



**PENERAPAN *GENERALIZED ESTIMATING EQUATION* (GEE)
BERBASIS WEB INTERAKTIF DENGAN R-SHINY UNTUK
RESPON MULTINOMIAL BERSKALA ORDINAL**

SKRIPSI

Oleh

**Lina Wijasih
NIM 111810101014**

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



**PENERAPAN *GENERALIZED ESTIMATING EQUATION* (GEE)
BERBASIS WEB INTERAKTIF DENGAN R-SHINY UNTUK
RESPON MULTINOMIAL BERSKALA ORDINAL**

SKRIPSI

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

Oleh

**Lina Wijiasih
NIM 1118101014**

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

PERSEMBAHAN

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

1. Ayahanda Trubus dan Ibunda Tentrem Kasiati yang selalu sabar memberi nasihat, bimbingan, doa, dan kasih sayang yang tiada putusnya;
2. Adikku Lukiya Wanda Risma dan semua keluarga besar yang selalu memberikan doa, dukungan dan motivasi;
3. guru-guruku sejak sekolah dasar hingga perguruan tinggi yang telah memberikan ilmu dan membimbing dengan penuh kesabaran;
4. Almamater Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

MOTTO

“Karena sesungguhnya sesudah kesulitan itu ada kemudahan.”
(Terjemahan Q.S. Alam Nasyrak: 6)^{*)}

“Pendidikan merupakan perlengkapan paling baik untuk hari tua.”
(Aristoteles)^{**)}

“Sesuatu yang belum dikerjakan, seringkali tampak mustahil, kita baru yakin kalau kita telah berhasil melakukannya dengan baik.”
(Evelyn Underhill)^{***)}

^{*)} Departemen Agama Republik Indonesia. 1989. *Al Qur'an dan Terjemahannya*. Semarang: CV Toha Putra

^{**)} <http://www.maribelajarbkk.web.id/2015/03/contoh-motto-terbaru-dalam-skripsi.html>
[11 November 2015]

^{***)} <http://www.maribelajarbkk.web.id/2015/03/contoh-motto-terbaru-dalam-skripsi.html>
[11 November 2015]

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

nama : Lina Wijiasih

NIM : 111810101014

menyatakan dengan sesungguhnya bahwa skripsi yang berjudul “Penerapan *Generalized Estimating Equation* (GEE) Berbasis Web Interaktif Dengan R-Shiny Untuk Respon Multinomial Berskala Ordinal” 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 Wijiasih

NIM 111810101014

SKRIPSI

**PENERAPAN *GENERALIZED ESTIMATING EQUATION* (GEE)
BERBASIS WEB INTERAKTIF DENGAN R-SHINY UNTUK
RESPON MULTINOMIAL BERSKALA ORDINAL**

Oleh:
Lina Wijiasih
NIM 1118101014

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 “Penerapan *Generalized Estimating Equation* (GEE) Berbasis Web Interaktif Dengan R-Shiny Untuk Respon Multinomial Berskala Ordinal” 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 1982022162006042002

Anggota I,

Anggota II,

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

M. Ziaul Arif, S.Si., M.Sc.
NIP 198501112008121002

Mengesahkan

Dekan,

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

RINGKASAN

Penerapan *Generalized Estimating Equation* (GEE) Berbasis Web Interaktif Dengan R-Shiny Untuk Respon Multinomial Berskala Ordinal; Lina Wijiasih, 111810101014; 2015; 55 halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Generalized Estimating Equation (GEE) merupakan salah satu metode statistika yang digunakan untuk menganalisa data berkorelasi salah satunya karena pengukuran berulang (*repeated measurement*). Data dengan respon berkorelasi disebut sebagai data longitudinal. Metode GEE dalam penelitian ini, diterapkan pada respon multinomial berskala ordinal.

Salah paket dalam R yang digunakan untuk analisis data dengan metode *Generalized Estimating Equation* (GEE) berskala ordinal adalah paket `multgee` dengan fungsi `ordLORgee()`. Namun, dalam penggunaannya paket tersebut tidak mudah terutama bagi peneliti yang kurang menguasai pemrograman yang dalam hal ini program R. Selain itu, untuk program GEE baik itu GEE binomial, GEE2 maupun GEE multinomial belum ada yang menggunakan sistem GUI. Sehingga, dalam penelitian ini akan dibuat program GEE multinomial berskala ordinal berbasis web interaktif menggunakan R-shiny. Program dibuat dalam bentuk tutorial yang meliputi ringkasan teori, aplikasi dan hasil analisis data.

Web interaktif program GEE multinomial berskala ordinal ini dapat diakses di alamat <http://statslab-rshiny.fmipa.unej.ac.id/JORS/MultGEEOrd/>. Analisis data menggunakan web interaktif program GEE multinomial berskala ordinal ini dapat dilakukan dengan pilihan data yang tersedia dalam menu atau mengimpor data milik pengguna. Untuk data respon multinomial dengan skala ordinal berkorelasi dipilih dua jenis struktur rasio odds lokal yaitu *uniform* dan *category exchangeability*. *Goodness of fit* untuk model GEE multinomial berskala ordinal dilihat berdasarkan

nilai *root mean square error* (RMSE) untuk memilih model yang lebih baik antara dua model dengan struktur berbeda. Untuk uji signifikansi parameter digunakan *p-value*.

Aplikasi data *Lapharoscopic Cholecystectomy* menggunakan GEE multinomial ordinal yaitu menganalisa tentang tingkat *Lapharoscopic Cholecystectomy* yang terdiri dari lima tingkatan. Variabel prediktornya meliputi terapi, jenis kelamin, umur, dan waktu pengukuran. Dari hasil analisis data diperoleh model dengan struktur rasio odds lokal *Uniform* merupakan model yang lebih baik dibandingkan dengan model dengan struktur rasio odds lokal *Category Exchangeability* dan dari uji signifikansi diperoleh model bahwa terapi tidak aktif (Terapi[TA]) dan waktu pengukuran keenam (Waktu[F]) signifikan terhadap respon. Bentuk persamaan yang diperoleh yaitu

$$g_1(\pi_1) = 0,76260 - 1,89322X_{i1}[TA] + 1,03980X_{i4}[F]$$

$$g_2(\pi_2) = 1,61849 - 1,89322X_{i1}[TA] + 1,03980X_{i4}[F]$$

$$g_3(\pi_3) = 2,59122 - 1,89322X_{i1}[TA] + 1,03980X_{i4}[F]$$

$$g_4(\pi_4) = 3,94952 - 1,89322X_{i1}[TA] + 1,03980X_{i4}[F]$$

dengan $g_i(\pi_i)$ adalah peluang respon pada kategori $i + 1$

PRAKATA

Puji syukur kehadiran Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Penerapan *Generalized Estimating Equation* (GEE) Berbasis Web Interaktif Dengan R-Shiny Untuk Respon Multinomial Berskala Ordinal”. Skripsi ini disusun untuk memenuhi salah satu syarat untuk 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 ingin 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 M. Ziaul Arif, S.Si., M.Sc selaku Dosen Penguji II yang telah memberi masukan berupa saran dan kritiknya terhadap skripsi ini;
3. Prof. Drs. Kusno, DEA., Ph.D selaku Dosen Pembimbing Akademik yang telah sabar membimbing penulis selama menjadi mahasiswa;
4. sahabat-sahabatku Ika Novitasari, Istiqomah, dan teman-teman kos Jawa VII No.10 atas do'a, dukungan dan semangatnya;
5. sahabat-sahabatku seperjuangan KRAMAT'11, khususnya Arum, Dewi, Kiki dan Lina Choiril yang telah memberikan do'a, 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

