

**GENERALIZED ESTIMATING EQUATIONS (GEE) PADA
DATA LONGITUDINAL BERSIFAT ORDINAL**

SKRIPSI

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

Oleh:

AZIZAH HIDAYATI

0810950029-95



**PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2013**

LEMBAR PENGESAHAN SKRIPSI

**GENERALIZED ESTIMATING EQUATIONS (GEE) PADA
DATA LONGITUDINAL BERSIFAT ORDINAL**

oleh:
AZIZAH HIDAYATI
08109500029-95

Setelah dipertahankan di depan Majelis Penguji
pada tanggal 15 Januari 2013
dan dinyatakan memenuhi syarat untuk memperoleh gelar
Sarjana Sains dalam bidang Statistika

Dosen Pembimbing I

Dosen Pembimbing II

Dr. Ir. M. Bernadetha Mitakda
NIP. 19520521 198103 2 001

Dr. Ir. Ni Wayan Surya W., MS
NIP. 19551102 198103 2 001

Mengetahui,
Ketua Jurusan Matematika
Fakultas MIPA
Universitas Brawijaya

Dr. Abdul Rouf Alghofari, M.Sc
NIP. 19670907 199203 1 001

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : AZIZAH HIDAYATI
NIM : 0810950029-95
Jurusan : Matematika
Program Studi : STATISTIKA
Skripsi berjudul :

GENERALIZED ESTIMATING EQUATIONS (GEE) PADA DATA LONGITUDINAL BERSIFAT ORDINAL

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 akan bersedia menanggung segala resiko.

Demikian pernyataan ini saya buat dengan segala kesadaran.

Malang, 15 Januari 2012

Yang menyatakan,

(AZIZAH HIDAYATI)

NIM. 0810950029-95

GENERALIZED ESTIMATING EQUATIONS (GEE) PADA DATA LONGITUDINAL BERSIFAT ORDINAL

ABSTRAK

Regresi Logistik Ordinal digunakan untuk mengetahui bentuk hubungan antar satu peubah respon bersifat ordinal dengan satu atau lebih peubah prediktor berupa kategorik maupun kontinyu. Banyak kategori peubah respon dikotom dan politom. Model ini diterapkan pada data longitudinal berupa hasil pengamatan beberapa unit *cross-sectional* selama beberapa periode waktu tertentu, karena adanya pengamatan berulang menyebabkan autokorelasi, sehingga perlu diatasi dengan metode *Generalized Estimating Equations* (GEE). Penelitian bertujuan untuk menerapkan GEE dan memilih struktur korelasi terbaik berdasarkan nilai QIC. Dari hasil analisis diperoleh model kadar eritrosit penderita demam berdarah $\text{logit}[P(Y \leq 1/x_i)] = 2,4473 - 10,3935 \text{ lama sakit1} + 4,0715 \text{ lama sakit2} + 3,0735 \text{ lama sakit3} + 0,1210 \text{ umur}$, dengan struktur korelasi *autoregressive* berdasarkan nilai QIC terkecil. Pada model tingkat penggunaan marijuana selama lima tahun $\text{logit}[P(Y \leq 1/x_i)] = -1,8060 + 2,3984 \text{ lama penggunaan1} + 1,1679 \text{ lama penggunaan2} + 0,4830 \text{ lama penggunaan3} + 0,2031 \text{ lama penggunaan4} + 0,2638 \text{ frekuensi}$, $\text{logit}[P(Y \leq 2/x_i)] = -0,1861 + 2,3984 \text{ lama penggunaan1} + 1,1679 \text{ lama penggunaan2} + 0,4830 \text{ lama penggunaan3} + 0,2031 \text{ lama penggunaan4} + 0,2638 \text{ frekuensi}$, dengan struktur korelasi *independent*.

