TESIS – SS14 2501

REGRESI POISSON MENGGUNAKAN GENERALIZED ESTIMATING EQUATION

(Studi Kasus: Data Longitudinal Frekuensi Terjadinya Banjir Di Jawa Timur Tahun 2011-2013)

ARIF SETIAWAN
NRP: 1315201719

Dosen Pembimbing:
Dr. Wahyu Wibowo, M.Si.
Prof. Nur Iriawan, M.Ikom., PhD.

PROGRAM MAGISTER
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2017

TESIS – SS14 2501

POISSON REGRESSION USING GENERALIZED ESTIMATING EQUATION

(Case Study: Flood Occurance Longitudinal Data in East Java Province Year 2011-2013)

ARIF SETIAWAN
NRP: 1315201719

Supervisor
Dr. Wahyu Wibowo, M.Si.
Prof. Nur Iriawan, M.Ikom., PhD.

PROGRAM OF MAGISTER
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCE
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2017

REGRESI POISSON MENGGUNAKAN *GENERALIZED ESTIMATING EQUATION*
(Studi Kasus: Data Longitudinal Frekuensi Terjadinya Banjir Di Jawa Timur Tahun 2011-2013)

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar Magister Sains (M.Si)
di
Institut Teknologi Sepuluh Nopember
Oleh:

ARIF SETIAWAN
NRP. 1315 201 7019

Tanggal Ujian : 4 Januari 2017
Periode Wisuda : Maret 2017

Disetujui oleh:

1. Dr. Wahyu Wibowo, M.Si. (Pembimbing I)
   NIP. 19740328 199802 1 001

2. Prof. Drs. Nur Iriawan, M.Ikom, Ph.D. (Pembimbing II)
   NIP. 196210151988031002

3. Dr. Bambang Widjanarko Otok, M.Si. (Penguji)
   NIP. 19681124 199412 1 001

4. Santi Wulan Purnami, M.Si, Ph.D. (Penguji)
   NIP 19720923 199803 2 001

5. Dr. Erni Tri Astuti, M. Math. (Penguji)
   NIP. 19671022 199003 2 002

an. Direktur Program Pascasarjana
Asisten Direktur

Prof. Dr. Ir. Tri Widjaja, M.Eng.
NIP. 19611021 198603 1 001

Direktur Program Pasca Sarjana,

Prof. Ir. Djauhar Manfaat, M.Sc., Ph.D.
NIP.19601202 198701 1 001

REGRESI POISSON MENGGUNAKAN GENERALIZED ESTIMATING EQUATION
(Studi Kasus: Data Longitudinal Frekuensi Terjadinya Banjir Di Jawa Timur Tahun 2011-2013)

Nama	: Arif Setiawan
NRP	: 1315201719
Pembimbing 1	: Dr. Wahyu Wibowo, M.Si.
Pembimbing 2	: Prof. Drs. Nur Iriawan, M.Ikom., Ph.D.

ABSTRAK

Provinsi Jawa Timur berada di peringkat kedua untuk kejadian banjir sejak tahun 1815 sampai dengan tahun 2016 menurut data BNPB. Regresi Poisson dapat digunakan untuk meneliti data count, dalam hal ini diaplikasikan untuk meneliti kejadian banjir di desa-desa di Jawa Timur dengan topografi, kebiasaan membuang sampah dan keberadaan pemukiman kumuh sebagai *covariate*. Banjir cenderung terjadi berulang-ulang pada daerah tertentu antar waktu sehingga analisis menggunakan data longitudinal dapat digunakan. Kelebihan analisis data longitudinal adalah mampu mengakomodasi korelasi antar waktu yang terjadi. *Generalized Estimating Equation* adalah pengembangan dari *Generalized Linear Model* untuk estimasi parameter pada analisis data longitudinal yang digunakan untuk variabel respon yang berautokorelasi. *Generalized Estimating Equation* digunakan dalam mengestimasi parameter longitudinal menggunakan *Quasi-Likelihood Under the Independece Information Criterion (QIC)* dimodelkan ke dalam Regresi Poisson. Estimator Parameter untuk Regresi Poisson menggunakan Generalized Estimating Equation adalah

$$\beta^{b+1} = \beta^{b} + \left[\sum_{i=1}^{N} \frac{\partial \mu_i^T}{\partial \beta} V_{\beta i}^{-1} Diag(1 - u^{(b)}) \frac{\partial \mu_i}{\partial \beta^T}\right]^{-1} S_\beta$$

dengan estimasi α untuk *Working Correlation Structure*-nya :

$$\alpha_1 = \frac{\frac{1}{N^*} \sum_{i=1}^{N} \sum_{s<t} \frac{(u_{is} - p_{is})(u_{it} - p_{it})}{\sqrt{p_{is}(1 - p_{is})\, p_{it}(1 - p_{it})}}}{\frac{1}{N_{tot}} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \left(\frac{u_{ij} - p_{ij}}{\sqrt{p_{ij}(1 - p_{ij})}}\right)^2} \Bigg|_{\gamma, u_{ij} = \gamma^{(b+1)}, u_{ij}^{(b)}}$$

Model dengan QIC terkecil adalah dengan *Working Correlation Structure* tipe *Unstructured* dan *Independent*. Variabel yang mempengaruhi terjadinya banjir di Timur adalah topografi dan keberadaan pemukiman di bantaran sungai.

Kata Kunci: Data Longitudinal, *General Estimating Equation*, *Poisson Regression*

POISSON REGRESSION USING GENERALIZED ESTIMATING EQUATION.
(Case Study Flood Occurance Longitudinal Data in East Java Province Year 2011-2013)

Name	: Arif Setiawan
Student Key	: 1315201719
Supervisor	: Dr. Wahyu Wibowo M.Si
Co-Supervisor	: Prof. Drs. Nur Iriawan M.IKom. Ph.D

ABSTRACT

ABSTRACT. East Java Province is the 2^{nd} flood occurance in Indonesia since 1815, according to the National Disaster Management Authority. Flood occurance frequently happen on a certain area. Longitudinal modeling couple with Generalized Estimating Equation accommodating its time correlation can cover this area. Poisson Regression will applied to study flood occurance on villages in East Java year 2011-2013 and topography, garbage disposal and slum existence on riverbanks as covariate. Generalized Estimating Equation is an extension Generalized Linear Model for correlated data parameter. Generalized Estimating Equation used Quasi-Likelihood Under the Independece Information Criterion (QIC) as model selection that depends on each Working Correlation Structure. Regresi Poisson. Parameter Estimator Estimator for Poisson Regression using Generalized Estimating Equation is :

$$\beta^{b+1} = \beta^{b} + \left[\sum_{i=1}^{N} \frac{\partial \mu_i^T}{\partial \beta} V_{\beta i}^{-1} Diag(1-u^{(b)}) \frac{\partial \mu_i}{\partial \beta^T}\right]^{-1} S_\beta$$

and α estimation for *Working Correlation Structure* :

$$\alpha_1 = \frac{\frac{1}{N^*}\sum_{i=1}^{N}\sum_{s<t}\frac{(u_{is}-p_{is})(u_{it}-p_{it})}{\sqrt{p_{is}(1-p_{is})p_{it}(1-p_{it})}}}{\frac{1}{N_{tot}}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left(\frac{u_{ij}-p_{ij}}{\sqrt{p_{ij}(1-p_{ij})}}\right)^2}\Bigg|_{\gamma,u_{ij}=\gamma^{(b+1)},u_{ij}^{(b)}}$$

Unstructured and Independent Working Correlation Matrix provide the best model according to QIC value. Topography and slum existence on the river bank affect on villages flood occurance in East Java.

Key words: *Longitudinal Data, General Estimating Equation, Poisson Regression.*

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah SWT karena atas berkat rahmat dan hidayah-Nya penulis diperkenankan menyelesaikan tesis yang berjudul **"REGRESI POISSON MENGGUNAKAN *GENERALIZED ESTIMATING EQUATION* (Studi Kasus: Data Longitudinal Frekuensi Terjadinya Banjir Di Jawa Timur Tahun 2011-2013)"** sesuai waktu yang diharapkan.

Pada kesempatan ini penulis ingin menyampaikan ucapan terimakasih dan penghargaan setinggi-tingginya kepada:

1. Badan Pusat Statistik (BPS) yang telah memberikan kesempatan serta beasiswa kepada penulis untuk melanjutkan program studi S2 di ITS.
2. Dr. Wahyu Wibowo, Prof. Drs. Nur Iriawan. M.Ikom, P.hD. atas segala bimbingan, saran, masukan serta motivasi yang diberikan selama penyusunan tesis ini.
3. Dr. Bambang Widjanarko Otok, M.Si., Santi Wulan, M.Si, P.hD. dan Dr. Erni Tri Astuti, M.Math. selaku dosen penguji yang banyak memberikan saran dan koreksi atas penulisan tesin ini.
4. Bapak Ketua Jurusan Statistika, Bapak Ketua Program Studi Pascasarjana Statistika ITS beserta jajarannya atas fasilitas yang disediakan dan arahan selama proses studi.
5. Dr. Purhadi, M.Sc selaku dosen wali, seluruh Bapak/Ibu dosen pengajar yang telah memberikan ilmu dan pengamalan yang bermanfaat kepada penulis.
6. Istri dan Anak-anakku serta keluarga besarku atas segala doa dan dukungannya sehingga penulis berhasil menyelesaikan studi dengan baik.
7. Teman-teman BPS angkatan 9: Ervin , Mbak Ayu Mbak Tiara, Mbak Kiki, Mbak Ika, Mbak Irva, Mbak Aty, Yuk Mety, Mbak Nunik, Mbak Risma, Mbak Lila, Mbak Dewi, Mas Agung, Bayu, Mas Dinu, Mas Leman, Mas Bambang, Bang Node dan Mas Suko. Terimakasih atas segala bantuan, Sangat senang bertemu kalian semua dan semoga kita bisa dipertemukan kembali pada kesempatan yang lebih baik.
8. Teman-teman reguler angkatan 2015. Senang bertemu kalian dan terimakasih untuk keseruannya.
9. Mas Rindang dan Syahrul. Terimakasih atas bantuan ilmu dan diskusi yang mencerahkan.
10. Semua pihak yang telah membantu penyelesaian tesis ini.

Akhir kata, dengan segala kerendahan hati, penulis menyadari bahwa tesis ini jauh dari sempurna, segala kritik dan saran yang membangun sangat diharapkan demi perbaikan tesis ini. Walaupun demikian, penulis berharap ilmu yang telah diperoleh menjadi barokah dan memberikan manfaat bagi pihak yang memerlukan. Semoga Allah SWT memberikan kebaikan untuk kita semua.

Surabaya, Januari 2017

Penulis

DAFTAR ISI

Halaman

DAFTAR TABEL

DAFTAR GAMBAR

BAB 1
PENDAHULUAN

1.1. Latar Belakang

Perubahan iklim global berdampak pada pemanasan global karena adanya efek rumah kaca *(green house effect)* sehingga dapat meningkatkan jumlah uap air di atmosfir, hal ini memicu peningkatan curah hujan. Peningkatan curah hujan yang tidak diantisipasi dengan baik menimbulkan kemungkinan terjadinya banjir.

Banjir biasanya melanda daerah yang mempunyai topografi relatif rendah dan daerah cekungan. Risiko kerugian akibat banjir akan meningkat pada daerah yang padat penduduknya. Selain itu, penutupan lahan dan penggunaan lahan juga sangat berpengaruh terhadap aliran air atau limpasan (*run off*) permukaan (Purwadhi, 2003).

Bencana banjir adalah masalah yang lazim terjadi di Indonesia, sebagai negara yang berada di sekitar khatulistiwa yang beriklim tropis. Berdasarkan data dari BNPB, bencana yang paling sering terjadi di Indonesia pada kurun waktu 1815-2013 adalah banjir, sebanyak 6.706 kejadian dan 918 kejadian banjir tersebut atau 13,69 persennya terjadi di Jawa Timur.

Provinsi Jawa Timur terdiri dari 29 Kabupaten dan 8 Kota dan mempunyai 28 sungai. Berdasarkan Indeks Kerawanan Bencana Banjir 2013, hanya 2 Kabupaten/Kota yang mempunyai Kelas Resiko Banjir Sedang di Jawa Timur, yaitu Kota Malang dan Kota Batu, sedangkan Kabupaten Kota lainya berada dalam Kelas Resiko Tinggi terhadap Banjir. Kabupaten yang paling sering mengalami banjir adalah Bojonegoro, Gresik, Tuban, Lamongan dan Pasuruan. Kabupaten ini dilewati oleh Sungai Bengawan Solo, terkecuali Pasuruan. Kabupaten Bojonegoro mengalami banjir sebanyak 105 kali. Menurut Liza (2015) Sungai Bengawan Solo dibagi menjadi tiga pembagian yakni, (1) Bengawan Solo Hulu, (2) Sub DAS Kali Madiun, dan (3) Sub DAS Bengawan Solo Hilir. Kabupaten Bojonegoro termasuk dalam Sub Bengawan Solo Hilir dengan kondisi topografi relatif datar dan sebagian daerahnya berada di dataran rendah, kemiringan landai, melalui dataran aluvial.

Indarto, Susanto dan Huda (2012) melakukan penelitian tentang Analisis Frekuensi Banjir (*Flood Frequency Analysis*) dengan metode Log Pearson III di 15 Daerah Aliran Sungai di Jawa Timur. Hasilnya adalah bahwa beberapa DAS memiliki kemiripan grafik frekuensi banjir yang disebabkan oleh kesamaan karakteristik hujan yang jatuh di DAS-DAS tersebut.

Kejadian banjir sering terjadi berulang dari tahun ke tahun di tempat yang sama, terutama pada daerah yang dipetakan rawan banjir. Pemerintah mempunyai kewajiban untuk mengatasi hal ini. Kebijakan penanggulangan banjir pada tiap tingkatan pemerintahan berbeda, tergantung cakupan wilayah dan wewenang masing-masing. Hal ini mendasari pemikiran untuk melakukan penelitian frekuensi banjir per tahun sebagai data longitudinal dan melihat hubunganya untuk wilayah Provinsi Jawa Timur. Kodoatie (2002) dikutip oleh Liza (2015) menyatakan bahwa pengendalian dan penanganan banjir setiap daerah berbeda-beda. Ketidaksamaan tersebut menyebabkan parameter penanganan banjir di suatu tempat tidak dapat dipakai sebagai acuan penanganan di tempat lain. Selain itu Kodoatie dan Sugiyanto (2002) dikutip oleh Rahmawati (2008) menyatakan bahwa faktor tindakan manusia juga mempengaruhi terjadinya banjir antara lain adalah kawasan kumuh di sepanjang sungai, dan perilaku membuang sampah.

