ANALISIS PEMODELAN VARIABEL BERKATEGORI DENGAN RESPON TIDAK BEBAS

SKRIPSI

Diajukan Untuk Memenuhi Persyaratan Program Sarjana Sains
Jurusan Matematika Fakultas Matematika Dan Ilmu Pengetahuan Alam
Universitas Jember

Oleh :

M. Yasin
NIM. 981810101018

DEPARTEMEN PENDIDIKAN NASIONAL
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
Juni 2002

Analisis Pemodelan Variabel Berkategori Dengan Respon Tidak Bebas.

SKRIPSI

Diajukan Untuk Memenuhi Persyaratan Penyelesaian Program Sarjana Sains Jurusan Matematika Fakultas Matematika Dan Ilmu Pengetahuan Alam Unversitas Jember

Oleh :

Nama	: M. Yasin
NIM	: 981810101018
Jurusan	: Matematika
Angkatan	: 1998
Asal Daerah	: Jember
Tempat, Tanggal Lahir	: Jember, 21 September 1980

Disetujui oleh :

Dosen Pembimbing Utama (DPU)

Drs. I Made Tirta, Dip.Sc, M.Sc, PhD
NIP. 131 474 500

Dosen Pembimbing Anggota. (DPA)

Yuliani S. Dewi, S.Si, M.Si.
NIP. 131 258 183

PENGESAHAN

Skripsi ini telah dipertahankan didepan Tim Penguji dan diterima oleh Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember pada :

Hari : Selasa

Tanggal : 18 Juni 2002

Tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Tim Penguji

Ketua	Sekretaris
Drs. I Made Tirta, Dip.Sc, M.Sc, PdD	Yuliani Sri Dewi, S.Si, M.Si.
NIP. 131 474 500	NIP. 132 258 183

Anggota :

Rita Ratih T., S.Si, M.Si	Alfian Futuhul Hadi, S.Si
NIP. 132 243 343	NIP. 132 278 621

Mengesahkan

Dekan FMIPA UNEJ

Ir. Sumadi, MS

NIP. 130 368 784

MOTTO

- *Tiada Jalan Keluar Yang Terbaik Kecuali Berserah Pada Tuhan.*
- *Tetapkanlah pikiranmu pada apa yang sudah menjadi cita – citamu, dan sanubarimu tidak akan salah menunjukkan jalan padamu untuk mencapainya.*
- *Makin banyak seseorang belajar, maka makin insaflah dia betapa sedikitnya yang dia ketahui.*

MOTTO

- *Tiada Jalan Keluar Yang Terbaik Kecuali Berserah Pada Tuhan.*
- *Tetapkanlah pikiranmu pada apa yang sudah menjadi cita – citamu, dan sanubarimu tidak akan salah menunjukkan jalan padamu untuk mencapainya.*
- *Makin banyak seseorang belajar, maka makin insaflah dia betapa sedikitnya yang dia ketahui.*

KUPERSEMBAHKAN SKRIPSI INI KEPADA

- *Ayahandaku " Harun Al Rosjid", Ibunda " Muzayanah ", Abi "H. Nasihuddin", Umi " Sundari" yang sangat ananda cintai dan ananda hormati, yang tiada sunyi menberikan doa, sebagai tanda bakti dan terima kasih atas segala ketulusan, kesabaran dan pengorbanannya,*
- *Kakak – kakakku yang tercinta, Om dan Tanteku yang aku hormati, Nenekku yang aku sayangi dan seluruh saudara – saudaraku, terima kasih atas segala dukungan yang diberikan,*
- *Teman – teman seperjuangan, dan*
- *Almamater dan Tanah Airku Tercinta.*

DEKLARASI

Skripsi ini berisi hasil kerja/ penelitian mulai bulan Februari 2002 sampai dengan bulan Mei 2002. Bersama ini saya nyatakan bahwa isi skripsi ini adalah hasil pekerjaan saya sendiri kecuali jika disebutkan sumbernya dan skripsi ini belum pernah diajukan pada institusi lain.

Jember, Mei 2002

Penulis.

(M. Yasin)

ABSTRAKSI

M. Yasin, Juni, 2002, " **Analisis Pemodelan Variabel Berkategori Dengan Respon Tidak Bebas**". Skripsi, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Pembimbing I : Drs. I Made Tirta, Dip.Sc, M.Sc, PhD.
Pembimbing II : Yuliani S. Dewi, S.Si, M.Si.

Kata Kunci : GEE, GLM, data berkorelasi, binomial, logit.

Penelitian ini bertujuan untuk mempelajari metode untuk menganalisis model dari suatu data yang tidak normal dan tidak bebas, khususnya data berdistribusi binomial melalui pendekatan regresi. Seperti contoh data yang diangkat dalam penelitian ini, merupakan data berkategori dengan respon yang saling berkorelasi. Suatu prosedure pembentukan model untuk data yang tidak normal dan tidak bebas adalah GLMM (*Generalized Linier Mixed Model*) yang merupakan pengembangan dari GLM (*Generalized Linier Model*) untuk data tidak bebas. Bentuk GLMM disusun atas fungsi densitas keluarga eksponensial, fungsi link, dan ketidak bebasan pada respon (korelasi). Pada bahasan ini digunakan pendekatan GEE (*Generalized Estimating Equation*) yang menjeneralisasi GLM dengan mengakomodasi model korelasi tanpa menggunakan bentuk *log-likelihood*. Bentuk umum GEE adalah

$$S(\beta) = \sum_{i=1}^{m} \left(\frac{\partial \eta_i}{\partial \beta} \right)^T Var(Y_i)^{-1} (Y_i - \mu_i) = 0 .$$

Dalam mengestimasi parameter digunakan alat bantu program komputer, yaitu salah satu program satistika untuk komputasi dan grafik yaitu paket R. Dalam Paket R terdapat pilihan distribusi, fungsi link, dan struktur korelasi untuk pembentukan model, dimana dalam kasus penelitian ini dicoba dengan menggunakan pilihan distribusi binomial, link *logit*, dan struktur korelasi *exchangeable* dan *independence*. Hasil perhitungan GEE dengan bantuan paket R diketahui bahwa struktur korelasi yang lebih baik dalam pembentukan model adalah struktur korelasi *exchangeable*. Terdapat hasil yang beda pada GEE antara model dengan mempertimbangkan korelasi (korelasi *exchangeable*) dan model yang mengabaikan korelasi (korelasi *independence*), hal ini ditunjukkan dengan membandingkan nilai probabilitas dan panjang interval kepercayaan koefisien model dari kedua struktur korelasi ini.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT, yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Sholawat serta salam semoga tetap tercurahkan kepada Nabi Muhammad SAW, dan orang – orang yang selalu berada dijalannya.

Selanjutnya penulis sampaikan terima kasih dan penghargaan yang setulusnya atas bantuan yang tidak ternilai kepada :

1. Ayahanda dan Ibunda yang saya hormati dan saya cintai, yang telah sabar mendidik, membimbing, memberi doa dan kasih sayang tanpa batas waktu.
2. Bapak Ir. Sumadi, MS. Selaku Dekan Fakultas MIPA Universitas Jember.
3. Bapak Drs. Kusno, DEA, PhD. Selaku Ketua Jurusan Matematika FMIPA Universitas Jember.
4. Bapak Drs. I Made Tirta, Dip.Sc, M Sc, Ph.D. Selaku Pembimbing Utama yang dengan penuh kesabaran dan ketulusan membimbing penulis dalam penyusunan skripsi ini.
5. Ibu Yuliani S. Dewi, S.Si, M.Si. Selaku Pembimbing Pendamping, yang telah banyak memberikan saran dan petunjuk bagi penulisan skripsi ini.
6. Ibu Rita Ratih T, S.Si, M.Si dan Bapak Alfian Fatuhul H, S.Si yang telah banyak memberikan bantuan penyediaan sarana Laboratorium Statistika.

7. Ayahanda dan Ibunda tercinta, Om dan Tanteku, Nenekku, Adikku " Arie" yang selalu menemani penulis, dan saudara – saudaraku yang selalu memberikan dorongan moril pada penulis.
8. Sahabat – sahabat terbaikku: Eni Kusuma, Indri, Tutut, Bahrul, Indah, Nonik, Agung, Toriq, Lena, Ninip, Adi, dan Dyah yang secara tulus memberikan bantuan, dorongan dan dukungan selama ini.
9. Teman – temanku Mahasiswa Math '98, teman – teman seperjuangan dalam penyusunan skripsi, semoga sukses senantiasa bersama kita.
10. Rekan – rekan dan semua pihak yang telah membantu dan memotivasi dalam penyusunan skripsi ini.

Semoga segala bantuan dan kebaikan yang telah diberikan kepada penulis akan mendapatkan imbalan yang setimpal dari Allah SWT. Amin ...

Akhir kata, penulis berharap semoga apa yang penulis tuangkan dalam skripsi yang sederhana ini bermanfaat bagi pembaca.

Jember, Juni 2002 Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1. Latar Belakang

Agresti (1990) menyatakan bahwa variabel berkategori adalah suatu variabel dengan skala pengukuran yang terdiri dari kumpulan kelompok – kelompok/ kategori – kategori. Skala berkategori lebih dulu umum dalam ilmu lingkungan, ilmu kesehatan umum, ilmu ekologi, pendidikan dan perdagangan dari pada ilmu – ilmu tingkat tinggi; seperti ilmu mesin, ilmu industri (kontrol kualitas produksi). Variabel berkategori memiliki beberapa tipe variabel diantaranya adalah variabel respon atau sering disebut sebagai variabel terikat dan *explanatory variable* yang sering disebut sebagai variabel bebas. Untuk selanjutnya, dalam bahasan ini *explanatory variable* tetap kita sebut sebagai variabel eksplanatori. Dalam struktur suatu model variabel respon melambangkan bagaimana distribusi dari suatu respon (jawaban) dan dilambangkan dengan huruf alphabet Y. Sedangkan suatu variabel eksplanatori didasarkan pada urutan level dan biasanya dinotasikan dengan huruf alphabet X.

Beberapa variabel respon berkategori yang hanya memiliki 2 kelompok/ golongan dalam pengamatan pada setiap subyeknya harus diklasifikasikan menjadi sukses atau gagal. Kondisi seperti ini berdistribusi

Binomial (Bernoulli), andaikata ada faktor eksplanatori tunggal yang mempunyai r kelompok dalam baris ke- i pada tabel r x 2, maka kedua peluang respun adalah $\pi_{1/i}$ dan $\pi_{2/i}$ dengan $\pi_{1/i} + \pi_{2/i} = 1$. Salah satu model untuk kondisi berdistribusi Binomial adalah model Logit (Agresti,1990). Model untuk respon yang tidak berdistribusi normal dan independen digunakan suatu prosedure yaitu GLM (*Generalized Linier Model*) (Tirta,2000 dan Tirta,2001)

Dalam kehidupan real, tidak kesemua masalah yang melibatkan variabel – variabel berkategori adalah independen. Yang dimaksud disini adalah hubungan antara kategori dalam variabel respon. Salah satu contoh kasus dengan variabel – variabel berkategori tersebut tidak bebas adalah artikel yang ditulis oleh Alan Agresti dan I-Ming Liu (1999). Dengan, memberikan data sample dari 262 peternak dari peternakan – peternakan Kansas dengan memberikan suatu pertanyaan " Apakah sumber informasi utama anda dalam bidang peternakan (kehewanan) ?", dengan pilihan jawaban yang sudah ditentukan atau dalam bentuk biner. Dengan jawaban dari pertanyaan itu dibagi dalam beberapa pilihan, yaitu :

A. Konsultan profesional.

B. Dokter hewan.

C. Jasa penerangan

D. Majalah.

E. Perusahaan makanan.

Dalam hal ini, para peternak dapat memilih beberapa pilihan jawaban yang sesuai bagi mereka dan disesuaikan dengan tingkat pendidikan mereka. Hal ini secara jelas ditunjukkan dalam Tabel. 1 dibawah ini.

Tabel 1. Sumber Informasi peternak dan kelompok level pendidikan dari 262 peternak

Level Pendidikan	Sumber Informasi					Total Respon	Total Subyek
	A	B	C	D	E		
High school	19	38	29	47	40	173	88
Vocational	2	6	8	8	4	28	16
2-year college	1	13	10	17	14	55	31
4-year college	19	29	40	53	29	170	113
Others	3	4	8	6	6	27	14
Total	44	90	95	131	93	453	262

Sumber : Agresti dan Liu (1999).

Table 1 diatas merupakan tabel kontingensi 5 x 5 dengan 453 jawaban positif (benar) dari 262 peternak. Dalam kasus ini, statistika biasa tidak tepat digunakan dalam pengambilan kesimpulan, sebab ada 453 masukan yang tidak bebas. Setiap level pendidikan terdapat 25 rangkaian jawaban yang mungkin (ya/tidak) menurut hasil pemilihan kelompok. Selanjutnya, jawaban lebih lengkap digambarkan dengan penggolongan campuran dari 5 pasang komponen. Variabel A diindikasikan sebagai jawaban Ya untuk sumber A, variabel B diindikasikan sebagai jawaban Ya untuk sumber B, dan demikian juga lainnya.

Dalam kasus ini, distribusinya adalah berdistribusi Binomial atau lebih umum adalah berdistribusi Multinomial. Selanjutnya apabila kita akan mencari model hubungan antara kategori – kategori dalam variabel respon dengan level – level pada variabel eksplanatori adalah tidak tepat bila kita menggunakan *Generalized Linier Model*, sebab hubungan kategori dalam variabel respon adalah tidak bebas. Selanjutnya, dalam skripsi ini penulis akan menyelesaikan masalah dengan menggunakan GEE (*Generalized Estimating Equation*).

1.2. Permasalahan .

Berdasarkan kerangka pemikiran diatas, maka masalah dalam studi ini dapat dirumuskan sebagai berikut :

1. Bagaimana model dari suatu data berkategori dengan variabel respon yang tidak bebas.
2. Bagaimana estimasi dan inferensi dengan menggunakan distribusi statistika.
3. Bagaimana efek model jika kebergantungan antara respon diabaikan .
4. Bagaimana hasil analisis dengan menggunakan paket statistika yang ada.

1.3. Tujuan Penelitian.

Studi kasus ini bertujuan untuk mengetahui :

1. Model dari suatu kasus yang melibatkan variabel berkategori dengan respon yang tidak bebas.
2. Estimasi dan inferensi dengan menggunakan distribusi statistika.
3. Efek model jika kebergantungn antara respon diabaikan.
4. Bentuk analisis komputer menggunakan sofwere komputasi statistika yang ada.

