.1085413
```

Lampiran 3. Pengujian Multikolinieritas

```
> #Input Data
> data=read.table("D:\\DATA KETIDAKLULUSAN TAHUN 2016.
txt", header=TRUE)
> Y=data[,1]
> X1=data[,2]
> X2=data[,3]
> X3=data[,4]
> X4=data[,5]
> X5=data[,6]
> a=cbind(Y,X1,X2,X3,X4,X5)
> b=data.frame(a)

> #pengujian multikolinieritas
> model=lm(Y~X1+X2+X3+X4+X5,data=b)
> summary(model)

Call:

lm(formula = Y ~ X1 + X2 + X3 + X4 + X5, data = b)
Residuals:
    Min      1Q  Median      3Q     Max
-1.4707 -0.2352 -0.0779  0.0105  8.1152

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.976677   0.671415  -2.944 0.004143 **
X1           0.904013   0.313764   2.881 0.004976 **
X2           0.030489   0.007708   3.956 0.000154 ***
X3          -0.003407   0.040878  -0.083 0.933765
X4           0.106912   0.900780   0.119 0.905792
X5           0.022398   0.056067   0.399 0.690505
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
Residual standard error: 0.9839 on 88 degrees of freedom
Multiple R-squared: 0.1579,      Adjusted R-squared: 0.11
F-statistic: 3.299 on 5 and 88 DF,  p-value: 0.008891

> library(car)
> vif=vif(model)
> vif
      X1       X2       X3       X4       X5
1.766648 1.920082 1.038006 1.275986 1.244988
```

Lampiran 4. Model Regresi Poisson

```
> #input data
> data=read.table("D:\\DATA KETIDAKLULUSAN TAHUN 2016.
txt",header=TRUE)
> Y=data[,1]
> X1=data[,2]
> X2=data[,3]
> X3=data[,4]
> X4=data[,5]
> X5=data[,6]
> a=cbind(Y,X1,X2,X3,X4,X5)
> b=data.frame(a)

> #Regresi Poisson
> library(car)
> model1=glm(formula=Y~X1+X2+X3+X4+X5, family="poisson",
data=b)
> summary(model1)

Call:
glm(formula = Y ~ X1 + X2 + X3 + X4 + X5, family =
"poisson",data = b)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1116  -0.5959  -0.4576  -0.3615   6.3243

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.923248   1.411693  -4.196 2.72e-05 ***
X1           1.627523   0.639732   2.544    0.011 *
X2           0.042986   0.007631   5.633 1.77e-08 ***
X3           0.093144   0.112218   0.830    0.407
X4          -2.959685   4.935407  -0.600    0.549
X5           0.129790   0.160411   0.809    0.418
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
(Dispersion parameter for poisson family taken to be 1)
    Null deviance: 114.170  on 93  degrees of freedom
Residual deviance:  82.049  on 88  degrees of freedom
AIC: 111.98
Number of Fisher Scoring iterations: 6
```

Lampiran 5. Pengujian *Overdispersi*