Data longitudinal sangat umum digunakan baik dalam studi observasional maupun studi eksperimental. Pada studi longitudinal, individu dalam penelitian diikuti selama periode waktu tertentu untuk setiap individu, sehingga data dikumpulkan pada beberapa titik waktu (Wu, 2010). Data longitudinal merupakan salah satu bentuk data berkorelasi. Pada data longitudinal, variabel respons diukur pada beberapa titik waktu untuk setiap subjek. Dalam studi longitudinal dimungkinkan untuk mempelajari perubahan respons antar waktu beserta faktor yang mempengaruhi perubahan tersebut, baik pada level populasi maupun level individu (Wu dan Zhang, 2006).

Data longitudinal dicirikan oleh fakta bahwa pengamatan berulang dalam subjek yang sama cenderung berkorelasi sehingga model-model untuk analisis data longitudinal harus memperhitungkan hubungan antara pengamatan berkala dalam subjek yang sama. Korelasi antar pengamatan berulang dapat dimodelkan secara eksplisit (melalui pola matriks kovarian), maupun secara implisit (melalui pengaruh acak).

Generalized Estimating Equation diperkenalkan oleh Liang dan Zeger pada tahun 1986 adalah pengembangan metode dari *Generalized Linear Model (GLM)* yang dapat digunakan untuk menduga parameter model yang berautokorelasi dan tidak berdistribusi normal. Metode ini termasuk semi parametrik karena estimating equations diturunkan tanpa spesifikasi penuh dari distribusi gabungan dari subyek observasinya. Pada metode GEE ini disyaratkan untuk memilih struktur korelasi yang tepat untuk menggambarkan korelasi tersebut, pemilihan struktur korelasi ini menggunakan *Quasi-Likelihood Under the Independece Information Criterion (QIC)*.

Cupal, Deev dan Linnertova (2015) menggunakan Regresi Poisson untuk memodelkan kejadian banjir di Praha Republik Ceko. Regresi Poisson merupakan model standar untuk *count data* dan termasuk dalam model regresi nonlinier. Regresi Poisson mengasumsikan keadaan yang equidispersi, namun sering terjadi kasus overdispersi yaitu nilai variansi lebih besar dari nilai *mean*. Penggunaan yang tidak tepat dari regresi Poisson pada data yang mengalami overdispersi dapat berakibat fatal dalam interpretasi model, khususnya parameter model karena diperoleh *standard error* yang *underestimate* dan dapat memberikan kesimpulan yang keliru tentang signifikan atau tidaknya parameter model regresi.

Seringkali model regresi Poisson menjadi tidak sesuai jika terdapat banyak data yang kosong (bernilai nol) atau jika asumsi mean sampel sama dengan variansinya tidak terpenuhi. Sementara itu pada data kejadian bencana untuk lingkup desa dalam provinsi sering dijumpai banyak data yang bernilai nol. Jika data yang bernilai nol atau kosong dijumpai pada data jenis count dan proporsinya besar (*zero inflation*), maka model Regresi *Zero Inflated Poisson* (ZIP) lebih disarankan (Lambert, 1992).

Model ZIP ini kadang tidak sesuai untuk kasus-kasus dimana terjadi *over/under dispersion*, yaitu variansi sampel lebih besar/lebih kecil dari mean sampel. Sementara itu, ada suatu model regresi count yang dapat mengatasi masalah over/under dispersion dalam keadaan data tidak terlalu banyak nol, yaitu model Negative Binomial (NB) dan Generalized Poisson (GP). Contohnya, model

GP yang digunakan Famoye, Wulu dan Singh (2004) dalam pemodelan data kecelakaan kendaraan ternyata lebih tepat menggambarkan keadaan data dibanding model Poisson. Oleh karena itu banyak para peneliti yang beralih dari model Poisson dan ZIP ke model lain yang dapat mengatasi *zero inflation* dan *over/under dispersion*. Ariani (2014) menyatakan bahwa model regresi *Zero Inflated Generalized Poisson (ZIGP)* menghasilkan nilai AIC (*Akaike Information Criterion*) yang lebih kecil dibandingkan model regresi *Zero Inflated Negative Binomial* pada kasus data dengan overdispersion dan zero inflation pada peubah respon.

1.2. Perumusan Masalah

Berdasarkan latar belakang di atas, masalah yang dapat dirumuskan adalah bagaimana membangun model Regresi Count dari frekuensi banjir di Jawa Timur menggunakan metode estimasi parameter *Generalized Estimating Equation*.

1.3. Tujuan Penelitian

1. Mengkaji estimasi parameter model Regresi Poisson kejadian banjir di Jawa Timur pada tahun 2011-2013 dengan menggunakan *Generalized Estimating Equation*
2. Mengkaji kecenderungan terjadinya banjir di desa-desa di Jawa Timur

1.4. Manfaat Penelitian

1. Melihat perbedaan kejadian banjir yang terjadi di desa-desa di Jawa Timur pada tahun 2011-2013
2. Memodelkan data longitudinal frekuensi banjir di desa-desa di Jawa Timur pada tahun 2014

1.5. Batasan Masalah

Batasan Masalah Pada Penelitian ini adalah

1. Data yang digunakan adalah frekuensi banjir di masing-masing desa di Jawa Timur pada tahun 2011-2013
2. Metode yang digunakan untuk mengestimasi Parameter *Generalized Estimating Equations*
3. Model yang digunakan adalah Regresi Poisson

BAB 2
TINJAUAN PUSTAKA

2.1. Data Longitudinal

Data longitudinal adalah data yang diperoleh melalui suatu pengamatan berulang yang dilakukan terhadap sejumlah objek yang sama. Data semacam ini banyak muncul di berbagai bidang misalnya kesehatan, pertanian dan ekonomi. Kebanyakan studi longitudinal dirancang untuk mengetahui nilai tengah respons sebagai fungsi dari waktu, dengan tetap memperhatikan peranan dari peubah penjelas. Saat ini terdapat beberapa metode untuk menduga nilai tengah respons, baik secara parametrik maupun secara non-parametrik.

Dalam analisis data longitudinal perlu dipertimbangkan adanya kemungkinan pengelompokan profil nilai tengah respon. Adanya pengelompokan dimungkinkan dengan adanya kesamaan nilai peubah penjelas yang mempengaruhi profil nilai tengah respons tersebut. Kesamaan nilai peubah penjelas sendiri mungkin bersifat alami, mungkin pula diadakan melalui pemberian perlakuan terhadap objek pengamatan. Dalam kondisi ini, pendekatan yang direkomendasikan oleh Diggle dkk (1995) untuk menentukan nilai tengah respons adalah dengan membentuk model yang terpisah dari beberapa kelompok data longitudinal. Dengan pengelompokan, diharapkan akan diperoleh penduga profil nilai tengah respons yang lebih homogen dengan tingkat akurasi yang tinggi

Studi longitudinal adalah suatu studi dimana suatu objek pengamatan diukur secara berulangkali dari waktu ke waktu. Studi semacam ini banyak muncul di berbagai bidang. Studi longitudinal sangat penting dalam epidemiologi, uji klinik dan evaluasi pengobatan. Walaupun studi longitudinal memerlukan upaya yang lebih besar daripada studi lintang potong (*cross-sectional*), namun terdapat beberapa keuntungan dari suatu studi longitudinal, yakni:

a. *Incident events recorded*: dapat mengamati terjadinya/ timbulnya suatu penyakit. Waktu permulaan dapat berkorelasi dengan perubahan-perubahan yang terjadi pada partisipan-partisipan yang mengalami paparan dalam jangka waktu lama (Fisher, 2004).

b. *Prospective ascertainment of exposure*: para partisipan dapat diketahui status paparannya dari catatan masing-masing yang berasal dari kunjungan ulang partisipan ke pusat pengobatan penelitian. Hal ini akan mengurangi *recall bias* bila partisipan diminta menyebutkan paparan-paparan yang dialaminya; sekaligus dapat diamati hubungan paparan dengan keluaran sesuai dengan berjalannya waktu (Fisher, 2004).

c. *Measurement of individual change in outcome*: dapat menilai perubahan dari paparan maupun keluaran pada tingkat individual. (Fisher, 2004; Hedeker, 2006).

d. *Separation of time effects*: *cohort, period, age*. Pada studi longitudinal, dapat dilakukan pengelompokan partisipan berdasarkan kohort subjek, berdasarkan periode pengamatan dan berdasarkan kelompok umur sesuai dengan berjalannya waktu pengamatan (Fisher, 2004; Hedeker, 2006).

e. *Control for cohort effects*: kelompok kohort dalam studi longitudinal merupakan kelompok yang *fixed*, sehingga tidak akan terpengaruh (*confounded*) oleh perbedaan karakteristik kelompok kohort lainnya (Fisher, 2004; Hedeker, 2006).

f. Setiap subjek dapat menjadi kontrol terhadap dirinya sendiri, misalnya pada studi eksperimental metode *crossover* (Hedeker, 2006).

g. Variabilitas intra subjek lebih kecil daripada inter subjek, sehingga hasil ujinya secara statistik lebih sensitif/ lebih *powerful* (Hedeker, 2006).

Studi longitudinal juga memiliki berbagai keterbatasan, antara lain:

a. *Participant follow-up:* Akan menyebabkan bias bila terjadi pemantauan yang tidak lengkap terhadap para partisipan, akibatnya kesimpulan penelitian tidak dapat merepresentasikan keadaan sebenarnya di populasi (Fisher, 2004; Hedeker, 2006)

b. *Analysis of correlated data:* memerlukan metode khusus yang dapat mencerminkan adanya korelasi intra-subjek dalam pengukuran respons setiap subjek. Dampaknya, akan sangat mengganggu validitas *statistical test* dan *confidence interval* pada waktu inferensi hasil penelitian ke populasi (Fisher, 2004)

c. *Time-varying covariates:* arah kausalitas dapat menjadi kompleks, akibat adanya kemungkinan efek pengaruh timbal balik antara *outcome* dan *exposure* sesuai dengan berjalannya waktu pengamatan (Fisher, 2004; Hedeker, 2006)

d. *Carry-over/ sequence effect* pada studi *crossover,* yakni kemungkinan bahwa respon terhadap obat yang lebih belakangan diberikan sebenarnya masih dipengaruhi oleh efek obat sebelumnya (Hedeker, 2006).

2.2. Generalized Linear Model

Menurut Agresti (2000) *Generalized Linear Models* merupakan perluasan regresi sederhana untuk mengatasi distribusi respon yang *non-normal* dan memodelkan fungsi dari rata-rata. Tiga komponen pembentuk sebuah *Generalized Linear Model* adalah: Sebuah komponen random yang mengidentifikasi variabel response dari Y beserta dengan distribusinya; sebuah *systematic component* yang menjelaskan variabel ekplanatori yang digunakan dalam fungsi prediktor linearnya; dan fungsi link yang menjelaskan fungsi dari ekspektasi Y yang disamadengankan dengan *systematic component* di atas.

Komponen random dari GLMs adalah variabel respon Y dengan nilai observasi independen yang distribusinya berasal dari keluarga eksponensial dengan *Probability Density Function* sebagai berikut:

$$f(y_i ; \theta_i) = a(\theta_i)\, b(y_i)\, \exp[y_i Q(\theta_i)] \qquad (2.1)$$

Komponen sistematik dari GLMs menghubungkan sebuah vektor terhadap variabel ekplanatori melalui sebuah model linier, vektornya dijelaskan sebagai berikut:

$$\eta_i = \sum_{j} \beta_j x_{ij}, \qquad i = 1, \ldots, N. \qquad (2.2)$$

Kombinasi Liner dari variabel eksplanatori ini disebut sebagai *linear predictor*.

Link Function adalah penghubung dari komponen random dan komponen sistematik. Model ini menghubungkan μ_i dan η_i dengan $\eta_i = g(\mu_i)$, dimana fungsi link g ini bersifat *monotomic*, yaitu fungsi yang bisa differensialkan. Jadi g menghubungkan $E(Y_i)$ dan variabel eksplanatori-nya melalui formula sebagai berikut :

$$g(\mu_i) = \underset{j}{} \beta_j x_{ij,} \qquad i = 1, \ldots, N. \tag{2.3}$$

GLM adalah model linear untuk nilai rata-rata variabel response yang ditransformasikan yang distribusinya keluarga eksponensial.

2.3. Generalized Estimating Equation

Generalized Estimating Equations (GEE) merupakan pendekatan regresi lainnya dari suatu studi longitudinal; pertama kali diperkenalkan oleh Liang & Zeger tahun 1986; digunakan untuk analisis data longitudinal dengan data yang berkorelasi (Twisk, 2003; Balingger, 2004; Fisher, 2004; Hedeker, 2006). Metode GEE merupakan perkembangan dari Generalized Linear Model (GLMs) dapat digunakan untuk menduga parameter model berdasarkan data yang mengandung autokorelasi dan data yang tidak menyebar normal

Komponen Utama GEE menurut Liang & Zeger (1986) adalah:

1. *Random component*: variabel dependen (Y) mengikuti distribusi *random* tertentu (normal, binomial, poisson) (Ballinger, 2004).
2. *Link function*: Fungsi tranformasi pada variabel dependen yang menghubungkan respon rata-rata dengan model linernya (Ballinger, 2004).
3. *Systematic component*: Variabel independent (X) dapat dikombinasikan dalam bentuk fungsi linier.

Dalam GEE ada 2 model yaitu:

1. Model regresi untuk melihat *mean response*. Menurut McCullagh & Nelder (1989), model regresi sangat fleksibel, dapat berupa model linier, regresi logistik, log-linier maupun bentuk *generalized linear model* (*GLMs*) (Fisher, 2004).
2. Model *within-subject correlation* (WSC). Ada dua kegunaan WSC yaitu untuk mendapatkan *weights* (*covariance inverse*) dan *model-based standard errors for the estimated coefficients* (Fisher, 2004; Balingger, 2004).