1.4. Manfaat Penelitian

Studi kasus ini dapat bermanfaat dalam berbagai keperluan diantaranya :

1. **Manfaat Teoritik.**

   Melalui studi kasus ini dapat digunakan sebagai suatu analisis pengembangan model – model dalam ilmu statistika, dengan fenomena – fenomena yang berkaitan dengan variabel berkategori. Asumsi kebebasan dan ketidak bebasan yang di ajukan dalam studi ini didasarkan pada asumsi - asumsi yang sangat mendasar. Kebenaran analisis model ini didasarkan pada asumsi yang diambil, apabila terbukti maka hipotesis ini memperkuat kebenaran pengembangan model. Namun apabila tidak terbukti maka diharapkan adanya perbaikan bagi analisis pengembangan model ini.

2. Manfaat Praktis.

Studi ini dapat digunakan untuk meningkatkan pemahaman penulis dalam bidang statistika serta diharapkan memacu semangat bagi mahasiswa dalam meningkatkan pola pemahaman ilmu statistika baik dalam statistika terapan maupun analisis statistika.

1.5. Penjelasan Istilah.

Dalam skripsi ini digunakan beberapa notasi dan istilah matematika, yaitu :

1. **Variabel**. Variabel dinotasikan dengan huruf alphabet dengan format miring. Variabel bebas (eksplanatori) dinotasikan dengan x atau X dan variabel terikat (respon) dengan notasi y atau Y. Seluruh notasi atau simbol dengan huruf alpahabet miring menyatakan sebuah variabel.
2. **Konstanta**. Konstanta di notasikan dengan bilangan atau huruf alphabet.
3. **Matrik dan Vektor**. Matrik dinotasikan dengan huruf tebal, contohnya **X**, **Y**, **Z**. Sebuah vektor dinyatakan dengan notasi yang sama dengan sebuah variabel namun terdapat identifikasi bagian, contohnya y_i, yang artinya sebuah komponen variabel y ke–i, $y_i = \{y_1, y_2, \ldots, y_n\}$, $i = 1, 2, \ldots, n$. ; n merupakan jumlah banyaknya variabel.
4. **Fungsi**. Sebuah fungsi f dengan variabel bebas x, dinotasikan dengan $f(x)$, contohnya $g(\mu)$, $l(\theta, y)$ dan sebagainya.
5. **Posisi iterasi** $\beta^{(m)}$ menyatakan nilai β pada iterasi ke- m

BAB II

TINJAUAN PUSTAKA

2.1 Model Linier Klasik dan Perkembangannya

Teori ini dijelaskan dalam diktat perkuliahan Pemodelan Matematika (Tirta,2000), dimana ringkasannya adalah sebagai berikut.

Pemodelan matematika terdiri dari dua macam yaitu pemodelan deterministik dan pemodelan stokastik. Pada pemodelan deterministik, variabel yang diamati dianggap tetap (*fixed*) dan tidak memiliki sebaran sehingga hubungan yang diperolah merupakan hubungan matematika yang bersifat fungsional murni, misalnya $y = f(x)$. Sedangkan pemodelan stokastik (statistik) merupakan suatu pemodelan yang menganggap harga variabel berubah – ubah dengan sebaran tertentu. Pemodelan stokastik memiliki bentuk umum :

$$y = f(x, \beta) + \varepsilon \tag{2.1}$$

Dalam hal ini ε merupakan kesalahan (*error*) yang diasumsikan merupakan variabel acak yang berasal dari distribusi tertentu, misalnya normal. x adalah variabel yang tetap dan β adalah parameter yang menentukan koefisisen dari variabel – variabel tetap. Komponen $f(x,\beta)$ disebut komponen tetap (*fixed component*) dan ε disebut sebagai komponen acak (*random component*) atau dalam hal ini secara khusus disebut komponen kesalahan (*error component*). Dari segi hubungan fungsi, yang paling sederhana adalah hubungan linier, sehingga model yang paling sederhana adalah **model linier**. Sedangkan dari segi komponen acaknya, yang paling sederhana adalah asumsi bahwa $\varepsilon \sim N(0,\sigma^2)$ dan respon saling independen satu sama lainnya. Selanjutnya asumsi ini melahirkan **model linier normal sederhana** atau **model linier klasik**. Bentuk dan asumsi model linier klasik adalah sebagai berikut :

- Model : $y_i = \sum_{j=0}^{p} \beta_j x_{ij} + \varepsilon_i$ atau dalam bentuk matrik $\mathbf{y} = \mathbf{x}\beta + \varepsilon$;
- Asumsi $\varepsilon_i \sim N(0,\sigma^2)$ dan ε_i independent dengan ε_{i^*} untuk setiap $i \neq i^*$

Dari asumsi di atas diperoleh bahwa secara keseluruhan ε berdistribusi multivariate normal (MVN) dengan koefisien varian-kovarian konstan, dinotasikan dengan

$$\varepsilon \sim MVN\left(\begin{bmatrix} 0 \\ \dots \\ 0 \end{bmatrix}\begin{bmatrix} \sigma^2 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \sigma^2 \end{bmatrix}\right) \text{ atau } \varepsilon \sim \mathbf{MVN}(\mathbf{0},\sigma^2 I). \qquad (2.2)$$

Model linier campuran merupakan perkembangan dari model linier klasik untuk suatu data dengan respon yang tidak independen. Dalam model ini hubungan antara respon satu dengan respon yang lain dianggap berasal dari pengaruh suatu variabel yang tidak kentara atau laten, sehingga komponen acak *f(x)* diuraikan lagi menjadi komponen efek acak dan komponen error (ε). Dengan demikian model ini memiliki dua komponen acak yaitu komponen error (ε) dan komponen efek acak yang biasa kita notasikan dengan u.

Kondisi lapangan yang tidak dapat ditangani dengan model linier klasik adalah adanya kenyataan bahwa tidak semua distribusi variabel respon adalah normal. Memang kondisi ini mungkin bisa ditanggulangi dengan mengadakan transformasi dari respon, namun ada kemungkinan syarat ketidak tergantungan menjadi tidak terpenuhi. Untuk mengatasi kondisi ini dimana respon yang ada tidak berdistribusi normal tapi masih saling bebas, selanjutnya model linier dikembangkan menjadi model linier tergeneralisasi (*Generalized Linier Model*, (GLM)), yang akan diuraikan pada sub bab 2.3.

Selanjutnya muncul suatu model akibat adanya tuntutan bahwa dilapangan sangat mungkin terjadi adanya respon yang tidak saja tidak berdistribusi normal, namun juga respon tidak saling independen. Model ini merupakan gabungan dari model linier campuran dan model linier tergeneralisasi. Model linier ini biasa disebut sebgai model linier campuran tergeneralisasi (*Generalized Linear Mixed Model*, (GLMM)).

2.2. Model – Model untuk Variabel Respon Biner(Berpasangan).

Teori ini dijelaskan dalam buku *Categorical Data Analysis* (Agresti, 1990), yang ringkasannya disampaikan pada bagian berikut.

2.2.1. Model Logit untuk Tabel Kontingensi.

Beberapa variabel respon berkategori yang memiliki 2 kelompok/ golongan. Pengamatan untuk setiap subyek harus diklasifikasikan menjadi sukses/ gagal. Penggambaran hasil yang mungkin adalah 1 dan 0. Distribusi Bernoulli untuk variabel acak berpasangan menggolongkan peluang $P(Y=1) = \pi$ dan $P(Y=0) = 1 - \pi$, yang mana $\pi = E(Y)$. Jika Y_i memiliki distribusi Bernoulli/ Binomial dengan parameter π_i, fungsi kepadatan peluang adalah

$$\begin{aligned} f(y_i;\pi_i) &= \pi_i^{y_i}(1-\pi_i)^{1-y_i} = (1-\pi_i)[\pi_i/(1-\pi_i)]^{y_i} \\ &= (1-\pi_i)\exp\left[y_i \log\left(\frac{\pi_i}{1-\pi_i}\right)\right] \end{aligned} \tag{2.3}$$

Distribusi di atas termasuk dalam keluarga eksponensial natural. Parameter natural $Q(\pi) = log[\pi / (1 - \pi)]$ merupakan log kemungkinan dari respon 1, hal ini disebut logit pada π. GLM yang menggunakan link logit disebut model logit

Pada tabel I x 2, ada faktor eksplanatori tunggal yang mempunyai I kategori/ kelompok. Kedua peluamg respon adalah $\pi_{1/i}$ dan $\pi_{2/i}$ dengan $\pi_{1/i} + \pi_{2/i} = 1$. Dalam model logit

$$\log\left(\frac{\pi_{1/i}}{\pi_{2/i}}\right) = \alpha + \beta_i \tag{2.4}$$

$\{\beta_i\}$ menggambarkan pengaruh respon dan merupakan deviasi dari rata – rata pada beris ke- i, β_i terbesar adalah logit terbesar pada baris ke- i dan nilai terbesar $\pi_{1/i}$, α adalah mean dari model. Untuk beberapa bagian $\{\pi_{1/i}>0\}$ada $\{\beta_i\}$seperti model di atas. Model itu memiliki beberapa parameter pengamatan Binomial dan hal ini dikatakan sebagai model penuh (*saturated*).

$$\log\left(\frac{\pi_{1/i}}{\pi_{2/i}}\right) = \alpha \tag{2.5}$$

Model di atas merupakan kasus khusus dimana $\beta_1 = \beta_2 = \ldots = \beta_I$. Selama kondisi ini ekuivalen dengan $\pi_{1/1} = \ldots = \pi_{1/I}$, model di atas adalah model statistika bebas pada respon dan faktor. Model (2.6) adalah kasus khusus dari (2.4) dengan $\beta_1 = \beta_2 \ldots = \beta_I$ dan ekivalen $\pi_{1/1} = \pi_{1/2} = \ldots = \pi_{1/I}$,

$$\log\left(\frac{\pi_{1/i}}{1-\pi_{1/i}}\right) = \beta_i \tag{2.6}$$

Ada model logit yang digeneralisasi ketika ada beberapa faktor berkategori. Andaikata, ada 2 faktor A dan B untuk respon berpasangan dan I menunjukkan bilangan level A dan J menunjukkan level B. Penunjukkan $\pi_{k/ij}$ sebagai peluang dari respon k ketika A pada level ke- i dan B pada level ke- j, sehingga $\pi_{1/ij} + \pi_{2/ij} = 1$. Model logit untuk i x j x 2 adalah

$$\log\left(\frac{\pi_{1/ij}}{\pi_{2/ij}}\right) = \alpha + \beta_i^A + \beta_j^B \tag{2.7}$$

$\{\beta_i^A\}$ menggambarkan pengaruh A dan $\{\beta_i^B\}$ menggambarkan pengaruh B. model ini memperlakukan $\{n_{ij+}\}$ dan $\{n_{ij1}\}$ variabel acak Binomial bebas dengan parameter $\{\pi_{1/ij}\}$.

2.2.2. Model Regresi Logistik

Penunjukan hasil dengan 1 atau 0 memberikan variabel acak Bernoulli dengan mean

$$E(Y) = 1 \text{ x } P(Y=1) + 0 \text{ x } P(Y=0) = P(Y=1)$$

Peluang ini kita gambarkan dengan $\pi(x)$, pencerminan tidak bebas pada $X = (X_1, \ldots, X_k)$. Selama

$$E(Y^2) = 1^2 \text{ x } \pi(X) + 0^2 \text{ x } \pi(X) = \pi(X)$$

Varian Y adalah

$$V(Y) = E(Y^2) - [E(Y)]^2 = \pi(X)[1 - \pi(X)]$$

Model regresi untuk respon berpasangan/ biner adalah

$$E(Y) = \pi(X) = \alpha + \beta x \quad (2.8)$$

yang disebut dengan model peluang linier. Selama pengamatan Y bebas, model ini adalah GLM dengan fungsi link identitas.

Karena terdapat masalah struktural dengan model peluang linier, maka lebih baik untuk dipelajari hubungan kurvalinier antara x dan $\pi(x)$. Bentuk kurva ini adalah bentuk asli dari kurva regresi, bentuk ini adalah

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \quad (2.9)$$

yang disebut fungsi logistik regresi.

Model logistik regresi untuk nilai $X = (x_1, \ldots, x_k)$ pada k variabel eksplanatori adalah

$$\log\left(\frac{\pi(X)}{1 - \pi(X)}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \quad (2.10)$$

Selanjutnya kita misalkan $X_i = (x_{i0}, x_{i1}, \ldots, x_{ik})$ untuk k variabel eksplanatori, $i = 1 \ldots r$, dengan $x_{i0} = 1$. Untuk N respon bebas dari variabel acak Bernoulli, model Logistik Regresi (2.9) menjadi :

$$\pi(X_i) = \frac{\exp\left(\sum_{j=0}^{k} \hat{\beta}_j x_{ij}\right)}{1 - \exp\left(\sum_{j=0}^{k} \hat{\beta}_j x_{ij}\right)} \quad \text{dimana} \quad \beta_0 = \alpha \quad (2.11)$$

Y_i adalah variabel acak Bernoulli dengan $E(Y_i)$ $n_i\ \pi(X_i)$, maka persamaan Likelihood adalah

$$\sum_i Y_i x_{ia} - \sum_i n_i \hat{\pi}_i x_{ia} = 0, \quad a = 0 \ldots k \tag{2.12}$$

dengan $\hat{\pi}_i = \dfrac{\exp\left(\sum_{j=0}^{k} \hat{\beta}_j x_{ij}\right)}{1 - \exp\left(\sum_{j=0}^{k} \hat{\beta}_j x_{ij}\right)}$ menunjukkan Pendugaan Maksimum Likelihood dari $\pi(X_i)$.

Kita sekarang mencari fungsi link untuk model (2.9) di atas, dimana untuk model ini kemungkinan mendapatkan respon 1 adalah

$$\frac{\pi(x)}{1-\pi(x)} = \exp(\alpha + \beta x) = e^{\alpha}\left(e^{\beta}\right)^{x}$$

Rumus ini bertujuan untuk menginterpretasikan tafsiran dasar untuk β. Log kemungkinan (*log odds*) memiliki hubungan linier

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x \tag{2.13}$$

Jadi ketepatan link adalah transformasi log kemungkinan logit.

2.3. Generalized Linear Models.

Teori ini dijelaskan dalam *Generalized Linier Models* (Mc Cullagh dan Nelder, 1989), Inferensi Pada Model – Model Linier Tergeneralisasi (Rita, 2000), dan dalam Tirta (2001), yang ringkasannya disampaikan pada bagian berikut.

Dalam banyak penelitian, seringkali dijumpai bahwa variabel respon adalah tidak berdistribusi normal, sehingga model linier sederhana yang ada tidak dapat digunakan untuk mengestimasi parameter. Maka dikembangkan suatu model linier umum yang melibatkan lebih dari satu variabel independen yang tidak hanya melibatkan variabel random yang berdistribusi normal, namun juga variabel random yang berdistribusi Poisson, Gamma, Binomial dan sebagainya.