```
> #Pengujian Overdispersi
> overdisper=function(b,model1){
+   pearson_chisquare=sum(residuals(model1,type="pearson
")^2)
+   n=nrow(b)
+   k=ncol(b)-1
+   critvalue=qchisq(0.05,n-k-1, lower.tail=F)
+   pvalue=1-pchisq(pearson_chisquare,n-k-1)
+   cbind(pearson_chisquare,critvalue,pvalue)
+ }
> d=overdisper(b,model1)
> d
     pearson_chisquare critvalue pvalue
[1,]            262.66   110.898      0
```

Lampiran 6. *Source Code* R Untuk Regresi ZIGP dengan Metode Bayesian

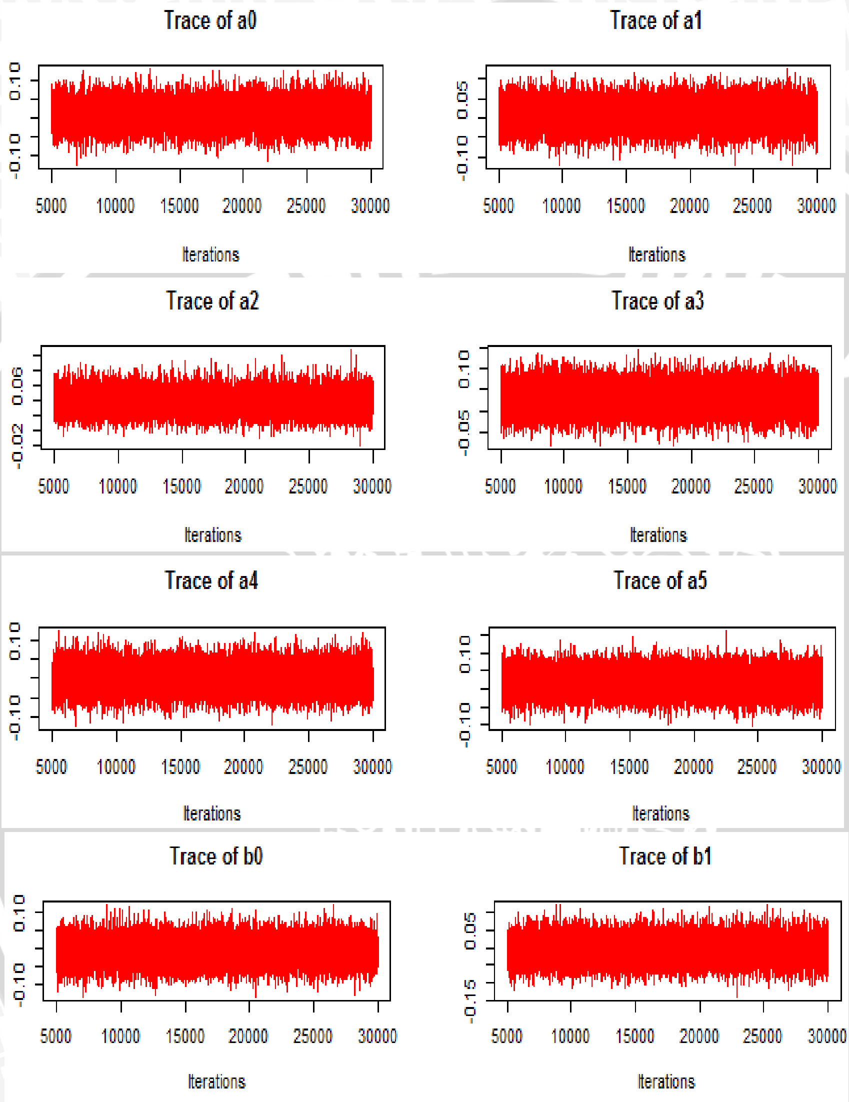
```
> #Input Data
> data=read.table("D:\\DATA KETIDAKLULUSAN TAHUN 2016.
txt",header=TRUE)

> #Input data dari directory
> data=as.list(data)
> data$N=nrow(data)