Karena GEE adalah pengembangan dari GLM untuk data yang berkorelasi, dengan *link function* sebagai berikut:

$$g\ \mu_{ij}\ = X_i'\beta \tag{2.4}$$

Penduga dari parameter β didapatkan dengan *Quasi-likelihood Estimating Equation* atau *Scorelike Equation* menurut Hedeker dan Gibbon (2006):

$$\beta = \sum_{j=1}^{m} X_j'V_j^{-1}X_j^{\ -1} \sum_{j=1}^{m} X_j'V_j^{-1}Y_j \tag{2.5}$$

dimana V_J adalah matriks varian kovarian dari Y_{ij} yang berukuran $n_j \times n_j$ pada obyek ke- i, yaitu :

$$V_j = \psi A_j^{\frac{1}{2}} R(\alpha)_{(n_j \times n_j)} A_j^{\frac{1}{2}}{}_{(n_j \times n_j)} \tag{2.6}$$

dimana ψ adalah parameter dispersi yang diduga dengan :

$$\varphi = \frac{1}{n-p} \sum_{i=1}^{n} \sum_{j=1}^{p} e_{ij}^2 \tag{2.7}$$

dimana *n* adalah jumlah sampelnya dan *p* adalah banyaknya parameter dan e_{ij} adalah residual Pearson, yaitu:

$$e_{ij} = \frac{Y_{ij} - \mu_{ij}}{\sqrt{var\ (\mu_{ij})}} \tag{2.8}$$

dan R(α) adalah matriks korelasi berukuran $n_i \times n_i$ yang berisi korelasi antar respon pada i obyek. Kemudian dilakukan iterasi Newton Rapson sehingga diperoleh parameter yang konvergen yaitu $\beta^t - \beta^{t-1} < 10^{-6}$ (Handayanti Mitakda 2015).

Rumus *Newton Raphson* :

$$\beta^{t+1} = \beta^t - \sum_{j=1}^{m} X_j'V_j^{-1}X_j^{\ -1} \sum_{j=1}^{m} X_j'V_j^{-1}Y_j \tag{2.9}$$

Observasi pada GEE bisa berkorelasi dan dianggap mengikuti *working correlation structure* (*WCS*) (Ballinger, 2004). Dikenal 5 bentuk WCS:

independent structure, exchangeable structure, stationary/ m-dependent structure, autoregressive correlation structure dan *unstructured correlation* (Twisk, 2003).

Dengan penjelasan sebagai berikut :

a. *Independent structure:* Korelasi antar pengukuran yang berurutan diasumsikan sama dengan 0. Pada analisis dengan GEE, WCS ini agak bertentangan, mengingat bahwa GEE memang digunakan pada data yang berkorelasi.
b. *Exchangeable structure:* Korelasi antar pengukuran yang berurutan diasumsikan sama, tanpa melihat interval waktu pengukuran.
c. *Stationary/ m-dependent structure:* Korelasi pada jarak *t* adalah sama, Korelasi pada $t+1$ adalah sama (pada $t=1$ s.d $t=m$); korelasi pada pengukuran > *t* diasumsikan sama dengan 0.
d. *Autoregressive correlation structure:* Korelasi pengukuran kedua diasumsikan sama dengan ρ, maka pengukuran ketiga sama dengan ρ^2, korelasi yang berjarak *t* adalah ρ^t.
e. *Unstructured correlation:* Seluruh korelasi dianggap berbeda.

Kemudian dilakukan pemilihan *working correlation structure* dilakukan dengan memilih model dengan *Quasi likelihood under the independence Information Criterion* (QIC). Model dengan QIC terkecil merupakan model dengan struktur korelasi terbaik. Rumus QIC adalah sebagai berikut:

$$QIC = -2Q\ \beta\ + 2(V_m^{-1}\ B\ V_e\ \beta\ ) \quad (2.10)$$

dimana $Q\ \beta$ adalah nilai dari *quasilikelihood* dari β untuk masing-masing struktur korelasi yang diasumsikan. $V_m\ \beta$ adalah matriks varian dari model dengan struktur korelasi independence dan $V_e\ \beta$ adalah hasil estimasi varian dari sandwich estimator dengan menggunakan struktur korelasi yang diasumsikan.

2.4. Distribusi Poisson

Suatu variabel random Y didefinisikan mempunyai distribusi Poisson jika densitas (fungsi peluangnya) diberikan sebagai berikut (Mood, Graybill dan Boes, 1974):

$$f_Y\ y\ =\ f_Y\ y;\mu\ =\ \frac{e^{-\mu}\mu^y}{y!}\qquad y = 0,1,2\ldots \tag{2.11}$$

$$0$$

dimana parameter μ memenuhi $\mu > 0$.

Persamaan di atas disebut juga sebagai fungsi peluang Poisson. Misalkan Y adalah suatu variabel random yang berdistribusi Poisson, maka mempunyai mean dan variansi yang sama yaitu μ.

Distribusi Poisson merupakan distribusi diskrit. Untuk nilai μ yang kecil maka distribusinya sangat menceng dan untuk nilai μ yang besar akan lebih mendekati distribusi normal. Untuk kasus yang jarang terjadi maka nilai μ akan kecil. Distribusi Poisson adalah suatu distribusi yang paling sederhana dalam pemodelan data yang berupa *count* (jumlah), tetapi bukan satu-satunya.

Distribusi Poisson sering digunakan dalam pemodelan kasus yang jarang terjadi (*rare event*), seperti pemodelan tentang kecelakaan, peperangan atau epidemi. Peristiwa terganggunya aktivitas seseorang karena sakit pada usia dewasa terutama yang masih aktif bekerja dan/atau melakukan kegiatan primer lainnya (sekolah, mengurus rumah tangga atau kegiatan sehari-hari lainnya) bisa dikatakan merupakan suatu peristiwa yang jarang, karena pada usia tersebut terutama kalangan usia muda cenderung masih melakukan aktivitas secara normal walaupun sakit (Lam, dkk, 2006).

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

2.5. Model Regresi Poisson

Regresi Poisson merupakan suatu bentuk analisis regresi yang digunakan untuk memodelkan data yang berbentuk *count* (jumlah), misalnya data tersebut dilambangkan dengan Y yaitu banyaknya kejadian yang terjadi dalam suatu periode waktu dan/atau wilayah tertentu. Regresi Poisson mengasumsikan bahwa variabel random Y berdistribusi Poisson dan logarithma dari nilai ekspektasi Y dapat dimodelkan dengan suatu kombinasi linear dari parameter-parameter yang tidak diketahui. Karena nilai mean *(μ)* harus bernilai positif, maka dibutuhkan suatu fungsi penghubung (*link function*) untuk parameter μ.

Model regresi Poisson merupakan Generalized Linear Model (GLM) dan data responnya (komponen random) diasumsikan berdistribusi Poisson (McCullagh and Nelder, 1989; Agresti, 2002). Pada model regresi Poisson, biasanya link function yang digunakan adalah log, sehingga log (μi) = ηi. Dengan demikian model regresi Poisson dapat dituliskan sebagai berikut:

$$\log \mu_i = \sum_{j=1}^{k} \beta_j x_{ij}, \quad i = 1,2, \ldots ., n \qquad (2.12)$$

Distribusi Poisson menampakkan tiga masalah utama dalam aplikasi analisis regresi yang mengikuti asumsi klasik. Pertama, distribusi Poisson bentuknya menceng sementara regresi tradisional mengasumsikan distribusi *error* yang simetris. Kedua, distribusi Poisson *non negative*, sementara regresi klasik mengasumsikan bisa bernilai negatif. Ketiga, variansi dari distribusi Poisson naik seiring dengan kenaikan *mean,* sementara regresi klasik mengasumsikan variansinya konstan (Ruru dan Barrios, 2003).

Suatu ciri dari distribusi Poisson adalah mean sama dengan variansi. Pada prakteknya, kadang-kadang ditemukan suatu kondisi dimana variasi data lebih besar dibanding mean. Kondisi seperti ini disebut *over dispersion*, dan model regresi Poisson yang dihasilkan akan menjadi tidak sesuai. Selain itu akan menghasilkan estimasi parameter yang bias (Ridout, dkk, 2001).

Masalah lain pada regresi Poisson adalah jika terdapat banyak data yang bernilai nol, sehingga lebih banyak data nol-nya dibanding regresi Poisson yang akan diprediksi. Jika hal ini terjadi, maka regresi Poisson menjadi tidak tepat menggambarkan data yang sebenarnya.

2.6. Penaksiran Parameter Regresi Poisson

Penaksiran Parameter Regresi Poisson dapat menggunakan metode Maximum Likelihood Estimation (MLE). Bentuk umum dari fungsi likelihood untuk regresi Poisson adalah :

$$
\begin{aligned}
L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} f(y_i, \boldsymbol{\beta}) \\
&= \prod_{i=1}^{n} \frac{[\mu(\boldsymbol{x}_i, \boldsymbol{\beta})]^{y_i} exp[-\mu(\boldsymbol{x}_i, \boldsymbol{\beta})]}{y_i!} \\
&= \frac{\prod_{i=1}^{n}[\mu(\boldsymbol{x}_i, \boldsymbol{\beta})]^{y_i}\ exp[-\sum_{i=1}^{n}\mu(x_i, \boldsymbol{\beta})]}{\prod_{i=1}^{n} y_i!}
\end{aligned}
$$

Kemudian dilakukan penurunan fungsi ln-likelihood dari persamaan tersebut terhadap β yaitu parameter yang ditaksir yang kemudian disamakan dengan nol. Fungsi tersebut adalah :

$$
\ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} y_i \ln[\mu(\boldsymbol{x}_i, \boldsymbol{\beta})] - \sum_{i=1}^{n} \mu(\boldsymbol{x}_i, \boldsymbol{\beta}) - \sum_{i=1}^{n} \ln(y_i!) \qquad (2.13)
$$

jika $\mu(\boldsymbol{x}_i, \boldsymbol{\beta}) = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})$ maka persamaan di atas akan menjadi persamaan berikut :

$$
\begin{aligned}
\ln L(\boldsymbol{\beta}) &= \sum_{i=1}^{n} y_i \ln[\exp(\boldsymbol{x}_i^T \boldsymbol{\beta})] - \sum_{i=1}^{n} \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) - \sum_{i=1}^{n} \ln(y_i!) \\
&= \sum_{i=1}^{n} [y_i(\boldsymbol{x}_i^T \boldsymbol{\beta}) - \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) - \ln(y_i!)]
\end{aligned}
$$

dimana turunan pertama dan kedua adalah

$$
\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} = \sum_{i=1}^{n} y_i \boldsymbol{x}_i - \sum_{i=1}^{n} \boldsymbol{x}_i \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})
$$

$$\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} = -\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i^T \exp(\boldsymbol{x}_i^T\boldsymbol{\beta})$$

Metode Iterasi Newton-Raphson diperlukan untuk mendapatkan nilai konvergennya. dengan langkah-langkah sebagai berikut :

1. Menentukan nilai taksiran awal parameter $\boldsymbol{\beta}_{(0)}$. Misalkan $\boldsymbol{\beta}_{(0)}\ 1 = y$ maka selanjutnya nilai taksiran untuk setiap *i* dapat diperoleh dari $\mu_i = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta})$
2. Membentuk vector gradien $\boldsymbol{g}\ \boldsymbol{\beta}$, dimana $\boldsymbol{g}\ \boldsymbol{\beta}$ diperoleh dari

   $$\boldsymbol{g}(\boldsymbol{\beta}_{\ m})_{\ k-1\ x1} = \left[\frac{\partial \ln L(\boldsymbol{\beta})}{\partial\beta_0}\frac{\partial \ln L(\boldsymbol{\beta})}{\partial\beta_1}\cdots\frac{\partial \ln L(\boldsymbol{\beta})}{\partial\beta_k}\right]^T_{\beta=\beta_{(m)}}$$

   dengak *k* adalah banyaknya parameter yang ditaksir
3. Setelah diperoleh $\boldsymbol{g}\ \boldsymbol{\beta}$ kemudian diuraikan menurut deret Taylor pada $\boldsymbol{\beta} = \boldsymbol{\beta}_{(\boldsymbol{m})}$ yaitu :

   $$\boldsymbol{g}\ \boldsymbol{\beta} = \boldsymbol{g}\ \boldsymbol{\beta}_{(\boldsymbol{m})} + \left.\frac{\partial\boldsymbol{g}\ \boldsymbol{\beta}}{\partial\boldsymbol{\beta}}\right|_{\beta-\beta\ m} \boldsymbol{\beta}-\boldsymbol{\beta}_{\ \boldsymbol{m}} + \cdots$$

   Apabila $\boldsymbol{\beta}_{(m+1)}$ merupakan solusi dari $\boldsymbol{g}\ \boldsymbol{\beta} = 0$, jika $\boldsymbol{g}\ \boldsymbol{\beta}_{(m+1)} = \boldsymbol{0}$ maka:

   $$\boldsymbol{g}\ \boldsymbol{\beta}_{(m+1)} = \boldsymbol{g}\ \boldsymbol{\beta}_{\ m} + \boldsymbol{H}\ \boldsymbol{\beta}_{\ m}\ \ \boldsymbol{\beta}_{\ m+1} - \boldsymbol{\beta}_{\ m} + \cdots \qquad (2.14)$$

   dimana **H** adalah matriks Hessian (*k*+1)x(*k*+1)

   Apabila persamaan 2.14 di atas diambil sampai suku yang kedua maka diperoleh :

   $$-\boldsymbol{g}\ \boldsymbol{\beta}_{\ m} = \boldsymbol{H}(\boldsymbol{\beta}_{\ m})(\boldsymbol{\beta}_{\ m+1} - \boldsymbol{\beta}_{\ m})$$

   bisa juga dituliskan sebaigai berikut :

   $$\boldsymbol{\beta}_{\ m+1} - \boldsymbol{\beta}_{\ m} = -\boldsymbol{H}^{-1}(\boldsymbol{\beta}_{\ m})\boldsymbol{g}\ \boldsymbol{\beta}_{\ m}$$

   $$\boldsymbol{\beta}_{\ m+1} = \boldsymbol{\beta}_{\ m} - \boldsymbol{H}^{-1}(\boldsymbol{\beta}_{\ m})\boldsymbol{g}\ \boldsymbol{\beta}_{\ m}$$
4. Proses iterasi dilakukan pada persamaan di atas, dimana $\boldsymbol{\beta}_{\ m}$ adalah penaksir parameter yang konvergen pada iterasi ke-m

5. Proses iterasi tersebut akan berhenti apabila $\boldsymbol{\beta}_{m+1} - \boldsymbol{\beta}_{m} < \delta$ dimana δ bilangan sangat kecil dengan :

$$\boldsymbol{\beta}_{m+1} - \boldsymbol{\beta}_{m} = \sqrt{\boldsymbol{\beta}_{m+1} - \boldsymbol{\beta}_{m}{}^{T}\ \boldsymbol{\beta}_{m+1} - \boldsymbol{\beta}_{m}}$$

2.9. Tinjauan Non Statistika

Definisi yang digunakan dalam penelitian ini mengambil dari Konsep Definisi Pendataan Potensi Desa 2014 yaitu sebagai berikut:

Banjir adalah peristiwa terbenamnya daratan karena volume air yang meningkat. Banjir dapat terjadi karena luapan air yang berlebihan di suatu tempat akibat hujan besar, luapan air sungai atau pecahnya bendungan air. Kejadian banjir yang selalu terjadi di suatu desa/kelurahan karena luapan sungai atau sistem drainase yang buruk, seperti yang terjadi di daerah Marunda, Jakarta Utara tetap dikategorikan sebagai banjir, selama warga di daerah tersebut merasa terganggu dan mengalami kerugian.

