



**PENGEMBANGAN WEB INTERAKTIF DENGAN R-SHINY  
UNTUK *GENERALIZED ESTIMATING EQUATIONS*  
ORDE2 (GEE2)**

**SKRIPSI**

Oleh

**Lina Choiril Oktafiani  
NIM 111810101009**

**JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS JEMBER  
2015**



**PENGEMBANGAN WEB INTERAKTIF DENGAN R-SHINY  
UNTUK *GENERALIZED ESTIMATING EQUATIONS*  
ORDE2 (GEE2)**

diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat  
untuk menyelesaikan Program Studi Matematika (S1)  
dan mencapai gelar Sarjana Sains

**SKRIPSI**

Oleh

**Lina Choiril Oktafiani  
NIM 111810101009**

**JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS JEMBER  
2015**

## PERSEMBAHAN

Alhamdulillahirobbil ‘Alamin, dengan puji syukur kehadirat Allah SWT, skripsi ini saya persembahkan untuk:

1. Ayahanda Muhammad Ali Wafa dan Ibunda Sistuning Agustin yang selalu sabar memberi nasihat, bimbingan, doa, dan kasih sayang yang tiada putusnya;
2. adik Feren Dwi Aprilia dan semua keluarga besar yang telah memberikan dukungan dan motivasi;
3. guru-guru sejak taman kanak-kanak hingga perguruan tinggi yang telah memberikan ilmu dan membimbing dengan penuh kesabaran;
4. Unit Kegiatan Mahasiswa Seni TITIK yang telah memberikan ilmu dan pengalaman;
5. Almamater Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

**MOTTO**

“Sesungguhnya Allah tidak akan merubah sesuatu kaum hingga mereka merubah keadaan yang ada pada diri mereka sendiri.”  
(Terjemahan Q.S. Ar-Ra’du: 11)<sup>\*)</sup>

“Banyak kegagalan dalam hidup ini dikarenakan orang-orang tidak menyadari betapa dekatnya mereka dengan keberhasilan saat mereka menyerah.”  
(Thomas Alva Edison)<sup>\*\*)</sup>

---

<sup>\*)</sup> Departemen Agama Republik Indonesia. 1989. *Al Qur'an dan Terjemahannya*. Semarang: CV Toha Putra

<sup>\*\*)</sup> <http://katakatabijak.com/tag/thomas-alva-edison>

**PERNYATAAN**

Saya yang bertanda tangan di bawah ini:

nama : Lina Choiril Oktafiani

NIM : 111810101009

menyatakan dengan sesungguhnya bahwa skripsi yang berjudul “Pengembangan Web Interaktif dengan R-Shiny untuk *Generalized Estimating Equations* Orde2 (GEE2)” adalah benar-benar hasil karya sendiri kecuali jika disebutkan sumbernya dan skripsi ini belum pernah diajukan pada institusi manapun serta bukan karya jiplakan. Saya bertanggung jawab atas keabsahan isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenar-benarnya, tanpa ada tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, November 2015

Yang menyatakan,

Lina Choiril Oktafiani

NIM 111810101009

**SKRIPSI**

**PENGEMBANGAN WEB INTERAKTIF DENGAN R-SHINY  
UNTUK *GENERALIZED ESTIMATING EQUATIONS*  
ORDE2 (GEE2)**

Oleh:  
Lina Choiril Oktafiani  
NIM 111810101009

Pembimbing:

Dosen Pembimbing Utama : Prof. Drs. I Made Tirta, M.Sc., Ph.D.  
Dosen Pembimbing Anggota : Dian Anggraeni, S.Si., M.Si.

**PENGESAHAN**

Skripsi berjudul “Pengembangan Web Interaktif dengan R-Shiny untuk *Generalized Estimating Equations* Orde2 (GEE2)” telah diuji dan disahkan pada:

hari, tanggal :

tempat : Fakultas MIPA Universitas Jember

Tim Penguji,

Ketua,

Sekretaris,

Prof. Drs. I Made Tirta, M.Sc., Ph.D.  
NIP 195912201985031002

Dian Anggraeni, S.Si., M.Si.  
NIP 198202162006042002

Penguji I

Penguji II

Dr. Alfian Futuhul Hadi, S.Si., M.Si.  
NIP 197407192000121001

Kosala Dwidja Purnomo, S.Si., M.Si.  
NIP 196908281998021001

Mengesahkan  
Dekan,

Prof. Drs. Kusno, DEA., Ph.D.  
NIP 196101081986021001

## RINGKASAN

**Pengembangan Web Interaktif dengan R-Shiny untuk *Generalized Estimating Equations Orde2 (GEE2)***; Lina Choiril Oktafiani; 111810101009; 2015; 54 halaman; Jurusan Matematika Fakultas MIPA Universitas Jember.

GEE2 merupakan perkembangan metode untuk menambah efisiensi dari GEE yang diperkenalkan oleh Zhao dan Prentice tahun 1990. Salah satu perbedaan GEE dan GEE2 adalah metode GEE digunakan dengan tidak memodelkan link korelasi, sedangkan GEE2 dengan memodelkan link korelasinya. Tujuan penelitian ini adalah mengembangkan program analisis GEE2 berbasis web interaktif menggunakan R-Shiny, mengaplikasikan dan menginterpretasi hasil, serta membandingkan hasil estimasi antara kedua link korelasi dalam GEE2.

Langkah awal pada penelitian ini yaitu pembuatan program GEE2 berbasis web menggunakan R-Shiny yang selanjutnya dapat bermanfaat bagi orang lain dalam menganalisis data menggunakan GEE2 tanpa menguasai skrip pada program R. Program yang telah dibuat, dapat diakses melalui <http://statslab-rshiny.fmipa.unej.ac.id/JORS/GEE2/>. Hasil analisis data yang telah diperoleh menggunakan program GEE2 berbasis web interaktif, selanjutnya akan dibandingkan nilai estimasi dari kedua link korelasi pada GEE2 yaitu link korelasi *Identity* dan *Fisherz*.

Uji coba program pada penelitian ini menggunakan data simulasi sebanyak 200 yang terbagi menjadi 40 level dengan 2 variabel jenis respon yaitu SHB (untuk respon kontinue) dan BS (untuk respon biner), serta 5 variabel prediktor yaitu sekolah, gender, tes ujian, nilai TPA, dan nilai UN. Data tersebut selanjutnya akan ditentukan pengaruh antara variabel prediktor terhadap variabel respon. Uji signifikansi yang digunakan yaitu dengan melihat nilai *p-value*.

Hasil analisis data simulasi yang telah diperoleh menunjukkan bahwa link korelasi berpengaruh pada hasil dari nilai estimasi koefisien korelasi, dimana link



korelasi *Identity* mempunyai nilai *standard error* yang lebih minimum dibandingkan dengan nilai *standard error* menggunakan link korelasi *Fisherz*. Hal tersebut dapat berarti bahwa link korelasi *Identity* merupakan model terbaik yang dapat digunakan dalam menganalisis data menggunakan GEE2.

Uji signifikansi dengan melihat nilai *p-value*  $\leq 0.05$  menunjukkan bahwa nilai TPA merupakan variabel yang signifikan, sedangkan variabel sekolah, gender, dan nilai UN tidak signifikan, baik untuk respon kontinue maupun respon biner. Persamaan model terbaik untuk respon kontinue dengan parameter signifikan yaitu  $E Y = 29,70922037 + 0,90977480X_{i3}$ , sebagai contoh untuk nilai TPA = 70, maka rata-rata nilai hasil belajar siswa-siswi suatu SMA adalah 93,3935. Persamaan model terbaik untuk respon biner dengan parameter signifikan yaitu  $g \pi = -3,46479092 + 0,05180844X_{i3}$ . Sebagai contoh untuk nilai TPA=70, maka penilaian afektif untuk kategori memuaskan dari siswa-siswi suatu SMA adalah 0,5404.

## PRAKATA

Puji syukur kehadiran Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Pengembangan Web Interaktif dengan R-Shiny untuk *Generalized Estimating Equations* Orde2 (GEE2)”. Skripsi ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan strata satu (S1) pada Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Penyusunan skripsi ini tidak lepas dari bantuan berbagai pihak, oleh karena itu penulis menyampaikan ucapan terima kasih kepada:

1. Prof. Drs. I Made Tirta, M.Sc., Ph.D. selaku Dosen Pembimbing Utama dan Dian Anggraeni, S.Si., M.Si. selaku Dosen Pembimbing Anggota yang telah meluangkan waktu, memberikan saran dan bimbingannya dalam penulisan skripsi ini;
2. Dr. Alfian Futuhul Hadi, S.Si., M.Si. selaku Dosen Penguji I dan Kosala Dwidja Purnomo, S.Si., M.Si. selaku Dosen Penguji II yang telah memberi masukan berupa saran dan kritiknya terhadap skripsi ini;
3. Prof. Drs. I Made Tirta, M.Sc., Ph.D. selaku Dosen Pembimbing Akademik yang telah sabar membimbing penulis selama menjadi mahasiswa;
4. Nuril Afandi, S.Si. yang selalu memberikan doa, saran, dan motivasinya;
5. sahabat-sahabat seperjuangan KRAMAT'11, khususnya Ulfa Hasanah, Hijratul Ilahiyah, Ivana Gabriella, Dyah Kiki, Lina Wijiasih dan teman-teman KKN yang telah memberikan motivasi, saran, dan semangatnya;
6. semua pihak yang tidak dapat disebutkan satu per satu.

Penulis juga menerima segala kritik dan saran dari semua pihak demi kesempurnaan skripsi ini. Akhirnya penulis berharap, semoga skripsi ini dapat bermanfaat bagi semua pihak.