> #Regresi ZIGP Bayesian
> library(R2OpenBUGS)
> model=function(){
   for(i in 1:94){
     D.X1[i]<-equals(X1[i],2)
     Y[i]~dpois(miu.star[i])
     miu.star[i]<-((1-omega[i])*miu[i]+omega[i]*Y[i])*
     (1-u[i])
     omega.cont[i]~dunif(0,1)
     omega[i]<-trunc(omega.cont[i])
     u[i]~dbern(p[i])
logit(p[i])<-a0+a1*D.X1[i]+a2*X2[i]+a3*X3[i]+a4*X4[i]+
a5*X5[i]
log(miu[i])<-b0+b1*D.X1[i]+b2*X2[i]+b3*X3[i]+b4*X4[i]+
b5*X5[i]
}
 a0~dnorm(0,1000)
 a1~dnorm(0,1000)
 a2~dnorm(0,1000)
 a3~dnorm(0,1000)
 a4~dnorm(0,1000)
 a5~dnorm(0,1000)
 b0~dnorm(0,1000)
 b1~dnorm(0,1000)
 b2~dnorm(0,1000)
 b3~dnorm(0,1000)
 b4~dnorm(0,1000)
 b5~dnorm(0,1000)
 w~dnorm(0,1000)
 }
> inits=function(){list(a0=0,a1=0,a2=0,a3=0,a4=0,a5=0,
b0=0,b1=0,b2=0,b3=0,b4=0,b5=0, w=0)}
> sim=bugs(data=data,inits=inits,parameters.to.save=c
("a0","a1","a2","a3","a4","a5","b0","b1","b2","b3","b4",
"b5",”w”),model.file=model,n.chains=1,n.thin=50,n.iter=
30000,n.burnin=5000,codaPkg=TRUE,debug=TRUE)

> #Membuat MCMC objek
> library(coda)
> sim.coda=read.bugs(sim)
> summary(sim.coda)
```

```
> nchain=nchain(sim.coda)
> npar=nvar(sim.coda)-1
> n=niter(sim.coda)

> #Monitoring konvergensi algoritma
> library(lattice)

> #Traceplot
> win.graph()
> par(mfrow=c(4,4))
> traceplot(sim.coda,col="red")

> #Autokorelasi
> win.graph()
> par(mfrow=c(4,4))
> autocorr.plot(sim.coda,lag.max=100,auto.layout=FALSE,
col="red")

> #Density plot
> win.graph()
> par(mfrow=c(4,4))
> densplot(sim.coda,col="red")

> #Ringkasan posterior
> param=as.matrix(sim.coda,iters=FALSE,chains=FALSE)
> library(mcmcse)
> MEAN=apply(param,2,function(x)mcse(x,method="bm",g=NULL)
$est)
> SE=apply(param,2,function(x)mcse(x,method="bm",g=NULL)$se)
> SD=apply(param,2,function(x)sd(x))
> Ringkasan=cbind(MEAN,SE,SD)

> #Credible interval
> q2.5=apply(param,2,function(x)quantile(x,c(0.025)))
> median=apply(param,2,function(x)quantile(x,c(0.5)))
> q97.5=apply(param,2,function(x)quantile(x,c(0.975)))
> quantile=cbind(q2.5,median,q97.5)
> cred_int=cbind(q2.5,q97.5)