Tempat sampah adalah tempat/wadah yang digunakan untuk menampung sampah yang berlokasi di sekitar halaman atau pagar bangunan dan terbuat dari tembok atau drum atau ember atau lubang besar dan sejenisnya, baik tertutup maupun terbuka.Tempat sampah, kemudian diangkut jika sampah ditampung sementara dalam wadah/tempat sampah yang kemudian sampah tersebut diangkut ke TPS atau langsung ke TPA.

Sedangkan jenis/cara membuang sampah menurut definisi dari Pendataan Potensi Desa 2014 adalah sebagai berikut:

1. Dalam lubang/dibakar jika sampah dibuang ke dalam lubang, baik lubang buatan maupun alamiah, atau sampah tersebut dibakar. Sungai/saluran irigasi/danau/laut jika sampah dibuang ke kali, sungai, saluran irigasi, danau, laut atau pinggir pantai
2. Drainase (got/selokan) jika sampah dibuang ke dalam saluran got/selokan yang pada dasarnya berfungsi sebagai saluran air.Lainnya misalnya sampah dikumpulkan kemudian dipakai sebagai bahan pembuatan kompos.

Topografi desa/kelurahan dilihat berdasarkan letak sebagian besar wilayah desa/ kelurahan, dibedakan menjadi:

1. Lereng adalah bagian dari gunung/bukit yang terletak di antara puncak sampai lembah. Lereng yang dimaksud juga mencakup punggung bukit dan puncak (bagian paling atas dari gunung).
2. Lembah adalah daerah rendah yang terletak di antara dua pegunungan atau dua gunung atau daerah yang mempunyai kedudukan lebih rendah dibandingkan daerah sekitarnya. Lembah di daerah pegunungan lipatan sering disebut sinklin. Lembah di daerah pegunungan patahan disebut graben atau slenk. Sedangkan lembah di daerah yang bergunung-gunung disebut lembah antar pegunungan.
3. Dataran adalah bagian atau sisi bidang tanah yang tampak datar, rata, dan membentang.

Informasi mengenai keberadaan permukiman di bantaran sungai yang mencakup banyaknya lokasi, bangunan rumah, dan keluarga yang bertempat tinggal di bantaran sungai.

Menurut PP No.38 tahun 2011,bantaran sungai adalah ruang antara tepi palung sungai dan kaki tanggul sebelah dalam yang terletak di kiri dan/atau kanan palung sungai. Garis sempadan adalah garis maya di kiri dan kanan palung sungai yang ditetapkan sebagai batas perlindungan sungai.

BAB 3
METODOLOGI PENELITIAN

3.1. Sumber Data

Data yang digunakan dalam penelitian ini adalah data frekuensi kejadian banjir, data Tempat Buang Sampah, data Topografi Wilayah dan data Keberadaan Pemukiman di Bantaran Sungai dari tiap desa di Jawa Timur dari Potensi Desa tahun 2014 di Provinsi Jawa Timur.

3.2. Variabel Penelitian

Variabel penelitian yang digunakan adalah :

1. Frekuensi terjadinya banjir pada tahun 2011 sampai dengan 2013 di masing-masing desa di Jawa Timur dari Potensi Desa tahun 2014 sebagai variabel Respon
2. Tempat Buang Sampah sebagai Variabel Prediktor
   dengan kategori sebagai berikut :
   1. Dalam Lubang/dibakar
   2. Selokan/Sungai
3. Topografi Wilayah sebagai Varibel Prediktor
   dengan kategori sebagai berikut :
   1. Lembah
   2. Lereng
   3. Dataran
4. Keberadaan Pemukiman di Bantaran Sungai sebagai Variabel Prediktor, menyatakan ada tidaknya pemukiman di bantaran sungai.

Struktur Data Longitudinal yang digunakan pada penelitian ini adalah sebagai berikut:

Tabel 3.1 Struktur Data Longitudinal

Desa	Observasi	Respon (y_{ij})	Tempat Buang Sampah (x_1)	Topografi (x_2)	Keberadaan Pemukiman di Bantaran Sungai (x_3)
1	1	y_{11}	x_{11}	x_{21}	x_{31}
1	2	y_{12}	x_{11}	x_{21}	x_{31}
1	3	y_{13}	x_{11}	x_{21}	x_{31}
2	1	y_{21}	x_{12}	x_{22}	x_{32}
2	2	y_{22}	x_{12}	x_{22}	x_{32}
2	3	y_{23}	x_{12}	x_{22}	x_{32}
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
n	1	y_{n1}	x_{n1}	x_{n2}	x_{n3}
n	2	y_{n2}	x_{n1}	x_{n2}	x_{n3}
n	3	y_{n3}	x_{n1}	x_{n2}	x_{n3}

3.3. Metode Penelitian

Sesuai dengan tujuan penelitian, yaitu untuk membentuk model regresi *count data* longitudinal. Langkah-langkah untuk mencapai tujuan pertama dalam penelitian ini adalah sebagai berikut :

1. Mengkaji Metode *Generalized Estimating Equation* untuk mengestimasi parameter dari Model Regresi Count dengan langkah-langkah sebagai berikut :
   a. Menyelesaikan persamaan sebagai berikut :

$$\sum_{i=1}^{N} \frac{\partial \mu_i^T}{\partial \beta} V_i^{-1}(\alpha,\beta)(Y_i - \mu_i) = 0$$

dimana

$$Y_i = (Y_{i1}, \ldots, Y_{in_i})^T$$

$$\mu_i = (\mu_{i1}, \ldots, \mu_{in_i})^T$$

$$V_i^{-1}(\alpha,\beta) = Var(Y_i) = A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}}$$

$$A_i = Diag[Var(Y_{ij})]_{j=1,\ldots,n_i}$$

b. Penentuan variabel laten u_{ij} dan estimasi u_{ij} dengan algoritma *Expectation Maximization (EM)*, misalkan untuk iterasi ke b, maka dapat dituliskan sebagai berikut :

$$u_{ij}^{(b)} = \Pr\left(u_{ij} = 1 \mid y_{ij}, \beta^{(b)}\right) = \frac{\Pr\left(u_{ij} = 1, y_{ij} = 0 \mid \mathrm{y}, \beta^{(b)}\right)}{\Pr\left(y_{ij} = 0 \mid \mathrm{y}, \beta^{(b)}\right)} \Big|_{y_{ij}=0}$$

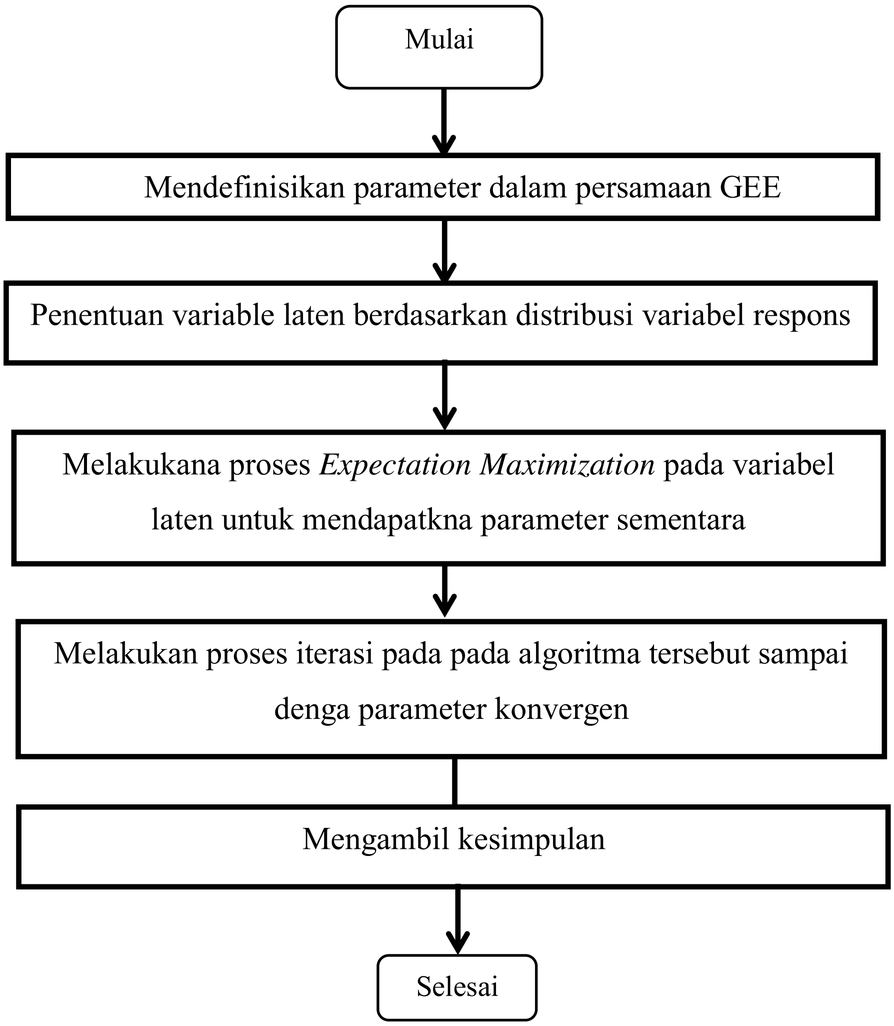
c. Menggunakan estimasi parameter iterasi ke b untuk memperbaharui β dalam persamaan berikut :

$$\beta^{(b+1)} = \beta^{(b)} + \left[\sum_{i=1}^{N} \frac{\partial \mu_i^T}{\partial \beta} \left(V_{\beta i}\right)^{-1} Diag(1 - u^{(b)}) \frac{\partial \mu_i}{\partial \beta^T}\right]^{-1} S_\beta$$

sampai dengan β konvergen

2. Pengambilan Kesimpulan

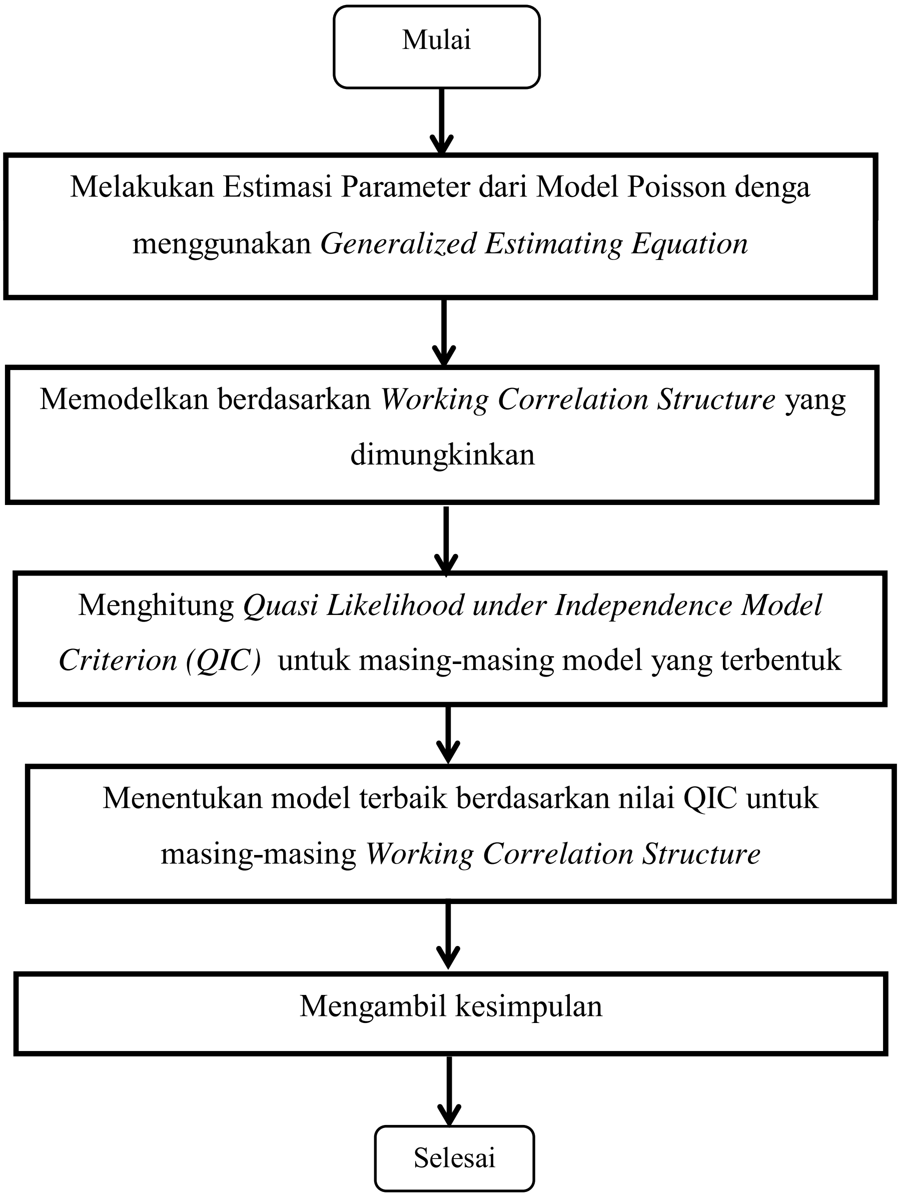
Tahapan pengkajian yang akan dilakukan untuk mencapai tujuan dari penelitian pertama adalah sebagai berikut:



Gambar 3.1 Bagan Tahapan Pengkajian

Adapun tahapan penelitian untuk mencapai tujuan kedua adalah sebagai berikut:

1. Melakukan *Exploratory Data Analysis* terhadap data kejadian banjir desa dari data PODES 2014
2. Menggunakan Metode *Generalized Estimating Equation* untuk mengestimasi parameter dari Model Regresi Count dengan langkah-langkah sebagai berikut :
   a. Melakukan Estimasi Parameter dari model regresi poisson
   b. Memodelkan berdasarkan masing-masing Working Correlation Stucture yang dimungkinkan :
      1. *Independent Structure*
      2. *Exchangeable Structure*
      3. *Stationary/m-dependent Structure*
      4. *Autoregressive Correlation Structure*
      5. *Unstructured Correlation*
   c. Menghitung nilai *Quasi Likelihood under Independence Model Criterion (QIC)* untuk masing-masing model berdasarkan *Working Correlation Stucture*
   d. Menentukan Model yang terbaik berdasarkan nilai *Quasi Likelihood under Independence Model Criterion (QIC)* yang dihasilkan.