Keluarga distribusi yang mencangkup kesemua distribusi ini adalah keluarga eksponensial. Keluarga densitas disebut keluarga eksponensial, bila ia dapat dinyatakan sebagai :

$$f(y;\theta;\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y;\phi)\right\} \qquad (2.14)$$

Pada model di atas kondisi khusus $a(\phi) = \phi$ yang merupakan parameter kanonik, dengan ϕ adalah parameter dispersi dan $b(\theta)$ adalah fungsi komulatif. Karakteristik tiap – tiap distribusi dibentuk pada Tabel 2.

Tabel 2 Karakteristik beberapa distribusi univariat dalam keluarga exponensial

	Notasi		Parameter Dispersi : ϕ	Fungsi Komulatif : $b(\theta)$	$c(y;\theta)$	$\mu(\theta)=$ $E(Y;\theta)$	Link kannonik : $\theta(\mu)$	Fungsi Varian : $V(\mu)$
Normal	$N(\mu,\sigma^2)$	$(-\infty,\infty)$	σ^2	$\theta^2/2$	$-\frac{1}{2}\left(\frac{y^2}{\phi}+\log(2\pi\phi)\right)$	θ	*identitas*	1
Poisson	$P(\mu)$	$0(1)\infty$	1	$\exp(\theta)$	$-\log y!$	$\exp(\theta)$	*log*	μ
Binomial	$B(m,\pi)/m$	$\frac{0(1)m}{m}$	$1/m$	$\log(1+e^{\theta})$	$\log\binom{m}{my}$	$e^{\theta}/(1+e^{\theta})$	*logit*	$\mu(1-\mu)$
Gamma	$G(\mu, v)$	$(0,\infty)$	v^{-1}	$-\log(-\theta)$	$v\log(vy)-\log y$ $-\log\Gamma(v)$	$-1/\theta$	*Resiprocal*	μ^2
Invers Gaussian	$IG(\mu,\sigma^2)$	$(0,\infty)$	σ^2	$-(-2\theta)^{1/2}$	$-\frac{1}{2}\left\{\log\left(2\pi\phi y^2\right)+\frac{1}{\phi y}\right.$	$(-2\theta)^{-1/2}$	$1/\mu^2$	μ^2

Sumber : Mc Cullagh dan Nelder (1990)

Model linier tergeneralisasi merupakan pengembangan model linier yang sering digunakan dalam analisis statistika, khususnya data kontinyu.

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \tag{2.11}$$

Model (2.) di atas adalah model linier klasik dengan $\mathbf{Y} = (y_1, \ldots y_N)$, $\varepsilon = (e_1, \ldots, e_N)^T$, $\boldsymbol{x}$ adalah sebuah matrik $N \times p$ dari veriabel eksplanatori dan $\beta = (\beta_1, \ldots \beta_N)$ (Mc. Cullagh dan Nelder, 1989); diasumsikan bahwa $e_i \sim$ NID$(0,\sigma^2)$ dan selama $\mathbf{Y} \sim N(E(\mathbf{Y}), \sigma^2\mathbf{I})$, dimana

$$E(Y_i) = \mu_i = x_i^T = \sum_{j=1}^{p} x_{ij}\beta_j .$$

Dalam model linier tergeneralisasi, asumsi pada model linier klasik kurang terbatas dan harus digeneralisasi dalam beberapa arah, yaitu

a. Y_i berdistribusi secara bebas dengan mean μ_i dari sebuah anggota keluarga exponensial.

b. Adanya fungsi η pada variabel eksplanatori sebagai sebuah prediktor linier pada variabel respon dan berhubungan dengan mean, yaitu

$$g(\mu_i) = \eta = x_i^T \beta$$

dengan g adalah fungsi kontinyu dan differensiabel. Fungsi g merupakan penghubung antara prodiktor linier (η_i) dan komponen acak (μ_i), yang selanjutnya disebut sebagai fungsi hubungan (*link function*)

Dapat disimpulkan bahwa komponen terdiri dari distribusi Y_i dari sebuah anggota keluarga eksponensial, adanya prediktor linier η , dan sebuah fungsi g yang menghubungkan antara prediktor linier η dengan komponen acak μ, $g(\mu) = \eta$.

Metode yang digunakan dalam mengestimasi parameter adalah metode maksimum likelihood dan kuadrat terkecil. Pada bahasan ini hanya akan dijelaskan mengenai metode maksimum likelihood. Penduga maksimum likelihood pada parameter θ dinotasikan dengan $\hat{\theta}$ yang mendefinisikan suatu ruang parameter (Ω) pada fungsi likelihood.

$\hat{\theta} \in \Omega$ adalah penduga ML jika $L(\hat{\theta}) \geq L(\theta), \forall \theta \in \Omega$

Penduga parameter pada GLM diperoleh dengan menggunakan metode estimasi maksimum likelihood. Fungsi likelihood untuk observasi y_1 ,$y_2, \ldots, y_n$ dari fungsi densitas keluarga exponensial adalah

$$L(y_i, \theta) = \prod_{i=1}^{N} f(y_i, \theta) \tag{2.15}$$

$$l(\theta, y) = \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y, \phi) \tag{2.16}$$

dengan fungsi link,

$$g(\mu_i) = x^T \beta = \sum_{j=1}^{p} x_{ij} \beta_j = \eta_i \tag{2.17}$$

g adalah fungsi link yang monoton dan diferensiabel, dan

$$E(y_i) = \mu_i = b'(\theta_i) \tag{2.18}$$

$$Var(y_i) = b''(\theta_i)a(\phi) = V(\mu_i)a(\phi) \tag{2.19}$$

Skor yang bersesuaian dengan parameter β_j didefisinikan sebagai

$$U_j = \frac{\partial l(\tilde{\theta}, \tilde{y})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \beta_j}$$

dengan

$$l_i = \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y, \phi) \tag{2.20}$$

Untuk mendapatkan U_j digunakan

$$\frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j}$$

Dari (2.20) diperoleh

$$\frac{\partial l_i}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)} \tag{2.21}$$

Dari (2.18) diperoleh $\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta)$ sehingga $\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{b''(\theta)}$ kemudian

dengan mensubtitusikan (2.19) diperoleh

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{a(\phi)}{Var(y_i)} \tag{2.22}$$

Dari (2.17) diperoleh

$$\frac{\partial \eta_i}{\partial \beta_j} = x_{ij} \text{ sehingga } \quad \frac{\partial \mu_i}{\partial \beta_j} = x_{ij} \frac{\partial \mu_i}{\partial \eta_i} \tag{2.23}$$

Maka dari (2.21), (2.22) dan (2.23) diperoleh

$$\frac{\partial l_i}{\partial \beta_j} = \frac{(y_i - \mu_i)}{a(\phi)} \frac{a(\phi)}{V(\mu_i)} x_{ij} \frac{\partial \mu_i}{\partial \eta_j}$$

$$= \left(\frac{(y_i - \mu_i) x_{ij}}{V(\mu_i)} \right) \left(\frac{\partial \mu_i}{\partial \eta_j} \right)$$

dan $\quad U_j = \sum \frac{(y_i - \mu_i) x_{ij}}{Var(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)$ $\quad$ dapat ditulis secara matrik

$$U_j = \mathbf{X}^T \mathbf{S}^T \mathbf{V}^{-1} (\mathbf{Y} - \mu) \text{ , dimana } \mathbf{S} = diag \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \tag{2.24}$$

dengan x_{ij} adalah elemen ke $-j$ dari x^T.

U_j adalah persamaan non linier yang harus diselesaikan secara numeris. Pendekatan ke- m dengan menggunakan metode Newton-Raphson adalah

$$b^{(m)} = b^{(m-1)} - \left(\frac{\partial^2 l}{\partial \beta_j \beta_k} \right) U^{(m-1)} \tag{2.25}$$

dengan $U^{(m-1)}$ adalah vektor U yang dihitung pada $\beta = b^{(m-1)}$, dan $b^{(m-1)}$ menunjukkan nilai b pada iterasi ke (m).

$$U'^{(m-1)} = \left(\frac{\partial^2 l}{\partial \beta_j \beta_k} \right)^{(m-1)} \tag{2.26}$$

adalah matrik turunan kedua l dihitung pada $\beta = b^{(m-1)}$.

Pada prakteknya digunakan metode yang lebih sederhana, yaitu metode scoring, dengan mengganti matrik pada (2.25) dengan matrik harga harapan

$$E\left(\frac{\partial l}{\partial \beta_j} \frac{\partial l}{\partial \beta_k} \right)$$

Dalam hal ini menurut Dobson(1990),

$$E\left(\frac{\partial l}{\partial \beta_j} \frac{\partial l}{\partial \beta_k} \right) = -E\left(\frac{\partial^2 l}{\partial \beta_j \beta_k} \right)$$

Jika $\Im = E(UU^T)$, maka

$$\Im_{jk} = E\left(\frac{\partial l}{\partial \beta_j} \frac{\partial l}{\partial \beta_k} \right) = -E\left(\frac{\partial^2 l}{\partial \beta_j \beta_k} \right)$$

Sehingga (2.26) menjadi :

$$b^{(m)} = b^{(m-1)} + \left[\Im^{(m-1)} \right]^{-1} U^{(m-1)} \tag{2.27}$$

atau

$$\Im^{(m-1)}b^{(m)} = \Im^{(m-1)}b^{(m-1)} + U^{(m-1)} \tag{2.28}$$

Untuk model linier tergeneralisasi elemen ke (j,k) dari $\Im$ adalah

$$\begin{aligned}\Im_{jk} &= E\left(\frac{\partial l}{\partial \beta_j}\frac{\partial l}{\partial \beta_k}\right) = -E\left(\frac{(y_i - \mu_i)^2 x_{ij}x_{ik}}{[Var(y_i)]^2}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\right)\\ &= \frac{x_{ij}x_{ik}}{[Var(y_i)]}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\\ \Im_{jk} &= \mathbf{X}^T\mathbf{S}^T\mathbf{V}^{-1}\mathbf{SX}\end{aligned} \tag{2.29}$$

dengan $\mathbf{S} = diag\left(\frac{\partial \mu_i}{\partial \eta_i}\right)$, $\mathbf{V} = diag(Var(Y_i))$

Persamaan iterasi (2.28) dapat ditulis menjadi :

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} + \left[\left(\mathbf{X}^T\mathbf{S}^T\mathbf{V}^{-1}\mathbf{S}\,\mathbf{X}\right)^{(m-1)}\right]^{-1}\left(\mathbf{X}^T\mathbf{S}^T\mathbf{V}^{-1}(\mathbf{Y}-\boldsymbol{\mu})\right)^{(m-1)} \tag{2.30}$$

Persamaan ini harus diselesaikan secara iterasi, prosedurnya adalah dengan mengambil harga pendekatan $\boldsymbol{b}^{(0)}$ untuk menghitung **S** dan **V**. Selanjutnya (2.30) diselesaikan dengan mengambil $\boldsymbol{b}^{(1)}$ yang diperoleh dari perhitungan **S** dan **V** sebelumnya. Proses ini diulang sampai diperoleh ***b*** yang konvergen atau ;

$\frac{\left|b^{(m+1)} - b^{(m)}\right|}{b^{(m)}} \leq \delta$; dimana δ adalah bilangan riil positif yang sangat kecil.

Secara umum untuk model linier tergeneralisasi, distribusi sampling diperoleh dari suatu pendekatan sample besar, sebagai berikut :

1. b_j merupakan estimator tak bias untuk β.
2. Dengan menggunakan teorema limit pusat, statistik $\frac{b_j - \boldsymbol{\beta}}{\sqrt{Var(b_j)}}$ berdistribusi N(0,1) atau $\frac{(b_j - \boldsymbol{\beta})^2}{Var(b_j)} \sim \chi^2_{(1)}$.

Inferensi untuk β dilakukan dengan menggunakan statistik Wald :

$$(\mathbf{b} - \boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{b} - \boldsymbol{\beta}) \sim \chi^2_{(p)}$$

Interval konfidensi (1 – α) untuk βj adalah

$$b_j - Z_{\alpha/2} Sd(b_j) < \beta_j < b_j + Z_{\alpha/2} Sd(b_j)$$

2.4. Model Marginal.

Terdapat beberapa model yang digunakan dalam analisis data, seperti model marginal. Model ini dijelaskan oleh Diggle *et. al* (1994) dalam sebuah buku *Analysis of Longitudinal Data*, yang ringkasannya adalah sebagai berikut.

Dalam sebuah model marginal, regresi dari respon pada variabel eksplanatori adalah merupakan pemodelan terpisah dalam unit korelasi. Dalam regresi tersebut, model expektasi marginal E(Y_{ij}) seperti sebuah fungsi variabel eksplanatori. Expektasi marginal tersebut adalah suatu model dalam mempelajari *cross-sectional*. Model marginal pada khususnya memiliki asumsi :

- Ekpektasi marginal respon, E(Y_{ij}) = μ_{ij}, tergantung pada variabel eksplanatori x_{ij} dengan $h(\mu_{ij}) = x_{ij}`\beta$, dimana h adalah fungsi link yang diketahui seperti logit untuk respon biner.
- Varian marginal tergantung pada mean marginal menurut $Var(Y_{ij}) = v(\mu_{ij})\phi$, dimana v adalah fungsi varian yang diketahui dan ϕ adalah skala parameter yang diperoleh dengan menduga.
- Korelasi antara Y_{ij} dan Y_{ik} adalah sebuah fungsi pada rataan marginal dan mungkin penjumlahan parameter – parameter α, $Corr(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}; \alpha)$ dimana $\rho(.)$ adalah sebuah fungsi yang diketahui.

Model marginal untuk data yang berkorelasi adalah analog alami dari GLM untuk data independent. Sebagai ilustrasi, sebuah model marginal logistik untuk mempertimbangkan masalah dalam menilai ketidak bebasan infeksi pernafasan pada vitamin A. Misal x_{ij} mengindikasikan ada atau tidak anak ke- i adalah vitamin A tidak sempurna (1- ya, 0-tidak) ke- j. Misal Y_{ij}

menunjukkan apakah anak mempunyai infeksi pernafasan (1- ya, 0-tidak) dan misal $\mu_{ij} = E(Y_{ij})$ dengan asumsi

- $\log it(\mu_{ij}) = \log \dfrac{\mu_{ij}}{1-\mu_{ij}} = \log \dfrac{\Pr(Y_{ij}=1)}{\Pr(Y_{ij}=0)} = \beta_0 + \beta_1 x_{ij}$
- $Var(Y_{ij}) = \mu_{ij}(1-\mu_{ij})$
- $Corr(Y_{ij}, Y_{ik}) = \alpha$

Disini, perubahan koefisien regresi *exp*(β_0) adalah rasio pada frekuensi menulari atau tidak antara subpopulasi pada anak dengan Vit A tidak sempurna. Parameter *exp*(β_1) adalah kemungkinan infeksi antara anak dengan Vit A tidak sempurna dan anak dengan Vit A sempurna. Dicatat bahwa *exp*(β_1) adalah rasio frekuensi populasi sehingga kita memilih parameter rataan populasi. Jika semua individu dengan x yang sama memiliki peluang yang sama, maka frekuensi populasi adalah peluang individu yang sama.