> #Ringkasan posterior full
> Ringkasan_full=cbind(MEAN,SD,SE,q2.5,median,q97.5)
> Ringkasan_full=as.matrix(Ringkasan_full)
> rownames(Ringkasan_full)=c("a0","a1","a2","a3","a4","a5",
"b0","b1","b2","b3","b4","b5","deviance",”w”)
> colnames(Ringkasan_full)=c("Rata-Rata","sd","MCError",
"2.5%","median","97.5%")
```

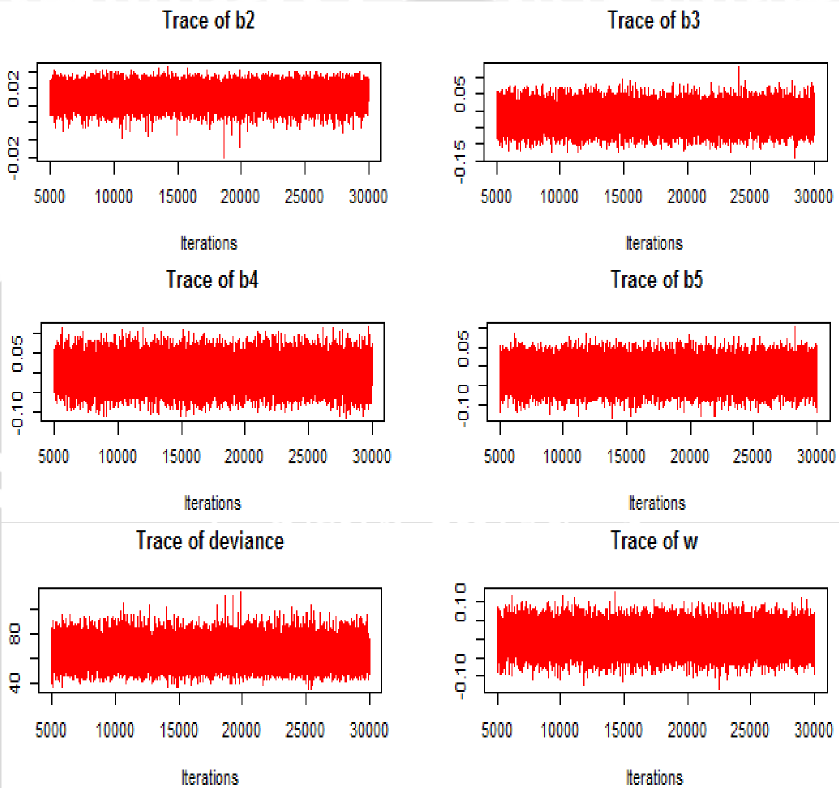
Lampiran 7. Ringkasan Distribusi Posterior Regresi ZIGP dengan Metode Bayesian

| Parameter | Rata-rata | Sd | MC Erorr | 2.5% | Median | 97.5% |
|---|---|---|---|---|---|---|