Gambar 3.2 Bagan Tahapan Analisis

BAB 4

HASIL DAN PEMBAHASAN

4.1. Estimasi Parameter Model Regresi Poisson menggunakan GEE

Fungsi Distribusi Poisson adalah sebagai berikut :

$$P(Y_{ij} = y) = \begin{cases} \dfrac{e^{-\mu}\mu^{y_{ij}}}{y_{ij}!} & y_{ij} = 0,1,2\ldots \\ 0 & lainya \end{cases} \tag{4.1}$$

Mean dan variasi model Poisson adalah

$E(Y_i|x_i) = \mu_{ij}$ dan

$V(Y_i|x_i) = \mu_{ij}$

Jika diasumsikan bahwa log $(\mu_{ij}) = x_{ij}^T\beta$ dan (i=1,2,….,n; j=1,2….,n_i) maka nilai $E(Y_{ij})$ bergantung pada parameter β untuk memperhitungkan korelasi pada obyek yang sama, digunakan matriks korelasi yang dilambangkan $R_i(\alpha)$ untuk subyek ke i, maka estimasi dari β dengan *menggunakan Generalized Estimating Equation* (GEE) (Hall&Zhang, 2004):

$$\sum_{i=1}^{N} \frac{\partial\mu_i^T}{\partial\beta} V_i^{-1}(\alpha,\beta)(Y_i - \mu_i) = 0 \tag{4.2}$$

dimana $Y_i = (Y_{i1},\ldots,Y_{in_i})^T$, $\mu_i = (\mu_{i1},\ldots,\mu_{in_i})^T, V_i^{-1}(\alpha,\beta) = Var(Y_i) = A_i^{\frac{1}{2}}R_i(\alpha)A_i^{\frac{1}{2}}$ dan $A_i = Diag[Var(Y_{ij})]_{j=1,\ldots,n_i}$.

jika u adalah variabel laten untuk β maka estimasi untuk β adalah sebagai berikut

$$\sum_{i=1}^{N} \frac{\partial\lambda_i^T}{\partial\beta} (V_{\beta i})^{-1} Diag(1 - u_i)(y_i - \mu_i) = 0 \tag{4.3}$$

Di sini $\mu_i^T = \mu_{i1}, \dots \mu_{in_i}$ dan μ_{ij} didapatkan dari $log\ \mu_{ij} = x_{ij}^T \beta$, $\mu_{ij} = exp\ x_{ij}^T \beta$ dan $\frac{\partial \mu_{ij}}{\partial \beta} = x_{ij} exp\ x_{ij}^T \beta$. $V_{\gamma i} = D_i^{\frac{1}{2}} R_{2i}\ \alpha_2\ D_i^{\frac{1}{2}}$ adalah matriks varian dari Y_i dimana $D_i = Diag\ \mu_{i1}\ 1 + \tau\mu_{i1}\ , \mu_{i2}\ 1 + \tau\mu_{i2}\ , \dots, \mu_{i3}\ 1 + \tau\mu_{i3}$. dan $R_{2i}\ \alpha_2$ adalah matriks korelasi untuk Y_i ketika Y_i mengikuti distribusi *Poisson*. Matrix diagonal $Diag\ 1 - u_i$ menandakan bahwa hanya Y_{ij} yang berdistribusi *Poisson* lah yang diperhitungkan dalam persamaan tersebut. dan Y_i dari distribusi yang sudah dihilangkan nilai nol nya $(u_{ij} = 1)$ tidak diperhitungkan dalam persamaan tersebut. karena u_{ij} adalah variabel laten maka u_{ij} perlu diestimasi untuk setiap iterasi dengan algoritma *Expectation Maximization (EM).*

$$u_{ij}^{(b)} = \Pr\ u_{ij} = 1 \mid y_{ij}, \beta^{\ b} = \frac{\Pr\ u_{ij} = 1, y_{ij} = 0 \mid \mathrm{y}, \beta^{\ b}}{\Pr\ y_{ij} = 0 \mid \mathrm{y}, \beta^{\ b}} \mid_{y_{ij}=0} \qquad (4.4)$$

dimana $\lambda_{ij}^{(b)} = exp\ x_{ij}^T \beta^{(b)}$ sehingga β dapat diperbaharui dengan formula yang sudah diiterasi menjadi berikut :

$$\beta^{\ b+1} = \beta^{\ b} + \sum_{i=1}^{N} \frac{\partial \gamma_i^T}{\partial \beta}\ V_{\gamma i}^{\ -1} Diag\ 1 - u^{(b)}\ \frac{\partial \mu_i}{\partial \beta^T}^{\ -1} S_\beta \qquad (4.5)$$

dimana $S_\beta = \sum_{i=1}^{N} \frac{\partial \mu_i^T}{\partial \beta}\ V_{\beta i}^{\ -1} Diag\ 1 - u^{(b)}\ y_i - \mu_i$ dan β, α_2 dari persamaan terakhir di atas diganti dengan $\beta^{(b)}, \alpha_2{}^{(b)}$

Berikutnya estimasi untuk α_1 dan α_2 :

$$U_{ist}\ \gamma = \frac{u_{is} - p_{is}\ \ u_{it} - p_{it}}{p_{is}\ 1 - p_{is}\ p_{it}\ 1 - p_{it}} \qquad (4.6)$$

dimana $U_{\gamma i} = (U_{i12}, U_{i13}, \dots, U_{in_i - 1, n_i})^T$ dan $\rho_{\gamma i}\ \alpha_1 = E\ U_{\gamma i} = (\rho_{i12,} \rho_{i13,}, \dots, \rho_{in_i - 1, n_i})^T$. sehingga $\alpha_1$bisa di estimasi dengan persamaan berikut :

$$\sum_{i=1}^{N} E_{\gamma i}^T W_{\gamma i}^{-1}\ U_{\gamma i} - \rho_{\gamma i}\ \alpha_1 = 0 \qquad (4.7)$$

Dimana $E_{\gamma i} = \frac{\partial \rho_{\gamma i}\ \alpha_1}{\partial \alpha_i^T}$, dan $W_{\gamma i} \cong Cov(U_{\gamma i})$. menurut Liang dan Zeger (1986) $W_{\gamma i}$ bisa dianggap sebagai matriks identitas, dan karena $R_{\gamma i}(\alpha_i)$ adalah struktur compound yang simetris maka α_i didapatkan sebagai berikut:

$$\alpha_1 = \frac{1}{N^{\cdot}} \sum_{i=1}^{N} \sum_{s<t} \frac{u_{is} - p_{is} \quad u_{it} - p_{it}}{\overline{p_{is}\ 1 - p_{is}\ p_{it}\ 1 - p_{it}}} \tag{4.8}$$

dimana $N^* = \sum_{i=1}^{N} \frac{n_i(n_i-1)}{2}$

alternatif lainnya untuk estimasi α_i adalah

$$\alpha_1 = \frac{\frac{1}{N^*} \sum_{i=1}^{N} \sum_{s<t} \frac{u_{is} - p_{is} \quad u_{it} - p_{it}}{\overline{p_{is}\ 1 - p_{is}\ p_{it}\ 1 - p_{it}}}}{\frac{1}{N_{tot}} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{u_{ij} - p_{ij}}{p_{ij}\ 1 - p_{ij}}^{2}} \tag{4.9}$$

dimana $N_{tot} = \sum_{i=1}^{N} n_i$

Untuk estimasi α_2 , kita misalkan :

$$U_{ist}\ \beta\ = \frac{y_{is} - \mu_{is} \quad y_{it} - \mu_{it}}{\overline{\mu_{is}\ 1 - \tau\mu_{is}\ \mu_{it}\ 1 - \tau\mu_{it}}}$$

dan $U_{\beta i} = (U_{i12}, U_{i13}, \dots, U_{in_i-1,n_1})^T$ dimana mempunyai nilai ekspektasi $\rho_{\beta i}\ \alpha_2\ = E\ U_{\beta i}\ = (\rho_{i12}, \rho_{i13}, \dots, \rho_{in_i-1,n_i})^T$ dimana Y_i adalah berdistribusi *Generalized Poisson*, sehingga α_2 bisa diestimasi menggunakan persamaan berikut :

$$\sum_{i=1}^{N} E_{\beta i}^T W_{\beta i}^{-1}\ \ U_{\beta i} - \rho_{\beta i}\ \alpha_2 \quad = 0 \tag{4.10}$$

dimana $E_{\beta i} = \frac{\partial \rho_{\beta i(\alpha_2)}}{\partial \alpha_2^T}$, $W_{\beta i}\ \cong Cov\ \ U_{\beta i}$ dan

$H_{\beta i} = Diag\{\ 1 - u_{i1}\quad 1 - u_{i2}\ , \dots, (1 - u_{in_1-1})(1 - u_{in_1})$ dan dalam hal ini $W_{\beta i}$ adalah matriks indentitas dan $R_{\beta i}(\alpha_2)$ adalah struktur compound yang simetris maka α_2 didapatkan sebagai berikut:

$$\alpha_2 = \frac{1}{N^{\cdot}} \sum_{i=1}^{N} \sum_{s<t} \frac{(1-u_{is})(1-u_{it})(y_{is}-\mu_{is})(y_{it}-\mu_{it})}{\phi \sqrt{\mu_{is}(1+\mu_{is})\mu_{it}(1+\tau\mu_{it})}} \tag{4.11}$$

dimana $N^{*} = \sum_{i=1}^{N} \sum_{s<t}(1-u_{is})(1-u_{it})$ maka alternatif estimasi dari α_2 adalah :

$$\alpha_2 = \frac{1}{N^{\cdot}} \sum_{i=1}^{N} \sum_{s<t} \frac{\dfrac{(1-u_{is})(1-u_{it})(y_{is}-\mu_{is})(y_{it}-\mu_{it})}{\sqrt{\mu_{is}(1+\tau\mu_{is})\mu_{it}(1+\tau\mu_{it})}}}{\dfrac{1}{N_{tot}} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \dfrac{(1-\mu_{ij})^2 (y_{ij}-\mu_{ij})^2}{\mu_{ij}(1+\tau\mu_{ij})}} \tag{4.12}$$

dimana $N_{tot} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} (1-u_{ij})^2$

Untuk mendapatkan estimasi final dari β, α_1dan α_2 diperlukan metode iterasi dari persamaan diatas sampai konvergen. berikut proses iterasi yang digunakan

langkah pertama : Memberikan nilai awal untuk estimasi parameter dari β, α_1 dan α_2 yang dituliskan sebagai $\beta^{(0)}, \alpha_1^{(0)}$, dan $\alpha_2^{(0)}$; b diset = 0

Langkah kedua : perbaharui variabel laten $u_{ij} = (i = 1, \ldots, N; j = 1, \ldots n_i)$ untuk iterasi ke b :

$$u_{ij}^{(b)} = \left(1 + \frac{\dfrac{e^{-\mu}\mu_{ij}^{y_{ij}}}{y_{ij}!}}{p_{ij}^{(b)}}\right)^{-1} \Big|_{y_{ij}=0} \tag{4.13}$$

Langkah ketiga : menggunakan estimasi parameter iterasi ke b untuk memperbaharui β dalam persamaan sebagai berikut :

$$\beta^{(b+1)} = \beta^{(b)} + \left(\sum_{i=1}^{N} \frac{\partial \mu_i^T}{\partial \beta} V_{\beta i}^{-1} Diag(1-u^{(b)}) \frac{\partial \mu_i}{\partial \beta^T}\right)^{-1} S_\beta \tag{4.14}$$

dimana $S_\beta = \sum_{i=1}^{N} \frac{\partial \mu_i^T}{\partial \beta} V_{\beta i}^{-1} Diag\left(1-u^{(b)}\right)(y_i - \mu_i)$ dan β, α_2 diganti dengan $(\beta^{(b)}, \alpha_2^{(b)})$

Langkah kempat : perbaharui α_1 dengan persamaan berikut :

$$\alpha_1^{(b+1)} = \alpha_1^{(b)} + \left[\sum_{i=1}^{n} E_{\gamma i}^{T} W_{\gamma i}^{-1} E_{\gamma i}\right]^{-1} \sum_{i=1}^{n} E_{\gamma i}^{T} W_{\gamma i}^{-1} \left(U_{\gamma i} - \rho_{\gamma i}(\alpha_i)\right) \Big|_{\gamma, u_{ij} = \gamma^{(b+1)}, u_{ij}^{(b)}} \tag{4.15}$$