Varian pada respon biner adalah fungsi yang diketahui pada rataan yang diberikan pada asumsi kedua. Korelasi antara 2 observasi pada anak – anak adalah diasumsikan menjadi α tanpa memperhatikan ekspektasi pengamatan. Pada data kontinyu bentuk korelasi ini diturunkan dari intersep acak dalam model linier. Hal ini hanya dapat menjadi aproksimasi pertama untuk hasil biner, karena korelasi antara 2 respon biner Y_1 dan Y_2 dengan μ_1 dan μ_2 diberikan oleh

$$Corr(Y_1, Y_2) = \frac{\Pr(Y_1 = 1, Y_2 = 1)\Pr(Y_1 = 0, Y_2 = 0)}{\{\mu_1(1-\mu_1)\mu_2(1-\mu_2)\}^{1/2}} \tag{2.31}$$

Peluang bersama, $\Pr(Y_1 = 1, Y_2 = 1)$ dipertimbangkan untuk memenuhi

$$\max(0, \mu_1 + \mu_2 - 1) < \Pr(Y_1 = 1, Y_2 = 1) < \min(\mu_1, \mu_2) \tag{2.32}$$

dilihat bahwa korelasi harus memenuhi sebuah pertimbangan yang tergantung dalam satu arah yang rumit susunannya pada rataan μ_1 dan μ_2. Model kelompok antara data biner menggunakan rasio kemungkinan (*odds ratio*)

$$OR(Y_1, Y_2) = \frac{\Pr(Y_1 = 1, Y_2 = 1)\Pr(Y_1 = 0, Y_2 = 0)}{\Pr(Y_1 = 1, Y_2 = 0)\Pr(Y_1 = 0, Y_2 = 1)} \tag{2.33}$$

yang tidak mempertimbangkan mean. Dalam contoh ilustrasi di atas kita dapat mengasumsikan rasio kemungkinan dari korelasi antara semua pasang respon untuk anak adalah suatu konstanta α yang tidak diketahui.

2.5. Generalized Estimating Equations (GEE).

Teori ini dijelaskan oleh Diggle *et. al* (1994)dengan ringkasannya adalah sebagai berikut.

Dalam model linier, fungsi likelihood mudah dievaluasi dan dimaksimumkan. Untuk GLM dengan fungsi link non linier, satu kesulitan adalah distribusi marginal pada Y_{i1} seringkali tidak dapat ditentukan dari distribusi kondisional $f(y_{ij}|y_{ij}-1)$ tanpa asumsi penjumlahan.

Sebuah alternatif sederhana untuk memaksimumkan likelihood kondisional pada $Y_{i2},\ldots Y_{in}$, diberikan Y_{i1}, yang mana berlaku dengan mengabaikan $f(y_{ij})$ dari persamaan sebelumnya di atas. Kondisional pendugaan maksimum likelihood dapat dicari menggunakan sofware standar GLM dengan memperlakukan fungsi respon sebelumnya pada variabel eksplanatori.

Jika asumsi penjumlahan dibuat, likelihood seringkali sulit dan melibatkan beberapa paremeter gangguan (*nuisance*) dalam penjumlahan α dan β yang harus diduga. Untuk alasan ini, pendekatan yang baik dalam mengatasi kesulitan ini adalah digunakan *Generalized Estimating Equations* (GEE) yang merupakan sebuah analog multivariat pada *quasi-likelihood*.

Pada fungsi likelihood, hal ini lebih baik untuk menduga/ mengestimasi β dengan menyelesaikan sebuah multivariat yang sama pada score-quasi :

$$S(\beta)=\sum_{i=1}^{m}\left(\frac{\partial\eta_i}{\partial\beta}\right)^T Var(Y_i)^{-1}(Y_i-\mu_i)=0 \qquad (2.34)$$

dimana, Y_i , μ_i adalah vektor dan Var(**Y**) merupakan matrik simetris. Dalam kasus multivariat, ada gabungan jumlah seperti S_β yang tergantung pada α maupun pada β selama $Var(Y_i) = Var(Y_i;\ \beta,\alpha)$. Ini dapat dikuasai oleh perubahan α dalam persamaan di atas dengan sebuah $m^{1/2}$- penduga konsisten $\hat{\alpha}(\beta)$. Dengan respon biner, parameter α dapat dirumuskan dan diestimasi dalam sebuah bilangan arah. Parametrik Var(Y_i) dalam hubungan korelasi dan digunakan pendugaan waktu untuk parameter tidak diketahui.

Solusi dari $S(\beta) = 0$, merupakan suatu pendekatan Gaussian dengan dua metode yang digunakan dalam menduga varian, yaitu

1. $\hat{\beta}$ hampir relatif efisien pada penduga β dengan maksimum likelihood; dalam kenyataanya, GEE adalah fungsi skor maksimum likelihood untuk multivariat Gaussian dan untuk data berpasangan (*binary*) dari suatu log-linier ketika $Var(Y_i)$ adalah kasus benar.
2. $\hat{\beta}$ adalah konsisten pada $m \rightarrow \sim$, meskipun struktur kovarian Y_i adalah suatu pengukuran yang tidak benar.

Nilai *Robust* pada inferensi terhadap β dapat dilihat dengan mencocokkan sebuah model akhir menggunakan asumsi kovarian yang berbeda, dan membandingkan 2 bagian pada penduganya dengan standar kesalahan *Robust*. Jika terdapat perbadaan yang nyata, maka bentuk kovarian model dianggap penting. (Diggle.*et.al*, 1994:145)

2.6. Algoritma Untuk Menghitung GEE.

Algoritma untuk menghitung GEE diuraikan oleh M.G. Kenward dan D. M. Smith (1931) dalam artikel *Computing The Generalized Estimating Equations for Repeated Measurments*, yang ringkasannya adalah sebagai berikut.

Iterasi persamaan kuadrat terkecil independen untuk GLM dapat ditulis sebagai berikut

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + \left(\mathbf{X}^T \left(\frac{\partial \mu}{\partial \eta}\right)^{(k)} \left(Var(\mu)^{(k)} \right)^{-1} \left(\frac{\partial \mu}{\partial \eta}\right) \right)^{-1} \mathbf{X}^T \left(\frac{\partial \mu}{\partial \eta}\right)^{(k)} \left(Var(\mu)^{(k)} \right)^{-1} (y-\mu)^{(k)}$$

Dimana indek k sama dengan iterasi ke- k, β adalah vektor koefisien, $\mathbf{X}$ adalah matrik variabel eksplanatori, μ adalah vektor nilai yang cocok, $\frac{\partial \mu}{\partial \eta}$ adalah penurunan invers dari fungsi ling dan η adalah prediktor linier.

Varian pada $\hat{\beta}_{k+1}$ adalah

$$\phi^{(k)} \left(\mathbf{X}^T \left(\frac{\partial \mu}{\partial \eta}\right)^{(k)} \left(Var(\mu)^{(k)} \right)^{-1} \left(\frac{\partial \mu}{\partial \eta}\right)^{T\,(k)} \mathbf{X} \right)^{-1}$$

Dimana ϕ_k adalah faktor skala (*scale factor*).

Perhitungan pokok dalam GEE adalah dengan mengembangkan GLM dalam 4 arah, yaitu :

1. Dengan menggunakan rataan pada tiap unit.

$$\left(\sum_{i=1}^{N} \mathbf{X}^T \left(\frac{\partial \mu}{\partial \eta}\right)^{(k)} \left(Var(\mu)^{(k)}\right)^{-1} \left(\frac{\partial \mu}{\partial \eta}\right)^{T(k)} \mathbf{X}\right)^{-1} \sum_{i=1}^{N} \mathbf{X}^T \left(\frac{\partial \mu}{\partial \eta}\right)^{(k)} \left(Var(\mu)^{(k)}\right)^{-1} (y-\mu)^{(k)}$$

Dimana N adalah jumlah unit yang terlibat, X_i adalah sebuah matrik t x p, dimana t adalah jumlah ulangan dan p adalah jumlah variabel eksplanatori.

2. Dengan mengubah vektor $\frac{\partial \mu_i}{\partial \eta_i}$ pada panjang t dengan sebuah matrik diagonal t x t.

3. Dengan menggunakan sebuah penduga *sandwich* pada varian β

$$Var_S(\beta) = \phi \mathbf{V}^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i^T \left(\frac{\partial \mu}{\partial \eta}\right) \left(Var(\mu)^{(k)}\right)^{-1} Var(y_i) \left(Var(\mu)^{(k)}\right)^{-1} \left(\frac{\partial \mu}{\partial \eta}\right)^T \mathbf{X}_i\right) \mathbf{V}^{-1}$$

$$\text{dimana } \mathbf{V} = \left(\sum_{i=1}^{N} \mathbf{X}_i^T \left(\frac{\partial \mu}{\partial \eta}\right) \left(Var(\mu)^{(k)}\right)^{-1} \left(\frac{\partial \mu}{\partial \eta}\right)^T \mathbf{X}_i\right)$$

penduga *naïve* pada varian β yaitu

$$Var_N(\beta) = \phi\, \mathbf{V}^{-1}$$

Varian Y diduga oleh $\mathbf{Res}_i\mathbf{Res}_i^T$, dimana $\mathbf{Res}_i$ adalah vektor residual $\frac{(y_i - \mu_i)}{\sqrt{Var(\mu_i)}}$ untuk subyek ke- i.

4. Dengan mengubah vektor $Var(\mu_i)$ dengan panjang t menjadi matrik t x t, yaitu

$$\phi\sqrt{Var(\mu_i)}\mathbf{R}(\alpha)\sqrt{Var(\mu_i)}$$

Dimana $\mathbf{R}(\alpha)$ adalah matrik korelasi antara y_i, α adalah sebuah parameter (atau vektor parameter) yang mendefinisikan **R**.

Implementasi metodologi GEE terdiri dari pendugaan faktor skala ϕ dan penduga $\mathbf{R}(\alpha)$. Faktor skala dan parameter korelasi (α) diduga dari elemen – elemen pada data berdasarkan matrik kovarian $\mathbf{C}_{res}$. Jadi untuk setiap iterasi pada proses GEE, pendugaan faktor skala dan struktur korelasi dari iterasi sebelumnya yang digunakan dalam perhitungan $\mathbf{C}_{res}$. Jika diasumsikan tidak ada data yang hilang

$$\mathbf{C}_{res} = \frac{\left(\sum_{i=1}^{N}\mathbf{Res}_i\mathbf{Res}_i^T\right)}{N}, \text{ misal } \mathbf{A}_i = \mathbf{Res}_i\mathbf{Res}_i^T$$

$$\text{maka } \mathbf{C}_{res} = \frac{\left(\sum_{i=1}^{N}\mathbf{A}_i\right)}{N}$$

Dalam prosedure penggambaran, ϕ adalah dapat berupa sebuah skalar atau sebuah matrik t x t. Faktor skala diduga dari elemen – elemen pada matrik varian-kovarian $\mathbf{C}_{res}$, yaitu

$$\mathbf{C}_{res} = \phi^{1/2}\mathbf{R}_{model}\phi^{1/2}$$

Jika ϕ adalah skalar, pendugaannya adalah

$$\hat{\phi} = \frac{trace\left(\mathbf{R}_{model}^{-1}\mathbf{C}_{res}\right)}{t}$$

Jika ϕ adalah sebuah matrik diagonal t x t, maka ada perbedaan faktor skala pada tiap operasi pada tiap saat. Faktor skala diduga seperti solusi suatu persamaan matrik

$$diagonal\left(\hat{\phi}^{-1/2}\mathbf{R}_{model}^{-1}\hat{\phi}^{-1/2}\mathbf{C}_{res}\right) = \mathbf{I} \tag{2.35}$$

Jika vektor $\hat{\mathbf{z}} = \hat{\phi}^{-1/2}$ (elemen diagonal) maka persamaan (2.35) menjadi

$$\hat{\mathbf{z}}\hat{\mathbf{z}}^T \times \mathbf{H}[\mathbf{1}] - [\mathbf{1}] = [\mathbf{0}]$$

Dimana [**0**] dan [**1**] adalah vektor nol dan vektor 1 dengan panjang t dan

$$\mathbf{H} = \mathbf{R}_{model}^{-1} \times \mathbf{C}_{res}$$

Dengan $\times$ adalah perkalian antara elemen.

Semua model untuk $\mathbf{R}(\alpha)$, dilakukan pendugaan pada α dari *residual* berdasarkan penggunaan matrik korelasi $\mathbf{R}_{res}$, yang mana merupakan solusi dari persamaan matrik

$$lower\ triangle\ of\left(\mathbf{R}_{res}^{-1} - \mathbf{R}_{res}^{-1}\phi^{-1/2}\mathbf{C}_{res}\phi^{-1/2}\mathbf{R}_{res}^{-1}\right) = 0 \tag{2.36}$$

subyek berdasarkan diagonal ($\mathbf{R}_{res}$) = **I**. Ini merupakan penyelesaian untuk $\mathbf{R}_{res}$ dengan prosedure iterasi dimana k menunjukkan iterasi

$$initial\ estimate\ \ \mathbf{R}_0 = \phi^{-1/2}C_{res}\phi^{-1/2}$$

penggambaran sebuah matrik korelasi, untuk $k = 1, \ldots$

$$\Lambda^{(k)} = diagonal\left[\mathbf{F}^{-1(k)}\left([1] - vec.diag\left(\phi^{-1/2}\mathbf{C}_{res}\phi^{-1/2}\right)\right)\right]$$

$$\text{dan } \mathbf{R}^{(k+1)} = \left(\phi^{-1/2}\mathbf{C}_{res}\phi^{-1/2} - \mathbf{R}^{(k)}\Lambda^{(k)}\mathbf{R}^{(k)}\right)$$

Dimana $\Lambda^{(k)}$ adalah sebuah matrik diagonal (t x t), $vec.diag\left(\phi^{1/2}\mathbf{C}_{res}\phi^{1/2}\right)$ adalah vektor dengan panjang t pada elemen diagonal, dan elemen ke- ij pada $\mathbf{F}^{(k)}$ adalah

$$\left(\mathbf{F}^{(k)}\right)_{ij} = \left(\mathbf{R}^{(k)}\right)^2_{ij}$$

Beberapa matrik korelasi untuk model yang berbeda, yaitu

- *Independent structure.*

  Untuk struktur independen $\mathbf{R}_{ind} = \mathbf{I}$. Dimana merupakan matrik diagonal t x t dengan nilai 1 pada diagonalnya.