Jember, November 2015

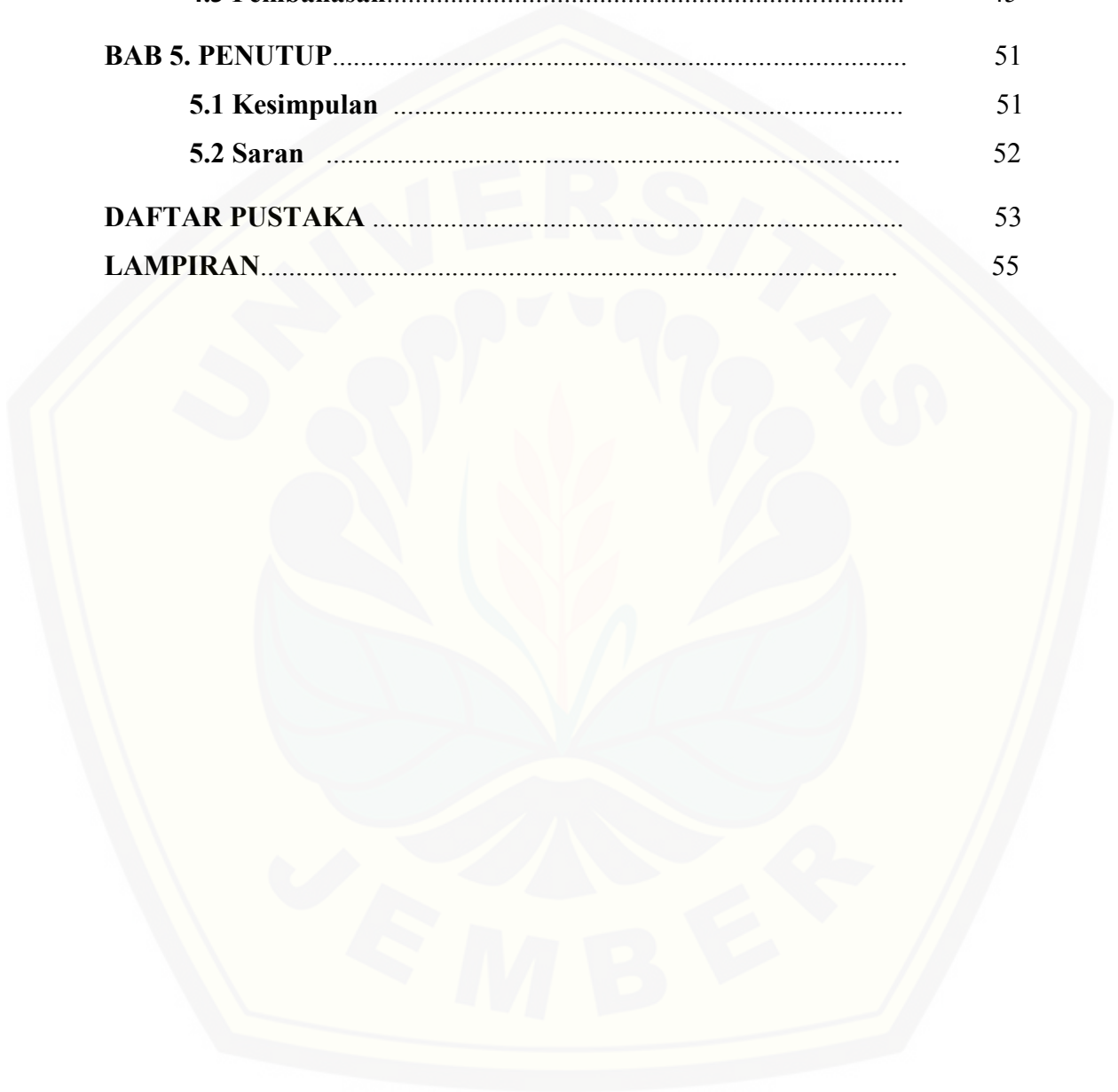
Penulis

**DAFTAR ISI**

	Halaman
<b>HALAMAN JUDUL</b> .....	ii
<b>HALAMAN PERSEMBAHAN</b> .....	iii
<b>HALAMAN MOTTO</b> .....	iv
<b>HALAMAN PERNYATAAN</b> .....	v
<b>HALAMAN BIMBINGAN</b> .....	vi
<b>HALAMAN PENGESAHAN</b> .....	vii
<b>RINGKASAN</b> .....	viii
<b>PRAKATA</b> .....	x
<b>DAFTAR ISI</b> .....	xi
<b>DAFTAR TABEL</b> .....	xiv
<b>DAFTAR GAMBAR</b> .....	xvi
<b>DAFTAR LAMPIRAN</b> .....	xvii
<b>BAB 1. PENDAHULUAN</b> .....	1
<b>1.1 Latar Belakang</b> .....	1
<b>1.2 Rumusan Masalah</b> .....	2
<b>1.3 Batasan Masalah</b> .....	3
<b>1.4 Tujuan</b> .....	3
<b>1.5 Manfaat</b> .....	3
<b>BAB 2. TINJAUAN PUSTAKA</b> .....	4

<b>2.1 Model Linier</b> .....	4
<b>2.2 Normal Linear Models (NLM)</b> .....	4
<b>2.3 Generalized Linear Model (GLM)</b> .....	4
<b>2.4 Model Marjinal</b> .....	6
<b>2.5 Generalized Estimating Equations (GEE)</b> .....	7
2.5.1 Distribusi dan Fungsi Link pada GEE.....	7
2.5.2 Struktur <i>Working Correlation</i> GEE .....	9
2.5.3 Estimasi Parameter GEE .....	11
2.5.4 Algoritma <i>Fisher Scoring</i> .....	12
<b>2.6 Generalized Estimating Equations Orde 2(GEE2)</b> .....	13
2.6.1 Model Probabilitas Marjinal GEE2 .....	14
2.6.2 Estimasi Parameter GEE2 .....	14
2.6.3 Struktur <i>Working Covariance</i> .....	15
<b>2.7 Goodness of Fit dari Model GEE</b> .....	17
2.7.1 <i>Quasi-likelihood Information Criterion (QIC)</i> .....	17
2.7.2 Uji Hipotesis.....	18
<b>2.8 Data Longitudinal</b> .....	18
<b>2.9 R-shiny</b> .....	19
<b>BAB 3. METODE PENELITIAN</b> .....	21
<b>3.1 Langkah-langkah Pembuatan Program GEE2 Berbasis     Web Interaktif</b> .....	21
<b>3.2 Ilustrasi</b> .....	24
<b>3.3 Metode Analisis</b> .....	26
<b>BAB 4. HASIL DAN PEMBAHASAN</b> .....	27
<b>4.1 Hasil dan Pembahasan Fitur Web</b> .....	27
4.1.1 Penyusunan Web.....	27
4.1.2 Tampilan Web.....	31
<b>4.2 Ilustrasi Data</b> .....	36

4.2.1 Hasil Estimasi dengan Respon Kontinue .....	37
4.2.2 Hasil Estimasi dengan Respon Biner .....	41
<b>4.3 Pembahasan</b> .....	45
<b>BAB 5. PENUTUP</b> .....	51
<b>5.1 Kesimpulan</b> .....	51
<b>5.2 Saran</b> .....	52
<b>DAFTAR PUSTAKA</b> .....	53
<b>LAMPIRAN</b> .....	55



DAFTAR TABEL

	Halaman
3.1 Identifikasi Variabel Respon Kontinue .....	25
3.2 Identifikasi Variabel Respon Biner .....	26
4.1 Stuktur Fungsi pada Paket <i>geepack</i> .....	35
4.2 Nilai Estimasi Koefisien Model ( $\beta$ ) dengan <i>Gaussian</i> , Struktur Korelasi <i>Exchangeable</i> dan Link Korelasi <i>Identity</i> .....	37
4.3 Nilai Estimasi Koefisien Model ( $\beta$ ) dengan <i>Gaussian</i> , Struktur Korelasi AR-1 dan Link Korelasi <i>Identity</i> .....	39
4.4 Nilai Estimasi Koefisien Model ( $\beta$ ) dengan <i>Gaussian</i> , Struktur Korelasi <i>Exchangeable</i> dan Link Korelasi <i>Fisherz</i> .....	39
4.5 Nilai Estimasi Koefisien Model ( $\beta$ ) dengan <i>Gaussian</i> , Struktur Korelasi AR-1 dan Link Korelasi <i>Fisherz</i> .....	40
4.6 Nilai Estimasi Koefisien Korelasi dengan <i>Gaussian</i> dan Link Korelasi <i>Identity</i> .....	41
4.7 Nilai Estimasi Koefisien Korelasi dengan <i>Gaussian</i> dan Link Korelasi <i>Fisherz</i> .....	41
4.8 Nilai Estimasi Koefisien Model ( $\beta$ ) dengan <i>Binomial</i> , Struktur Korelasi <i>Exchangeable</i> dan Link Korelasi <i>Identity</i> .....	42
4.9 Nilai Estimasi Koefisien Model ( $\beta$ ) dengan <i>Binomial</i> , Struktur Korelasi AR-1 dan Link Korelasi <i>Identity</i> .....	42
4.10 Nilai Estimasi Koefisien Model ( $\beta$ ) dengan <i>Binomial</i> , Struktur Korelasi <i>Exchangeable</i> dan Link Korelasi <i>Fisherz</i> .....	43
4.11 Nilai Estimasi Koefisien Model ( $\beta$ ) dengan <i>Binomial</i> , Struktur Korelasi AR-1 dan Link Korelasi <i>Fisherz</i> .....	44

4.12 Nilai Estimasi Koefisien Korelasi dengan <i>Binomial</i> dan Link Korelasi <i>Identity</i> .....	44
4.13 Nilai Estimasi Koefisien Korelasi dengan <i>Binomial</i> dan Link Korelasi <i>Fisherz</i> .....	45
4.14 Nilai <i>Standard Error</i> Koefisien Model ( $\beta$ ) dengan Respon Kontinue.....	45
4.15 Nilai <i>Standard Error</i> Koefisien Model ( $\beta$ ) dengan Respon Biner.....	46
4.16 Nilai <i>Standard Error</i> Koefisien Korelasi ( $\alpha$ ) dengan Respon Kontinue.....	46
4.17 Nilai <i>Standard Error</i> Koefisien Korelasi ( $\alpha$ ) dengan Respon Biner.....	46
4.18 Nilai <i>P-value</i> dengan Respon Kontinue.....	48
4.19 Nilai <i>P-value</i> dengan Respon Biner.....	49