Karena $W_{\gamma i}$ adalah matrix indentitas, dan $R_{\gamma i}(\alpha_i)$ punya *compound symmetric structure*, maka α_1 bisa diestimasi sebagai berikut :

$$\alpha_1 = \frac{\frac{1}{N^*} \sum_{i=1}^{N} \sum_{s<t} \frac{(u_{is} - p_{is})(u_{it} - p_{it})}{\sqrt{p_{is}(1 - p_{is})\, p_{it}(1 - p_{it})}}}{\frac{1}{N_{tot}} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \left(\frac{u_{ij} - p_{ij}}{\sqrt{p_{ij}(1 - p_{ij})}}\right)^2} \Bigg|_{\gamma, u_{ij} = \gamma^{(b+1)}, u_{ij}^{(b)}} \tag{4.16}$$

Langkah kelima : perbaharui α_2 dengan persamaan sebagai berikut :

$$\alpha_2^{(b+1)} = \alpha_2^{(b)} + \left[\sum_{i=1}^{N} E_{\beta i}^{T} W_{\beta i}^{-1} H_{\beta i} E_{\beta i}\right]^{-1} S_\beta \tag{4.17}$$

dimana $S_\beta = \sum_{i=1}^{N} E_{\beta i}^{T} W_{\beta i}^{-1} H_{\beta i} \left(U_{\beta i} - \rho_{\beta i}(\alpha_2)\right)$ dan $(\beta, \gamma, \tau, u_{ij})$ pada sisi persamaan sebelah kana diganti dengan $(\beta^{(b)}, \gamma^{(b)}, \tau^{(b)}, u_{ij}^{(b)})$

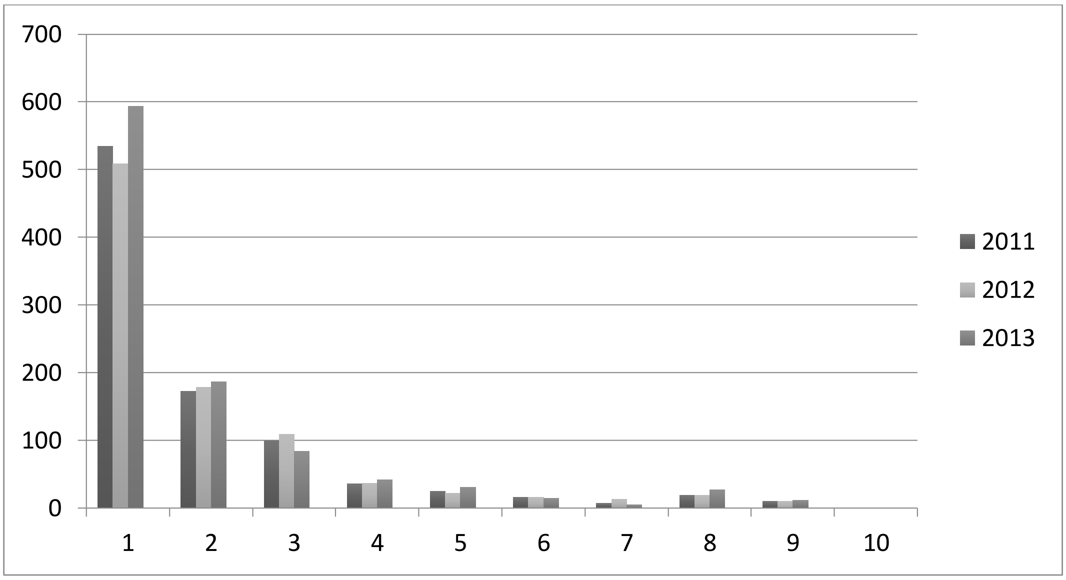
Langkah keenam : ulangi langkah ke dua sampai dengan langkah ke 5 sampai β dan γ konvergen.

4.2. Analisis Deskriptif

Provinsi Jawa Timur terdiri atas 29 Kabupaten dan 9 Kotamadya, dialiri oleh 2 sungai besar yaitu Sungai Brantas(290 km) dan Bengawan Solo(548 km). Dari segi topografi, 7309 desa di Jawa Timur berada di daerah dataran, 104 desa berada di daerah lembah dan sisanya 1089 desa berada di daerah lereng bukit dan pegunungan. Kebiasaan membuang sampah sebagian besar keluarga menurut hasil podes 2014 adalah 1156 desa menyatakan sebagian besar keluarga di desa tersebut mempunyai tempat sampah khusus dan diangkut ke TPS (Tempat Pembuangan Sementara) atau TPA (Tempat Pembuangan Akhir), 6878 desa menyatakan membuang ke lubang tanah baik alami maupun buatan atau dibakar, 173 desa menyatakan membuang sampah ke sungai/danau, 22 desa menyatakan membuat ke got (drainase/selokan) dan 273 desa menyatakan tidak tahu/lainya. 1433 Desa menyatakan adanya keberadaan pemukiman di bantaran sungai, dan 5540 desa menyatakan tidak adanya keberadaan pemukiman di bantaran sungai, sedangkan sisanya 1529 menyatakan tidak ada sungai yang melewati desa tersebut. Ada 8502 desa di Jawa Timur, tahun 2011 ada 2932 kejadian banjir yang dialami oleh 921 desa, 2012 ada 2916 kejadian banjir yang dialami oleh 914 desa dan tahun 2013 ada 3010 kejadian banjir yang dialami oleh 997 desa. jumlah ini menyatakan hampir sepertiga jumlah desa di Jawa Timur mengalami kejadian banjir pada tahun 2011 sampai 2013, hal ini menjadi petunjuk banyaknya nilai observasi nol dalam data kejadian banjir di Jawa Timur. Kabupaten-kabupaten yang mengalami banjir paling banyak adalah Kabupaten Bojonegoro, Gresik, Pasuruan dan Mojokerto

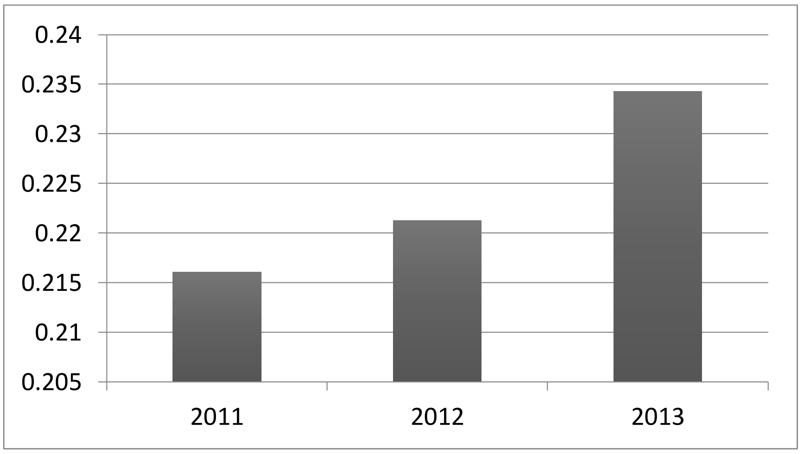
Tabel 4.1 Jumlah Kejadian Banjir di Kabupaten Kota di Jawa Timur 2011-2013

Kabupaten	Tahun		
	2011	2012	2013
BANGKALAN	16	10	15
BANYUWANGI	20	17	19
BATU	0	0	8
BLITAR	45	45	49
BOJONEGORO	256	265	268
BONDOWOSO	34	38	47
GRESIK	194	179	204
JEMBER	33	33	56
JOMBANG	106	97	93
KEDIRI	26	28	13
LAMONGAN	61	56	58
LUMAJANG	18	18	20
MADIUN	25	22	25
MAGETAN	4	5	6
MALANG	29	35	46
MOJOKERTO	124	120	125
NGANJUK	53	52	85
NGAWI	56	58	64
PACITAN	37	73	46
PAMEKASAN	16	21	32
PASURUAN	254	255	243
PONOROGO	60	58	45
PROBOLINGGO	57	59	49
SAMPANG	60	68	85
SIDOARJO	83	82	78
SITUBONDO	14	27	43
SUMENEP	11	17	26
SURABAYA	8	9	14
TRENGGALEK	71	69	61
TUBAN	34	35	44
TULUNGAGUNG	32	30	25
Total	**1837**	**1881**	**1992**



Gambar 4.1 Jumlah Kumulatif Banjir Masing-Masing Desa di Jawa Timur

Pada Gambar 4.1 terlihat bahwa jumlah kejadian banjir sebagian besar ada pada nilai 1 atau 1 kali kejadian banjir dalam satu tahun, dan jumlah desa yang terkena banjir lebih dari itu menjadi semakin sedikit di setiap tahunnya, dan berhenti di angka/jumlah 9 kejadian banjir setiap tahunya. tidak ada desa yang mengalami kejadian banjir sebanyak 10 kali atau lebih di Provinsi Jawa Timur selama tahun 2011 sampai dengan 2013



Gambar 4.2 Rata-Rata Jumlah Kejadian Banjir per Desa di Jawa Timur pada Tahun 2011-2013

Gambar 4.2 menggambarkan kenaikan rata-rata kejadian banjir per desa di Jawa Timur pada tahun 2011 sampai dengan 2013. disini terlihat ada tren kenaikan rata-rata kejadian banjir per desa di Jawa Timur pada tahun 2011 sampai 2013. pada tahun 2011 rata-rata kejadian banjir per desa adalah 0,216. kemudian meningkat menjadi 0,221 pada tahun 2012 dan pada tahun 2013 rata-rata tiap-tiap desa mengalami kejadian banjir sebanyak 0,234 kali.

Berdasarkan Chi Square Pearson test didapatkan bahwa dengan *p-value* 0.3223 data variabel respon dalam penelitian ini mengikuti distribusi Poisson. Dari hasil *overdispersion test*, ternyata hasilnya adalah gagal tolak H_0 sehingga bisa disimpulkan bahwa data kejadian banjir di desa di Jawa Timur 2011-2013 mengalami overdispesi, dengan nilai dispersinya 3,262677

4.3. Generalized Linear Model

Pembentukan Model dengan Generalized Linear Model menggunakan Program R package "glm". menghasilkan output sebagi berikut :

Model yang terbentuk adalah :

$$Log\ \ \pi_i\ = 0.574848 - 0.152662 topo3 + 0.010011 jmlbansu$$

Dari output menyatakan bahwa topografi suatu desa mempunyai pengaruh signifikan terhadap kecenderungan terjadinya banjir di desa tersebut. angka negatif dalam parameter tersebut bisa disimpulkan sebagai semakin kecil angka variabel topografi maka kecenderungan terjadinya banjir akan semakin besar. Sesuai dengan tinjauan definisi variabel topografi, bahwa angka 1 menyatakan lereng bukit/gunung, angka 2 menyatakan lembah, dan angka 3 menyatakan dataran rendah. desa yang terletak di lereng bukit/gunung mempunyai kecenderungan lebih rendah untuk terjadinya banjir dibandingkan dengan desa yang terletak di lembah dan dataran rendah. sebaliknya, Desa yang terletak di dataran rendah mempunyai kecenderungan mengalami banjir dibandingkan dengan desa di lereng bukit/gunung, dan lembah.

Variabel "jmlbansu" adalah keberadaan pemukiman di bantaran sungai. Jumlah pemukiman di bantaran sungai mempunyai pengaruh signifikan terhadap kecenderungan terjadinya banjir di suatu desa. angka positif dalam parameter tersebut menyatakan semakin tinggi nilai *variabel* ini, atau semakin banyak jumlah pemukiman di bantaran sungai maka kecenderungan banjir akan semakin tinggi di desa tersebut.

4.4. Pemodelan Regresi Poisson dengan GEE

Proses Estimasi parameter dalam penelitian ini menggunakan metode Generalized Estimating Equation yang pengolahanya dilakukan dengan program R package “gee”. menghasilkan parameter sebagai berikut.

Tabel 4.2 Hasil Estimasi Parameter GEE Berdasarkan Tipe *Working Correlation Structure*

WCS	Intercept	β_{11}	β_{12}	β_3
Exchangeable	-1.929929***	0.753414*	0.431713*	0.003610*
Independent	-1.929929***	0.753414*	0.431713*	0.003610***
AR1	-1.96492***	0.75605**	0.47553**	0.00365***
Unstructured	-1.947364***	0.766886*	0.446673**	0.003617***

simbol * menyatakan tingkat signifikansi

Dari tabel di atas, hasil dari estimasi parameter dengan Generalized Estimating Equation didapatkan bahwa faktor topografi (x_1), untuk (x_{11}) dan (x_{12}) serta pemukiman di bantaran sungai (x_2) dan mempunyai pengaruh yang signifikan terhadap kecenderungan terjadinya banjir di masing-masing desa di Jawa Timur. β_0 (*intercept*) untuk semua Working Correlation Structure memiliki pengaruh signifikan terhadap terjadinya banjir di desa-desa di Jawa Timur, artinya jika variabel dependen semua tidak diperhitungkan, maka kecenderungan positif terjadi banjir untuk setiap pertambahan tahun. artinya ada kecenderungan peningkatan rata-rata banjir tiap tahun di desa di Jawa Timur.