- *Unstructure.*

  Untuk tidak terstruktur $\mathbf{R}_{uns} = \mathbf{R}_{res}$, yang merupakan matrik t x t berdasarkan pada residual.

- *Exchangeable structure.*

  Untuk Exchangeable *structure*

$$\mathbf{R}_{exc} = \begin{bmatrix} 1 & & & & \\ \alpha & 1 & & & \\ \alpha & \alpha & 1 & & \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \alpha & 1 \end{bmatrix}$$

parameter α diduga dengan

$$\hat{\alpha} = \frac{\sum r_{res\ i,k}}{\frac{t(t-1)}{2} - \frac{p}{n}} \quad \text{dimana penjumlahan untuk } k < i$$

- *Autoregressive structure.*

  Struktur *autoregressive* adalah

$$\mathbf{R}_{aut} = \begin{bmatrix} 1 & & & & \\ \alpha & 1 & & & \\ \alpha^2 & \alpha & 1 & & \\ \dots & \dots & \dots & \dots & \dots \\ \alpha^{t-1} & \dots & \dots & \alpha & 1 \end{bmatrix}$$

- *Dependence structure.*

  Struktur *dependence* order 2

$$\mathbf{R}_{dep} = \begin{bmatrix} 1 & & & & \\ \alpha_1 & 1 & & & \\ \alpha_2 & \alpha_1 & 1 & & \\ 0 & \alpha_2 & \alpha_1 & \dots & \dots \\ \dots & 0 & \alpha_2 & \alpha_1 & 1 \end{bmatrix}$$

parameter α_1, α_2 *diduga* dengan

$$\hat{\alpha} = \frac{\sum_{i=j+1}^{t} r_{res\ i,i-j}}{t-j} \qquad j = 1....$$

BAB III

METODOLOGI

3.1. Identifikasi Variabel dan Distribusi.

Penggunaan notasi – notasi Matematis sama dengan notasi yang ada pada teori yang digunakan. Notasi **X** untuk variabel eksplanatori (subyek) dan **Y** untuk variabel respon atau dapat diuraikan sebagai berikut :

Individual (subject)	Response					Correspond. Vector/ Matrices	
	1	...	j	...	n_i	**Y**	**X**
1	y_{11}	...	y_{1j}	...	y_{1n1}	$\mathbf{y}_1^T$	$\mathbf{x}_1^T = [x_{11}, \ldots, x_{1j}, \ldots x_{in1}]$
:	:	:	:	:	:		:
i	y_{i1}	...	yij	...	Y_{1nj}	$\mathbf{y}_i^T$	$\mathbf{x}_1^T = [x_{i1}, \ldots, x_{ij}, \ldots x_{ini}]$
:	:	:	:	:	:		:
m	y_{m1}	....	y_{mj}	...	Y_{mnm}	$\mathbf{y}_m^T$	$\mathbf{x}_m^T = [x_{m1}, \ldots, x_{mj}, \ldots x_{mnm}]$

Model linier yang digunakan adalah model linier secara umum, yaitu :

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \text{ .}$$

Dalam bentuk matrik adalah sebagai berikut:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_i \\ \vdots \\ \mathbf{y}_m \end{bmatrix}; \mu = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_i \\ \vdots \\ \boldsymbol{\mu}_m \end{bmatrix}; \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \\ \vdots \\ \beta_q \end{bmatrix}; \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_i \\ \vdots \\ \mathbf{X}_m \end{bmatrix}; \mathbf{X_i} = \begin{bmatrix} \mathbf{x}_{i1}^T \\ \mathbf{x}_{i2}^T \\ \vdots \\ \mathbf{x}_{ij}^T \\ \vdots \\ \mathbf{x}_{in_i}^T \end{bmatrix} \text{ untuk } \mathbf{x}_{ij} = \begin{bmatrix} x_{ij1} \\ x_{ij2} \\ \vdots \\ x_{ijk} \\ \vdots \\ x_{ijq} \end{bmatrix}$$

dengan $i = 1, \ldots, m$ dan vektor pengamatan $\mathbf{y_i}$ adalah bentuk dari $\mathbf{y_i}^T = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})$ dengan $j = 1, \ldots, n_i$ yang mengindikasikan banyaknya observasi yang berulang, vektor $\mathbf{y}$ disini berdistribusi Binomial. Pengaruh tetap (*fixed effect*) ditunjukkan oleh vektor $\beta^T = (\beta_1, \ldots, \beta_k, \ldots, \beta_q)$.

3.2. Langkah - Langkah Pemodelan.

Langkah - langkah pemodelan sangat manunjang proses pemodelan itu sendiri. Disini akan diuraikan langkah - langkah dalam menyelesaikan model yang disesuaikan dengan permasalahan yang akan diungkapkan :

1. Berdasarkan teori – teori yang diungkapkan, kita ikuti penyelesaian model berdasarkan teori tersebut. Dimana tolak ukur penyelesaian didasarkan pada data lapangan yang diungkapkan pada latar belakang permasalahan.

2. Model yang digunakan identik dengan model pada GLM, yaitu berupa model linier. Dimana dalam penyelesaian model kita gunakan prosedur GEE (*Generalized Linear Model*).
3. Dalam proses pemodelan, hal yang paling penting adalah mengubah data dalam bentuk tabel kontingensi tersebut menjadi bentuk data matrik yang sesuai dengan persamaan umum dari model linier yaitu, $\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$.
4. Dari data yang telah ditransformasi dalam bentuk matrik – matrik di atas maka selanjutnya kita masukkan data tersebut dalam proses iterasi GEE yang telah ada pada software R, namun sebelumnya diterapkan data tersebut dalam fungsi GLM sebagai langkah awal perbandingan jika faktor korelasi diabaikan.
5. Langkah terakhir adalah menjawab semua permasalahan yang diungkapkan sesuai dengan tujuan penelitian ini.

Dalam pemodelan, disini kita asumsikan $\phi(\alpha)$ diketahui, sehingga kita gunakan bentuk *GEE* untuk estimasi β yaitu :

$$S(\beta) = \sum_{i=1}^{m} \left(\frac{\partial \eta_i}{\partial \beta} \right)' Var(Y_i)^{-1} (Y_i - \mu_i) = 0$$

Dalam pemodelan GEE terdapat beberapa struktur matrik varian kovarian yang dapat digunakan antara lain, struktur independen, struktur tidak berstruktur, struktur *exchangeable*, struktur *autoregressive*, dan struktur dependen.

Sedangkan struktur varian-kovarian yang mungkin digunakan dalam mengestimasi parameter adalah dalam bentuk matrik yang disebut matrik ***compound symmetry*** atau struktur *exchangeable* atau korelasi seragam dan struktur independen, yaitu matrik simetris varian-kovarian yang paling sederhana. Kedua struktur tersebut digunakan dengan pertimbangan bahwa tidak ada kondisi khusus hubungan antara variabel respon dan eksplanatori pada kasus dalam penelitian ini seperti adanya perubanhan terhadap waktu, bobot dan sebagainya. Bentuk matrik varian kovarian adalah sebagai berikut:

$$\mathrm{var}(y) = \sigma^2 \begin{bmatrix} 1 & \dots & \rho \\ \dots & \dots & \dots \\ \rho & \dots & 1 \end{bmatrix}$$ merupakan struktur *exchangeable* dan

$$\mathrm{var}(y) = \sigma^2 \begin{bmatrix} 1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & 1 \end{bmatrix}$$ merupakan struktur independen.

3.3. Metodologi Analisis.

3.3.1. Program Statistika R (Paket R).

3.3.1.1. Paket R.

Salah satu proses dari pemodelan adalah penggunaan alat bantu komputer untuk mengestimasi parameter – parameter yang ada serta untuk studi kasus ini, maka akan ditunjukkan bagaimana penggunaan program komputasi statistika pada salah satu paket komputer. Paket yang digunakan adalah R, yaitu sebuah sistem untuk statistika komputasi dan grafik.

Paket R adalah sebuah bahasa yang mirip dengan bahasa S yang dibuat oleh AT & T Bell Laboratories oleh Rick Beeker, John Chambers dan Allan Wilks. R adalah sebuah sistem untuk komputasi dan grafik statistika. R terdiri dari sebuah bahasa ditambah sebuah prosedur running dengan grafik, *debugger*, akses suatu fungsi, dan kemampuan untuk menjalankan program dalam file script.

3.3.1.2. Struktur Fungsi GEE Pada Paket R.

Struktur fungsi yang digunakan dalam paket R adalah

```
gee(formula, id,
    data, subset, na.action,
    R = NA, b = NA,
    tol = 0.001, maxiter = 25,
    family = gaussian, corstr = "independence",
    Mv = 1, silent = TRUE, contrasts = NULL,
    scale.fix = FALSE, scale.value = 1,
    v4.4compat = FALSE)
```

Keterangan detail tentang bagian – bagian fungsi GEE ini selengkapnya pada Lampiran 5.

Langkah – langkah dalam prosedur ini adalah sebagai berikut:

1. Memasukkan data pengamatan dalam bentuk matrik itu dalam paket R yang selanjutnya digunakan sebagai `data` pada program.
2. Selanjutnya kita tetapkan beberapa ketentuan yang dibutuhkan R dalam proses iterasi seperti `family`, `fungsi link`, `fungsi varian`, `id`, `toleransi` dan lainnya.
3. Langkah selanjutnya memasukkan semua informasi di atas pada prosedur GEE dalam R, yang kemudian kita jalankan (*run*) paket tersebut untuk memperoleh informasi yang kita butuhkan mengenai GEE tersebut.

3.3.2. Analisis Data.

3.3.2.1. Penyusunan Data

Proses pertama yang diperlukan dalam penelitian ini adalah menyusunan data yang dijadikan sebagai salah satu contoh kasus suatu variabel respon yang berkorelasi, yaitu data dari artikel yang ditulis oleh Alan Agresti dan I-Ming Liu (1999) tentang data sampel dari 262 peternak dari peternakan – peternakan Kansas, dimana data yang ada merupakan data biner untuk masing – masing respon. Sehingga dari data tersebut perlu dimodifikasi atau disusun agar dapat diterjemahkan sebagai input program yang sesuai dengan paket R, langkah – langkah penyusunan data adalah :

1. Data pada Tabel 1 dicacah dalam bentuk variabel – variabel yang sesuai dengan kondisi penelitian. Dimana, yang merupakan variabel respon (Y) adalah sumber – sumber informasi dan variabel explanatori (X) adalah jenis – jenis level pendidikan.

   $\mathbf{Y} = \{ y_1, y_2, y_3, y_4, y_5 \}$, y_1 = Sumber dari Konsultan

   y_2 = Sumber dari Dokter Hewan

   y_3 = Sumber dari Jasa Penerangan

   y_4 = Sumber dari Majalah

   y_5 = Sumber dari Perusahaan Makanan

   Dan

   $\mathbf{X} = \{ x_1, x_2, x_3, x_4, x_5 \}$ x_1 = Level *High School*

   x_2 = Level *Vocational*

   x_3 = Level *2 – year college*

   x_4 = Level *4 – year college*

   x_5 = Level *Others*

2. Dalam kasus ini terdapat 5 kategori respon yang tidak bebas. Dimana tiap subyek dapat memilih lebih dari satu sumber informasi dan tiap sumber informasi memiliki 2 kemungkinan jawaban, yaitu "Ya" dan "Tidak". Untuk jawaban "Ya" sesuai dengan bilangan biner dinotasikan dengan angka 1 atau memiliki nilai 1 dan untuk jawaban "Tidak" dinotasikan dengan angka 0 atau

memiliki nilai 0. Maka, pada variabel respon (**Y**) terdapat 2^5 kombinasi kategori respon yang mungkin dan bersesuaian dengan level – level pendidikan, sehingga terdapat 32 kombinasi y_i (i = 1,2,3,4,5) untuk tiap level pendidikan (x_i, i = 1,2,3,4,5), kombinasi kategori respon dan bentuk data variabel explanatori ditunjukkan pada Lampiran 1.

Pada atiap level pendidikan memiliki kombinasi pemilihan jawaban (0 atau 1) yang sama, dimana penyebaran jumlah subyek untuk tiap variabel explanatori (level pendidikan) digambarkan pada Lampiran 1.

Total seluruh subyek sama dengan nilai atau jumlah subyek pada tabel kontingensi (Tabel 1), selanjutnya dari data di atas, digunakan sebagai data masukan dalam komputasi menggunakan software Paket R.

3. Data yang telah ditransformasi dalam bentuk matrik baik variabel respon maupun variabel explanatori dimasukkan dalam struktur perintah yang sesuai dengan paket R (contoh pada Lampiran 5). Struktur perintahnya adalah :

```
Skripsi <- structure(.Data=list(id=c(entry
data kode respon),
Jawaban=c(entry data variabel respon,
X1=c(entry data variabel X1),
X2=c(entry data variabel X2),
X3=c(entry data variabel X3),
X4=c(entry data variabel X4),
X5=c(entry data variabel
X5)),row.names=c(entry urutan data),class=
"data.frame")
```

Keterangan :

- `skripsi` = merupakan nama dari data, yaitu " skripsi " yang selanjutnya terbaca sebagai obyek dalam R.
- id = merupakan identifikasi dari kategori pada variabel respon (Y).
- jawaban = merupakan entry data kombinasi variabe respon yang tersusun manjadi satu kolom, sehingga data ini terbentuk menjadi 160 baris.
- X_1 = merupakan identifikasi entry data untuk variabel explanatori pada level High School, dimana data ini tersusun menjadi satu kolom atau berupa matrik 160 x 1 atau $(x_{11}, x_{12}, x_{13}, x_{14}, x_{15})^T$.
- X_2 = entry data untuk level Vocational, dimana strukturnya sama dengan X_1.

- X_3 = entry data untuk level 2-Year College , dimana strukturnya sama dengan X_1.
- X_4 = entry data untuk level 4-Year College, dimana strukturnya sama dengan X_1.
- X_5 = entry data untuk level Other, dimana strukturnya sama dengan X_1.
- Row.names = merupakan penamaan urutan data ("1",….,"160")

Script program data tersebut ditulis pada Notepad yang selanjutnya kita identifikasi sebagai obyek dalam R, langkah – langkah identifikasi ini digambarkan pada Lampiran 2.