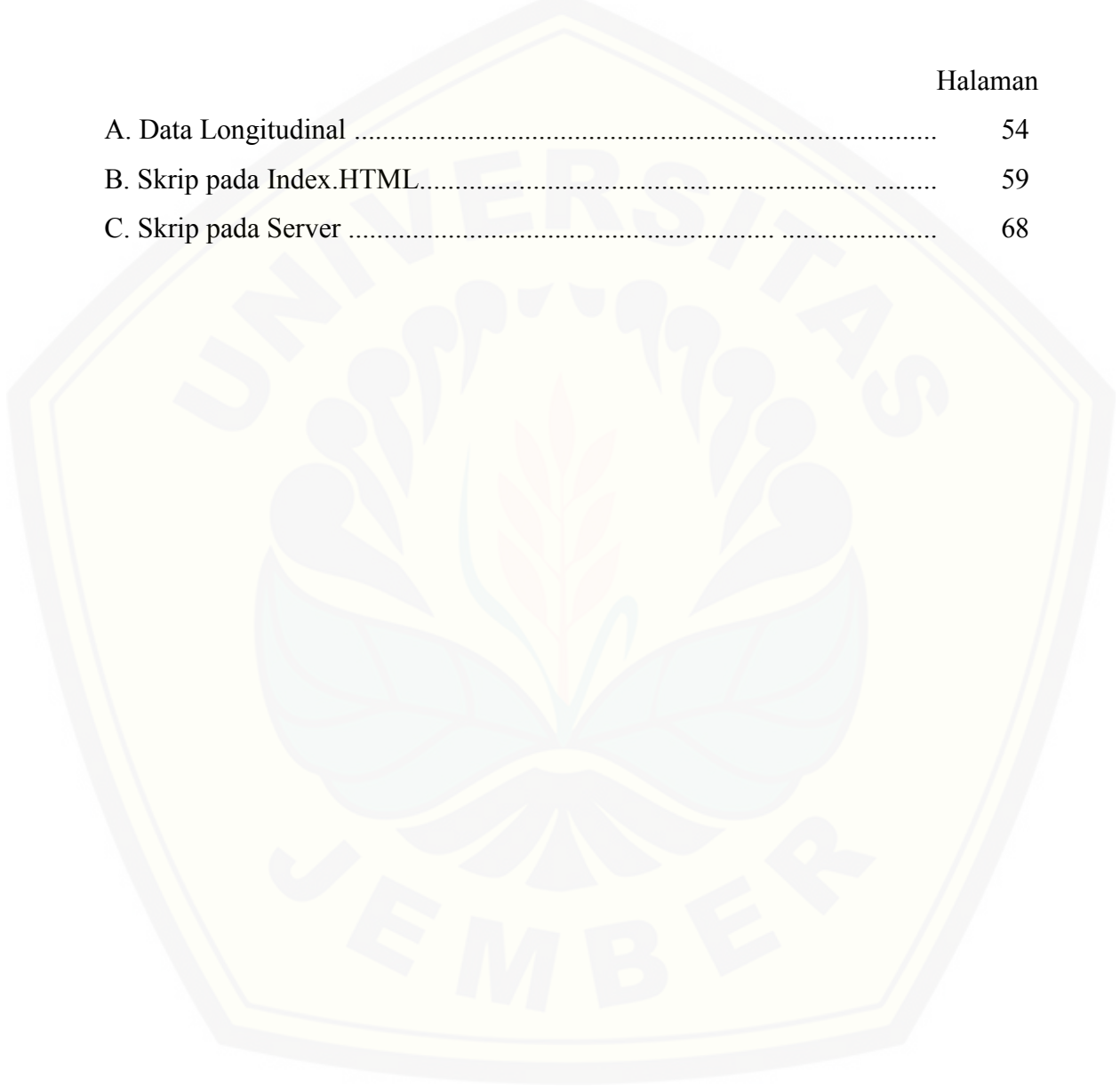
**DAFTAR GAMBAR**

	Halaman
2.1 Struktur Komponen R-shiny .....	20
3.1 Skema Langkah-langkah Pengolahan Data .....	24
4.1 Tampilan Awal Program.....	32
4.2 Tampilan Pilihan Data.....	33
4.3 Tampilan Input Program.....	33
4.4 Tampilan Output Umum.....	34
4.5 Tampilan Pilihan Output Khusus.....	35



**DAFTAR LAMPIRAN**

	Halaman
A. Data Longitudinal .....	54
B. Skrip pada Index.HTML.....	59
C. Skrip pada Server .....	68



## BAB 1. PENDAHULUAN

### 1.1 Latar Belakang

Ilmu matematika merupakan salah satu ilmu pengetahuan yang mempunyai banyak peranan dalam menyelesaikan permasalahan di berbagai bidang, misalnya dalam bidang kesehatan, sosial, ekonomi, pendidikan, dan sebagainya. Salah satu bidang dalam matematika yang dapat digunakan untuk menyelesaikan berbagai permasalahan dalam bidang-bidang tersebut yaitu bidang statistika yang mempunyai banyak model didalamnya, diantaranya model linier.

Perkembangan model linier dimulai dengan perkembangan analisis regresi pada abad 19 oleh Pearson (Tirta,2009). Analisis regresi linier merupakan salah satu analisis pada statistika yang menjelaskan tentang hubungan antara satu variabel respon ( $Y$ ) dengan satu atau lebih variabel prediktor ( $X_i$ ). Analisis regresi linier biasanya digunakan untuk menganalisis variabel respon dari data kontinue yang berdistribusi normal dan saling bebas. Model yang digunakan dalam menganalisis data tersebut yaitu model linier normal sederhana atau *Normal Linear Models* (disingkat dengan NLM).

Namun dalam perkembangannya, masih terdapat data yang tidak berdistribusi normal tetapi responnya saling bebas, maka model linier yang dapat digunakan untuk data tersebut adalah model linier tergeneralisasi (*Generalized Linear Models* atau disingkat dengan GLM) yang dipelopori oleh Nelder dan Wedderburn tahun 1972. *Generalized Linear Models* (GLM) menggunakan asumsi bahwa variabel responnya berdistribusi eksponensial yaitu distribusi yang sifatnya lebih umum seperti distribusi Normal, Gamma, Poisson.

Kondisi lain yang juga terjadi di lapang seiring dengan perkembangannya, terdapat pula data yang tidak berdistribusi normal sekaligus tidak saling bebas. Liang dan Zeger pada tahun 1986 memperkenalkan metode yang disebut sebagai

*Generalized Estimating Equations* (GEE) yang merupakan generalisasi multivariat dari *quasi-likelihood*. GEE juga merupakan perluasan dari GLM.

Tambun (2012) telah menganalisis data yang menghasilkan model dengan korelasi *Independence* merupakan model yang lebih baik dibandingkan dengan model korelasi *Exchangeable* dan menurut Uji Wald diperoleh intersep serta variabel  $X_1$  signifikan terhadap respon. Hidayati (2013) juga telah menganalisis data dengan menggunakan *Generalized Estimating Equations* orde 2 (GEE2). Penelitian tersebut menghasilkan nilai AIC untuk metode GEE2 jauh lebih minimum dibandingkan dengan GEE biasa, sehingga model dengan metode GEE2 lebih baik daripada GEE biasa. Penelitian-penelitian tersebut menggunakan skrip manual program R dalam penyelesaiannya, sehingga mengharuskan seseorang mengetahui dan memahami skrip program dari suatu analisis data menggunakan GEE ataupun GEE2 dan belum membandingkan kedua link korelasi dalam metode GEE2. Namun, dalam penelitian ini penulis akan menganalisis data menggunakan metode GEE2 dengan membuat program berbasis web interaktif menggunakan R-Shiny yang dapat memudahkan seseorang dalam menganalisis data menggunakan metode GEE2 ini tanpa menguasai skrip dalam program R serta membandingkan kedua link korelasinya. Selain itu, juga akan dilakukan uji signifikansi antara variabel prediktor dan variabel respon.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dibahas di atas, dapat dirumuskan permasalahan sebagai berikut :

- a. bagaimana membuat program GEE2 berbasis web interaktif menggunakan R-Shiny yang memiliki kemampuan sama dengan pilihan menggunakan skrip?
- b. bagaimana penggunaan program GEE2 berbasis web interaktif menggunakan R-Shiny untuk menganalisis data simulasi dan interpretasi hasilnya?

### 1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah paket R yang digunakan untuk metode analisis GEE2 dibatasi dengan menggunakan paket `geepack`.

### 1.4 Tujuan

Berdasarkan rumusan masalah yang telah dibahas, dapat diketahui tujuan dari penelitian ini adalah :

- a. membuat program GEE2 berbasis web interaktif menggunakan R-Shiny yang memiliki kemampuan sama dengan pilihan menggunakan skrip.
- b. menggunakan program GEE2 berbasis web interaktif untuk menganalisis data simulasi dan menginterpretasi hasilnya.

### 1.5 Manfaat

Manfaat yang diperoleh dari penelitian ini adalah memperoleh solusi dari hasil analisis regresi menggunakan program *Generalized Estimating Equation* orde 2 (GEE2) berbasis web interaktif menggunakan R-Shiny. Selain itu, program GEE2 berbasis web interaktif dapat membuat metode GEE2 lebih mudah digunakan untuk belajar dan latihan menganalisis data menggunakan GEE2.

## BAB 2. TINJAUAN PUSTAKA

Bab ini akan membahas dasar-dasar teori yang digunakan untuk menganalisis data menggunakan program *Generalized Estimating Equations* orde 2 (GEE2) berbasis web interaktif menggunakan R-Shiny. Dasar-dasar teori tersebut antara lain.