Setelah mendapatkan hasil pengolahan dengan berbagai Working Correlation Structure, maka langkah berikutnya adalah menentukan model terbaik dengan QIC (*Quasilikelihood Index Criterion*), dimana model dengan Working

Correlation Structure dengan nilai QIC paling kecil berarti model terbaik. Hasil dari pengolahanya adalah sebagai berikut :

Tabel 4.3 Hasil Penghitungan QIC Untuk Masing-Masing *Working Correlation Structure*

	Working Correlation Structure			
	Exchangeable	Ar1	Independent	Unstructured
QIC	28148.4	28148.8	28148.4	28148.9

Dari tabel di atas terlihat bahwa model GEE dengan *Working Correlation Structure* jenis “Unstructured” dan “Independent” menghasilkan nilai QIC yang paling kecil, walaupun perbedaan relatif kecil diantara semua *Working Correlaiton Structure* yang terbentuk. Hal ini menunjukkan bahwa Model GEE yang terbaik adalah Model GEE dengan *Working Correlation Structure* “ Unstructured”

Sehingga disimpulkan bahwa model GEE terbaik dalam penelitian ini adalah

$$Log\ \mu_i = -1.929929 + 0.753414topo2 + 0.431713topo3 + 0.003610pembansu$$

Dari model ini dapat disimpulkan bahwa jumlah kejadian banjir di Jawa Timur dipengaruhi oleh variabel yang digunakan dalam penelitian ini, yaitu topografi, dan keberadaan pemukiman di bantaran sungai. Topografi jenis 2 yaitu lembah mempunyai kecenderungan mempengaruhi rata-rata terjadinya banjir di desa-desa di jawa timur dibandingkan dengan topografi jenis 1 yaitu lereng, dengan nilai odds ratio 2.08 maka berarti peluang desa dengan topografi jenis lembah 2.08 kali lebih besar dibandingkan dengan desa dengan topografi lereng. .Begitu juga dengan topografi jenis 3 yaitu dataran terhadap lereng, dengan nilai odds ratio 1.53 maka desa dengan topografi dataran mempunyai peluang terjadi banjil lebih banyak 1.53 kali dibanding dengan desa dengan topografi lereng. Keberadaan pemukiman di bantaran sungai mempunyai kecenderungan terjadinya banjir dibandingkan dengan desa yang tidak ada pemukiman di bantaran sungainya. Dengan odds ratio 1.003 maka berarti desa yang ada pemukiman dibantaran sungai mempunyai 1.003 peluang lebih besar mengalami banjir dibandingkan dengan desa yang tidak ada pemukiman di bantaran sungai.

.

BAB 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan analisis dan pembahasan pada bab sebelumnya, dapat disimpulkan bahwa :

1. Estimator Parameter untuk Regresi Poisson menggunakan *Generalized Estimating Equation* adalah

$$\beta^{\,b+1} = \beta^{\,b} + \left(\sum_{i=1}^{N} \frac{\partial \mu_i^T}{\partial \beta} V_{\beta i}^{-1} Diag(1-u^{(b)}) \frac{\partial \mu_i}{\partial \beta^T} \right)^{-1} S_\beta$$

dengan estimasi α untuk Working Correlation Structurenya :

$$\alpha_1 = \frac{\frac{1}{N^*} \sum_{i=1}^{N} \sum_{s<t} \frac{u_{is}-p_{is} \quad u_{it}-p_{it}}{p_{is} \; 1-p_{is} \; p_{it} \; 1-p_{it}}}{\frac{1}{N_{tot}} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{u_{ij}-p_{ij}}{p_{ij} \; 1-p_{ij}}^{2}} \Big|_{\gamma, u_{ij} \,=\, \gamma^{(b+1)}, u_{ij}^{(b)}}$$

2. Rata-rata Jumlah Kejadian banjir di desa-desa di jawa timur dipengaruhi secara signifikan oleh karakteristik desa sebagai berikut : Topografi, dan Jumlah Pemukiman di Bantaran Sungai. model yang terbentuk berdasarkan nilai QIC maka WCS yang terbaik adalah "*Unstructured"* dan *"Independent*"

5.2. Saran

Berdasakan penelitian yang dilakukan, saran yang dapat diberikan adalah sebagai berikut :

1. Penelitian ini menggunakan Metode Generalized Estimating Equation sebagai estimasi parameter untuk model regresi count poisson, untuk berikutnya bisa dipertimbankan penggunaan metode estimasi parameter yang lain, GEE2, Pendekatan Bayesian dan lain lain
2. Regresi Count untuk data yang mengandung dispersi dan Zero-Inflated diatasi dengan model Zero Inflated Generalized Poisson dan dapat

dipertimbangkan penggunaan metode lain seperti Zero Inflated Negative Binomial, Regresi Hurdle dan lain lain

3. Penelitian terhadap variabel-variabel lain yang mempengaruhi jumlah kejadian banjir
4. Penggunaan jenis *Working Correlation Structure* lainya untuk penelitian menggunakan *Generalized Estimating Equation* selanjutnya.
5. Pengujian dan inferensia statistik dengan pengujian hipotesis dapat dilakukan dalam penelitian selanjutnya.

DAFTAR PUSTAKA

Agresti, A. (2002), *Categorical Data Analysis Second Edition*, John Wiley & Sons, New York.

Ariani, N. (2014). *Perbandingan Regresi Zero Inflated Generalized Poisson (ZIGP) dan Regresi Zero Inflated Negative Binomial (ZINB) pada data Overdispersion*. Universitas Brawijaya

Chang Y.C. (2000). *Residuals analysis of the generalized linear models for longitudinal data*. *Statistics in Medicine* 19(10):1277–1293.

Cupal, M., Deev, dan O., Linnertova, D. (2015). *The Poisson Regression Analysis for Occurrence of Flood*. *Procedia Economic and Finance* 23 (2015) 1499-1502

Czado, C. dan Min, A. (2006). *Testing for Zero Modification in Count Regression Models*. Paper 474 dari SFB 386 (http://epub.ub.uni-munchen.de/).

Diggle, P. J. dan Kenward, M. G. (1994). *Informative drop-out in longitudinal data analysis (with discussion)*. Applied Statistics 43, 49–93

Erhardt, V. (2008), *Zero Inflated Generalized Poisson (ZIGP) Regression Models*, The Comprehensice R Archive Network 7 Oktober 2008, http://cran.r-project.org

Erhardt, Vincenz dan Czado, C, *Generalized Estimating Equations for Longitudinal Generalized Poisson Count Data with Regression Effect on the Mean and Dipersion Level*. Lehrstuhl fur Mathematische Statistik. Technische Universitat Munchen. Garching

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.

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) 117-130.

Fisher, LD. (2004). *Biostatistics*. *A Methodology for the health sciences*, Second ed., John Wiley& Sons, Inc., New Jersey, pp 728-764.

Handayanti, K.W. dan Mitakda, M.B. (2015). Kajian Metode *Generalized*

Estimating Equation (GEE) dalam Pendugaan Parameter Model Regresi Multilevel. Universitas Brawijaya. Malang

Hedeker, D. dan Gibbons, R. D. (2006). *Longitudinal data analysis*. Hoboken, NJ: John Wiley & Sons.

Hilbe, J. M. (2007). *Negative Binomial Regression*, vol. 1. Cambridge University Press, New York

Indarto. Susanto, Boedi. Huda, Hisbullah. (2012), *Studi Tentang Karakteristik Fisik dan Frekuensi Banjir Pada 15 DAS di Jawa Timur*, Universitas Jember.

Kodoatie, R.J dan R. Sjarief. *Hidrolika Terapan Aliran Pada Saluran Terbuka Dan Pipa*. Yogyakarta: Andi. (2002).

Kodoatie, R.J dan Sugiyanto, 2002. *Banjir: Beberapa Penyebab dan Metode Pengendaliannya, Dalam Perspektif Lingkungan*. Yogyakarta. Pustaka Pelajar

Lam, K. F., Xue, H., dan Cheung, Y.B. (2006), *Semiparametric Analysis of Zero Inflated Count Data*. Biometrics 62, 996 1003.

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

Liza, Mirna. (2015), *Pengelolaan Dampak Banjir Secara Komprehensif di Kabupaten Bojonegoro Jawa Timur*. Universitas Gajah Mada.

Liang, K.Y. dan Zeger, S. L. (1986*). Longitudinal Data Analysis Using Generalized Linear Models*. Biometrika 73, 13–22.

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

Mood, A.M., Graybill, F.A. dan Boes, D.C. (1974), *Introduction to The Theory of Statistics* Third Edition, McGraw-Hill, Singapura.

Rahmawati, Ima Puspita. (2008), *Sistem Pengendalian Banjir Sungai Sengkarang (Normalisali Sungai)*. Universitas Diponegoro

Ruru, Y. dan Barrios, E.B. (2003), *Poisson Regression Models of Malaria Incidence in Jayapura*, Indonesia, The Philippine Statistician, Vol. 52, Nos.1-4, pp. 27-38.

Ridout, M., Demetrio, C. G. B. dan Hinde, J. (1998). *Models for count data with* *many zeros, International Biometric Conference*, Cape Town, Desember 1998

Ridout, M., Hinde, J. dan Demetrio, C.G.B. (2001), *A Score Test for Testing A Zero-Inflated Poisson Regression Models Againts Zero-Inflated Negatif Binomial Alternatives*, Biometrics 57: 219-223.

Twisk, J. (2003), *Applied Longitudinal Data Analysis for Epidemiology*, Cambridge University Press, New York.

Wu, L. 2010. *Mixed Effect Models for Complex data*. New York: CRC Press, Taylor dan Francis Group.

Wu, H., dan Zhang, J.T. 2006. *Nonparametric Regression Methods for Longitudinal Data Analysis*. New Jersey: John Wiley dan Sons, Inc.

Lampiran 1

Pearson's Chi-squared test

```
data:  dtt3$banjir and dtt3$jumlah
X-squared = 19.918, df = 18, p-value = 0.3374
```

Overdispersion test

```
data:  m.glm
z = 18.858, p-value < 2.2e-16
alternative hypothesis: true dispersion is greater than 1
sample estimates:
dispersion
  3.262677
```

```
#GEEPOISSON tanpa 0#
geeex<-geeglm(formula = jumlah ~ topo + buangs + jmlbansurt,
 family = "poisson", data = BANJIR, id = ID, corstr =
 "exchangeable")
geear1<-geeglm(formula = jumlah ~ topo + buangs +
 jmlbansurt, family = "poisson", data = BANJIR, id = ID,
 corstr = "ar1")
geein<-geeglm(formula = jumlah ~ topo + buangs + jmlbansurt,
 family = "poisson", data = BANJIR, id = ID, corstr =
 "independence")
geeun<-geeglm(formula = jumlah ~ topo + buangs + jmlbansurt,
 family = "poisson", data = BANJIR, id = ID, corstr =
 "unstructured")
B.glm<-
 glm(jumlah~topo+buangs+jmlbansu,data=BANJIR,family=poisson)
summary(geeex)
summary(geear1)
summary(geein)
summary(geeun)
summary(B.glm)

#GLM MLE POISSON#
B.glm2<-glm(jumlah~topo+jmlbansu,data=BANJIR,family=poisson)
summary(B.glm2)

#QIC
QIC = function(model.R) {
  library(MASS)
  model.indep = update(model.R, corstr = "independence")
  # Quasilikelihood
  mu.R = model.R$fitted.values
  y = model.R$y
```

```
  type = family(model.R)$family
  quasi.R = switch(type,
                   poisson = sum((y*log(mu.R)) - mu.R),
                   gaussian = sum(((y - mu.R)^2)/-2),
                   binomial = sum(y*log(mu.R/(1 - mu.R)) +
 log(1 - mu.R)),
                   Gamma = sum(-y/(mu.R - log(mu.R))),
                   stop("Error: distribution not
 recognized"))
  # Trace Term (penalty for model complexity)
  omegaI = ginv(model.indep$geese$vbeta.naiv) # Omega-hat(I)
 via Moore-Penrose generalized inverse of a matrix in MASS
 package
  #AIinverse = solve(model.indep$geese$vbeta.naiv) # solve
 via indenity
  Vr = model.R$geese$vbeta
  trace.R = sum(diag(omegaI %*% Vr))
  px = length(mu.R) # number non-redunant columns in design
 matrix
  # QIC
  QIC = 2*(trace.R - quasi.R)
  #QICu = (-2)*quasi.R + 2*px    # Approximation assuming
 model structured correctly
  output = c(QIC, quasi.R, trace.R, px)
  names(output) = c('QIC', 'Quasi Lik', 'Trace', 'px')
  return(output)
}


sapply(list(geeex, geeun, geein, geear1), QIC)


#ZIP
ZIP<-
 zeroinfl(formula=jumlah~topo+buangs+jmlbansu,dist="poisson"
 ,data=dtt3)
summary(ZIP)

#GEEPOISSON dengan 0 #
gee0ex<-geeglm(formula = jumlah ~ topo + buangs +
 jmlbansurt, family = "poisson", data = dtt3, id = ID,
 corstr = "exchangeable")
gee0ar1<-geeglm(formula = jumlah ~ topo + buangs +
 jmlbansurt, family = "poisson", data = dtt3, id = ID,
 corstr = "ar1")
gee0ind<-geeglm(formula = jumlah ~ topo + buangs +
 jmlbansurt, family = "poisson", data = dtt3, id = ID,
 corstr = "independence")
gee0un<-geeglm(formula = jumlah ~ topo + buangs +
 jmlbansurt, family = "poisson", data = dtt3, id = ID,
 corstr = "unstructured")
```

```
summary(gee0ex)
summary(gee0ar1)
summary(gee0ind)
summary(gee0un)

#GEEPOISSON dengan 0 variabel signifikan#
gee0ex2<-geeglm(formula = jumlah ~ topo + jmlbansurt, family
 = "poisson", data = dtt3, id = ID, corstr = "exchangeable")
gee0ar12<-geeglm(formula = jumlah ~ topo + jmlbansurt,
 family = "poisson", data = dtt3, id = ID, corstr = "ar1")
gee0ind2<-geeglm(formula = jumlah ~ topo + jmlbansurt,
 family = "poisson", data = dtt3, id = ID, corstr =
 "independence")
gee0un2<-geeglm(formula = jumlah ~ topo + jmlbansurt, family
 = "poisson", data = dtt3, id = ID, corstr = "unstructured")

summary(gee0ex2)
summary(gee0ar12)
summary(gee0ind2)
summary(gee0un2)

#QIC
QIC = function(model.R) {
  library(MASS)
  model.indep = update(model.R, corstr = "independence")
  # Quasilikelihood
  mu.R = model.R$fitted.values
  y = model.R$y
  type = family(model.R)$family
  quasi.R = switch(type,
                   poisson = sum((y*log(mu.R)) - mu.R),
                   gaussian = sum(((y - mu.R)^2)/-2),
                   binomial = sum(y*log(mu.R/(1 - mu.R)) +
 log(1 - mu.R)),
                   Gamma = sum(-y/(mu.R - log(mu.R))),
                   stop("Error: distribution not
 recognized"))
  # Trace Term (penalty for model complexity)
  omegaI = ginv(model.indep$geese$vbeta.naiv) # Omega-hat(I)
 via Moore-Penrose generalized inverse of a matrix in MASS
 package
  #AIinverse = solve(model.indep$geese$vbeta.naiv) # solve
 via indenity
  Vr = model.R$geese$vbeta
  trace.R = sum(diag(omegaI %*% Vr))
  px = length(mu.R) # number non-redunant columns in design
 matrix
  # QIC
  QIC = 2*(trace.R - quasi.R)
  #QICu = (-2)*quasi.R + 2*px    # Approximation assuming
 model structured correctly
  output = c(QIC, quasi.R, trace.R, px)
```

```
  names(output) = c('QIC', 'Quasi Lik', 'Trace', 'px')
  return(output)
}


sapply(list(gee0ex2, gee0un2, gee0ind2, gee0ar12), QIC)
```