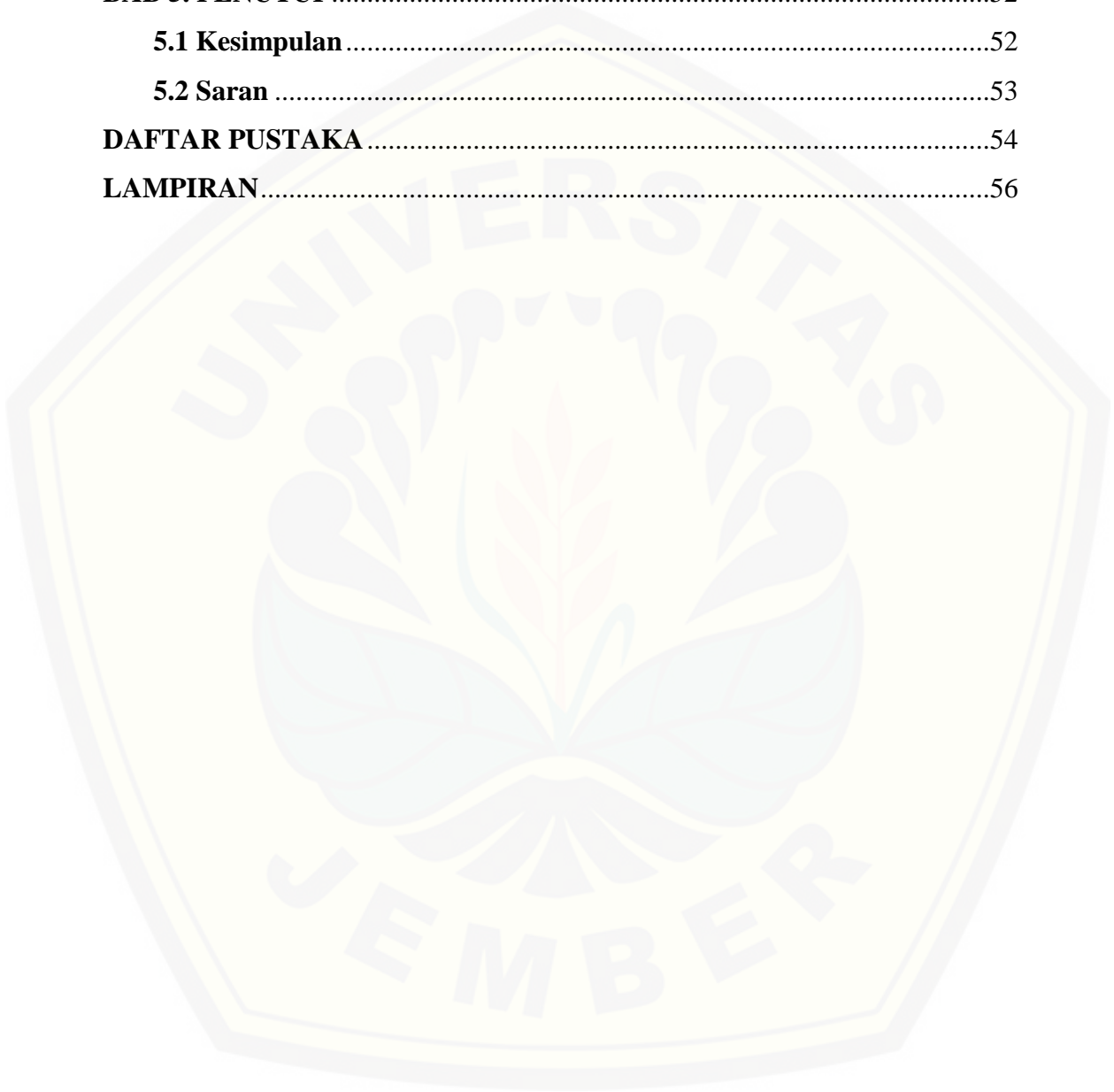
Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
PERSEMBAHAN.....	ii
MOTTO	iii
PERNYATAAN.....	iv
HALAMAN BIMBINGAN.....	v
PENGESAHAN	vi
RINGKASAN	vii
PRAKATA	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN	xv
BAB 1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan.....	3
1.4 Manfaat.....	4
BAB 2. TINJAUAN PUSTAKA.....	5
2.1 Analisis Regresi	5
2.2 Model Linier	5
2.3 Data Longitudinal	6
2.4 Distribusi Multinomial	6
2.5 Regresi Logistik Multinomial	7
2.6 Pendekatan GEE.....	7

2.6.1 Struktur <i>Working Correlation</i>	8
2.6.2 Estimasi Parameter	10
2.6.3 Algoritma Fischer Scoring	11
2.7 GEE untuk Respon Multinomial	12
2.7.1 Model Marjinal untuk Respon Multinomial Berskala Ordinal	12
2.7.2 Model Marjinal untuk Respon Multinomial Berskala Nominal....	13
2.8 Struktur Dasar GEE untuk Respon Multinomial	13
2.8.1 Estimasi Parameter	14
2.8.2 Sifat-sifat Asimtotik dari Estimator GEE.....	15
2.9 Rasio Odds Lokal untuk Pendekatan GEE	16
2.9.1 <i>Iterative Proportional Fitting Procedure (IPFP)</i>	17
2.9.2 Struktur Rasio Odds Lokal	18
2.9.3 Paket untuk Implementasi Pendekatan GEE Lokal Odds Rasio	20
2.10 R-Shiny	20
BAB 3. METODE PENELITIAN	22
3.1 Langkah-Langkah Pembuatan Web Interaktif	22
3.2 Langkah – Langkah Ilustrasi Data	24
BAB 4. HASIL DAN PEMBAHASAN	27
4.1 Hasil dan Pembahasan Fitur Web	28
4.1.1 Penyusunan Web Interaktif Berbasis R-Shiny	28
4.1.2 Tampilan Web untuk Model GEE Multinomial Ordinal dengan R-Shiny	32
4.2 Hasil dan Pembahasan Ilustrasi Data	39
4.2.1 Ilustrasi Data GEE Multinomial Ordinal dengan Korelasi <i>Uniform</i>	41
4.2.2 Ilustarsi Data GEE Multinomial Ordinal dengan Korelasi <i>Category Exchangeability</i>	44

4.2.3 Uji Kecocokan Model	47
4.2.4 Prediksi Peluang Model	48
BAB 5. PENUTUP	52
5.1 Kesimpulan	52
5.2 Saran	53
DAFTAR PUSTAKA	54
LAMPIRAN	56



DAFTAR TABEL

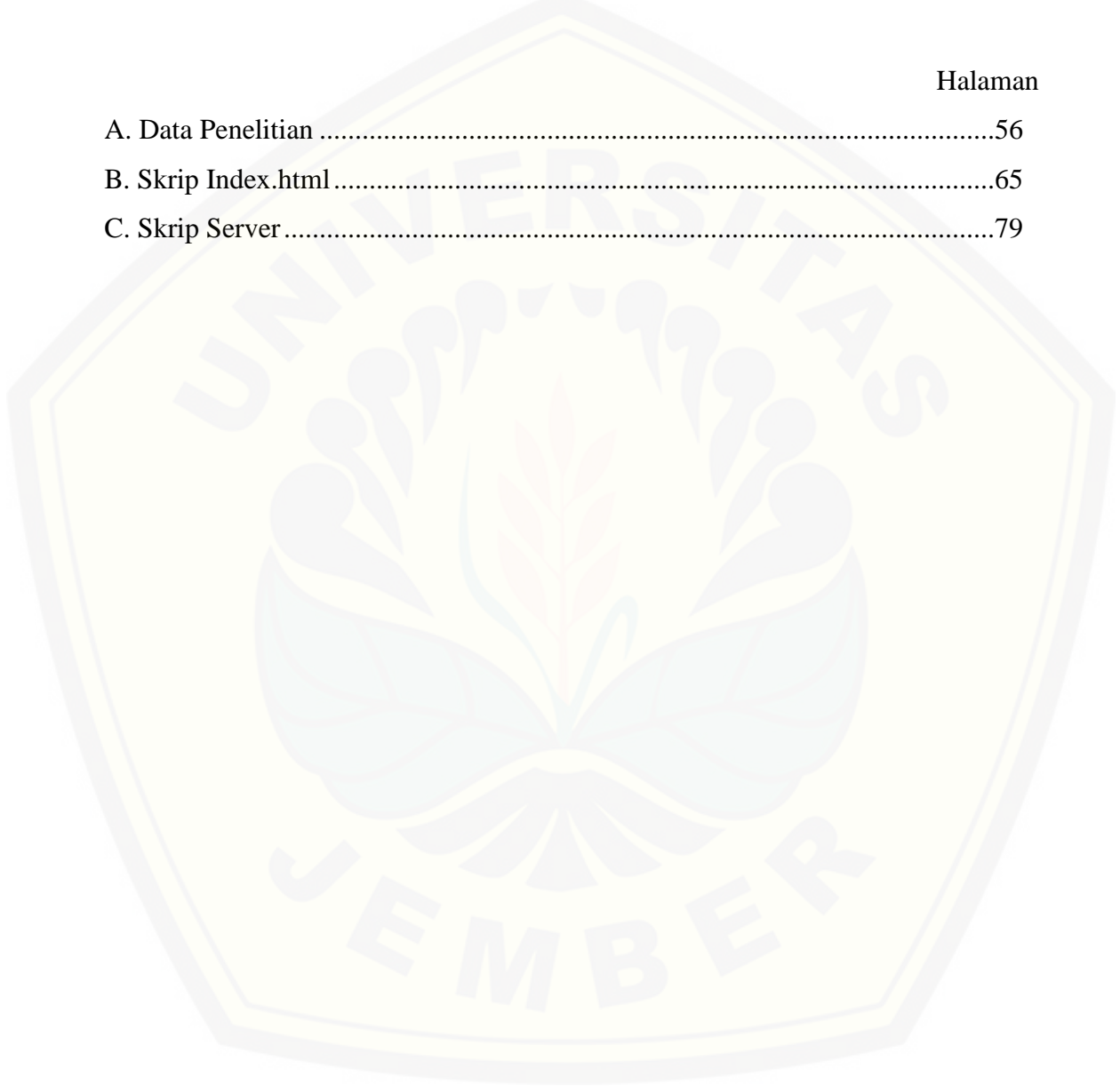
	Halaman
4.1 Struktur Fungsi R Skrip dan Web Interaktif.....	38
4.2 Hasil Ilustrasi dengan Struktur Rasio Odds Lokal <i>Uniform</i>	41
4.3 Hasil Ilustrasi dengan Parameter Signifikan untuk Struktur <i>Uniform</i>	43
4.4 Hasil Ilustrasi dengan Struktur Rasio Odds Lokal <i>Category Exchangeability</i>	45
4.5 Hasil Ilustrasi dengan Parameter Signifikan untuk Struktur <i>Category</i> <i>Exchangeability</i>	46
4.6 Nilai RMSE untuk Korelasi <i>Uniform</i> dan <i>Category Exchangeability</i>	48