### 2.1 Model Linier

Perkembangan model linier dimulai dengan perkembangan analisis regresi pada abad 19 oleh Pearson. Teori regresi menjadi dasar perkembangan teori model linier. Perkembangan model linier tidak bisa dilepaskan dengan perkembangan teori matriks atau aljabar linier.

Pemodelan statistika dimulai dari yang sederhana, yang secara matematis mudah diselesaikan, kemudian berkembang ke arah yang lebih realistis. Hal tersebut dapat diterapkan dengan berbagai asumsi yang berbeda terhadap distribusi kesalahan dalam model yang digunakan. Prinsip seperti ini telah berkembang dari model yang paling sederhana (klasik), ke model hirarkis tergeneralisasi (Tirta, 2009).

### 2.2 Normal Linear Models (NLM)

*Normal Linear Models* (NLM) adalah model linier yang pertama kali muncul pada abad ke-19. Model ini menganalisis hubungan antar variabel, sehingga harus melibatkan lebih dari satu variabel dalam percobaannya. Variabel respon pada NLM diasumsikan berdistribusi normal dan saling independen antar satu respon dengan respon yang lain.

### 2.3 Generalized Linear Models (GLM)

Model regresi linier berkembang seiring adanya data-data yang tidak memenuhi karakteristik model linier sederhana seperti halnya data yang tidak

berdistribusi normal. Hal yang dapat digunakan untuk menangani pengamatan yang responnya tidak berdistribusi normal, tetapi masih saling independen, maka dapat menggunakan model linier yang disebut dengan *Generalized Linear Models* (GLM) yang dipelopori oleh Nelder dan Wedderburn tahun 1972. Model ini mengasumsikan bahwa respon memiliki distribusi keluarga eksponensial, seperti Normal, Gamma, Poisson (Tirta, 2009).

Menurut Nelder dan Wedderburn (1972), *Generalized Linear Models* (GLM) memiliki tiga komponen yaitu:

1. komponen acak yang menetapkan distribusi bersyarat dari variabel respon,  $Y_i$  (untuk  $i$  dari  $n$  sampel pengamatan yang independen), memberi nilai dari variabel penjelas dalam model. Persamaan asli dari Nelder dan Wedderburn, distribusi pada  $Y_i$  merupakan anggota dari keluarga eksponensial, seperti Gaussian (normal), binomial, Poisson, gamma, atau distribusi keluarga invers Gaussian. Selanjutnya, GLM diperluas menjadi keluarga eksponensial multivariat (seperti distribusi multinomial), keluarga eksponensial tertentu (distribusi binomial negatif dua parameter), dan beberapa situasi dalam distribusi  $Y_i$  yang tidak ditetapkan seutuhnya.
2. prediktor linier merupakan fungsi linier dari regresi :

$$\eta_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$$

seperti pada model linier, regresi  $X_{ij}$  adalah fungsi *prespecified* dari variabel penjelas dan boleh memasukkan banyak variabel penjelas, bentuk dari banyak variabel penjelas, regresi polinomial, regresi dummy, interaksi dan sebagainya. Salah satu keuntungan dari GLM adalah struktur prediktor linier yang terkenal pada struktur model linier.

3. fungsi penghubung (*link function*), yang mengubah dugaan dari variabel respon,  $\mu_i = E(Y_i)$ , pada prediktor linier:

$$g(\mu_i) = \eta_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$$

Karena fungsi penghubung dibalik, maka dapat juga ditulis:

$$\mu_i = g^{-1} \eta_i = g^{-1}(\alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik})$$

dengan demikian, GLM boleh dianggap sebagai model linier untuk transformasi respon yang diharapkan atau sebagai model regresi nonlinier untuk respon. Invers penghubung  $g^{-1}(\cdot)$  juga disebut fungsi mean. Perhatikan bahwa identitas penghubung sederhana menghasilkan argumen yang tek berubah, yaitu  $\eta_i = g \mu_i = \mu_i$ , dengan demikian  $\mu_i = g^{-1} \eta_i = \eta_i$ .

#### 2.4 Model Marjinal

Model marjinal merupakan model untuk data yang tidak berdistribusi normal juga tidak saling bebas, dengan kata lain  $Y_i$  bukan respon tunggal tetapi vektor respon dengan  $Y_i = (Y_{i1}, Y_{i2}, Y_{ij}, Y_{im})^T$ . Fitzmaurice, *et al* (2009) menjelaskan bahwa dalam model marjinal, regresi dari respon pada variabel eksplanatori adalah pemodelan terpisah dalam unit korelasi. Dalam regresi tersebut model ekspektasi marjinal  $E(Y_{ij})$  seperti sebuah fungsi variabel eksplanatori. Model marjinal khususnya memiliki asumsi-asumsi seperti berikut ini:

1. ekspektasi marjinal  $E Y_{ij} = \mu_{ij}$ , bergantung pada vektor peubah eksplanatori  $x_{ij}$  dengan hubungan  $g(\mu_{ij}) = x_{ij}\beta$ , dimana  $g(\cdot)$  adalah fungsi link yang diketahui.
2. varians marjinal, tergantung pada rataan atau ekspektasi marjinal menurut hubungan  $\text{Var } Y_{ij} = \phi v(\mu_{ij})$ , dimana  $v(\cdot)$  adalah fungsi varians yang diketahui dan  $\phi$  adalah parameter skala yang mungkin perlu diduga.
3. korelasi antara  $Y_{ij}$  dan  $Y_{ik}$  adalah fungsi dari rataan marjinal dan parameter-parameter tambahan, yaitu  $\text{Corr } Y_{ij}, Y_{ik} = \psi(\mu_{ij}; \mu_{ik}; \alpha)$ , dimana  $\psi(\cdot)$  adalah fungsi yang diasumsikan diketahui (Diggle, *et al.* 2002).

## 2.5 Generalized Estimating Equations (GEE)

*Generalized Estimating Equations* (GEE) merupakan model pendekatan yang dapat digunakan untuk data yang tidak saling bebas dan berdistribusi Binomial, Poisson, atau Gamma. GEE pertama kali diperkenalkan oleh Liang dan Zeger pada tahun 1986. GEE merupakan sebuah generalisasi multivariat dari quasi-likelihood.

Menurut Tirta (2009), pendekatan GEE fokus pada model linier untuk mean dari pengamatan berkorelasi dalam kluster tanpa spesifikasi distribusi bersama secara penuh pada pengamatan. Sehingga GEE merupakan model marjinal yang tidak didasarkan atas bentuk *Likelihood* lengkap dari respon, tetapi hanya berdasarkan hubungan antar mean (momen pertama) dan ragamnya (momen kedua) serta bentuk matriks korelasinya.

GEE menyediakan sebuah kerangka dasar untuk menganalisis sekumpulan data dengan pengamatan berkorelasi. Kerangka dasar ini membutuhkan solusi dari sebuah persamaan estimasi yang menyertakan turunan matriks dan kovarian. GEE merupakan perluasan *quasi-likelihood* yang hanya mengharuskan mempunyai struktur mean dan kovarian yang jelas sehingga menghasilkan estimasi yang konsisten untuk koefisien regresi dan standar errornya. Bagaimanapun persamaan regresi merupakan hal yang pokok pada GEE sedangkan korelasi diperlakukan sebagai *nuisance* (parameter gangguan). Maka ada perbedaan pada efisiensi dalam asumsi korelasi yang berbeda. Jika variabel hasil menunjukkan sebuah kondisi seperti berdistribusi normal, maka GEE dapat dipandang sama seperti atau identik dengan maksimum *Likelihood* (Lane, 2007).

### 2.5.1 Distribusi dan Fungsi Link pada GEE

Menurut Rao & Toutenburg (1999), distribusi beserta fungsi link yang dapat digunakan pada metode GEE diantaranya.



a. Distribusi *Gaussian*

Distribusi *Gaussian* digunakan saat respon kontinue, hubungan antara varian dan mean bersifat bebas. Bentuk fungsi kepadatan peluang dari variabel respon  $Y$  yang berdistribusi Normal atau *Gaussian* adalah

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

dengan  $-\infty < y < \infty, -\infty < \mu < \infty$ , dan  $\sigma > 0$ , dimana  $\sigma$  adalah parameter *nuisance*. Oleh karena itu, untuk data kontinue yang berdistribusi normal, mengikuti distribusi normal dengan fungsi link *identity* dan struktur mean:

$$y_i \sim N(\mu_i, \sigma^2), \quad E(y_i) = \mu_i, \quad \mu_i = \eta_i = x'_{ij}\beta$$

b. Distribusi *Binomial*

Distribusi *Binomial* digunakan saat respon biner, hubungan antar varian dan mean bersifat linier. Misalkan respon  $Y_i = (y_{i1}, \dots, y_{im})'$  merupakan vektor respon dari kluster ( $i = 1, \dots, N$ ) dengan matriks kovarian  $\text{Cov}(y_i)$  dan  $x_{ij}$  merupakan vektor ( $p \times 1$ ) dari  $y_{ij}$ . Variabel  $y_{ij}$  diasumsikan biner dengan nilai 1 dan 0, dan asumsi  $P(y_{ij} = 1) = \pi_{ij}$ . Kemudian  $\mu_{ij} = \pi_{ij}$  dan  $\pi'_i = \pi_{i1}, \dots, \pi_{im}$ . Model GEE untuk distribusi binomial dengan fungsi  $g(\cdot)$  dari link logit adalah

$$g(\pi_{ij}) = x'_{ij}\beta$$

$\eta(\cdot)$  merupakan invers fungsi, yaitu

$$\mu_{ij} = \pi_{ij} = \eta(x'_{ij}\beta)$$

untuk link kanonik

$$\text{logit}(\pi_{ij}) = \ln \frac{\pi_{ij}}{1 - \pi_{ij}} = g(\pi_{ij}) = x'_{ij}\beta$$

maka

$$\pi_{ij} = \eta(x'_{ij}\beta) = \frac{\exp(x'_{ij}\beta)}{1 + \exp(x'_{ij}\beta)}$$

c. Distribusi *Poisson*

Misalkan  $y_i (i = 1, 2, \dots, N)$  berdistribusi *Poisson* dengan  $E y_i = \mu_i$  dan fungsi kepadatan

$$P y_i = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \text{ untuk } y_i = 0, 1, 2, \dots$$

Fungsi link dapat di pilih sebagai  $\ln \mu_i = x'_{ij} \beta$ .

c. Distribusi *Gamma*

Distribusi *Gamma* digunakan saat respon kontinue dengan rentang  $0 < x < \infty$ , hubungan antar varian dan mean bersifat kuadratik. Variabel respon  $Y$  yang berdistribusi *Gamma* memiliki kepadatan peluang

$$f y = \frac{\mu_i (y \mu_i)^{\phi-1} e^{-y \mu_i}}{\Gamma(\phi)}, y > 0$$

dimana  $\phi$  adalah parameter *nuisance*. Maka,  $E y_i = \phi / \mu_i$  dan  $\text{Var } y_i = \phi / \mu_i^2$ . Oleh karena itu untuk data kontinue yang berdistribusi *Gamma*, mempunyai fungsi link  $\eta_i = 1 / \mu_i$ .