OUTPUT

```
Call:
geeglm(formula = jumlah ~ factor(topo) + buangs +
 jmlbansurt,
    family = "poisson", data = BANJIR, id = ID, corstr =
 "exchangeable")

 Coefficients:
               Estimate   Std.err   Wald Pr(>|W|)
(Intercept)    0.6573973  0.1684969 15.222 9.56e-05 ***
factor(topo)2  0.0531249  0.2212208  0.058    0.810
factor(topo)3 -0.1722873  0.1332154  1.673    0.196
buangs        -0.0240812  0.0548999  0.192    0.661
jmlbansurt    -0.0004632  0.0003788  1.496    0.221
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
 1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)    1.897  0.1012

Correlation: Structure = exchangeable  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha   0.8074 0.01577
Number of clusters:   1218   Maximum cluster size: 3
> summary(geear1)

Call:
```

```
geeglm(formula = jumlah ~ factor(topo) + buangs +
 jmlbansurt,
    family = "poisson", data = BANJIR, id = ID, corstr =
 "ar1")

 Coefficients:
                Estimate   Std.err  Wald Pr(>|W|)
(Intercept)     0.625459  0.171297 13.33  0.00026 ***
factor(topo)2  0.054520  0.221304  0.06  0.80541
factor(topo)3 -0.133372  0.135788  0.96  0.32599
buangs        -0.023844  0.054833  0.19  0.66368
jmlbansurt    -0.000396  0.000367  1.16  0.28063
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
 1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     1.89   0.101

Correlation: Structure = ar1  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.859  0.0118
Number of clusters:   1218   Maximum cluster size: 3
> summary(geein)

Call:
geeglm(formula = jumlah ~ factor(topo) + buangs +
 jmlbansurt,
    family = "poisson", data = BANJIR, id = ID, corstr =
 "independence")

 Coefficients:
                Estimate   Std.err  Wald Pr(>|W|)
(Intercept)     0.657397  0.168497 15.22  9.6e-05 ***
factor(topo)2  0.053125  0.221221  0.06     0.81
factor(topo)3 -0.172287  0.133215  1.67     0.20
buangs        -0.024081  0.054900  0.19     0.66
jmlbansurt    -0.000463  0.000379  1.50     0.22
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
 1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)      1.9   0.101

Correlation: Structure = independenceNumber of clusters:
 1218   Maximum cluster size: 3
> summary(geeun)
```

```
Call:
geeglm(formula = jumlah ~ factor(topo) + buangs +
 jmlbansurt,
    family = "poisson", data = BANJIR, id = ID, corstr =
 "unstructured")

 Coefficients:
                 Estimate   Std.err  Wald Pr(>|W|)
(Intercept)      0.652736  0.168608 14.99  0.00011 ***
factor(topo)2    0.053369  0.220685  0.06  0.80891
factor(topo)3   -0.166623  0.133355  1.56  0.21149
buangs          -0.024043  0.054849  0.19  0.66114
jmlbansurt      -0.000453  0.000376  1.45  0.22852
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
 1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)      1.9   0.101

Correlation: Structure = unstructured  Link = identity

Estimated Correlation Parameters:
          Estimate Std.err
alpha.1:2    0.817  0.0216
alpha.1:3    0.787  0.0213
alpha.2:3    0.818  0.0210
Number of clusters:   1218   Maximum cluster size: 3
> summary(B.glm)

Call:
glm(formula = jumlah ~ factor(topo) + buangs + jmlbansu,
 family = poisson,
    data = BANJIR)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.990  -0.619  -0.451   0.370   4.173

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      0.63849    0.06891    9.27  < 2e-16 ***
factor(topo)2    0.05440    0.11072    0.49  0.62316
factor(topo)3   -0.16853    0.04835   -3.49  0.00049 ***
buangs          -0.02505    0.02080   -1.20  0.22852
jmlbansu         0.01017    0.00451    2.25  0.02420 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
 1
```

```
(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 5978.5  on 3653  degrees of freedom
Residual deviance: 5958.4  on 3649  degrees of freedom
AIC: 12830

Number of Fisher Scoring iterations: 5

>
> #GLM MLE POISSON#
> B.glm2<-
 glm(jumlah~factor(topo)+jmlbansu,data=BANJIR,family=poisson
 )
> summary(B.glm2)

Call:
glm(formula = jumlah ~ factor(topo) + jmlbansu, family =
 poisson,
    data = BANJIR)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.997  -0.636  -0.454   0.367   4.123

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.57485    0.04468   12.87   <2e-16 ***
factor(topo)2  0.06520    0.11037    0.59   0.5547
factor(topo)3 -0.15266    0.04661   -3.28   0.0011 **
jmlbansu       0.01001    0.00452    2.21   0.0269 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 5978.5  on 3653  degrees of freedom
Residual deviance: 5959.9  on 3650  degrees of freedom
AIC: 12829

Number of Fisher Scoring iterations: 5

>
> sapply(list(geeex, geeun, geein, geear1), QIC)
             [,1]    [,2]    [,3]  [,4]
QIC        6330.0  6330.0  6330.0  6331
Quasi Lik -3151.1 -3151.1 -3151.1 -3152
Trace        13.9    13.9    13.9    14
px         3654.0  3654.0  3654.0  3654
>

Call:
```

```
geeglm(formula = jumlah ~ factor(topo) + buangs +
 jmlbansurt,
    family = "poisson", data = dtt3, id = ID, corstr =
 "exchangeable")

 Coefficients:
                Estimate   Std.err   Wald Pr(>|W|)
(Intercept)    -1.940156  0.183519 111.77   <2e-16 ***
factor(topo)2  0.753250  0.329919   5.21   0.0224 *
factor(topo)3  0.433299  0.163960   6.98   0.0082 **
buangs         0.004450  0.069169   0.00   0.9487
jmlbansurt     0.003613  0.000391  85.21   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
 1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     3.29   0.183

Correlation: Structure = exchangeable  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.887  0.0598
Number of clusters:   8502   Maximum cluster size: 3
> summary(gee0ar1)

Call:
geeglm(formula = jumlah ~ factor(topo) + buangs +
 jmlbansurt,
    family = "poisson", data = dtt3, id = ID, corstr =
 "ar1")

 Coefficients:
                Estimate   Std.err   Wald Pr(>|W|)
(Intercept)    -1.972654  0.187012 111.27   <2e-16 ***
factor(topo)2  0.755930  0.330345   5.24   0.0221 *
factor(topo)3  0.476734  0.166811   8.17   0.0043 **
buangs         0.003362  0.069128   0.00   0.9612
jmlbansurt     0.003658  0.000392  86.98   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
 1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     3.29   0.197

Correlation: Structure = ar1  Link = identity

Estimated Correlation Parameters:
```

```
      Estimate Std.err
alpha     0.92   0.045
Number of clusters:   8502   Maximum cluster size: 3
> summary(gee0ind)

Call:
geeglm(formula = jumlah ~ factor(topo) + buangs +
 jmlbansurt,
    family = "poisson", data = dtt3, id = ID, corstr =
 "independence")

 Coefficients:
              Estimate   Std.err   Wald Pr(>|W|)
(Intercept)   -1.940156  0.183519 111.77   <2e-16 ***
factor(topo)2  0.753250  0.329919   5.21   0.0224 *
factor(topo)3  0.433299  0.163960   6.98   0.0082 **
buangs         0.004450  0.069169   0.00   0.9487
jmlbansurt     0.003613  0.000391  85.21   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
 1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     3.29   0.183

Correlation: Structure = independenceNumber of clusters:
 8502   Maximum cluster size: 3
> summary(gee0un)

Call:
geeglm(formula = jumlah ~ factor(topo) + buangs +
 jmlbansurt,
    family = "poisson", data = dtt3, id = ID, corstr =
 "unstructured")

 Coefficients:
             Estimate  Std.err   Wald Pr(>|W|)
(Intercept)   -1.95934  0.18455 112.72   <2e-16 ***
factor(topo)2  0.76670  0.33161   5.35   0.0208 *
factor(topo)3  0.44851  0.16532   7.36   0.0067 **
buangs         0.00522  0.06920   0.01   0.9399
jmlbansurt     0.00362  0.00039  86.16   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
 1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     3.31   0.189

Correlation: Structure = unstructured  Link = identity
```

```
Estimated Correlation Parameters:
          Estimate Std.err
alpha.1:2    0.873  0.0613
alpha.1:3    0.877  0.0617
alpha.2:3    0.913  0.0629
Number of clusters:   8502   Maximum cluster size: 3
>
> #GEEPOISSON dengan 0 variabel signifikan#
> gee0ex2<-geeglm(formula = jumlah ~ factor(topo) +
 jmlbansurt, family = "poisson", data = dtt3, id = ID,
 corstr = "exchangeable")
> gee0ar12<-geeglm(formula = jumlah ~ factor(topo) +
 jmlbansurt, family = "poisson", data = dtt3, id = ID,
 corstr = "ar1")
> gee0ind2<-geeglm(formula = jumlah ~ factor(topo) +
 jmlbansurt, family = "poisson", data = dtt3, id = ID,
 corstr = "independence")
> gee0un2<-geeglm(formula = jumlah ~ factor(topo) +
 jmlbansurt, family = "poisson", data = dtt3, id = ID,
 corstr = "unstructured")
>
> summary(gee0ex2)

Call:
geeglm(formula = jumlah ~ factor(topo) + jmlbansurt, family
 = "poisson",
    data = dtt3, id = ID, corstr = "exchangeable")

 Coefficients:
               Estimate   Std.err   Wald Pr(>|W|)
(Intercept)   -1.929929  0.167468 132.81   <2e-16 ***
factor(topo)2  0.753414  0.329277   5.24    0.022 *
factor(topo)3  0.431713  0.171873   6.31    0.012 *
jmlbansurt     0.003610  0.000389  86.09   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
 1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     3.29    0.17

Correlation: Structure = exchangeable  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.887  0.0579
Number of clusters:   8502   Maximum cluster size: 3
> summary(gee0ar12)

Call:
```

```
geeglm(formula = jumlah ~ factor(topo) + jmlbansurt, family
 = "poisson",
    data = dtt3, id = ID, corstr = "ar1")

 Coefficients:
              Estimate  Std.err   Wald Pr(>|W|)
(Intercept)   -1.96492  0.17018 133.32   <2e-16 ***
factor(topo)2  0.75605  0.32975   5.26   0.0219 *
factor(topo)3  0.47553  0.17444   7.43   0.0064 **
jmlbansurt     0.00365  0.00039  87.87   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
 1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     3.29   0.181

Correlation: Structure = ar1  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha     0.92  0.0433
Number of clusters:   8502   Maximum cluster size: 3
> summary(gee0ind2)

Call:
geeglm(formula = jumlah ~ factor(topo) + jmlbansurt, family
 = "poisson",
    data = dtt3, id = ID, corstr = "independence")

 Coefficients:
               Estimate   Std.err   Wald Pr(>|W|)
(Intercept)   -1.929929  0.167468 132.81   <2e-16 ***
factor(topo)2  0.753414  0.329277   5.24    0.022 *
factor(topo)3  0.431713  0.171873   6.31    0.012 *
jmlbansurt     0.003610  0.000389  86.09   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
 1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     3.29    0.17

Correlation: Structure = independenceNumber of clusters:
 8502   Maximum cluster size: 3
> summary(gee0un2)

Call:
geeglm(formula = jumlah ~ factor(topo) + jmlbansurt, family
 = "poisson",
```

```
    data = dtt3, id = ID, corstr = "unstructured")

 Coefficients:
               Estimate   Std.err   Wald Pr(>|W|)
(Intercept)   -1.947364  0.168878 132.97   <2e-16 ***
factor(topo)2  0.766886  0.330962   5.37   0.0205 *
factor(topo)3  0.446673  0.173233   6.65   0.0099 **
jmlbansurt     0.003617  0.000388  87.00   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
 1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     3.31   0.175

Correlation: Structure = unstructured  Link = identity

Estimated Correlation Parameters:
          Estimate Std.err
alpha.1:2    0.873  0.0594
alpha.1:3    0.877  0.0597
alpha.2:3    0.913  0.0610
Number of clusters:   8502   Maximum cluster size: 3
>
> sapply(list(gee0ex2, gee0un2, gee0ind2, gee0ar12), QIC)
              [,1]     [,2]     [,3]     [,4]
QIC        28148.4  28148.8  28148.4  28149.9
Quasi Lik -14063.1 -14063.2 -14063.1 -14063.7
Trace         11.1     11.2     11.1     11.2
px         25506.0  25506.0  25506.0  25506.0
```

BIOGRAFI PENULIS

Penulis dilahirkan di Yogyakarta pada tanggal 26 Maret 1983 dan putri pertama dari pasangan suami istri Bapak Nyono Sunarso dan Ibu Sumaryanti. Saat ini penulis telah berkeluarga dengan istri bernama Mariatul Kiftiyah dan telah dikaruniai dengan satu putri Mahya Aretya Diah Garini dan satu orang putra Satrio Pinandito .

Riwayat pendidikan penulis diawali dari SDN Pujokusuman, SMPN 2 Yogyakarta SMUN 6 Yogyakarta dan Sekolah Tinggi Ilmu Statistik (STIS) Jakarta jurusan Statistik Ekonomi. Setelah menamatkan pendidikan D-IV STIS, Pembaca yang ingin memberikan kritik, saran dan pertanyaan mengenai penelitian ini dapat menghubungi penulis melalui email arifsetiawan1982@gmail.com