DAFTAR GAMBAR

	Halaman
2.1 Struktur Komponen R-Shiny.....	21
3.1 Diagram Alir Metode Penelitian	26
4.1 Tampilan Awal Program	33
4.2 Tampilan Pilihan Data.....	34
4.3 Tampilan Ilustrasi Data.....	34
4.4 Tampilan Menu Fungsi Link.....	35
4.5 Tampilan Menu Struktur Rasio Odds Lokal	35
4.6 Tampilan Hasil Ilustrasi Data	36
4.7 Tampilan Nilai <i>Root Mean Square Error</i>	36
4.8 Tampilan Pilihan Hasil Ilustrasi.....	37

DAFTAR LAMPIRAN

	Halaman
A. Data Penelitian	56
B. Skrip Index.html	65
C. Skrip Server	79



BAB 1. PENDAHULUAN

1.1 Latar Belakang

Pada mulanya, model linier diasumsikan berdistribusi normal dan saling bebas. Namun pada kenyataannya, asumsi tersebut sering tidak terpenuhi. Sehingga, kemudian berkembang model linier untuk mengatasi data yang tidak harus berdistribusi normal tetapi harus saling bebas. Model linier tersebut yaitu *Generalized Linear Models* (GLM) atau yang biasa dikenal dengan Model Linier Tergeneralisasi.

Generalized Linear Models (GLM) merupakan salah satu metode statistika yang menangani masalah data tidak berdistribusi normal tetapi termasuk dalam keluarga eksponensial. Salah satu asumsi GLM yaitu respon saling bebas dari keluarga eksponensial (Tirta, 2009). Tetapi dalam kenyataannya banyak ditemukan hubungan respon yang tidak saling bebas atau respon yang saling berkorelasi salah satunya karena pengukuran berulang (*repeated measurement*). Sehingga, apabila respon berkorelasi maka hal itu dapat melanggar asumsi GLM. Data dengan respon berkorelasi disebut sebagai data longitudinal.

Data longitudinal merupakan sekumpulan pengukuran dari unit percobaan yang sama, yang biasanya menunjukkan subjek atau individu dari waktu ke waktu (Noguchi *et al*, 2012). Data longitudinal diamati secara berurutan dari waktu ke waktu yang dikumpulkan dari rancangan percobaan atau studi observasi dimana variabel respon berhubungan dengan urutan sebuah peristiwa yang tercatat dititik waktu tertentu dalam masa studi (Zeger *et al*, 1988). Untuk menganalisa data longitudinal digunakan analisis *Generalized Estimating Equation* (GEE).

Tambun (2012) menerapkan GEE untuk model bivariat dengan respon biner. Keragaman pemilihan respon dalam model bivariat dengan respon biner mengakibatkan ketidakbebasan antar peubah respon. Hidayati (2013) mengembangkan penelitian Tambun (2012) dengan menggunakan metode yang

diperkenalkan oleh Prentice (1988) dan Zhao&Prentice (1990) yang disebut dengan *Generalized Estimating Equation* orde dua (GEE2). GEE2 diperkenalkan untuk mengatasi masalah pada GEE yang menganggap korelasi sebagai *nuisance* dengan menerapkan analisa mean dan struktur korelasi secara bersamaan, selain itu pada GEE2 diterapkan estimasi bersama untuk koefisien regresi dan koefisien korelasi dengan memodelkan link korelasi yang tidak diterapkan pada GEE biasa.

Beberapa penelitian yang disebutkan di atas, semuanya masih menggunakan data dengan respon biner, tetapi belum dapat diterapkan untuk data dengan respon multinomial (ordinal dan nominal). Dalam penelitian Lumley (1996) telah dibahas GEE untuk data ordinal dengan parameter rasio odds global. Touloumis, *et.al.*, (2013) memperkenalkan GEE yang diterapkan untuk data respon multinomial dengan menggunakan parameter rasio odds lokal untuk untuk matriks *working* kovariannya. Penelitian Touloumis, *et.al.*, (2013) dapat diterapkan untuk respon ordinal dan respon nominal. Dalam penelitian ini akan dibahas aplikasi pendekatan GEE respon multinomial untuk data ordinal menggunakan parameter lokal odds rasio.

Dalam menganalisis GEE untuk respon multinomial, Touloumis (2014) dan Touloumis (2015) telah memperkenalkan paket dalam program R yaitu paket *multgee*. Namun, dalam penggunaannya paket tersebut tidak mudah terutama bagi peneliti yang kurang paham tentang pemrograman yang dalam hal ini program R. Selain itu, untuk program GEE baik itu GEE binomial, GEE2 maupun GEE multinomial belum ada yang menggunakan sitem GUI. Sehingga, dalam tugas akhir ini paket *multgee* dalam pengaplikasian GEE respon multinomial untuk data ordinal akan dibuat *interface* web interaktif dengan menggunakan R-shiny. Hal ini dimaksudkan untuk mempermudah seseorang yang kurang mengerti pemrograman R menjadi lebih mudah untuk melakukan analisis GEE multinomial pada data ordinal secara online hanya dengan memasukkan datanya dan memilih menu tanpa harus menginstall program R dan melakukannya secara *offline* dalam komputernya dan tidak diwajibkan bagi pengguna web untuk menguasai program terlebih dahulu.

Selanjutnya, dilakukan analisis data menggunakan program web interaktif untuk GEE multinomial berskala ordinal dengan membandingkan dua struktur rasio odds lokal yaitu *uniform* dan *category exchangeability* menggunakan data *Laparoscopic Cholecystectomy* dari penelitian Jorgensen, *et.al.*, (1995) dalam Lumley (1996).

1.2 Rumusan Masalah

Dari beberapa keterangan diatas, maka didapat beberapa permasalahan yang akan dibahas dalam penelitian ini

1. Bagaimana aplikasi analisis GEE repon multinomial untuk data berskala ordinal berbasis web interaktif ?
2. Apa saja kelebihan dan kekurangan program GEE repon multinomial untuk data berskala ordinal berbasis web interaktif ?
3. Bagaimana model dari respon multinomial berskala ordinal dengan pendekatan GEE untuk data *Laparoscopic Cholecystectomy*?
4. Bagaimana inferensi model suatu respon multinomial dengan metode GEE data *Laparoscopic Cholecystectomy*?

1.3 Tujuan

Dari permasalahan diatas maka tujuan dalam penelitian ini yaitu

1. Membuat program yang dapat mengaplikasikan pendekatan GEE untuk respon multinomial pada data ordinal secara online.
2. Mengetahui kelebihan dan kekurangan program GEE repon multinomial untuk data berskala ordinal berbasis web interaktif.
3. Mengetahui model dari suatu respon multinomial berskala ordinal dengan menggunakan pendekatan GEE untuk data *Laparoscopic Cholecystectomy*.
4. Mengetahui inferensi model suatu respon multinomial dengan metode GEE untuk data *Laparoscopic Cholecystectomy*.

1.4 Manfaat

Dari tujuan di atas maka manfaat dari tugas akhir ini adalah

1. Dapat membantu pengguna yang tidak mengerti pemrograman R menjadi lebih mudah untuk melakukan pendekatan GEE untuk respon multinomial dalam hal ini pada data ordinal hanya dengan memasukkan data dan memilih menu dalam program *online* tersebut dan mampu menginterpretasikan hasil analisisnya.
2. Memperkaya dan memperluas pengetahuan tentang analisis data dengan respon multinomial berskala ordinal yang berkorelasi.

BAB 2. TINJAUAN PUSTAKA

2.1 Analisis Regresi

Menurut Tirta (2009) analisis regresi merupakan model statistika yang tidak sekedar mempelajari hubungan asosiasi dua peubah atau lebih melainkan mempelajari hubungan fungsionalnya. Dalam analisis regresi satu peubah atau lebih yang disebut sebagai peubah respon diuji hubungan fungsionalnya dengan beberapa peubah lainnya yang dinamakan peubah penjelas. Rumus umum dari regresi adalah

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

dengan

Y = peubah respon

X = peubah prediktor yang mempengaruhi respon

2.2 Model Linier

Perkembangan pemodelan stokastik, terutama model linier, dapat dikatakan dimulai pada abad ke-19 yang ditandai dengan adanya kajian oleh Galton (1822-1911) tentang hubungan tinggi banda ayah dengan anaknya. Teori regresi menjadi dasar pengembangan teori model linier. Perkembangan model linier tidak bisa dilepaskan dari teori matriks atau aljabar linier. Melalui teori matriks (determinan, invers, perkalian matriks) pembahasan model linier dapat didekati secara umum.