2.5.2 Struktur *Working Correlation* GEE

Menurut Kim & Shults (2010), struktur *working correlation* yang umum digunakan untuk GEE sebagai berikut.

a. *The First-order Autoregressive* (AR(1))

Struktur ini mengasumsikan bahwa

$$\text{Corr } y_{ij}, y_{ik} = \begin{cases} 1 & j = k \\ \alpha^{|j-k|} & j \neq k \end{cases}$$

Model korelasi ini juga sering disebut dengan model korelasi serial dengan struktur seperti:

$$R_{\alpha} = \begin{matrix} 1 & \alpha & \alpha^2 & \dots & \alpha^p \\ \alpha & \ddots & \vdots & \ddots & \vdots \\ \alpha^2 & \dots & 1 & \dots & \alpha^2 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \alpha^p & \dots & \alpha^2 & \dots & 1 \end{matrix}$$

b. *Unstructured*

Struktur ini digunakan untuk model dengan korelasi tanpa struktur. Model *unstructured* juga disebut sebagai model multivariat penuh. Struktur ini mengasumsikan bahwa:

$$\text{Corr } y_{ij}, y_{ik} = \begin{matrix} 1 & j = k \\ \alpha_{jk} & j \neq k \end{matrix}$$

sehingga struktur *working correlation*-nya seperti berikut:

$$R_{\alpha} = \begin{matrix} 1 & \alpha_{1,2} & \dots & \alpha_{1,t} \\ \alpha_{1,2} & 1 & \dots & \alpha_{2,t} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1,t} & \alpha_{2,t} & \dots & 1 \end{matrix}$$

c. *Independence*

Struktur ini mengasumsikan bahwa respon saling bebas. Model *independence* ini identik dengan model  $\text{glm}(\cdot)$ . Asumsi dari struktur ini adalah:

$$\text{Corr } y_{ij}, y_{ik} = \begin{matrix} 1 & j = k \\ 0 & j \neq k \end{matrix}$$

dengan struktur korelasi sebagai berikut:

$$R_{\alpha} = \begin{matrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{matrix}$$

d. *Exchangeable*

Struktur ini mengasumsikan adanya korelasi seragam/uniform atau lebih dikenal sebagai *compound symmetry*. Asumsi dari struktur ini adalah:

$$\text{Corr } y_{ij}, y_{ik} = \begin{cases} 1 & j = k \\ \alpha & j \neq k \end{cases}$$

dengan struktur korelasinya seperti berikut:

$$R_{\alpha} = \begin{pmatrix} 1 & \alpha & \dots & \alpha \\ \alpha & 1 & \dots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \dots & 1 \end{pmatrix}$$

Pemodelan marginal parameter mean dan perhitungan dalam GEE bergantung pada unit sebuah kluster dengan menentukan struktur *working correlation* untuk respon observasi. Model yang paling populer untuk menganalisis data dengan respon berkorelasi dalam *Generalized Linear Models* (GLM) adalah model *Generalized Estimating Equation* (GEE). GEE didesain fokus pada mean marginal dan struktur korelasinya sebagai *nuisance*. Sehingga implementasi dari GEE biasanya terbatas pada jumlah struktur *working correlation*-nya (seperti AR-1, *exchangeable*, *independence*, dan *unstructured*) dimana struktur tersebut konstan pada seluruh kluster. Keuntungan dari GEE yaitu parameter mean diestimasi secara konsisten terlepas dari apakah struktur korelasinya ditentukan dengan benar, selama mean juga ditentukan dengan benar. Akan tetapi, efisiensi tersebut dapat hilang ketika struktur *working correlation*-nya salah (Crespi *et al*, 2009).

### 2.5.3 Estimasi Parameter GEE

Menurut Abdy pada tahun 2009 dalam Hidayati (2013), *Quasi-Likelihood* diperkenalkan Wedderburn pada tahun 1974 merupakan pendekatan model linier umum untuk respon univariat. Persamaan *quasi-Likelihood* yang dimaksud adalah:

$$S \beta = \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta} V_i^{-1} Y_i - \mu_i \beta = 0$$

dimana  $Y_i = (y_{i1}, \dots, y_{im})'$  adalah vektor  $m \times 1$  yang menyatakan variabel respon,  $X_i = (x_{i1}, \dots, x_{im})'$  adalah matriks  $m \times p$  yang merupakan nilai dari variabel prediktor untuk subjek ke- $i$  ( $i = 1, \dots, n$ ;  $j = 1, \dots, m$ ) dan  $\mu_{ij}$  menyatakan ekspektasi bersyarat

dari respon  $Y_{ij}$ , diberikan matriks  $X_{ij}$ , yaitu  $\mu_{ij} = E(Y_{ij}|X_{ij})$ , serta diasumsikan hubungan  $l \mu_{ij} = X_{ij}\beta$  dengan  $l$  adalah fungsi link.

Kemudian, misalkan  $\text{var } Y_{ij} = \phi V(\mu_{ij})$  dimana  $V$  adalah fungsi dari  $\mu_{ij}$  yang diketahui dan  $\phi$  adalah parameter dispersi. Jika varian  $Y_i$  pada persamaan *quasi-Likelihood* diatas diganti dengan matriks kovarian  $Y_i$ , yaitu matriks yang berbentuk  $V_i = \phi A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}}$ , dimana  $A_i = \text{diag}\{v \mu_{i1}, \dots, v \mu_{ij}\}$  dan elemen ke-( $j,k$ ) dari  $R_i$  adalah  $\text{Corr}(Y_{ij}, Y_{ik})$ . Pengembangan *quasi-Likelihood* ke bentuk multivariat tersebut dinamakan GEE. Bentuk persamaan untuk  $\beta$  dalam model marjinal yaitu:

$$S \beta = \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta}{}' V_i^{-1} Y_i - \mu_i \beta = 0$$

dengan:

$$V_i = \phi A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}}$$

$$A_i = \text{diag}\{v \mu_{i1}, \dots, v \mu_{ij}\}$$

$R_i(\alpha)$  = matriks korelasi

$\phi$  = parameter dispersi

$\alpha$  =  $\text{Corr}(Y_{ij}, Y_{ik}), j \neq k$

$A_i^{\frac{1}{2}}$  merupakan matriks diagonal:

$$A_i^{\frac{1}{2}} = \begin{matrix} \sqrt{v(\mu_1)} & 0 & \dots & 0 \\ 0 & \sqrt{v(\mu_2)} & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{v(\mu_j)} \end{matrix}, A_i = \begin{matrix} v(\mu_1) & 0 & \dots & 0 \\ 0 & v(\mu_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & v(\mu_j) \end{matrix}$$

#### 2.5.4 Algoritma Fisher Scoring

Langkah-langkah untuk menyelesaikan kasus GEE dengan menggunakan algoritma *Fisher Scoring* yaitu:

- berikan nilai awal untuk  $\hat{\beta}^{(m)}$
- hitung *working correlation*  $R_i(\alpha)$

$$\phi = \frac{\sum_{i=1}^n \sum_{j=1}^s (r_{ij})^2}{NS}$$

$$r_{ij} = \frac{Y_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{ij}(1 - \hat{\pi}_{ij})}$$

c. hitung matriks *working covariance*

$$V_i = \phi A_i^2 R_i(\alpha) A_i^2$$

d. hitung

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + \sum_{i=1}^n D_i^m ' V_i^{-1} D_i^m \sum_{i=1}^n D_i^m ' V_i^{-1} Y_i - \mu_i$$

e. iterasi sampai konvergen  $\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)}$

## 2.6 Generalized Estimating Equations Orde 2 (GEE2)

Tahun 1991, Zhao & Prentice telah memperkenalkan sebuah variasi pengembangan dari GEE biasa yang disebut dengan GEE2. GEE2 dapat digunakan untuk menambah efisiensi dari GEE. Jika pada GEE struktur mean merupakan kepentingan utama dan struktur korelasi dianggap sebagai parameter gangguan (*nuisance*), sedangkan pada GEE2 struktur korelasi juga dianggap penting. Struktur mean diperlukan untuk menyesuaikan struktur korelasi untuk kovariat. Pada GEE2 diterapkan analisis mean dan struktur korelasi secara bersamaan (Kastner & Ziegler, 1999).

Menurut Lane (2007), GEE menyediakan kerangka kerja yang cocok untuk estimasi yang konsisten dari koefisien regresi dan standar errornya ketika mempunyai data pengamatan yang merupakan data berkorelasi. Seperti yang telah dijelaskan sebelumnya bahwa GEE menganggap korelasi sebagai sebuah *nuisance*, sedangkan GEE2 mengatasi masalah ini dengan memperkenalkan persamaan estimasi kedua untuk parameter kovarian yang diselesaikan sekaligus dengan persamaan estimasi pertama.