Kata Kunci: Regresi Logistik Ordinal, Data Longitudinal, GEE

GENERALIZED ESTIMATING EQUATIONS (GEE) OF ORDINAL LONGITUDINAL DATA

ABSTRACT

Ordinal Logistic regression was used to determine the pattern of relationship between the ordinal response variable with one or more categorical or continuous predictor variables. Number of categories of response variable could be dichotomous and politom. This model is applied to longitudinal data, obtained from observations at several cross-sectional units over a period of time. Repeated observations lead to autocorrelation, this needs to be solved by Generalized Estimating Equations (GEE). The study aims to apply the GEE and select the best correlation structure based on the value of QIC. Model for erythrocyte levels dengue patient is $\text{logit}[P(Y \leq 1|x_i)] = 2.4473 - 10.3935 \text{ 1st day} + 4.0715 \text{ 2nd day} + 3.0735 \text{ 3rd day} + 0.1210 \text{ age}$, with autoregressive correlation structure based on the smallest QIC value. Model for level of marijuana use over the past five years is $\text{logit}[P(Y \leq 1|x_i)] = -1.8060 + 2.3984 \text{ 1st years} + 1.1679 \text{ 2nd years} + 0.4830 \text{ 3rd years} + 0.2031 \text{ 4rd years} + 0,2638 \text{ frequency}$ $\text{logit}[P(Y \leq 2|x_i)] = -0.1861 + 2.3984 \text{ 1st years} + 1.1679 \text{ 2nd years} + 0.4830 \text{ 3rd years} + 0.2031 \text{ 4rd years} + 0,2638 \text{ frequency}$, only independent correlation structure available.

Keywords: Ordinal logistic regression, Longitudinal Data, GEE.

KATA PENGANTAR

Alhamdulillah, puji syukur kehadirat Allah SWT atas rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi yang berjudul ” *GENERALIZED ESTIMATING EQUATIONS (GEE) PADA DATA LONGITUDINAL BERSIFAT ORDINAL* ”. Dalam penyusunan skripsi ini, banyak pihak yang telah memberikan bantuan kepada penulis. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih kepada :

1. Ibu Dr. Ir. M. Bernadetha Mitakda selaku dosen pembimbing I yang telah memberikan motivasi, bimbingan, dan pengarahan hingga skripsi ini terselesaikan dengan baik.
2. Ibu Dr. Ir. Ni Wayan Surya Wardhani, MS selaku dosen pembimbing II yang telah memberikan bimbingan dan nasehat hingga skripsi ini terselesaikan dengan baik.
3. Ibu Eni Sumarminingsih, S.Si., MM, selaku dosen penguji yang telah memberikan saran dan masukan hingga skripsi ini terselesaikan dengan baik.
4. Bapak Dr. Abdul Rouf Alghofari, M.Sc. selaku Ketua Jurusan Matematika Fakultas MIPA Universitas Brawijaya Malang.
5. Keluargaku terutama Ayah, Ibu, Mama, Mbah uti, mbak Emma, mas Nyoman yang telah memberikan doa dan dorongan semangat hingga terselesainya skripsi ini.
6. Gusti Virmandi S. yang telah memberikan bantuan, semangat, dukungan serta perhatiannya selama ini.
7. Sahabat-sahabatku (Amri, Virgi, Cira, Nensi, Nonis, Rahma, Yeeryzkhe, Ineke, Yeni, Mendha dan Septi) yang senantiasa selalu memberi dukungan, motivasi serta indahny kebersamaan dalam menjalani kuliah di sini.
8. Teman-teman Statistika dari berbagai angkatan, terutama angkatan 2008.
9. Semua pihak yang telah membantu menyelesaikan penyusunan skripsi ini yang tidak dapat disebutkan satu per satu. Terima

kasih atas bantuan yang telah diberikan dan semoga Allah memberikan balasan dengan kebaikan.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak kekurangan, oleh karena itu saran dan kritik yang membangun demi kesempurnaan penulisan selanjutnya sangat penulis harapkan. Semoga skripsi ini dapat memberikan manfaat bagi semua pihak yang membutuhkan.

Malang, Januari 2013

Penulis



DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN	iii
ABSTRAK/ ABSTRACT	iv
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR LAMPIRAN	xiii
BAB I. PENDAHULUAN	
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah	2
1.4. Tujuan Penelitian	2
1.5. Manfaat Penelitian	2
BAB II. TINJAUAN PUSTAKA	
2.1. Data Longitudinal	3
2.2. <i>Generalized Linear Models (GLM)</i>	4
2.3. Regresi Logistik	5
2.4. Regresi Logistik Ordinal	6
2.4.1. <i>Proportional Odds</i>	7
2.4.2. Pendugaan Parameter Model Regresi Logistik Ordinal	8
2.4.3. Pengujian Parameter	10
2.5. <i>Generalized Estimation Equations (GEE)</i>	12
2.5.1 Pendugaan Parameter	12
2.5.2 Pendugaan Matriks Ragam-Peragam Parameter	14
2.5.3 <i>Generalized Wald Test</i> dan <i>Wald Test</i>	14
2.5.4 Pemilihan Struktur Korelasi.....	15