Model linier telah digunakan dalam analisis statistik terutama untuk menganalisa data yang berdistribusi kontinu. Tehnik analisisnya didasarkan pada asumsi pada distribusi normal pada komponen acaknya dan adanya hubungan linier antara nilai tengah dengan peubah penjelas (komponen sistematiknya) (Tirta, 2009).

2.3 Data Longitudinal

Data longitudinal merupakan sekumpulan pengukuran dari unit percobaan yang sama, yang biasanya menunjukkan subjek atau individu dari waktu ke waktu (Noguchi *et al*, 2012). Data longitudinal diamati secara berurutan dari waktu ke waktu yang dikumpulkan dari rancangan percobaan atau studi observasi dimana variabel respon berhubungan dengan urutan sebuah peristiwa yang tercatat dititik waktu tertentu dalam masa studi (Zeger *et al*, 1988).

Karakteristik dan cakupan utama dari penelitian longitudinal meliputi :

- a. Data dikumpulkan untuk setiap variabel pada dua atau lebih periode waktu tertentu.
- b. Subjek atau kasus yang dianalisis sama, atau setidaknya dapat diperbandingkan antara satu periode dengan periode berikutnya.
- c. Analisis melibatkan perbandingan data yang sama dalam satu periode dengan antar metode yang berbeda.

Tujuan dari studi longitudinal salah satunya adalah untuk menyelidiki perubahan suatu respon dapat diatasi dalam waktu tertentu dengan kovariat. Atau dapat dikatakan sebagai gambaran suatu hasil y_{it} untuk subjek ke- i dalam waktu t sebagai fungsi dari suatu kovariat. Pengamatan berulang-ulang untuk suatu objek dalam data longitudinal akan cenderung berkorelasi. Ada tiga metode yang sering digunakan untuk menganalisa data longitudinal, yaitu model marjinal, model efek tetap dan model transisi.

2.4 Distribusi Multinomial

Distribusi multinomial merupakan suatu distribusi yang sering digunakan dalam analisis data kategori. Apabila terdapat j kategori respon maka peluangnya dinotasikan dengan $\{\pi_1, \pi_2, \dots, \pi_j\}$ dengan $\sum_j \pi_j = 1$. Untuk n sampel, peluang multinomial bahwa n_1 termasuk kategori 1, n_2 termasuk kategori 2, \dots , n_j termasuk kategori j dengan $\sum_j \pi_j = n$ adalah

$$P(n_1, n_2, \dots, n_j) = \left(\frac{n!}{n_1! n_2! \dots n_j!} \right) \pi_1^{n_1} \pi_2^{n_2} \dots \pi_j^{n_j}$$

(Agresti, 1990)

2.5 Regresi Logistik Multinomial

Regresi logistik multinomial adalah regresi logistik yang menggunakan peubah respon lebih dari dua kategori berskala nominal dan ordinal. Data berskala nominal merupakan data dengan angka yang diberikan kepada objek mempunyai arti sebagai label dan tidak menunjukkan tingkatan apapun sedangkan data dengan skala ordinal merupakan data dengan angka yang diberikan kepada objek mempunyai arti sebagai sebuah tingkatan (order) untuk respon objek tersebut. Apabila terdapat j kategori maka model logistik yang akan terbentuk sebanyak $j - 1$. Menurut Agresti (1990), model umum regresi logistik multinomial untuk p banyaknya peubah prediktor yang dinyatakan dalam vektor x_1 serta probabilitas kategori respon ke- j sebagai berikut:

$$\pi_j(x_i) = P(y = j | x_i) = \frac{\exp(g_j(x_i))}{\sum_{k=0}^{j-1} \exp(g_k(x_i))}$$

2.6 Pendekatan GEE

Liang & Zeger (1986) menyatakan bahwa GEE merupakan perluasan dari Model Linier Tergeneralisasi (GLM) untuk mengatasi data longitudinal. Pendekatan GEE yang diusulkan oleh Liang & Zeger (1986) merupakan analogi atau generalisasi multivariat dari pendekatan quasi likelihood untuk model linier umum yang diperkenalkan oleh Wedderburn (1974) dalam Lipsitz & Fitzmourice (2009). Pendekatan GEE fokus pada model linier untuk mean dari pengamatan data berkorelasi dalam kluster tanpa spesifikasi distribusi bersama secara penuh. Sehingga GEE merupakan model marjinal yang tidak didasarkan atas bentuk *likelihood* lengkap

dari respon, tetapi hanya berdasarkan hubungan antar nilai tengah (momen pertama) dan ragamnya (momen kedua) serta bentuk matriks korelasinya.

GEE merupakan perluasan dari quasi-*likelihood* yang hanya mengharuskan mempunyai struktur mean dan kovarian yang jelas sehingga menghasilkan estimasi yang konsisten untuk koefisien regresi dan *standart error*. Persamaan regresi merupakan hal yang pokok dalam GEE sedangkan korelasi merupakan perlakuan yang dianggap sebagai *nuisance* atau parameter pengganggu. Maka ada perbedaan pada efisiensi dalam asumsi korelasi yang berbeda. Jika variabel hasil menunjukkan sebuah kondisi seperti distribusi normal, maka GEE dapat dipandang sama seperti atau identik dengan maksimum *likelihood*.

2.6.1 Struktur *Working Correlation* GEE

Menurut Tirta (2009), struktur *working correlation* yang umum digunakan untuk GEE adalah sebagai berikut

a. *Independence*

Untuk struktur ini diasumsikan responnya saling bebas. Sehingga model ini identik dengan model `glm()`.

Struktur ini mengasumsikan

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

Sehingga struktur *working correlation*nya yaitu:

$$\mathbf{R}(\boldsymbol{\alpha}) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

b. *The First-Order Autoregressive (AR(1))*

Model korelasi ini juga sering disebut sebagai model korelasi serial. Struktur ini mengasumsikan bahwa

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

Sehingga struktur *working correlation*-nya menjadi

$$\mathbf{R}(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^p \\ \alpha & \ddots & \vdots & \ddots & \vdots \\ \alpha^2 & \ddots & 1 & \ddots & \alpha^2 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \alpha^p & \dots & \alpha^2 & \dots & 1 \end{pmatrix}$$

c. *Exchangeable*

Struktur ini mengasumsikan adanya korelasi seragam/*uniform* atau biasa dikenal dengan *coumpound symetry*.

Struktur ini mengasumsikan bahwa

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

Sehingga matriks korelasinya adalah

$$\mathbf{R}(\alpha) = \begin{pmatrix} 1 & \alpha & \dots & \alpha \\ \alpha & 1 & \dots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \dots & 1 \end{pmatrix}$$

Pada GEE pemodelan marjinal mean dan perhitungannya bergantung pada unit pada sebuah *cluster* dengan menentukan struktur *working correlation* untuk respon observasi. GEE difokuskan pada mean marjinal dan memperlakukan struktur korelasi sebagai *nuisance*. Sehingga implementasi dari GEE biasanya terbatas pada jumlah struktur *working correlation*-nya, dan struktur tersebut konstan pada seluruh *cluster*/kelompok. Keuntungan dari pendekatan GEE yaitu parameter mean diestimasi secara konsisten terlepas dari apakah struktur korelasi yang ditentukan benar atau tidak selama mean mempunyai spesifikasi yang benar. Walau bagaimanapun efisiensi tersebut bisa hilang ketika struktur *working correlation* salah (Crespi *et al.*,2009).

2.6.2 Estimasi Parameter

Tujuan utama GEE yaitu untuk pengukuran berulang (data longitudinal). Misalkan Y_{it} merupakan variabel respon dan misalkan x_{it} adalah $p \times 1$ vektor kovariat ke- i pada waktu t , untuk $i = 1, \dots, N$ dan $t = 1, \dots, T$.

Kita asumsikan bahwa hasil mean dideskripsikan dengan

$$g(\mu_{it}) = \mathbf{x}_{it}^T \boldsymbol{\beta} \quad (2.1)$$

dimana $\mu_{it} = E(Y_{it} | x_{it}, \boldsymbol{\beta})$ dan $\boldsymbol{\beta}$ adalah $p \times 1$ vektor parameter. Disini difokuskan pada pembuatan fungsi $g(\cdot)$ dalam persamaan yang dinamakan sebagai fungsi link. Parameter regresi $\boldsymbol{\beta}$ termasuk dalam ekspektasi bersyarat ditekankan bahwa mean yang benar juga mengimplikasikan bahwa

$$\mu_{it} = E(Y_{it} | x_{it}, \boldsymbol{\beta}) = E(Y_{it} | x_i, \boldsymbol{\beta})$$

dimana $x_i = (x_{i1}, \dots, x_{it})$. Hal tersebut adalah kondisi marginal mean μ_{it} sesuai hanya dalam nilai covariat dalam waktu t .