### 2.6.1 Model Probabilitas Marjinal GEE2

Dasar memperoleh GEE2 adalah model eksponensial kuadrat, perluasan sederhana dari model GLM yang digunakan di bawah GEE. *Generalized Estimating Equation* Orde 2 (GEE2) sangat mirip dengan GEE biasa, yaitu masih dalam konteks memecahkan persamaan GEE, namun persamaan estimasi kedua untuk kovarian diperkenalkan dan diselesaikan secara bersamaan terlebih dahulu. Model probabilitas marjinal menurut model eksponensial kuadrat adalah sebagai berikut:

$$P_i(y_i, \mu_i, \sigma_i) = \tau_i^{-1} \exp(y_i' \theta_i + \omega_i' \lambda_i + c_i \cdot y_i) \quad (2.1)$$

dimana:

$\tau_i$  =  $\tau_i(\sigma_i, \lambda_i, c_i \cdot)$  adalah *normalizing constant*

$\omega_i'$  =  $(y_{i1}^2, y_{i1}, y_{i2}, \dots, y_{i2}^2, y_{i2}, y_{i3}, \dots)$

$c_i \cdot$  = fungsi *shape*

$\theta_i'$  =  $\theta_i'(\mu_i, \sigma_i) = (\theta_{i1}, \dots, \theta_{im_i})$

$\lambda_i'$  =  $\lambda_i'(\mu_i, \sigma_i) = (\lambda_{i11}, \lambda_{i12}, \dots, \lambda_{i22}, \lambda_{i23}, \dots)$

dengan  $\theta_i$  dan  $\lambda_i$  adalah fungsi dari parameter mean dan varian, dan  $\mu_i = \mu_i(\beta)$ ,  $\sigma_i = \sigma_i(\beta, \alpha)$  untuk  $i = 1, \dots, n$ . Disini setiap kluster akan terdiri dari  $n_i$  individu dan 2 kluster itu variannya independen ( $y_i, y_i' = 0$ ).

Seperti yang telah disebutkan oleh Zhao & Prentice tahun 1991, setiap anggota kluster, model sepenuhnya ditentukan oleh fungsi *shape*  $c_i \cdot$  untuk respon multivariat diskrit ataupun kontinu yang hanya melibatkan mean dan kovarian. Ketika  $c_i \cdot = 0$ , maka persamaan (2.1) berubah menjadi distribusi multivariat normal.

### 2.6.2 Estimasi Parameter GEE2

Estimasi parameter dapat ditunjukkan dalam 'score' persamaan untuk  $\beta$  dan  $\alpha$  dengan menggunakan GEE2 yaitu

$$K^{1/2} \sum_{i=1}^n D_i' V_i^{-1} f_i = 0 \quad (2.2)$$

dimana

$$D_i = \begin{matrix} \frac{\partial \mu_i}{\partial \beta} & 0 \\ \frac{\partial \sigma_i}{\partial \beta} & \frac{\partial \sigma_i}{\partial \alpha} \end{matrix}$$

$$V_i = \begin{matrix} V_{111} & V_{112} \\ V_{121} & V_{222} \end{matrix} = \begin{matrix} \text{var}(y_i) & \text{cov}(y_i, s_i) \\ \text{cov}(s_i, y_i) & \text{var}(s_i) \end{matrix}$$

$$f_i = \begin{matrix} y_i - \mu_i \\ s_i - \sigma_i \end{matrix}$$

dengan  $s_i' = (s_{i11}, s_{i12}, \dots, s_{im_i m_i})$  dan  $s_{ijk} = y_{ij} - \mu_{ij} (y_{ik} - \mu_{ik})$  menjadi kovarian empiris (diperoleh dari hasil observasi) antara individu  $j$  dan  $k$ .

Turunan matriks untuk GEE2, yaitu pada persamaan (2.2) yang menunjukkan bahwa ada sebuah submatriks yang melibatkan turunan dari mean sehubungan dengan parameter mean. Persamaan estimasi kedua untuk kovarian dapat dilihat pada submatriks  $\frac{\partial \sigma_i}{\partial \beta}$  dan  $\frac{\partial \sigma_i}{\partial \alpha}$ , sehingga bentuk GEE2 sebenarnya sama dengan GEE biasa, tetapi secara bersamaan memecahkan mean dan kovarian.

Apabila diberikan estimasi awal  $(\beta_0, \alpha_0)$  dengan menetapkan  $\mu_i(\beta_0)$  dan  $\sigma_i(\beta_0, \alpha_0)$ , dengan menggunakan prosedur iterasi untuk menghitung  $\theta_i$  dan  $\lambda_i$ , maka akan mengarah pada perhitungan langsung dari  $V_i$ . Selanjutnya estimasi diubah menjadi

$$\begin{matrix} \beta_1 \\ \alpha_1 \end{matrix} = \begin{matrix} \beta_0 \\ \alpha_0 \end{matrix} + \sum_{i=1}^n D_{i0}' V_{i0}^{-1} D_{i0} \begin{matrix} -1 \\ \end{matrix} \sum_{i=1}^n D_{i0}' V_{i0}^{-1} f_{i0}$$

### 2.6.3 Struktur Working Covariance

Misalkan  $y_i$  adalah vektor dari variabel acak biner. Struktur mean dapat dimisalkan dengan  $E y_{ij} x_{ij} = \mu_{ij} = g x_{ij}' \beta_0$ . Fungsi varian dari distribusi binomial yang dipilih yaitu  $\sigma_{ij}^2 = \mu_{ij}(1 - \mu_{ij})$ . Kemudian  $\rho_{ijk} = \text{Corr } y_{ij}, y_{ik} / \sigma_{ij}, \sigma_{ik}$  digunakan untuk model hubungan antara  $\rho_{ijk}$  dan  $\alpha$ . Link hubungan fisherz-nya yaitu:



$$\text{Corr } y_{ij}, y_{ik} = \frac{\exp \{c x_{ij}, x_{ik}\}^\alpha - 1}{\exp \{c x_{ij}, x_{ik}\}^\alpha + 1}$$

dimana  $c$  adalah sebuah fungsi yang menggambarkan hubungan antara variabel eksplanatori untuk struktur korelasi  $x_{ij}$  dan  $x_{ik}$  serta koefisien korelasi. Kemudian pilih sebuah spesifikasi matriks *working covariance* untuk orde momen ketiga dan keempat. Berikut beberapa struktur *working covariance* yang diberikan.

a. *Independence matriks working covariance*

Struktur ini mengasumsikan pengamatan  $y_i$  adalah independen. Kemudian  $\text{Cov } y_i, s_i = 0$  dan  $\text{Var}(s_i)$  adalah diagonal dengan elemen  $\text{Var } s_{ijk} = y_{ij} - \mu_{ij}^2 y_{ik} - \mu_{ik}^2 = \sigma_{ij}^2 \sigma_{ik}^2$  untuk  $i = j$ . Matriks *working covariance* sepenuhnya ditentukan dengan menggunakan nilai distribusi normal untuk  $\text{Var } s_{ijk} = \text{Var } y_{ij} - \mu_{ij}^4 = 2\sigma_{ij}^2$ .

b. Matriks *working covariance* untuk aplikasi

Struktur ini memiliki matriks *working covariance*  $\text{Cov } y_i, s_i = 0$  dan  $\text{Var } s_i = I$ . Matriks *working covariance* ini mempunyai dua keuntungan. Pertama, yaitu menjamin sifat beraturan untuk bagian turunan dari matriks *working covariance* sehingga menghindari masalah kekonvergenan dari algoritma GEE2. Kedua, momen orde ketiga dan keempat tidak membutuhkan estimasi, yang mana dapat menambah kecepatan dari algoritma GEE2.

c. *Common working correlation*

*Common working correlation* ini untuk momen orde ketiga dan keempat. *Common working correlation* ini merupakan model dasar dari matriks *working covariance* dibawah normal. Misalkan

$$\text{Cov } y_{ij}, s_{ilm} = E \{ y_{ij} - \mu_{ij} \ y_{il} - \mu_{il} \ y_{im} - \mu_{im} \} = y_{jlm} \sigma_{ij}^2 \sigma_{il}^2 \sigma_{im}^2^{1/2}$$

$$\begin{aligned} \text{COV } S_{ijk}, S_{ilm} &= E (y_{ij} - \mu_{ij} \quad y_{ik} - \mu_{ik} \quad y_{il} - \mu_{il} \quad y_{im} - \mu_{im}) - \sigma_{ijk}\sigma_{ilm} \\ &= \sigma_{ijl}\sigma_{ikm} + \sigma_{ijm}\sigma_{ikl} + \delta_{jklm} \overline{\sigma_{ij}^2\sigma_{ik}^2\sigma_{il}^2\sigma_{im}^2} \end{aligned}$$

dengan penambahan parameter  $\gamma_{jlm}$  dan  $\delta_{jklm}$  yang dapat diestimasi konsisten menggunakan mean. Versi yang mudah dari matriks kovarian ini dapat diperoleh dengan menyamakan spesifikasi parameter  $\gamma_{jlm}$  dan  $\delta_{jklm}$ . Hal tersebut mungkin terjadi jika elemen dari  $y_i$  mengikuti struktur *exchangeable*.

d. *Independence* dengan struktural tak nol

Struktur ini memiliki matriks *working covariance*  $\text{COV } y_{ij}, S_{ijk} = 1 - 2\mu_{ij} \quad 1 - 2\mu_{ik} \quad \sigma_{ijk}$ ,  $\text{Var } S_{ijk} = \sigma_{ij}^2\sigma_{ik}^2 + 1 - 2\mu_{ij} \quad 1 - 2\mu_{ik} \quad \sigma_{ijk}$ ,  $\text{COV } S_{ijk}, S_{ijl} = \sigma_{ij}^2\sigma_{ik}^2 - \sigma_{ijk}\sigma_{ijl}$ . Perhatian utama terhadap penggunaan struktur *working covariance* ini dalam pengaplikasiannya yaitu mengakibatkan ketunggalan matriks *working covariance* untuk dua alasan. Pertama, keduanya  $\text{Var } S_{ijk}$  dan  $\text{COV } S_{ijk}, S_{ijl}$  termasuk hasil dari dua varian dan hasilnya dapat mendekati 0. Kedua, suatu istilah yang dikurangi dari keduanya  $\text{Var } S_{ijk}$  dan  $\text{COV } S_{ijk}, S_{ijl}$  dapat pula mendekati nilai 0. Selain itu, karena kestabilan numerik, estimasi dari  $\text{Var } S_{ijk}$  dapat menjadi negatif (Ziegler, 2011).