BAB III. METODE PENELITIAN

3.1. Sumber Data	17
3.2. Metode Penelitian	18

BAB IV. HASIL DAN PEMBAHASAN

4.1. Pemodelan Regresi Logistik	21
4.1.1. Kadar Eristrosit Penderita Demam Berdarah ...	21
4.1.2. Tingkat Penggunaan Marijuana Selama Lima Tahun	21
4.2. Pemilihan Struktur Korelasi Terbaik.....	22
4.2.1. Kadar Eristrosit Penderita Demam Berdarah ...	22
4.2.2. Tingkat Penggunaan Marijuana Selama Lima Tahun.....	23
4.3. Pemodelan <i>Generalized Estimating Equations</i>	24
4.3.1. Kadar Eristrosit Penderita Demam Berdarah ...	24
4.3.2. Tingkat Penggunaan Marijuana Selama Lima Tahun.....	27

BAB V. PENUTUP

5.1. Kesimpulan.....	33
5.2. Saran.....	33

DAFTAR PUSTAKA	35
-----------------------------	----

LAMPIRAN	37
-----------------------	----

DAFTAR GAMBAR

Halaman

Gambar 3.1. Diagram Alir Prosedur Analisis

19

UNIVERSITAS BRAWIJAYA



DAFTAR TABEL

	Halaman
Tabel 2.1. Struktur Data Longitudinal Seimbang	3
Tabel 2.2. Struktur Data Longitudinal Tidak Seimbang.....	3
Tabel 3.1. Sumber Data	17
Tabel 4.1. Pengujian Parameter Secara Parsial Kadar Eritrosit Penderita Demam Berdarah	21
Tabel 4.2. Pengujian parameter secara parsial Tingkat Penggunaan Marijuana selama Lima Tahun	22
Tabel 4.3. Nilai QIC Kadar Eritrosit Penderita Demam Berdarah.....	23
Tabel 4.4. Nilai Penduga Parameter Awal Kadar Eritrosit.. Penderita Demam Berdarah.....	24
Tabel 4.5. Nilai Penduga Parameter GEE Kadar Eritrosit... Penderita Demam Berdarah	24
Tabel 4.6. Hasil Pengujian Statistik Uji <i>Wald</i> Kadar Eritrosit Penderita Demam Berdarah.....	25
Tabel 4.7. Hasil Pengujian Statistik Uji <i>Generalized Wald</i> Kadar Eritrosit Penderita Demam Berdarah.....	26
Tabel 4.8. Hasil Pendugaan Rasio Odds Kadar Eritrosit Penderita Demam Berdarah	26
Tabel 4.9. Nilai Penduga Parameter Awal Tingkat Penggunaan Marijuana selama Lima Tahun	27
Tabel 4.10. Nilai Penduga Parameter GEE Tingkat Penggunaan Marijuana selama Lima Tahun.....	28
Tabel 4.11. Hasil Pengujian Statistik Uji <i>Wald</i> Tingkat Penggunaan Marijuana selama Lima Tahun.....	28
Tabel 4.12. Hasil Pengujian Statistik Uji <i>Generalized Wald</i> Tingkat Penggunaan Marijuana selama Lima Tahun	29
Tabel 4.13. Hasil Pendugaan Rasio Odds Tingkat Penggunaan Marijuana selama Lima Tahun.....	30

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Kadar Eritrosit Pasien Penderita Demam Berdarah.	37
Lampiran 2. Tingkat Penggunaan Marijuana Selama Lima Tahun.....	38
Lampiran 3. Syntax SAS Kadar Eritrosit Pasien Penderita Demam Berdarah.....	39
Lampiran 4. Syntax SAS Tingkat Penggunaan Marijuana Selama Lima Tahun.....	40
Lampiran 5. Output SAS Kadar Eritrosit Pasien Penderita Demam Berdarah.....	41
Lampiran 6. Output SAS Tingkat Penggunaan Marijuana Selama Lima Tahun.....	44
Lampiran 7. Output SPSS Nilai QIC Struktur Matriks Korelasi Kadar Eritrosit Pasien Penderita Demam Berdarah.....	48