Quasi *likelihood* yang diperkenalkan Wedderburn (1974) merupakan pendekatan model linier umum untuk respon univariat (Tambun, 2012). Persamaan quasi likelihood untuk mengestimasi parameter adalah

$$S(\boldsymbol{\beta}) = \sum_{i=1}^N \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} \right)' \mathbf{V}_i^{-1} [\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})] = 0 \quad (2.2)$$

dimana $\mathbf{Y}_i = (y_{i1}, \dots, y_{im})'$ adalah vektor $m \times 1$ yang menyatakan variabel respon, $\mathbf{X}_i = (x_{i1}, \dots, x_{im})'$ adalah matriks $m \times p$ yang merupakan nilai dari variabel eksplanatori untuk subjek ke- i ($i = 1, \dots, N$) dan $j = 1, \dots, m$ dan μ_{ij} menyatakan ekspektasi bersyarat dari respon Y_{ij} , diberikan matriks \mathbf{X}_i yaitu $\mu_{ij} = E(Y_{ij} | \mathbf{X}_i)$ serta diasumsikan hubungan $g(\mu_{ij}) = X_{ij} \boldsymbol{\beta}$ dimana g merupakan fungsi link. Selanjutnya, misalkan $\text{var}(\mathbf{Y}_{ij}) = \boldsymbol{\Phi} \mathbf{V}(\mu_{ij})$ dimana \mathbf{V} fungsi dari μ_{ij} yang diketahui dan $\boldsymbol{\Phi}$ merupakan parameter dispersi. Jika varian \mathbf{Y}_i pada persamaan quasi-likelihood diatas

diganti dengan matriks kovarian \mathbf{Y}_i , yaitu matriks yang berbentuk $\mathbf{V}_i = \phi \mathbf{A}_i^{\frac{1}{2}} \mathbf{R}_i(\alpha) \mathbf{A}_i^{\frac{1}{2}}$, untuk $\mathbf{A}_i = \text{diag}\{V(\mu_{i1}), \dots, V(\mu_{ij})\}$ dan elemen ke- (j, k) dari \mathbf{R}_i adalah $\text{Corr}(Y_{ij}, Y_{ik})$.

Pengembangan quasi – *likelihood* ke bentuk multivariat dinamakan *Generalized Estimating Equation* (GEE). Bentuk persamaan untuk β dalam model marjinal yaitu

$$S(\beta) = \sum_{i=1}^N \left(\frac{\partial \mu_i}{\partial \beta} \right)' \mathbf{V}_i^{-1} [\mathbf{Y}_i - \boldsymbol{\mu}_i(\beta)] = 0 \quad (2.3)$$

dengan

$$\mathbf{V}_i = \phi \mathbf{A}_i^{\frac{1}{2}} \mathbf{R}_i(\alpha) \mathbf{A}_i^{\frac{1}{2}},$$

$$\mathbf{A}_i = \text{diag}\{V(\mu_{i1}), \dots, V(\mu_{ij})\},$$

$$\mathbf{R}_i(\alpha) = \text{Matriks Korelasi}$$

$$\phi = \text{Parameter Dispersi, dan}$$

$$\alpha = \text{Corr}(Y_{ij}, Y_{ik}).$$

$\mathbf{A}_i^{\frac{1}{2}}$ merupakan matriks diagonal dalam bentuk:

$$\mathbf{A}_i^{\frac{1}{2}} = \begin{bmatrix} \sqrt{v(\mu_1)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & v(\mu_j) \end{bmatrix}, \mathbf{A}_i = \begin{bmatrix} v(\mu_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & v(\mu_j) \end{bmatrix}$$

2.6.3 Algoritma *Fischer Scoring*

Langkah langkah untuk menyelesaikan kasus GEE dengan menggunakan algoritma *Fischer Scoring* yaitu

1. Berikan nilai awal untuk $\hat{\beta}^m$
2. Hitung *working correlation* $\mathbf{R}_i(\alpha)$

$$\hat{\phi} = \sum_{i=1}^N \sum_{j=1}^2 \frac{(r_{ij})^2}{2N}$$

untuk

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

3. Hitung matriks kovarian $V_i = \Phi A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}}$

4. Hitung

$$\hat{\beta}^{(m+1)} = \hat{\beta}^m + \left\{ \left[\sum_{i=1}^N (\mathbf{D}_i^{(m)})' \mathbf{V}_i^{-1} (\mathbf{D}_i^{(m)}) \right]^{-1} \sum_{i=1}^N (\mathbf{D}_i^{(m)})' \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mu_i) \right\}$$

5. Iterasi hingga konvergen $\hat{\beta}^{(m+1)} = \hat{\beta}^m$

2.7 GEE untuk Respon Multinomial

Liang & Zeger (1986) memperkenalkan GEE untuk menganalisis data longitudinal. Persamaan yang digunakan adalah quasi-likelihood. GEE ini difokuskan pada hubungan kovariat dengan peluang respon, dimana korelasi antar respon dianggap sebagai gangguan (*nuisance*). Saat respon biner maka model tersebut dinamakan GEE binomial. Sedangkan saat respon multinomial maka model tersebut dinamakan GEE multinomial. Model GEE untuk model multinomial dibagi dua yaitu GEE multinomial berskala ordinal dan GEE multinomial berskala nominal. Kedua skala GEE tersebut dibedakan oleh model marjinal yang digunakan untuk masing-masing skala respon tersebut.

2.7.1 Model Marjinal Untuk Respon Multinomial Berskala Ordinal.

Pilihan dari model marjinal tergantung skala respon. Untuk respon multinomial skala ordinal, model marjinal yang digunakan adalah model *cumulative link model*

$$F^{-1}[P(Y_{it} \leq j | x_i)] = \beta_{0j} + \beta_*^T x_{it}$$

atau *adjacent category logit model*

$$\log\left(\frac{\pi_{itj}}{\pi_{it(j+1)}}\right) = \beta_{0j} + \beta_*^T x_{it}$$

dimana $\{\beta_{0j}: j = 1, \dots, I - 1\}$ adalah intersep kategori spesifik dan β_* adalah parameter vektor kategori ke- j (Touloumis, 2015).

2.7.2 Model Marjinal Untuk Respon Multinomial Berskala Nominal.

Untuk respon multinomial skala nominal, model marjinal yang digunakan adalah *baseline-category logit model*

$$\log\left(\frac{\pi_{itj}}{\pi_{it(j+1)}}\right) = \beta_{0j} + \beta_j^T x_{it}$$

dimana $\{\beta_{0j}: j = 1, \dots, I - 1\}$ adalah intersep kategori spesifik dan β_j adalah parameter vektor kategori ke- j . Koefisien parameter regresi dari kovariat x_{it} adalah juga merupakan kategori spesifik dalam model *baseline-category logit* (Touloumis, 2015).

2.8 Struktur Dasar GEE untuk Respon Multinomial

Touloumis *et al.* (2013) memberikan notasi dan struktur dasar pendekatan GEE untuk respon multinomial dengan memisalkan $Y_{it} \in \{1, 2, \dots, I > 2\}$ sebagai respon multinomial untuk subjek i ($i = 1, \dots, N$) dalam waktu t ($t = 1, \dots, T$), memiliki nilai $j = 1, 2, \dots, I$. Didefinisikan vektor dari respon untuk subjek i

$$Y_i = (Y_{i11}, \dots, Y_{i1(I-1)}, \dots, Y_{i21}, \dots, Y_{i2(I-1)}, \dots, Y_{iT1}, \dots, Y_{iT(I-1)})^T \quad (2.4)$$

dimana Y_{itj} merupakan Y untuk subjek i pada waktu t pada kategori j . Misalkan $x_i = (x_{i1}^T, \dots, x_{iT}^T)^T$ merupakan matriks kovariat untuk subjek ke- i maka $\pi_{itj} = E(Y_{itj} | x_i) = P(Y_{itj} | x_i) = P(Y_{it} = j | x_i)$ untuk $j = 1, \dots, I - 1$ dan $t = 1, \dots, T$ merupakan peluang kategori respon j untuk subjek i dan pada waktu t , dan misalkan $\pi_i = (\pi_{i1}^T, \dots, \pi_{iT}^T)^T$ adalah mean vektor Y_i , dimana $\pi_{it} = (\pi_{it1}, \dots, \pi_{it(I-1)})^T$. Sehingga $Y_{itj} = 1 - \sum_{j=1}^{I-1} Y_{itj}$ dan $\pi_{itj} = 1 - \sum_{j=1}^{I-1} \pi_{itj}$ (Touloumis, 2015).

2.8.1 Estimasi Parameter

Dengan pemetaan $\mathbf{g}:(0, 1)^{I-1} \rightarrow \mathfrak{R}^{I-1}$ yang merupakan sebuah vektor variat $(I - 1)$ dari fungsi link dan berdasarkan *Generalized Linear Model* (GLM) multinomial digunakan untuk memodelkan ekspektasi vektor marjinal untuk subjek i pada waktu t

$$g[E(\mathbf{Y}_{it} | x_i)] = \mathbf{g}(\boldsymbol{\pi}_{it}) = x_{it}\boldsymbol{\beta} \quad (2.5)$$

dimana $\boldsymbol{\beta}$ adalah vektor p -variabel regresi. Link vektor \mathbf{g} yang dipilih harus mencerminkan skala respon. Pilihan untuk link vektor \mathbf{g} yang dapat digunakan untuk kategori respon ordinal adalah *cumulative link* model atau *adjacent category model* (Touloumis, *et al.*, 2013).