## 2.7 Goodness of Fit dari Model GEE

*Goodness of Fit* atau uji kelayakan model dari model GEE diuji untuk mengetahui ketepatan model yang terbaik. Berikut metode yang dapat digunakan untuk memperoleh hasil tentang ketepatan model yang terbaik, diantaranya.

### 2.7.1 Quasi-likelihood Information Criterion (QIC)

QIC (*Quasi-likelihood Information Criterion*) merupakan modifikasi atau perluasan dari AIC (*Akaike Information Criterion*). Menurut Pan (2001), QIC

merupakan uji pemilihan model terbaik diantara beberapa korelasi yang didefinisikan dengan

$$QIC R = -2Q \hat{\beta} R ; I, \mathcal{D} + 2\text{trace}(\hat{\Omega}_I \hat{V}_R)$$

dimana

$\hat{\beta} R$  : solusi GEE untuk struktur korelasi R.

$\hat{V}_R$  : estimasi Robust varian untuk struktur korelasi R.

$\mathcal{D}$  :  $y_i x_{ij}$  dengan  $y_i$  : vektor respon dan  $x_{ij}$  : vektor kovariat; serta

$$\hat{\Omega}_I : - \frac{\partial^2 Q \hat{\beta} ; I, \mathcal{D}}{\partial \beta \partial \beta'} \Big|_{\beta = \hat{\beta}(R)}$$

Untuk mendapatkan QIC dapat digunakan bantuan paket R.

### 2.7.2 Uji Hipotesis

Menurut Fitzmaurice, *et al* (2009) pengujian hipotesis dilakukan setelah memperoleh model GEE2 untuk respon. Pengujian dilakukan terhadap signifikansi parameter terhadap variabel respon dengan uji Wald. Secara umum pola uji signifikansi  $H_0$  dan  $H_1$  adalah

$$H_0: \beta_{it} = 0$$

$$H_1: \beta_{it} \neq 0$$

dengan  $i = 1, \dots, N$  dan  $t = 1 \dots T$ . Statistik ujinya adalah

$$Z_i = \frac{\hat{\beta}_{it}}{SE(\hat{\beta})_{it}}$$

## 2.8 Data Longitudinal

Data longitudinal merupakan sekumpulan pengukuran dari unit percobaan yang sama dan biasanya menunjukkan subjek atau individu dari waktu ke waktu. Data longitudinal banyak ditemui dalam bidang biologi, kedokteran, sosiologi, psikologi, dan sering muncul dalam aturan faktorial (Noguchi *et al*, 2012). Data longitudinal dapat menjelaskan dua macam informasi yaitu informasi *cross-section* pada perbedaan antar subjek dan informasi *time series* yang merefleksikan perubahan

pada subjek waktu. Ketika kedua informasi tersebut tersedia, maka analisis data longitudinal dapat digunakan

Menurut Ruspini dalam Nurdini (2006), data longitudinal memiliki beberapa karakteristik sebagai berikut:

1. data yang dikumpulkan untuk setiap variabel dalam dua atau lebih periode waktu tertentu.
2. subjek atau studi kasus yang dianalisis sama, atau setidaknya dapat dibandingkan antara satu periode dengan periode setelahnya.
3. analisis melibatkan perbandingan data yang sama dalam satu periode.

## 2.9 R-shiny

R-shiny merupakan paket dalam R yang berbasis web dan digunakan sebagai situs pembelajaran virtual. Menurut Tirta (2014), R-shiny bekerja di server (belakang layar) sedangkan pengguna disugahi menu yang ramah dan interaktif berbasis web, sehingga pengguna tidak disyaratkan menguasai program R. Dengan menggunakan program sebagai mesin pengolah, simulasi dan visualisasi dapat dilakukan pada semua aspek analisis data (seperti simulasi data, estimasi, dan pemeriksaan *goodness of fit*).

Secara umum, komponen pada program Shiny dibedakan menjadi dua kelompok besar, yaitu:

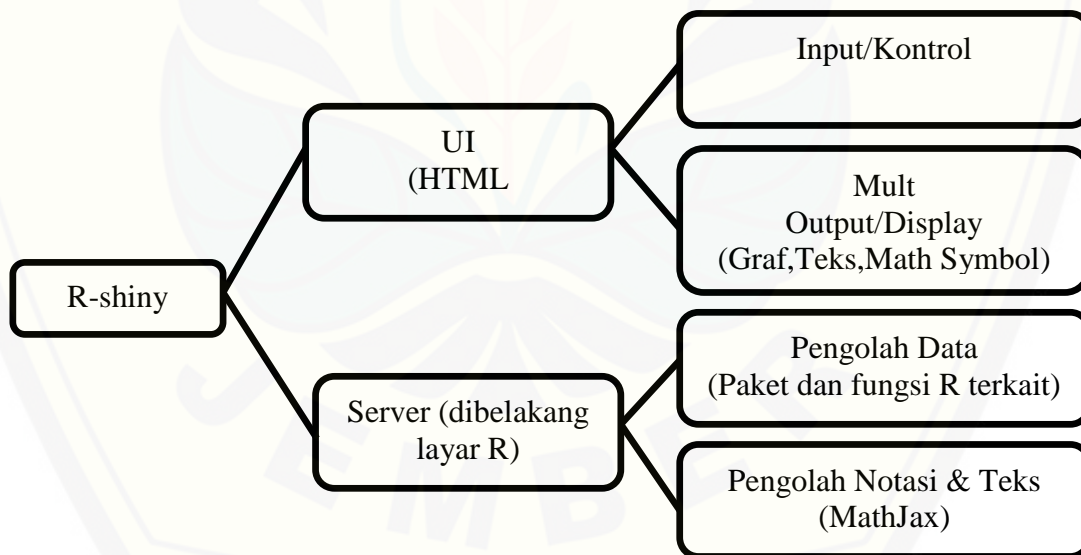
1. *user interface*. Bagian ini bermanfaat untuk :
  - a. panel kontrol adalah panel untuk mengontrol input berupa data, variabel, model, tergantung kompleksitas modul. Tampilan kontrol dapat berupa slider, radio button, check-box, dan lain-lain.
  - b. pemasukan permintaan nilai input yaitu berupa data dengan berbagai jenis variabel yang diperlukan, pemilihan model, jenis dan kriteria uji statistika.
  - c. penyajian output yaitu terkait hasil atau uji. Hasil output dapat berupa grafik (histogram, diagram pencar, dan lain-lain), bentuk angka atau teks bisa

berbentuk asli (verbatim) maupun dalam bentuk tabel, dan teks khusus bernotasi matematika dengan format LaTeX.

Untuk mengakomodasi berbagai jenis luaran tadi, shiny juga menyediakan berbagai format output seperti `plotOutput`, `textOutput`, `verbatimTextOutput`, `tableOutput`, dan lain-lain. Bagian ini dapat disajikan pada file khusus `ui.r`, dan dapat juga disajikan penuh melalui file HTML, misalnya `index.html`.

2. *server*. Bagian yang merupakan otak dari program yang bertugas melakukan simulasi, berbagai analisis data sesuai pilihan pengguna dan selanjutnya mengirim hasilnya ke bagian output. Bagian ini didukung oleh berbagai prosedur analisis data yang pada umumnya telah tersedia pada berbagai paket R. Bagian ini disimpan dalam file yang diberi nama `server.r`.

Struktur umum komponen pemrograman dengan R-shiny dapat dilihat pada Gambar 2.1 berikut:



Gambar 2.1. Struktur Komponen R-shiny

## BAB 3. METODE PENELITIAN

### 3.1 Langkah-langkah Pembuatan Program GEE2 Berbasis Web Interaktif

Penelitian ini menggunakan program R i386 3.1.3 dengan bantuan R-Shiny untuk membuat program berbasis web interaktif. Adapun langkah-langkah pembuatan program GEE2 berbasis web interaktif menggunakan R-Shiny ini dibagi menjadi dua tahap.

a. Tahap I, penyusunan web interaktif menggunakan R-Shiny yaitu:

1. menentukan input dan output program. Input dalam penelitian ini berisi menu untuk pilihan data longitudinal dan formula dari model GEE2 . Sedangkan output program yang akan dihasilkan berupa *summary* dari hasil analisis menggunakan metode GEE2 berupa hasil estimasi, *standard error* dan *p-value*, serta hasil *Goodness of Fit*. Paket yang digunakan adalah paket `geepack` dengan fungsi `geese()` untuk analisis data menggunakan GEE2 dan `geeglm()` untuk menentukan *Goodness of Fit* dengan nilai QIC ditentukan menggunakan paket `MuMIn`. Struktur fungsi `geese()` dan `geeglm()` pada paket `geepack` disajikan dalam persamaan berikut :

```
geese(formula = formula(data), sformula = ~1, id, waves =
NULL,data = parent.frame(), subset = NULL, na.action =
na.omit,contrasts = NULL, weights = NULL, zcor = NULL, corp
= NULL,control = geese.control(...), b = NULL, alpha =
NULL, gm = NULL,family = gaussian(), mean.link = NULL,
variance = NULL,cor.link = "identity", sca.link =
"identity", link.same = TRUE,scale.fix = FALSE, scale.value
= 1, corstr = "independence", ...)
```

```
geeglm(formula, family = gaussian, data=parent.frame(),
weights, subset, na.action, start = NULL, etastart,
mustart, offset,control = geese.control(...), method =
"glm.fit", x = FALSE, y = TRUE,contrasts = NULL, id,
waves=NULL, zcor=NULL,corstr = "independence",scale.fix =
FALSE,scale.value =1, std.err="san.se",...)
```