3.3.2.2. Analisis Data Dengan Software R.

Setelah melalui proses pembentukan data, yaitu hasil transformasi data lapangan ke bentuk data yang sesuai dengan prosedur dalam R. Proses selanjutnya adalah menganalisis data dengan menggunakan prosedur GLM dan GEE yang sudah ada dalam R, yaitu dengan memasukkan perintah sesuai dengan format fungsi GLM dan GEE pada Window Console R. Terlebih dahulu dilakukan pengestimasian parameter menggunakan fungsi GLM. Hal ini dilakukan karena, dalam fungsi GLM menggunakan fungsi GLM. Hal ini dilakukan karena, dalam fungsi GLM kita dapat menentukan/ mengetahui model terbaik dengan melihat nilai AIC (*Akaike's Information Criteria*)(Chambers dan Hastie,1997), yaitu

$$AIC = D + 2p\phi$$

Dengan D adalah devian, p adalah derajat bebas kecocokan dan ϕ adalah penduga parameter dispersi.

Dalam GEE kita tidak dapat mengetahui model mana yang terbaik. Model terbaik memiliki nilai AIC yang terendah, selanjutnya dapat dibentuk suatu model $f(x, \beta)$.

Format fungsi GLM adalah sebagai berikut :

```
> glm(jawaban ~ x1+x2+x3+x4+x5, data=skripsi,
  family=binomial)
```

dan untuk menspesifikasi koefisien model digunakan format GLM sebagai berikut

```
> glm(jawaban ~ x1+x2+x3+x4+x5-1, data=skripsi,
  family=binomial)
```

Selain itu digunakan suatu pembobot model (*weighted*) dalam fungsi GLM, dimana format fungsi adalah

```
> summary(glm(jawaban~x1+x2+x3+x4+x5,
  data=skripsi, family= binomial,weights = id))
```


dan

```
> summary(glm(jawaban~x1+x2+x3+x4+x5-1,
  data=skripsi, family= binomial,weights = id))
```

Bentuk atau format fungsi GEE adalah sebagai berikut :

```
> gee(jawaban ~ x1+x2+x3+x4+x5, id=id,
  data=skripsi, family=binomial,
  corstr="exchangeable")
```

Bentuk matrik korelasi yang digunakan adalah *Exchangeable*, yaitu bentuk matrik korelasi yang homogen. Untuk menampilkan summary dengan menambahkan perintah tersebut pada perintah GEE.

```
> summary(gee(jawaban ~ x1+x2+x3+x4+x5, id=id,
  data=skripsi,family=binomial,
  corstr="exchangeable"))
```

Untuk menspesifikasikan koefisien model maka digunakan format perintah :

```
> summary(gee(jawaban ~ x1+x2+x3+x4+x5 - 1,
  id=id, data=skripsi, family=binomial,
  corstr="exchangeable"))
```

Format fungsi dengan menggunakan bentuk Korelasi "Independence" adalah


```
> summary(gee(jawaban ~ x1+x2+x3+x4+x5, id=id,
  data=skripsi, family=binomial,
  corstr="independence"))
```

Untuk menspesifikasikan koefisien output model maka digunakan format perintah :

```
> summary(gee(jawaban ~ x1+x2+x3+x4+x5 - 1,
  id=id, data=skripsi, family=binomial,
  corstr="independence"))
```

Langkah – langkah analisis data menggunakan Paket R selengkapnya terdapat pada Lampiran 2 dan script fungsi dan output program GLM terdapat pada Lampiran 3, dan script fungsi dan output program GEE terdapat pada Lampiran 4.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan.

Dari pembahasan pada bab sebelumnya dapat diambil kesimpulan sebagai berikut :

1. Dalam pemodelan suatu kasus atau data berkategori dengan variabel respon tidak bebas digunakan prosedure GEE. pada data kasus dari 262 peternak Kansas tentang pemilihan kategori sumber informasi yang dilakukan pada 5 level pendidikan; pembentukan model dilakukan dengan prosedure GEE dengan menggunakan model korelasi *exchangeable.*
2. Model matematis menggunakan metode GEE yang merupakan hasil estimasi bentuk GEE adalah sebagai berikut :

$$logit\ Pr\left(Y_{ij}=1\right) = 0.1046X_{i1} + 0.1754X_{i2} + 0.0574X_{i3} - 0.1364X_{i4} + 0.106X_{i5}$$

dengan matrik korelasi R(α) adalah sebagai berikut

$$R(\alpha) = \begin{bmatrix} 1 & -0.30125 & -0.30125 & \ldots\ldots -0.30125 \\ -0.30125 & 1 & -0.30125 & \ldots\ldots -0.30125 \\ -0.30125 & -0.30125 & 1 & \ldots\ldots -0.30125 \\ \vdots & \vdots & \vdots & \vdots \\ -0.30125 & -0.30125 & -0.30125 & \ldots\ldots \quad 1 \end{bmatrix}$$

yang konvergen pada iterasi ke 5, dengan penduga skala parameter (ϕ) adalah 1,033476. Pemilihan model yang lebih baik diantara struktur korelasi *exchangeable* (korelasi seragam) dan *independence* (korelasi independen) didasarkan pada panjang interval kepercayaan dan nilai probabilitas koefisien model dari kedua struktur korelasi tersebut.

3. Dalam pembetukan model GEE ini dilakukan dengan alat bantu komputer, yaitu paket R. inferensi output paket R meliputi koefisien model, *naïve SE, Naïve Z, Robust SE, robust Z,* dan *scale parameter. Naïve S.E* merupakan standar kesalahan dalam penduga *naïve* menggunakan penduga *log-likelihood* pada varian β, *Naïve z* merupakan nilai Z untuk penduga *Naïve* pada varian β. *Robust S.E* merupakan standar kesalahan dalam penduga *Robust* menggunakan penduga kuadrat terkecil yang terboboti pada varian β, *Robust z* merupakan nilai Z untuk penduga *Robust* pada varian β. Parameter Skala (ϕ) adalah sebuah skalar atau sebuah matrik yang diduga dari elemen matrik varian kovarian.

4. Peggunaan GLM mengasumsikan bahwa korelasi pada variabel respon diabaikan, model stokastik yang dihasilkan berdasarkan inferensi output dari paket R. Nilai probabilitas koefisien model GEE dengan korelasi seragam lebih kecil jika dibandingkan dengan nilai probabilitas pada GLM. Sehingga, jelas bahwa penggunaan GEE pada

data dengan variabel respon tidak bebas adalah tepat dan lebih baik dari pada digunakan GLM dalam pembentukan model matematis dari suatu data berkategori.

5.2. Saran.

Beberapa saran – saran bagi peneliti lain, yaitu :

1. Peneliti lain diharapkan dapat menemukan dan menggunakan kasus – kasus lain yang melibatkan variabel berkategori sehingga prosedure penyelesaian dengan menggunkan GEE lebih bervariasi, terutama dalam menggunakan struktur korelasi.
2. Diharapkan peneliti lain dapat menggunakan Paket Komputer Statistika yang lain untuk menganalisis data dengan prosedure GEE, selain paket R yang digunakan penulis.

DAFTAR PUSTAKA

Agresti. A. (1990). *Categorical Data Analysis.* AWiley-Interscience Publication.

Agresti, A dan I-Ming Liu. (1999), *Modeling a Categorical Variable Allowing Arbitrarily Many Category Choice,* Biometrics 55, 936-934.

Battacharyya, G.K dan johnson, R.A. (1970). *Statistical Concepts and Methods*, John Wiley & Sons.

Cox, D.R dan Snell, E.J. (1989). *Analysis of Binary Data,* Monograph on Statistics and Applied Probability, Chapman and Hall.

Chambers, J.M dan Hastie, T.J. (1997). *Statistical Models In S*, Chapman & Hall Computer Science Series, AT&T Bell Laboratories.

Diggle *et. al.* (1994). Analysis *of Longitudinal Data,* Oxford Science Publication.

Dobson, A.J. (1983). *Introduction To Statistical Modelling.* London, Chapman & Hall. 1st Edition.

Kenward. K.M. and Smith. D.M. (1995). *Computing the generalized estimating equations for repeated measurement.* Genstat Newsletter.

Montegomery, D.c. dan Pect. E.A. (1992), *Introduction To Linier regression Analyisis*, John Wiley & Sons.NIC. 2nd Edition.

McCullagh. P. danNelder, J.A. (1990). *Generalized Linier Models*, London : Chapman and Hall, 2nd Edition.

Ratih, R. (2000). Inferensi pada *Model – Model Linier Tergeneralisasi, Majalah Matematika dan Statistika*, Volume 1. FMIPA Universitas Jember.

Ripley *et. al.* (1997). *Software R* , GNU General Public License (www.r-project.org).

Tirta. I.M. (2000). *Pemodelan Statistika. Diktat perkuliahan, Laboratorium Statistika, Jurusan Matematika*, FMIPA, Universitas Jember.

Tirta I.M. (2001). *Generalized Linier Models, Diktat perkuliahan Laboratorium Statistika*, Jurusan Matematika, FMIPA, Universitas Jember.

LAMPIRAN 1

STRUKTUR DATA

STRUKTUR DATA

Struktur data yang digunakan ini adalah struktur data yang sesuai dengan data kasus yang menjadi obyek permasalahan dalam penelitian ini. Struktur kombinasi jawaban pada tiap ketegori dalam variabel respon adalah sebagai berikut :

Tabel 1. Kombinasi Jawaban Untuk Tiap Kategori Respon

NO	Y1	Y2	Y3	Y4	Y5
1.	1	1	1	1	1
2.	1	1	1	1	0
3.	1	1	1	0	1
4.	1	1	0	1	1
5.	1	0	1	1	1
6.	0	1	1	1	1
7.	1	1	1	0	0
8.	1	1	0	0	1
9.	1	0	0	1	1
10.	0	0	1	1	1
11.	1	1	0	1	0
12.	1	0	1	1	0
13.	0	1	1	1	0
14.	1	1	0	0	0
15.	1	0	0	0	1
16.	0	0	0	1	1
17.	1	0	0	1	0
18.	0	0	1	1	0
19.	1	0	1	0	0
20.	0	1	1	0	0
21.	1	0	0	0	0
22.	0	1	0	0	0
23.	0	0	1	0	0
24.	0	0	0	1	0
25.	0	0	0	0	1
26.	0	1	0	1	0
27.	1	0	1	0	1
28.	0	1	1	0	1
29.	0	1	0	1	1
30.	0	1	0	0	1
31.	0	0	1	0	1
32.	0	0	0	0	0

Struktur data pada tiap level pendidikan dari variabel pendidikan adalah sebagai berikut :

Tabel 2.

Sebaran Data Jumlah Subyek Untuk Level Pendidikan *High School*

NO	Y1	JML	Y2	JML	Y3	JML	Y4	JML	Y5	JML	TOT
1.	**1**	7	**1**	7	**1**	7	**1**	7	**1**	7	7
2.	**1**	0	**1**	0	**1**	0	**1**	0	**0**	0	0
3.	**1**	0	**1**	0	**1**	0	**0**	0	**1**	0	0
4.	**1**	2	**1**	2	**0**	0	**1**	2	**1**	2	2
5.	**1**	0	**0**	0	**1**	0	**1**	0	**1**	0	0
6.	**0**	0	**1**	3	**1**	3	**1**	3	**1**	3	3
7.	**1**	0	**1**	0	**1**	0	**0**	0	**0**	0	0
8.	**1**	1	**1**	1	**0**	0	**0**	0	**1**	0	1
9.	**1**	0	**0**	0	**0**	0	**1**	0	**1**	0	2
10.	**0**	0	**0**	0	**1**	3	**1**	3	**1**	3	3
11.	**1**	1	**1**	1	**0**	0	**1**	0	**0**	0	1
12.	**1**	0	**0**	0	**1**	0	**1**	0	**0**	0	0
13.	**0**	0	**1**	2	**1**	2	**1**	2	**0**	0	2
14.	**1**	1	**1**	1	**0**	0	**0**	0	**0**	0	1
15.	**1**	2	**0**	0	**0**	0	**0**	0	**1**	2	2
16.	**0**	0	**0**	0	**0**	0	**1**	3	**1**	3	3
17.	**1**	2	**0**	0	**0**	0	**1**	2	**0**	0	2
18.	**0**	0	**0**	0	**1**	2	**1**	2	**0**	0	2
19.	**1**	0	**0**	0	**1**	0	**0**	0	**0**	0	0
20.	**0**	0	**1**	2	**1**	2	**0**	0	**0**	0	2
21.	**1**	3	**0**	0	**0**	0	**0**	0	**0**	0	3
22.	**0**	0	**1**	10	**0**	0	**0**	0	**0**	0	10
23.	**0**	0	**0**	0	**1**	3	**0**	0	**0**	0	3
24.	**0**	0	**0**	0	**0**	0	**1**	19	**0**	0	19
25.	**0**	0	**0**	0	**0**	0	**0**	0	**1**	9	9
26.	**0**	0	**1**	2	**0**	0	**1**	2	**0**	0	2
27.	**1**	0	**0**	0	**1**	0	**0**	0	**1**	0	0
28.	**0**	0	**1**	3	**1**	3	**0**	0	**1**	3	3
29.	**0**	0	**1**	0	**0**	0	**1**	2	**1**	2	2
30.	**0**	0	**1**	2	**0**	0	**0**	0	**1**	2	2
31.	**0**	0	**0**	0	**1**	4	**0**	0	**1**	4	4
32.	**0**	0	**0**	0	**0**	0	**0**	0	**0**	0	0
	x_{11}	19	x_{12}	38	x_{13}	29	x_{14}	47	x_{15}	40	88

Tabel 3.

Sebaran Data Jumlah Subyek Untuk Level Pendidikan *Vocational*

NO	Y1	JML	Y2	JML	Y3	JML	Y4	JML	Y5	JML	TOT
1.	1	0	1	0	1	0	1	0	1	0	0
2.	1	1	1	1	1	1	1	1	0	0	1
3.	1	0	1	0	1	0	0	0	1	0	0
4.	1	0	1	0	0	0	1	0	1	0	0
5.	1	0	0	0	1	0	1	0	1	0	0
6.	0	0	1	0	1	0	1	0	1	0	0
7.	1	1	1	1	1	1	0	0	0	0	1
8.	1	0	1	0	0	0	0	0	1	0	0
9.	1	0	0	0	0	0	1	0	1	0	0
10.	0	0	0	0	1	1	1	1	1	1	1
11.	1	0	1	0	0	0	1	0	0	0	0
12.	1	0	0	0	1	0	1	0	0	0	0
13.	0	0	1	1	1	1	1	1	0	0	1
14.	1	0	1	0	0	0	0	0	0	0	0
15.	1	0	0	0	0	0	0	0	1	0	0
16.	0	0	0	0	0	0	1	0	1	0	0
17.	1	0	0	0	0	0	1	0	0	0	0
18.	0	0	0	0	1	1	1	1	0	0	1
19.	1	0	0	0	1	0	0	0	0	0	0
20.	0	0	1	1	1	1	0	0	0	0	1
21.	1	0	0	0	0	0	0	0	0	0	0
22.	0	0	1	1	0	0	0	0	0	0	1
23.	0	0	0	0	1	2	0	0	0	0	2
24.	0	0	0	0	0	0	1	4	0	0	4
25.	0	0	0	0	0	0	0	0	1	2	2
26.	0	0	1	0	0	0	1	0	0	0	0
27.	1	0	0	0	1	0	0	0	1	0	0
28.	0	0	1	0	1	0	0	0	1	0	0
29.	0	0	1	0	0	0	1	0	1	0	0
30.	0	0	1	1	0	0	0	0	1	1	1
31.	0	0	0	0	1	0	0	0	1	0	0
32.	0	0	0	0	0	0	0	0	0	0	0
	x_{21}	2	x_{22}	6	x_{23}	8	x_{24}	8	x_{25}	4	16

Tabel 4.