Persamaan quasi likelihood untuk mengestimasi parameter adalah

$$U(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^N \mathbf{D}_i \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\pi}_i) = 0$$

dimana $\mathbf{D}_i = \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\beta}}$ dan $\mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$, $\mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$ merupakan $T_i(I - 1) \times T_i(I - 1)$ matriks pembobot yang biasanya berpengaruh sebagai “*working*” matriks kovarian (Touloumis *et al.*, 2013). Bentuk dari matriks kovarian \mathbf{V}_i adalah

$$\mathbf{V}_i = \begin{bmatrix} V_{i11} & V_{i12} & \cdots & V_{i1T_i} \\ V_{i21} & V_{i22} & \cdots & V_{i2T_i} \\ \vdots & \vdots & \cdots & \vdots \\ V_{iT_i1} & V_{iT_i2} & \cdots & V_{iT_iT_i} \end{bmatrix}$$

dimana $\mathbf{V}_{itt'}$ adalah matriks $(I - 1) \times (I - 1)$ dengan elemen kovarian ke (j, j') antara Y_{itj} dan $Y_{itj'}$ dengan x_i untuk semua t, t' dan j, j' . Dalam model marjinal terdapat x_{it} dan $\boldsymbol{\beta}$, dan $Y_{it} \sim \text{MN}(1, \boldsymbol{\pi}_{it})$ yaitu distribusi multinomial dengan ukuran 1 dan vektor peluang $\boldsymbol{\pi}_{it}$. Oleh sebab itu, matriks diagonal $\mathbf{V}_{itt'}$ korespondensi multinomial matriks kovarian

$$\text{cov}(Y_{it}|x_i, \beta) = \text{diag}(\pi_{it}) - \pi_{it}\pi_{it}^T$$

disini, $\text{diag}(\pi_{it})$ merupakan sebuah matriks diagonal $(I - 1) \times (I - 1)$ dengan elemen ke- j elemen diagonal π_{itj} dan 0 ditempat lain. Dengan demikian, kita butuh untuk memodelkan $V_{itt'}$ untuk semua $t < t'$, selama dibatasi $V_{itt'} = V_{it't}^T$ yang diperlukan untuk V_i merupakan matriks simetris (Touloumis, 2011).

2.8.2 Sifat-sifat Asimtotik dari Estimator GEE

Untuk suatu $\hat{\alpha}$, penarikan kesimpulan dari β didasarkan fakta bahwa $\sqrt{N}(\hat{\beta}_G - \beta) \sim N(0, V_G)$ secara asimtotik dimana

$$V_G = \Sigma_G \lim_{N \rightarrow \infty} N \Sigma_0^{-1} \Sigma_1 \Sigma_0^{-1} \quad (2.6)$$

dengan $\Sigma_0 = \sum_i D_i' V_i^{-1} D_i$;

$$\Sigma_1 = D_i' V_i^{-1} \text{cov}(Y_i) V_i^{-1} D_i ; \text{ dan}$$

$$\text{cov}(Y_i) = \text{cov}(Y_i|x_i)$$

$\text{cov}(Y_i) = \text{cov}(Y_i|x_i)$ merupakan matriks kovarian $T_i(I - 1) \times T_i(I - 1)$ untuk subjek i . Elemen dari $\text{cov}(Y_i)$ adalah

$$\text{cov}(Y_{itj}, Y_{it'j'}|x_i) = \begin{cases} \pi_{itj}(1 - \pi_{itj}) & \text{jika } t = t' \text{ dan } j = j' \\ -\pi_{itj}\pi_{it'j'} & \text{jika } t = t' \text{ dan } j \neq j' \\ \pi_{itjt'j'} - \pi_{itj}\pi_{it'j'} & \text{jika lainnya} \end{cases}$$

dimana $\pi_{itjt'j'} = E(Y_{itj}Y_{it'j'} = 1|x_i) = \Pr(Y_{it} = j, Y_{it'} = j'|x_i) > 0$ (Touloumis *et al.*, 2013). Untuk ukuran sampel terbatas, \hat{V}_G diestimasi dengan mengabaikan limit dalam (2.6) dan mengganti β dengan $\hat{\beta}_G$ dan $\text{cov}(Y_i)$ dengan $(Y_i - \hat{\pi}_i)(Y_i - \hat{\pi}_i)^T$, \hat{V}_G/N sering dimasukkan sebagai “sandwich” atau “robust” matriks kovarian $\hat{\beta}_G$ (Touloumis, 2015).

2.9 Rasio Odds Lokal untuk Pendekatan GEE

Touloumis *et al.* (2013) menyarankan untuk mengikuti dua step pendekatan GEE untuk masalah teoritis yang dirasa rumit. Langkah pertama, kita definisikan dan estimasi α dengan memaksimalkan fungsi pseudo likelihood. Misalkan $L = T(T - 1)/2$ pasangan waktu dimana $T = \max\{T_1, \dots, T_N\}$ pada setiap pasangan waktu, dan misalkan G merupakan kelompok variabel dengan level order pasangan L . Untuk masing – masing pasangan waktu (t, t') abaikan kovariat dan pengelompokan respon antar subjek untuk membentuk sebuah tabel kontingensi $I \times I$ untuk total baris merupakan total observasi pada waktu t dan total kolom adalah total observasi pada waktu t' , dan misalkan $\theta_{tjt'j'}$ merupakan rasio odds lokal pada titik potong (j, j') berdasarkan ekspektasi frekuensi $\{f_{tjt'j'} : j, j' = 1, \dots, I\}$. Misalkan A dan B berturut turut merupakan variabel baris dan kolom. Mengasumsikan skema sampel Poisson ke himpunan L dari tabel kontingensi $I \times I$, mencocokkan RC-G(1) tipe model (Becker dan Clogg 1989) dalam (Touloumis, 2015)

$$\log f_{tjt'j'} = \lambda + \lambda_j^A + \lambda_{j'}^B + \lambda_{(t,t')}^G + \lambda_{j(t,t')}^{AG} + \lambda_{j'(t,t')}^{BG} + \phi^{(t,t')} \mu_j^{(t,t')} \mu_{j'}^{(t,t')} \quad (2.7)$$

dimana $\{\mu_j^{(t,t')} : j = 1, \dots, I\}$ adalah nilai parameter untuk kategori respon I pada pasangan waktu (t, t') . Setelah memperkenalkan batasan batasan dalam parameter regresi dalam (2.7), struktur log rasio odds lokal diberikan dengan

$$\log \theta_{tjt'j'} = \phi^{(t,t')} \left(\mu_j^{(t,t')} - \mu_{j+1}^{(t,t')} \right) \left(\mu_{j'}^{(t,t')} - \mu_{j'+1}^{(t,t')} \right) \quad (2.8)$$

pada masing-masing pasangan waktu, sehingga persamaan (2.8) merupakan struktur rasio odds lokal berkaitan dengan nilai parameter I dan parameter intrinsik $\phi^{(t,t')}$ yang mengukur rata rata korelasi dari tabel kontingensi marjinal. Selama nilai parameter tidak perlu diperbarui, struktur rasio odds lokal dapat diaplikasikan untuk keduanya baik respon multinomial nominal maupun ordinal (Touloumis, 2015).

Touloumis *et al.* (2013) mendefinisikan α sebagai vektor parameter yang berisi struktur rasio odds lokal marjinal

$$\alpha = \left(\theta_{1121}, \dots, \theta_{1(I-1)2(I-1)}, \dots, \theta_{(T-1)1T1}, \dots, \theta_{(T-1)(I-1)T(I-1)} \right)^T \quad (2.9)$$

dimana $\theta_{tjt'j'}$ adalah rasio odds lokal pada titik potong (j, j') dalam tabel kontingensi marjinal berkorespondensi ke pembagian waktu (t, t') dalam persamaan 2.8

Untuk skala respon ordinal, Touloumis *et.al.*, (2013) mengajukan menggunakan nilai parameter jarak antar unit yang umum $(\mu_{ij}^{(t,t')} = j)$ dan/atau parameter intrinsik yang umum $(\phi^{(t,t')} = \emptyset)$. Untuk skala respon nominal, nilai parameter harus diestimasi dan batasan tambahan diperlukan. Hal ini dilakukan dengan penambahan batasan kehomogenan nilai parameter $(\mu_{ij}^{(t,t')} = \mu_j)$ dan batasan yang digunakan adalah $\sum_{j=1}^I \mu_j = 0$ dan $\sum_{j=1}^I \mu_j^2 = 1$ (Touloumis, *et al* 2013).