Keterangan :

`formula` : menyatakan hubungan antara variabel respon ( $y$ ) dan variabel prediktor ( $x$ ) dalam bentuk formula

`sformula = ~1` : sebuah ekspresi formula dari bentuk `~prediktor`, sedangkan respon diabaikan. Ekspresi ini menyatakan prediktor linier dari model dispersi

`id` : panjang vektor yang mengidentifikasi kluster. Panjang vektor harus sama dengan jumlah observasi yang dilakukan

`data` : data yang digunakan dalam observasi yang menafsirkan variabel dalam formula

`family` : menyatakan jenis distribusi dan link yang digunakan pada model

`mean.link` : karakter *string* yang mengkhususkan fungsi link untuk mean

`variance` : karakter *string* yang mengkhususkan fungsi varian untuk mean

`cor.link` : karakter *string* yang mengkhususkan fungsi link untuk koefisien korelasi, link yang disediakan yaitu "identity" dan "fisherz"

`corstr` : karakter *string* yang mengkhususkan pada struktur korelasi. Struktur korelasi yang disediakan yaitu "independence", "exchangeable", "ar1", "unstructure".

Struktur fungsi yang biasa digunakan dalam menganalisis data menggunakan metode

GEE2 dengan fungsi `geese()` yaitu `formula, sformula = ~1, id, family = gaussian(), mean.link = NULL, variance = NULL, cor.link = "identity", corstr = "independence"`. Sedangkan fungsi `geeglm()` yang digunakan pada penelitian ini hanya untuk menentukan nilai QIC dan *quasilikelihood* dengan paket `MuMIn` yaitu:

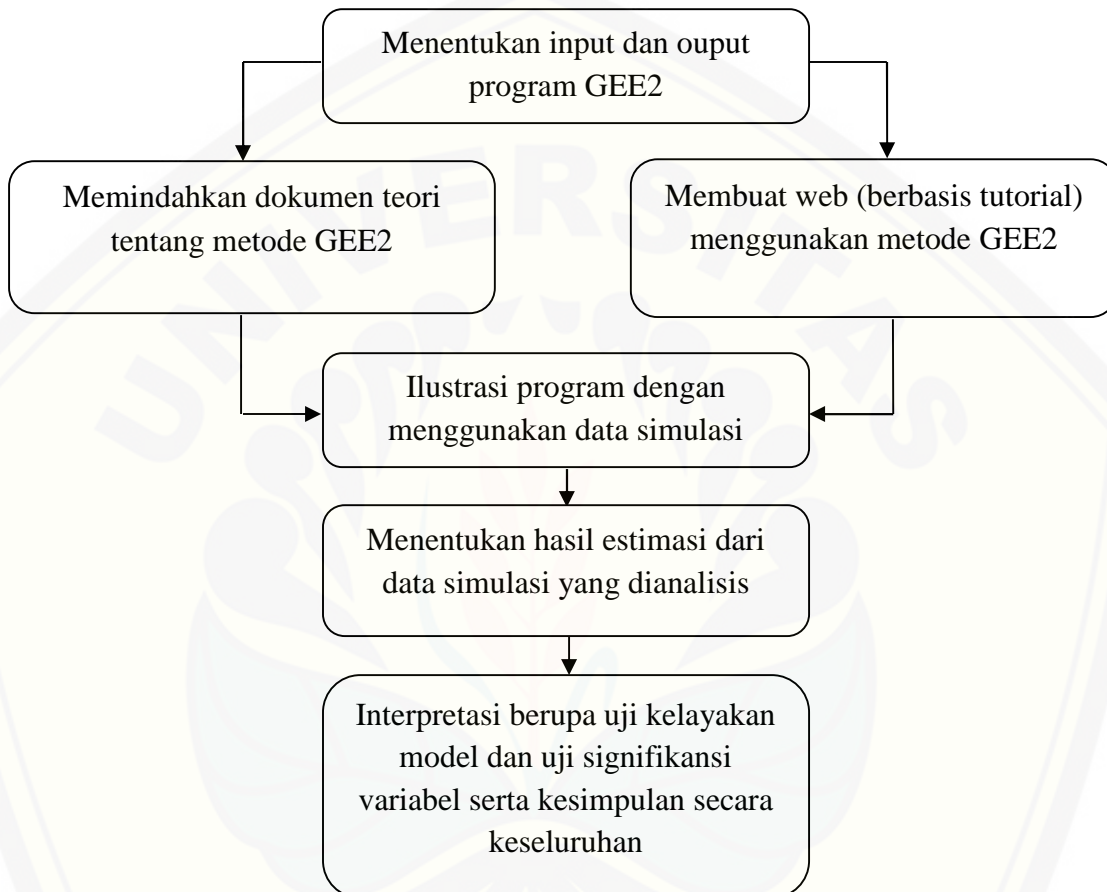
```
QIC(object, ..., typeR = FALSE)
quasiLik(object, ...)
```

2. meringkas dan mentransfer dokumen teori tentang metode GEE2 kedalam format LaTeX agar dapat ditampilkan secara online.

3. membuat web (berbasis tutorial) dengan data simulasi menggunakan program GEE2 berbasis web interaktif ini. Program menggunakan R-shiny berupa *user interface* dan *server*. *User interface* berupa file HTML yang digunakan untuk memasukkan nilai input dan menampilkan output. Sedangkan *server* merupakan bagian untuk mengaplikasikan program GEE2 untuk data longitudinal dengan paket `geepack` yang telah dijelaskan pada poin sebelumnya.
- b. Tahap II, ilustrasi penggunaan program web interaktif yaitu:
  4. menganalisis data simulasi dengan menggunakan program GEE2 berbasis web interaktif yang sudah jadi untuk mengetahui program R-shiny berjalan dengan baik.
  5. menentukan hasil estimasi data menggunakan GEE2 berbasis web interaktif ini berupa model link korelasi, estimasi parameter, *standard error*, *p-value*, dan nilai *goodness of fit*.
  6. interpretasi berupa hasil analisis data menggunakan metode GEE2 berbasis web interaktif yang kemudian akan dibandingkan hasil antara link korelasinya berdasarkan nilai *standard error*-nya dan menentukan uji signifikansi dari hasil analisis data dengan melihat nilai *p-value*. Sedangkan kesimpulan berupa hasil keseluruhan dari interpretasi model GEE2 berbasis web interaktif.



Langkah-langkah pembuatan program GEE2 berbasis web interaktif menggunakan R-Shiny yang disajikan dalam skema dapat dilihat pada Gambar 3.1 berikut ini :



Gambar 3.1 Skema Langkah-langkah Pengolahan Data

### 3.2 Ilustrasi

Data yang digunakan dalam penelitian menggunakan GEE2 adalah data longitudinal yaitu data yang diukur lebih dari satu kali. Sedangkan data yang digunakan untuk uji coba menggunakan program GEE2 berbasis web interaktif menggunakan R-Shiny ini merupakan data simulasi. Data simulasi ini sebanyak 200 yang terbagi dalam 40 level dengan variabel-variabel berikut.

a. Variabel Respon ( $Y$ )

Variabel respon  $Y$  yang terdapat pada data simulasi ini terdiri dari SHB dan BS. SHB merupakan skor atau nilai hasil belajar siswa-siswi suatu SMA tiap semester, dari semester 1 sampai semester 5. Variabel respon SHB ini merupakan variabel respon kontinu. Sedangkan BS merupakan penilaian afektif siswa-siswi SMA yang ditunjukkan dengan angka 0 (kurang memuaskan) dan 1 (memuaskan). Variabel respon BS merupakan variabel respon biner.

b. Variabel Prediktor ( $X$ )

Variabel Prediktor ( $X$ ) pada data terdiri dari Sekolah ( $X_1$ ), Gender ( $X_2$ ), STPA ( $X_3$ ), dan SUN( $X$ ). Sekolah terdiri atas 4 macam yaitu A, B, C, dan D. Gender terdiri atas L (laki-laki) dan P (perempuan). STPA merupakan skor atau nilai TPA (Tes Potensial Akademik) dan SUN merupakan skor atau nilai UN (Ujian Nasional) ketika SMP.

Data simulasi diatas dapat dijelaskan dalam bentuk tabel. Berikut penjelasan identifikasi variabel pada data simulasi yang disajikan dalam Tabel 3.1:

Tabel 3.1 Identifikasi Variabel Respon Kontinue

Subject	id	Tes					$X_1$	$X_2$	$X_3$	$X_4$
		$Y_1$ (T1)	$Y_2$ (T2)	$Y_3$ (T3)	$Y_4$ (T4)	$Y_5$ (T5)				
1	1	$y_{11}$	$y_{12}$	$y_{13}$	$y_{14}$	$y_{15}$	A	L	$x_{31}$	$x_{41}$
2	2	$y_{21}$	$y_{22}$	$y_{23}$	$y_{24}$	$y_{25}$	A	L	$x_{31}$	$x_{41}$
3	$\vdots$	$y_{31}$	$y_{32}$	$y_{33}$	$y_{34}$	$y_{35}$	A	L	$\vdots$	$\vdots$
$\vdots$	5	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$x_{31}$	$x_{41}$
26	1	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	A	P	$x_{32}$	$x_{42}$
$\vdots$	2	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$x_{32}$	$x_{42}$
51	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	B	L	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n$	5	$y_{n1}$	$y_{n2}$	$y_{n3}$	$y_{n4}$	$y_{n5}$	D	P	$x_{3n}$	$x_{4n}$