| a0 | 0.0112 | 0.0316 | 0.000188 | -0.0505 | 0.0109 | 0.0726 |
| a1 | 0.0092 | 0.0314 | 0.000196 | -0.0522 | 0.0095 | 0.0701 |
| a2 | 0.0364 | 0.0134 | 0.000076 | 0.0117 | 0.0358 | 0.0643 |
| a3 | 0.0295 | 0.0304 | 0.000193 | -0.0299 | 0.0296 | 0.0887 |
| a4 | 0.0019 | 0.0316 | 0.000207 | -0.0603 | 0.0019 | 0.0635 |
| a5 | 0.0172 | 0.0308 | 0.000197 | -0.0425 | 0.0172 | 0.0779 |
| b0 | -0.0074 | 0.0315 | 0.000190 | -0.0691 | -0.0075 | 0.0539 |
| b1 | -0.0057 | 0.0317 | 0.000228 | -0.0677 | -0.0059 | 0.0567 |
| b2 | 0.0165 | 0.0049 | 0.000032 | 0.0059 | 0.0169 | 0.0254 |
| b3 | -0.0198 | 0.0296 | 0.000195 | -0.0782 | -0.0198 | 0.0381 |
| b4 | -0.0010 | 0.0317 | 0.000234 | -0.0632 | -0.0010 | 0.0605 |
| b5 | -0.0198 | 0.0295 | 0.000190 | -0.0781 | -0.0196 | 0.0372 |
| W | 0.0005 | 0.0314 | 0.000195 | -0.0609 | 0.0008 | 0.0613 |

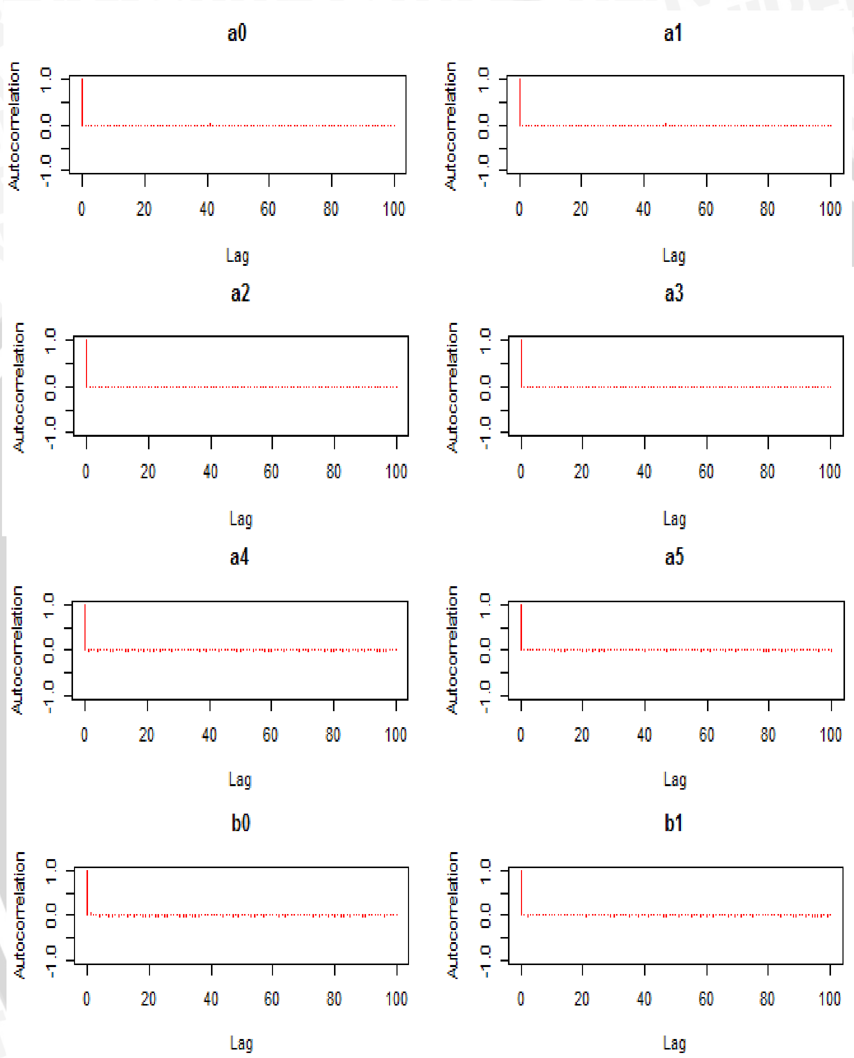
Lampiran 8. *Trace Plot*



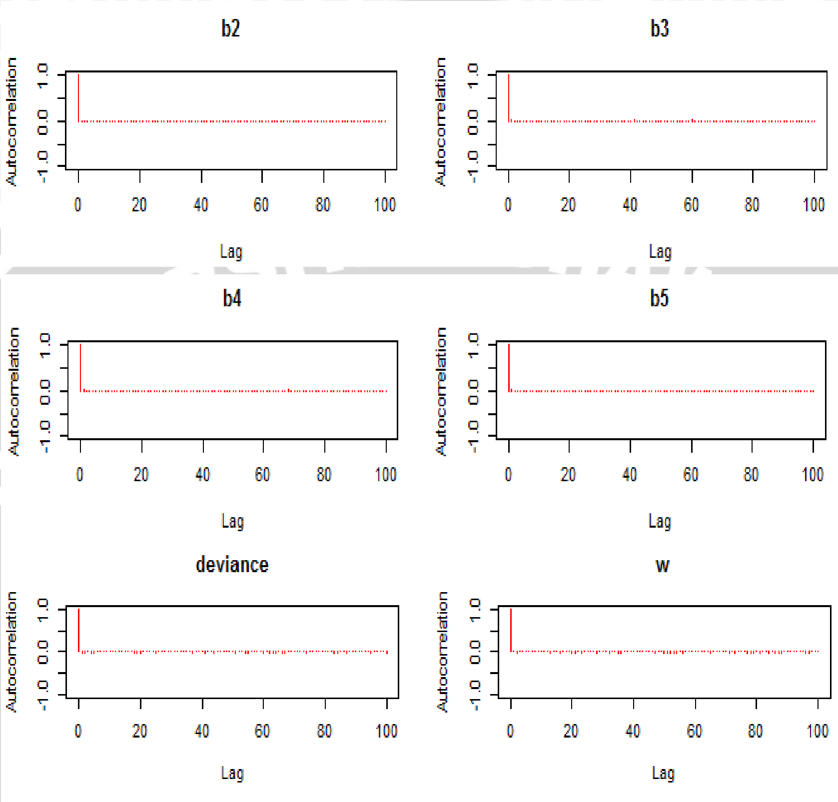
Lampiran 8. (Lanjutan)



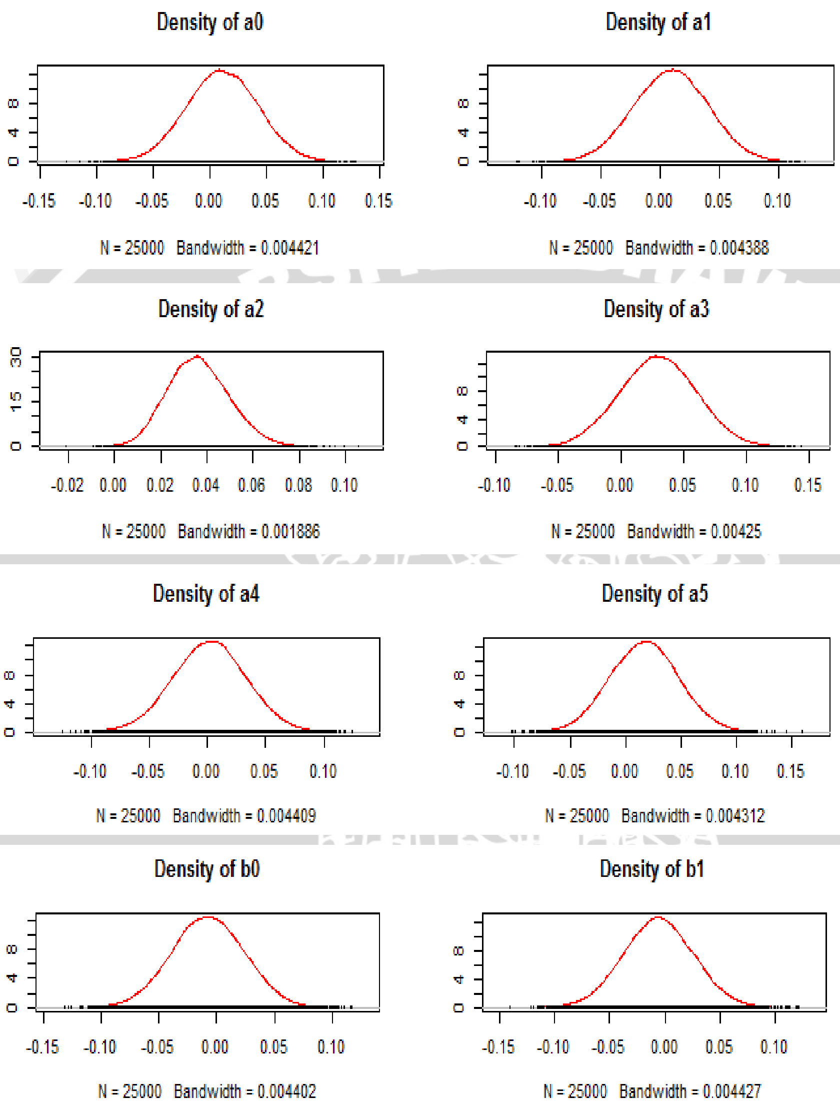
Lampiran 9. Autokorelasi



Lampiran 9. (Lanjutan)



Lampiran 10. *Density Plot*



Lampiran 10. (Lanjutan)