Untuk mengestimasi α metode maximum likelihood digunakan untuk memperlakukan L tabel kontingensi marjinal sebagai independen. Dalam mengestimasi struktur rasio odds lokal marjinal $\hat{\alpha}$ dan spesifikasi model marjinal pada waktu t dan t' , $P(Y_{it} = j, Y_{it'=j'} | x_i)$ diperoleh dengan *Iterative Proportional Fitting Procedure* (IPFP) (Deming and Stephan 1940) dalam Touloumis, *et.al*, (2013). Oleh karena itu, V_i dapat dihitung dan estimasi sehingga persamaan (2.9) dapat diselesaikan.

2.9.1 *Iterative Proportional Fitting Procedure (IPFP)*

Dalam $I \times J$ tabel kontingensi, misalkan π_{i+} dan π_{+j} merupakan baris dan kolom probabilitas marjinal berturut-turut untuk $i = (1, \dots, I)$ dan $j = (1, \dots, J)$. Diasumsikan bahwa probabilitas tersebut pasti dan diketahui dan yang akan diestimasi adalah probabilitas sel π_{ij} subjek dari marjinal tersebut. Untuk menyelesaikan permasalahan tersebut, Deming dan Stephen (1940) dalam Touloumis (2011) memperkenalkan *Iterative Proportional Fitting Procedure* (IPFP). Algoritma IPFP mengikuti aturan dibawah ini

1. Mempunyai nilai awal n_{ij} untuk sel ke- (i, j) menciptakan nilai awal

$$n_{ij}^{(0)} = \frac{n_{ij}}{n}, \forall i, j$$

dimana $n = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$ adalah total ukuran sampel.

2. Pada langkah ke-2m memberikan

$$\pi_{ij}^{(2m-1)} = \pi_j^{2m-2} \frac{\pi_{i+}}{\pi_{i+}^{(2m-2)}}, \forall i, j$$

3. Pada langkah ke-(2m+1)

$$\pi_j^{(2m)} = \pi_{ij}^{(2m-1)} \frac{\pi_{+j}}{\pi_{+j}^{(2m-1)}}, \forall i, j$$

4. Mengiterasi langkah 2 dan 3 hingga konvergen.

2.9.2 Struktur Rasio Odds Lokal

Tabel marginal L merupakan tabel persegi, untuk selanjutnya dapat ditetapkan model korelasi dengan menggunakan $\log \theta_{tjt'j'}$. Misalkan $\log \theta_{tjt'j'}$ adalah rasio odds lokal (j, j') pada pasangan waktu (t, t') , $j, j' = 1, \dots, I - 1$ dan $t, t' = 1, \dots, L$. Dinotasikan dengan

$$\Theta_{tt'} = \begin{pmatrix} \theta_{11tt'} & \theta_{12tt'} & \theta_{13tt'} & \dots & \theta_{1(I-1)tt'} \\ \theta_{21tt'} & \vdots & \vdots & \ddots & \vdots \\ \theta_{31tt'} & \vdots & \theta_{33tt'} & \vdots & \theta_{3(I-1)tt'} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \theta_{(I-1)1tt'} & \dots & \theta_{(I-1)3tt'} & \dots & \theta_{(I-1)(I-1)tt'} \end{pmatrix}$$

merupakan matriks rasio odds lokal pada pasangan waktu t, t' . Himpunan matriks $\{\Theta_{tt'}, tt' = 1, \dots, L\}$ mengindikasikan struktur korelasi. Dalam matriks struktur korelasi dapat dibentuk seperti

$$\begin{pmatrix} 0 & \Theta_1 & \Theta_2 & \dots & \Theta_{T-1} & \Theta_T \\ \Theta_1^T & 0 & \Theta_{T+1} & \dots & \Theta_{2T-2} & \Theta_{2T-1} \\ \Theta_2^T & \Theta_{T+1}^T & 0 & \dots & \Theta_{3T-3} & \Theta_{3T-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \Theta_{T-1}^T & \Theta_{2T-2}^T & \Theta_{3T-3}^T & \dots & 0 & \Theta_L \\ \Theta_T^T & \Theta_{2T-1}^T & \Theta_{3T-2}^T & \dots & \Theta_L^T & 0 \end{pmatrix}$$

Untuk data dengan respon multinomial skala respon ordinal struktur yang digunakan adalah

- a. *Uniform Structure* , $\log \theta_{tjt'j'} = \emptyset$, dapat diperoleh dengan mencocokkan ke homogenan model asosiasi. Struktur ini mengasumsikan keseragaman dari pasangan waktu dan pasangan antar kategori.
- b. Struktur *Category Exchangeability*, $\log \theta_{tjt'j'} = \emptyset_{tt'}$, dapat diperoleh dengan mencocokkan ke heterogenan model asosiasi. Struktur ini mengasumsikan lokal odds rasio yang sama pada setiap pasangan waktu, tetapi mengizinkan pasangan waktu berbeda untuk memiliki asosiasi yang berbeda.
- c. Struktur *Time Exchangeability*, $\log \theta_{tjt'j'} = \emptyset(\mu_j - \mu_{j+1})(\mu_{j'} - \mu_{j'+1})$, dapat diperoleh dengan mengatur $\emptyset_{tt'} = \emptyset$ dan $\mu_{tj}^{tt'} = \mu_{t'j'}^{tt'} = \mu_j$ untuk semua j dan $t < t'$ tetapi memperlakukan nilai kategori sebagai parameter yang tidak diketahui dibandingkan tetap. Struktur ini tidak mengasumsikan dependensi setiap waktu dan ini menyiratkan lokal odds rasio sama pada kategori titik potong (j, j') dan j', j ketika $j \neq j'$.
- d. Struktur RC, $\log \theta_{tjt'j'} = \emptyset_{tt'}(\mu_j^{tt'} - \mu_{j+1}^{tt'})(\mu_{j'}^{tt'} - \mu_{j'+1}^{tt'})$, adalah sebuah perluasan dari struktur sebelumnya yang juga mengizinkan dependensi waktu. Struktur ini memiliki lebih banyak parameter daripada yang lain, disarankan untuk menggunakannya hanya ketika struktur asosiasi yang sebenarnya terlihat sangat rumit (Touloumis, *et al.*, 2013).

Struktur *Time Exchangeability* dan struktur RC dapat diaplikasikan untuk skala respon ordinal dan nominal, tetapi *Uniform Structure* dan *Category Exchangeability* hanya untuk kategori respon ordinal. Struktur *Time Exchangeability* dan struktur RC dapat diaplikasikan untuk skala respon ordinal tetapi dengan tidak meyertakan batasan tambahan batasan kehomogenan nilai parameter $(\mu_{tj}^{(t,t')} = \mu_j)$ dan batasan yang digunakan adalah $\sum_{j=1}^I \mu_j = 0$ dan $\sum_{j=1}^I \mu_j^2 = 1$ (Touloumis *et al.*, 2013)

2.9.3 Paket untuk Implementasi Pendekatan GEE Rasio Odds Lokal

Sebuah paket dalam software R yaitu `multgee` telah dikembangkan untuk menerapkan pendekatan GEE yang diusulkan. Tipe `ordLORgee` dirancang untuk kategori respon ordinal menggunakan struktur *uniform*, *category exchangeability*, *time exchangeability* dan *RC* sebagai pilihan struktur rasio odds lokal. Rincian lebih lanjut tentang paket `multgee` dapat ditemukan di Touloumis (2014). Dalam pengaplikasiannya paket `multgee` pada Program R offline akan dikembangkan dalam basis web interaktif dengan menggunakan R-Shiny dan MathJax dalam bentuk HTML.

2.10 R-Shiny

Menurut Tirta (2014) R-Shiny mempunyai komponen yang dibagi menjadi dua kelompok yaitu seperti berikut ini.

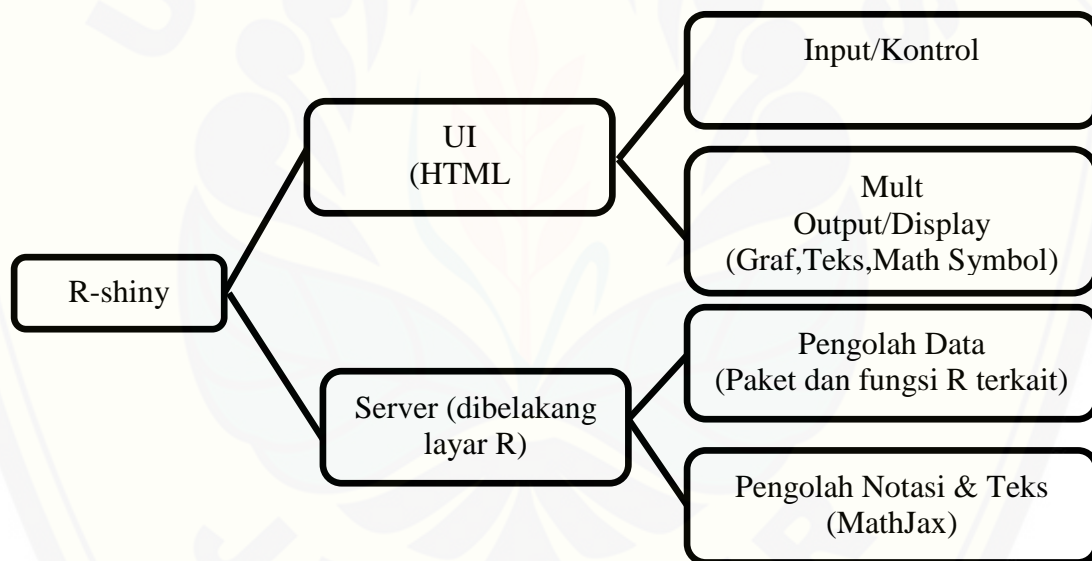
1. *User Interface*. Manfaat dari *user interface* adalah untuk
 - a. Panel Kontrol adalah panel untuk mengontrol input berupa data, variabel, model, tergantung kompleksitas modul. Tampilan kontrol dapat berupa slider, radiobutton, check-box dan lain-lain.
 - b. Pemasukan permintaan nilai input (data dengan berbagai jenis variabel yang diperlukan, pemilihan model, jenis, dan kriteria uji statistika).
 - c. Penyajian output terkait hasil analisis/uji. Hasil output dapat berupa
 - i. Grafik (histogram, diagram pencar, dan lain-lain)
 - ii. Bentuk angka/teks bisa berbentuk asli (verbatim) maupun dalam bentuk tabel
 - iii. Teks khusus dengan notasi matematika dengan format LaTeX.