Sebaran Data Jumlah Subyek Untuk Level Pendidikan *2-Year College*

NO	Y1	JML	Y2	JML	Y3	JML	Y4	JML	Y5	JML	TOT
1.	1	0	1	0	1	0	1	0	1	0	0
2.	1	0	1	0	1	0	1	0	0	0	0
3.	1	0	1	0	1	0	0	0	1	0	0
4.	1	1	1	1	0	0	1	1	1	1	1
5.	1	0	0	0	1	0	1	0	1	0	0
6.	0	0	1	2	1	2	1	2	1	2	2
7.	1	0	1	0	1	0	0	0	0	0	0
8.	1	0	1	0	0	0	0	0	1	0	0
9.	1	0	0	0	0	0	1	0	1	0	0
10.	0	0	0	0	1	1	1	1	1	1	1
11.	1	0	1	0	0	0	1	0	0	0	0
12.	1	0	0	0	1	0	1	0	0	0	0
13.	0	0	1	1	1	1	1	1	0	0	1
14.	1	0	1	0	0	0	0	0	0	0	0
15.	1	0	0	0	0	0	0	0	1	0	0
16.	0	0	0	0	0	0	1	2	1	2	2
17.	1	0	0	0	0	0	1	0	0	0	0
18.	0	0	0	0	1	0	1	0	0	0	0
19.	1	0	0	0	1	0	0	0	0	0	0
20.	0	0	1	0	1	0	0	0	0	0	0
21.	1	0	0	0	0	0	0	0	0	0	0
22.	0	0	1	4	0	0	0	0	0	0	4
23.	0	0	0	0	1	4	0	0	0	0	4
24.	0	0	0	0	0	0	1	7	0	0	7
25.	0	0	0	0	0	0	0	0	1	3	3
26.	0	0	1	1	0	0	1	1	0	0	1
27.	1	0	0	0	1	0	0	0	1	0	0
28.	0	0	1	1	1	1	0	0	1	1	1
29.	0	0	1	2	0	0	1	2	1	2	2
30.	0	0	1	1	0	0	0	0	1	1	1
31.	0	0	0	0	1	1	0	0	1	1	1
32.	0	0	0	0	0	0	0	0	0	0	0
	x_{31}	1	x_{32}	13	x_{33}	10	x_{34}	17	x_{35}	14	31

Tabel 5.

Sebaran Data Jumlah Subyek Untuk Level Pendidikan *4-Year College*

NO	Y1	JML	Y2	JML	Y3	JML	Y4	JML	Y5	JML	TOT
1.	1	0	1	0	1	0	1	0	1	0	0
2.	1	0	1	0	1	0	1	0	0	0	0
3.	1	0	1	0	1	0	0	0	1	0	0
4.	1	2	1	2	0	0	1	2	1	2	2
5.	1	0	0	0	1	0	1	0	1	0	0
6.	0	0	1	0	1	0	1	0	1	0	0
7.	1	3	1	3	1	3	0	0	0	0	3
8.	1	0	1	0	0	0	0	0	1	0	0
9.	1	2	0	0	0	0	1	2	1	2	2
10.	0	0	0	0	1	4	1	4	1	4	4
11.	1	2	1	2	0	0	1	2	0	0	2
12.	1	2	0	0	1	2	1	2	0	0	2
13.	0	0	1	3	1	3	1	3	0	0	3
14.	1	2	1	2	0	0	0	0	0	0	2
15.	1	0	0	0	0	0	0	0	1	0	0
16.	0	0	0	0	0	0	1	3	1	3	3
17.	1	2	0	0	0	0	1	2	0	0	2
18.	0	0	0	0	1	4	1	4	0	0	4
19.	1	0	0	0	1	0	0	0	0	0	0
20.	0	0	1	0	1	0	0	0	0	0	0
21.	1	1	0	0	0	0	0	0	0	0	1
22.	0	0	1	17	0	0	0	0	0	0	17
23.	0	0	0	0	1	19	0	0	0	0	19
24.	0	0	0	0	0	0	1	29	0	0	29
25.	0	0	0	0	0	0	0	0	1	13	13
26.	0	0	1	0	0	0	1	0	0	0	0
27.	1	3	0	0	1	3	0	0	1	3	3
28.	0	0	1	0	1	0	0	0	1	0	0
29.	0	0	1	0	0	0	1	0	1	0	0
30.	0	0	1	0	0	0	0	0	1	0	0
31.	0	0	0	0	1	2	0	0	1	2	2
32.	0	0	0	0	0	0	0	0	0	0	0
	x_{41}	19	x_{42}	29	x_{43}	40	x_{44}	53	x_{45}	29	113

Tabel 6.

Sebaran Data Jumlah Subyek Untuk Level Pendidikan *Others*

NO	Y1	JML	Y2	JML	Y3	JML	Y4	JML	Y5	JML	TOT
1.	1	0	1	0	1	0	1	0	1	0	0
2.	1	1	1	1	1	1	1	1	0	0	1
3.	1	0	1	0	1	0	0	0	1	0	0
4.	1	0	1	0	0	0	1	0	1	0	0
5.	1	0	0	0	1	0	1	0	1	0	0
6.	0	0	1	0	1	0	1	0	1	0	0
7.	1	0	1	0	1	0	0	0	0	0	0
8.	1	0	1	0	0	0	0	0	1	0	0
9.	1	0	0	0	0	0	1	0	1	0	0
10.	0	0	0	0	1	1	1	1	1	1	1
11.	1	0	1	0	0	0	1	0	0	0	0
12.	1	0	0	0	1	0	1	0	0	0	0
13.	0	0	1	0	1	0	1	0	0	0	1
14.	1	0	1	0	0	0	0	0	0	0	0
15.	1	0	0	0	0	0	0	0	1	0	0
16.	0	0	0	0	0	0	1	2	1	2	2
17.	1	0	0	0	0	0	1	0	0	0	0
18.	0	0	0	0	1	1	1	1	0	0	1
19.	1	1	0	0	1	1	0	0	0	0	1
20.	0	0	1	1	1	1	0	0	0	0	1
21.	1	0	0	0	0	0	0	0	0	0	0
22.	0	0	1	2	0	0	0	0	0	0	2
23.	0	0	0	0	1	1	0	0	0	0	1
24.	0	0	0	0	0	0	1	1	0	0	1
25.	0	0	0	0	0	0	0	0	1	1	1
26.	0	0	1	0	0	0	1	0	0	0	0
27.	1	1	0	0	1	1	0	0	1	1	1
28.	0	0	1	0	1	0	0	0	1	0	0
29.	0	0	1	0	0	0	1	0	1	0	0
30.	0	0	1	0	0	0	0	0	1	0	0
31.	0	0	0	0	1	1	0	0	1	1	1
32.	0	0	0	0	0	0	0	0	0	0	0
	x_{51}	3	x_{52}	4	x_{53}	8	x_{54}	6	x_{55}	6	14

LAMPIRAN 2

SCRIPT DATA PEMODELAN

I. SCRIPT DATA KASUS

Data pada Lampiran 1 harus disusun dalam bentuk script program. Script program ini merupakan script program data yang ditulis pada software NOTEPAD pada WINDOWS, dan selanjutnya digunakan sebagai objek dalam Software R. Penulisan script program dan langkah – langkah pembuatannya adalah sebagai berikut :

1. Penulisan Script program untuk data kasus yang akan digunakan dalam orientasi pemodelan menggunakan GEE (*Generalized Estimating Equations*), dilakukan pada window NOTEPAD yang hampir sama cara penulisannya pada Microsoft Word.

```
dataskripsi - Notepad
File Edit Search Help
skripsi<-
structure(.Data=list(id=c(
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,
5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5),
jawaban = c(
1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0,
1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1,
1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0,
1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1, 0),
x1=c(
7, 0, 0, 2, 0, 3, 0, 1, 0, 3, 1, 0, 2, 1, 2, 3, 2, 2, 0, 2, 3, 10, 3, 19, 9, 2, 0, 3, 2, 2, 4, 0,
7, 0, 0, 2, 0, 3, 0, 1, 0, 3, 1, 0, 2, 1, 2, 3, 2, 2, 0, 2, 3, 10, 3, 19, 9, 2, 0, 3, 2, 2, 4, 0,
7, 0, 0, 2, 0, 3, 0, 1, 0, 3, 1, 0, 2, 1, 2, 3, 2, 2, 0, 2, 3, 10, 3, 19, 9, 2, 0, 3, 2, 2, 4, 0,
7, 0, 0, 2, 0, 3, 0, 1, 0, 3, 1, 0, 2, 1, 2, 3, 2, 2, 0, 2, 3, 10, 3, 19, 9, 2, 0, 3, 2, 2, 4, 0,
7, 0, 0, 2, 0, 3, 0, 1, 0, 3, 1, 0, 2, 1, 2, 3, 2, 2, 0, 2, 3, 10, 3, 19, 9, 2, 0, 3, 2, 2, 4, 0),
x2=c(
0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 2, 4, 2, 0, 0, 0, 0, 1, 0, 0,
0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 2, 4, 2, 0, 0, 0, 0, 1, 0, 0,
0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 2, 4, 2, 0, 0, 0, 0, 1, 0, 0,
0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 2, 4, 2, 0, 0, 0, 0, 1, 0, 0,
0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 2, 4, 2, 0, 0, 0, 0, 1, 0, 0),
x3=c(
```


2. Script program data kasus yang telah ditulis dalam NOTEPAD disimpan dalam bentuk ekstensi TXT (ex. Dataskripsi.txt) agar dapat di *source* dari software R dan digunakan sebagai object data dalam R, yang selanjutnya menjadi data dalam prosedure fungsi GEE.

3. Script Program data kasus ini adalah sebagai berikut :

```
skripsi<-
structure(.Data=list(id=c(
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1,
2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
2, 2, 2, 2, 2, 2, 2, 2,
3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,
3, 3, 3, 3, 3, 3, 3, 3,
4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,
4, 4, 4, 4, 4, 4, 4, 4,
5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5,
5, 5, 5, 5, 5, 5, 5, 5),
jawaban = c(
1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0,
0, 0, 1, 0, 0, 0, 0, 0,
1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0,
0, 1, 0, 1, 1, 1, 0, 0,
1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0,
0, 0, 1, 1, 0, 0, 1, 1,
1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1,
0, 1, 0, 0, 1, 0, 0, 0,
1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,
1, 0, 1, 1, 1, 1, 1, 0),
x1=c(
7, 0, 0, 2, 0, 3, 0, 1, 0, 3, 1, 0, 2, 1, 2, 3, 2, 2, 0, 2, 3, 10, 3,
19, 9, 2, 0, 3, 2, 2, 4, 0,
7, 0, 0, 2, 0, 3, 0, 1, 0, 3, 1, 0, 2, 1, 2, 3, 2, 2, 0, 2, 3, 10, 3,
19, 9, 2, 0, 3, 2, 2, 4, 0,
7, 0, 0, 2, 0, 3, 0, 1, 0, 3, 1, 0, 2, 1, 2, 3, 2, 2, 0, 2, 3, 10, 3,
19, 9, 2, 0, 3, 2, 2, 4, 0,
7, 0, 0, 2, 0, 3, 0, 1, 0, 3, 1, 0, 2, 1, 2, 3, 2, 2, 0, 2, 3, 10, 3,
19, 9, 2, 0, 3, 2, 2, 4, 0,
7, 0, 0, 2, 0, 3, 0, 1, 0, 3, 1, 0, 2, 1, 2, 3, 2, 2, 0, 2, 3, 10, 3,
19, 9, 2, 0, 3, 2, 2, 4, 0),
x2=c(
0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 2, 4,
2, 0, 0, 0, 0, 1, 0, 0,
0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 2, 4,
2, 0, 0, 0, 0, 1, 0, 0,
0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 2, 4,
2, 0, 0, 0, 0, 1, 0, 0,
0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 2, 4,
2, 0, 0, 0, 0, 1, 0, 0,
0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 2, 4,
2, 0, 0, 0, 0, 1, 0, 0),
x3=c(
0, 0, 0, 1, 0, 2, 0, 0, 0, 1, 0, 0, 1, 0, 0, 2, 0, 0, 0, 0, 0, 4, 4, 7,
3, 1, 0, 1, 2, 1, 1, 0,
0, 0, 0, 1, 0, 2, 0, 0, 0, 1, 0, 0, 1, 0, 0, 2, 0, 0, 0, 0, 0, 4, 4, 7,
3, 1, 0, 1, 2, 1, 1, 0,
```

```
0, 0, 0, 1, 0, 2, 0, 0, 0, 1, 0, 0, 1, 0, 0, 2, 0, 0, 0, 0, 0, 4, 4, 7,
3, 1, 0, 1, 2, 1, 1, 0,
0, 0, 0, 1, 0, 2, 0, 0, 0, 1, 0, 0, 1, 0, 0, 2, 0, 0, 0, 0, 0, 4, 4, 7,
3, 1, 0, 1, 2, 1, 1, 0,
0, 0, 0, 1, 0, 2, 0, 0, 0, 1, 0, 0, 1, 0, 0, 2, 0, 0, 0, 0, 0, 4, 4, 7,
3, 1, 0, 1, 2, 1, 1, 0),
x4=c(
0, 0, 0, 2, 0, 0, 3, 0, 2, 4, 2, 2, 3, 2, 0, 3, 2, 4, 0, 0, 1, 17, 19,
29, 13, 0, 3, 0, 0, 0, 2, 0,
0, 0, 0, 2, 0, 0, 3, 0, 2, 4, 2, 2, 3, 2, 0, 3, 2, 4, 0, 0, 1, 17, 19,
29, 13, 0, 3, 0, 0, 0, 2, 0,
0, 0, 0, 2, 0, 0, 3, 0, 2, 4, 2, 2, 3, 2, 0, 3, 2, 4, 0, 0, 1, 17, 19,
29, 13, 0, 3, 0, 0, 0, 2, 0,
0, 0, 0, 2, 0, 0, 3, 0, 2, 4, 2, 2, 3, 2, 0, 3, 2, 4, 0, 0, 1, 17, 19,
29, 13, 0, 3, 0, 0, 0, 2, 0,
0, 0, 0, 2, 0, 0, 3, 0, 2, 4, 2, 2, 3, 2, 0, 3, 2, 4, 0, 0, 1, 17, 19,
29, 13, 0, 3, 0, 0, 0, 2, 0),
x5=c(
0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 2, 0, 1, 1, 1, 0, 2, 1, 1,
1, 0, 1, 0, 0, 0, 1, 0,
0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 2, 0, 1, 1, 1, 0, 2, 1, 1,
1, 0, 1, 0, 0, 0, 1, 0,
0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 2, 0, 1, 1, 1, 0, 2, 1, 1,
1, 0, 1, 0, 0, 0, 1, 0,
0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 2, 0, 1, 1, 1, 0, 2, 1, 1,
1, 0, 1, 0, 0, 0, 1, 0,
0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 2, 0, 1, 1, 1, 0, 2, 1, 1,
1, 0, 1, 0, 0, 0, 1, 0)),
row.names = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "11",
"12", "13", "14",
"15", "16", "17", "18", "19", "20", "21", "22", "23","24", "25", "26",
"27", "28", "29",
"30", "31", "32", "33", "34", "35","36", "37", "38", "39", "40", "41",
"42", "43", "44",
"45", "46", "47", "48", "49", "50", "51", "52", "53", "54", "55", "56",
"57", "58", "59",
"60", "61", "62", "63", "64", "65", "66", "67", "68", "69", "70", "71",
"72", "73", "74",
"75", "76", "77", "78", "79", "80", "81", "82", "83", "84", "85", "86",
"87", "88", "89",
"90", "91", "92", "93", "94", "95", "96", "97", "98", "99", "100",
"101", "102", "103",
"104", "105", "106", "107", "108", "109", "110", "111", "112", "113",
"114", "115",
"116", "117", "118", "119", "120", "121", "122", "123", "124", "125",
"126", "127", "128", "129", "130", "131", "132", "133", "134", "135",
"136", "137", "138", "139", "140", "141", "142", "143", "144", "145",
"146", "147", "148", "149", "150", "151", "152", "153", "154", "155",
"156", "157", "158", "159", "160"),
class="data.frame")
```