Untuk mengakomodasi berbagai jenis luaran tadi, shiny juga menyediakan berbagai format output seperti `plotOutput`, `textOutput`, `tableOutput`, `verbatimTextOutput`, dan lain-lainnya. Bagian user interface ini dapat

disajikan pada file khusus `ui.r`, dapat juga disajikan penuh melalui file HTML, misalnya `index.html`

2. *Server*. Bagian ini 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 dan 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



Gambar 2.1. Struktur Komponen R-Shiny

BAB 3. METODE PENELITIAN

Dalam pemodelan GEE untuk multinomial respon ordinal struktur rasio odds lokal yang umum digunakan yaitu struktur *uniform* dan *category exchangeability* dengan fungsi link (peluang marjinal) yang digunakan adalah model marjinal kumulatif logit. Dalam langkah-langkah penelitian ini dilakukan pembuatan program web interaktif berbasis tutorial untuk penerapan paket `multgee` untuk skala ordinal (`OrdLORgee()`) dengan menggunakan MathJax dan R-Shiny untuk analisis data secara *online*. Selanjutnya dilakukan ilustrasi data menggunakan program web interaktif tersebut menggunakan data sekunder yaitu data *Laparoscopic Cholecystectomy* dari penelitian Jorgensen, *et.al.*, (1995) dalam Lumley (1996).

3.1 Langkah-Langkah Pembuatan Web Interaktif

Langkah-langkah pada pembuatan program web interaktif untuk GEE respon multinomial skala ordinal adalah sebagai berikut:

1. Menentukan *input* dan *output* program. *Input* dalam hal ini akan berisi menu untuk data yang akan digunakan untuk analisis GEE untuk respon multinomial yaitu berupa data dengan ciri berupa data kategorik, berskala ordinal dan berupa data longitudinal, selain itu input untuk program ini adalah pemilihan variabel (x dan y), id, link dan struktur lokal odds rasio. *Output* program akan menghasilkan *summary* hasil estimasi, *root mean square error* (RMSE), *p-value* serta tampilan opsi detail hasil keseluruhan.
2. Mentransfer dokumen teori kedalam format LaTeX sehingga dapat ditampilkan secara *online*.
3. Membuat program R-shiny yang berupa *user interface* dan *server*. *User interface* berupa file HTML yang digunakan untuk memasukkan nilai *input* dan penyajian

output. *Server* merupakan otak dari program sehingga pada bagian ini dilakukan pengaplikasian program GEE respon multinomial untuk data berskala ordinal dalam hal ini digunakan paket `multgee` pada R kedalam program R-Shiny.

Paket `multgee` merupakan salah satu paket yang secara khusus menangani masalah data berkorelasi nominal atau ordinal respon multinomial (Touloumis, 2014 dan Touloumis, 2015). Paket ini terdiri dari dua fungsi yang mencocokkan model GEE untuk korelasi respon multinomial `ordLORgee()` untuk skala respon ordinal dan `nomLORgee()` untuk skala respon nominal. Dalam program ini data yang digunakan adalah data dengan skala respon ordinal dengan mengasumsikan *cumulatif link models* atau *adjacent categories logit models* untuk peluang marginal. Struktur pada paket `multgee` yaitu

```
ordLORgee(formula, data, id =id, repeated = NULL, link =  
          "logit", LORstr = "category.exch", LOrem = "3way)
```

Keterangan:

`formula` : sebuah ekspresi rumus seperti regresi lainnya untuk respon multinomial dalam bentuk `respon~predictors`.

`data` : data frame yang mengandung variabel yang disediakan dalam `formula`, `id`, dan `repeated`

`id` :vektor yang mengidentifikasi cluster. Panjang `id` harus sama dengan jumlah observasi.

`repeated` : vektor opsional yang mengidentifikasi order dari observasi antara satu dengan yang lain.

`link` : karakter string yang menetapkan fungsi link. Pilihannya termasuk "logit", "probit", "cauchit", "cloglog" atau "acl".

LORstr : karakter string yang mengindikasikan struktur marjinal lokal odds rasio. Pilihannya meliputi "uniform", "category.exch", "time.exch", atau "RC".

LORem : karakter string yang mengindikasikan jika struktur marjinal lokal odds rasio adalah ditaksir secara bersamaan ("3way") atau dengan mandiri disetiap taraf pasangan dari pengulangan repeated ("2way").

(Touloumis, 2014)

4. Menggabungkan dokumen teori dan program *online* dalam *server* yang sudah jadi dan menjadi satu bentuk tutorial *online* GEE untuk respon multinomial berskala ordinal.
5. Uji coba dengan beberapa data untuk mengetahui program R-shiny berjalan dengan baik.

3.2 Langkah – Langkah Ilustrasi Data

Data yang digunakan untuk ilustrasi data dalam uji coba data yaitu data *Laparoscopic Cholecystectomy* dari penelitian Jorgensen, *et.al.*, (1995). Data ini mempunyai respon berdistribusi multinomial berskala ordinal yaitu tingkat *Laparoscopic Cholecystectomy* sebanyak lima tingkatan dengan variabel prediktor berupa umur, jenis kelamin (Laki-laki (L) dan Perempuan (P)) dan terapi (Aktif (A) dan Tidak Aktif (TA)). Data tersebut diukur berulang sebanyak enam kali (Waktu = A, B, C, D, E, F). Data *Laparoscopic Cholecystectomy* ini dalam Lumley (1996) dianalisis menggunakan GEE multinomial ordinal dengan parameter rasio odds global untuk matriks kovariannya. Langkah-langkah dalam ilustrasi data *Laparoscopic Cholecystectomy* dengan program web interaktif untuk GEE respon multinomial skala ordinal adalah sebagai berikut:

1. Menginputkan data longitudinal berskala ordinal ke dalam program online. Data data yang diinputkan adalah data longitudinal berskala ordinal. Variabel-variabel yang digunakan dalam uji coba program ini adalah

a. Variabel Respon

Untuk data respon multinomial data longitudinal berskala ordinal maka variabel respon (Y) lebih dari satu untuk setiap responden (N).

b. Variabel Prediktor

Variabel prediktor merupakan variabel yang mempengaruhi respon yang bersifat tetap untuk jangka waktu pengukuran data. Variabel prediktor dinotasikan dengan (X).

2. Eksplorasi data dengan memilih variabel x , variabel y , id, link dan struktur rasio odds lokal dalam menu program web interaktif.
3. Mengestimasi parameter dengan memilih menu program `multgee` online untuk skala ordinal. Dalam langkah ini, program akan menampilkan hasil estimasi, *standart error*, *p-value* dan nilai *root mean square error* (RMSE)-nya.
4. Melakukan uji signifikansi dengan menggunakan nilai *p-value*. Taraf signifikansi yang digunakan sebesar $\alpha = 0,05$. Variabel dikatakan berpengaruh terhadap variabel respon apabila variabel tersebut memiliki $p - value \leq 0,05$.
5. Uji kecocokan model dengan membandingkan nilai RMSE dari masing-masing struktur rasio odds lokal yang berbeda untuk memilih model terbaik dari dua struktur rasio odds lokal yang berbeda.
6. Interpretasi Hasil
Dalam hasil ini kita interpretasikan hasil estimasi yang berupa hasil uji hipotesis dan struktur rasio odds lokal. Pengaruh prediktor akan signifikan untuk taraf signifikansi 5%.

Langkah-langkah pada pembuatan program web interaktif untuk GEE respon multinomial skala ordinal dan langkah-langkah dalam ilustrasi data *Laparoscopic Cholecystectomy* dengan program web interaktif untuk GEE respon multinomial skala ordinal adalah sebagai berikut langkah – langkah yang disajikan dalam gambar 3.1.