4. Nama script data adalah " dataskripsi.txt ", jawaban merupakan seluruh kombinasi jawaban dari kategori variabel respon (sumber informasi). X1 adalah data untuk level pendidikan "High School", X2 adalah data untuk level pendidikan "Vocational", X3 adalah data untuk level pendidikan "2-Year

College", X4 adalah data untuk level pendidikan "4-Year College", dan X5 adalah data untuk level pendidikan "Others". Id adalah data pengkategorian dari variabel respon.

II. PEMODELAN

Prosedure pemodelan ini menggunakan *software* R, yang merupakan salah satu *software* statistik untuk komputasi, grafik dan analisis data. Paket R ini secara mudah dapat di download dari internet dengan website http://www.r-project.org/.

Paket R adalah sebuah bahasa yang tidak berbeda dengan bahasa S yang dibuat oleh AT & T Bell Laboratories oleh Rick Beeker, John Chambers dan Allan Wilks. R adalah sebuah sistem untuk komputasi dan grafik statistika. R terdiri dari sebuah bahasa ditambah sebuah prosedure running dengan grafik, *debugger*, akses suatu fungsi, dan kemampuan untuk menjalankan program dalam file script.

Langkah awal dalam pemodelan adalah menganalisis data, yaitu dengan menggunakan script data sebagai data obyek dalam paket R. Data ini di input dari script program yang telah dibuat seperti pada lampiran 1 dengan menggunakan menu **File – Source R Code**, langkah langkah selanjutnya adalah memasukkan fungsi GLM dan GEE pada Window Console. Langkah ini merupakan langkah akhir dalam analisis data, dimana akan ditampilkan output dari prosedure GLM maupun GEE.

LAMPIRAN 3

FUNGSI DAN OUT PUT GLM

FUNGSI DAN OUTPUT GLM

```
R : Copyright 2000, The R Development Core Team
Version 1.1.1  (August 15, 2000)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type    "?license" or "?licence" for distribution details.

R is a collaborative project with many contributors.
Type    "?contributors" for a list.

Type    "demo()" for some demos, "help()" for on-line help, or
        "help.start()" for a HTML browser interface to help.
Type    "q()" to quit R.

[Previously saved workspace restored]

> source("C:/My Documents/skripsi/dataskripsi.txt")

> summary(glm(jawaban~x1+x2+x3+x4+x5,data=skripsi,family=binomial))

Call:
glm(formula = jawaban ~ x1 + x2 + x3 + x4 + x5, family = binomial,
    data = skripsi)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.389  -1.221   0.980   1.108   1.909

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.30081    0.22528   1.335    0.182
x1           0.02530    0.09055   0.279    0.780
x2           0.08195    0.37563   0.218    0.827
x3           0.05068    0.26242   0.193    0.847
x4          -0.09979    0.07277  -1.371    0.170
x5          -0.21655    0.31284  -0.692    0.489

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

```
    Null deviance: 221.78  on 159  degrees of freedom
Residual deviance: 212.49  on 154  degrees of freedom
AIC: 224.49

Number of Fisher Scoring iterations: 3


> summary(glm(jawaban~x1+x2+x3+x4+x5-1,data=skripsi,family=binomial))

Call:
glm(formula = jawaban ~ x1 + x2 + x3 + x4 + x5 - 1, family = binomial,
    data = skripsi)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3159  -1.1337   0.9836   1.1774   1.9090

Coefficients:
   Estimate Std. Error z value Pr(>|z|)
x1  0.06779    0.08633   0.785    0.432
x2  0.13192    0.36885   0.358    0.721
x3  0.05224    0.26100   0.200    0.841
x4 -0.11963    0.07123  -1.679    0.093 .
x5 -0.04897    0.28304  -0.173    0.863
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1


(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 221.81  on 160  degrees of freedom
Residual deviance: 214.28  on 155  degrees of freedom
AIC: 224.28

Number of Fisher Scoring iterations: 3
```



```
> summary(glm(jawaban~x1+x2+x3+x4+x5,data=skripsi,family=binomial,
  weights=id))

Call:
glm(formula = jawaban ~ x1 + x2 + x3 + x4 + x5, family = binomial,
    data = skripsi, weights = id)

Deviance Residuals:
   Min       1Q   Median       3Q      Max
-3.243   -1.842   1.035    1.871    3.372

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.13473    0.13016   1.035   0.3006
x1           0.04216    0.05211   0.809   0.4185
x2           0.10719    0.21436   0.500   0.6170
x3           0.40288    0.15912   2.532   0.0113 *
x4          -0.18127    0.04290  -4.226 2.38e-05 ***
x5          -0.07268    0.17994  -0.404   0.6863
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 665.35  on 159  degrees of freedom
Residual deviance: 633.81  on 154  degrees of freedom
AIC: 645.8

Number of Fisher Scoring iterations: 4


> summary(glm(jawaban~x1+x2+x3+x4+x5-1,data=skripsi,family=binomial,
  weights=id))

Call:
glm(formula = jawaban ~ x1 + x2 + x3 + x4 + x5 - 1, family = binomial,
    data = skripsi, weights = id)

Deviance Residuals:
   Min       1Q   Median       3Q      Max
-3.141   -1.811   1.060    1.938    3.267
```



```
Coefficients:
    Estimate Std. Error z value Pr(>|z|)
x1  0.060779   0.048983   1.241   0.2147
x2  0.128432   0.211784   0.606   0.5442
x3  0.398820   0.157587   2.531   0.0114 *
x4 -0.189453   0.041879  -4.524 6.07e-06 ***
x5  0.005134   0.162548   0.032   0.9748
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 665.42  on 160  degrees of freedom
Residual deviance: 634.88  on 155  degrees of freedom
AIC: 644.88

Number of Fisher Scoring iterations: 3
```

LAMPIRAN 4

FUNGSI DAN OUT PUT GEE

FUNGSI DAN OUTPUT GEE

```
R : Copyright 2000, The R Development Core Team
Version 1.1.1  (August 15, 2000)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type    "?license" or "?licence" for distribution details.

R is a collaborative project with many contributors.
Type    "?contributors" for a list.

Type    "demo()" for some demos, "help()" for on-line help, or
        "help.start()" for a HTML browser interface to help.
Type    "q()" to quit R.

[Previously saved workspace restored]

> source("C:/My Documents/skripsi/dataskripsi.txt")

> library(gee)

> summary(gee(jawaban~x1+x2+x3+x4+x5-
  1,id=id,data=skripsi,family=binomial,corstr="exchangeable"))

[1] "Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27"
[1] "running glm to get initial regression estimate"
[1]  0.06778558  0.13191587  0.05223651 -0.11962535 -0.04897059

 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Exchangeable

Call:
gee(formula = jawaban ~ x1 + x2 + x3 + x4 + x5 - 1, id = id,
    data = skripsi, family = binomial, corstr = "exchangeable")

Summary of Residuals:
       Min         1Q     Median         3Q        Max
-0.6202380 -0.5000000  0.3247016  0.4738721  0.8307881

Coefficients:
      Estimate Naive S.E.    Naive z Robust S.E.   Robust z
x1  0.10460689 0.08506626  1.2297107   0.0780630  1.3400317
x2  0.17538515 0.37284956  0.4703912   0.3612874  0.4854450
x3  0.05746314 0.27120549  0.2118805   0.4429712  0.1297221
x4 -0.13640135 0.07151431 -1.9073294   0.0977039 -1.3960685
x5  0.10595941 0.25684995  0.4125343   0.2585069  0.4098901
```

```
Estimated Scale Parameter:  1.033476
Number of Iterations:  5

Working Correlation
              [,1]        [,2]        [,3]    . . . .    [,32]
 [1,]  1.00000000 -0.03012539 -0.03012539 . . .  -0.03012539
 [2,] -0.03012539  1.00000000 -0.03012539 . . .  -0.03012539
 [3,] -0.03012539 -0.03012539  1.00000000 . . .  -0.03012539
   :        :           :           :        :         :
   :        :           :           :        :         :
[32,] -0.03012539 -0.03012539 -0.03012539           1.00000000


> summary(gee(jawaban~x1+x2+x3+x4+x5-
   1,id=id,data=skripsi,family=binomial,corstr="independence"))

[1] "Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27"
[1] "running glm to get initial regression estimate"
[1]  0.06778558  0.13191587  0.05223651 -0.11962535 -0.04897059

 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Independent

Call:
gee(formula = jawaban ~ x1 + x2 + x3 + x4 + x5 - 1, id = id,
    data = skripsi, family = binomial, corstr = "independence")

Summary of Residuals:
       Min         1Q     Median         3Q        Max
-0.5792569 -0.4741033  0.3835518  0.5000000  0.8383116

Coefficients:
      Estimate Naive S.E.    Naive z Robust S.E.   Robust z
x1  0.06778564 0.08775939  0.7724032  0.06680193  1.0147257
x2  0.13191596 0.37490336  0.3518666  0.37335458  0.3533262
x3  0.05223642 0.26527040  0.1969177  0.44143752  0.1183325
x4 -0.11962542 0.07240090 -1.6522642  0.10610531 -1.1274216
x5 -0.04897056 0.28767801 -0.1702270  0.24920389 -0.1965080

Estimated Scale Parameter:  1.032838
Number of Iterations:  1

Working Correlation
      [,1] [,2] [,3] [,4] [,5] [,6]        [,32]
 [1,]    1    0    0    0    0    0            0
 [2,]    0    1    0    0    0    0            0
 [3,]    0    0    1    0    0    0            0
 [4,]    0    0    0    1    0    0            0
 [5,]    0    0    0    0    1    0            0
 [6,]    0    0    0    0    0    1            0
   :     :    :    :    :    :    :     :      :
:        :    :    :    :    :    :     :      :
[32,]    0    0    0    0    0    0     0      1
```

```
> summary(gee(jawaban~x1+x2+x3+x4+x5,id=id,data=skripsi,family=binomial,
  corstr="exchangeable"))

[1] "Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27"
[1] "running glm to get initial regression estimate"
[1]  0.30080904  0.02530442  0.08194752  0.05067899 -0.09978641 -0.21654986

 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Exchangeable


Call:
gee(formula = jawaban ~ x1 + x2 + x3 + x4 + x5, id = id, data = skripsi,
    family = binomial, corstr = "exchangeable")

Summary of Residuals:
       Min         1Q     Median         3Q        Max
-0.6187958 -0.5253864  0.3812042  0.4584488  0.8383090


Coefficients:
               Estimate Naive S.E.    Naive z Robust S.E.   Robust z
(Intercept)  0.30119927 0.17097101  1.7616979   0.2063101  1.4599344
x1           0.02530457 0.09375581  0.2698987   0.0626019  0.4042140
x2           0.08194609 0.38889281  0.2107164   0.3963605  0.2067464
x3           0.05068565 0.27167906  0.1865644   0.4527306  0.1119554
x4          -0.09978321 0.07532753 -1.3246579   0.1189050 -0.8391842
x5          -0.21655641 0.32385651 -0.6686801   0.2830538 -0.7650716

Estimated Scale Parameter:  1.040031
Number of Iterations:  2

Working Correlation
              [,1]        [,2]        [,3] . . .          [,32]
 [1,]  1.00000000 -0.03036106 -0.03036106 . . .   -0.03036106
 [2,] -0.03036106  1.00000000 -0.03036106 . . .   -0.03036106
 [3,] -0.03036106 -0.03036106  1.00000000 . . .   -0.03036106
 [4,] -0.03036106 -0.03036106 -0.03036106 . . .    1.00000000
   :          :           :           :     :           :
   :          :           :           :     :           :

[32,] -0.03036106 -0.03036106 -0.03036106 . . .   1.00000000
```

```
> summary(gee(jawaban~x1+x2+x3+x4+x5,id=id,data=skripsi,family=binomial,
  corstr="independence"))

[1] "Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27"
[1] "running glm to get initial regression estimate"
[1]  0.30080904  0.02530442  0.08194752  0.05067899 -0.09978641 -0.21654986

 GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
 gee S-function, version 4.13 modified 98/01/27 (1998)
```

```
Model:
 Link:                        Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:      Independent

Call:
gee(formula = jawaban ~ x1 + x2 + x3 + x4 + x5, id = id, data = skripsi,
    family = binomial, corstr = "independence")

Summary of Residuals:
       Min         1Q     Median         3Q        Max
-0.6187025 -0.5252874  0.3812975  0.4585437  0.8383796


Coefficients:
               Estimate Naive S.E.    Naive z Robust S.E.   Robust z
(Intercept)  0.30080907 0.22975576  1.3092558   0.2048815  1.4682103
x1           0.02530446 0.09236716  0.2739551   0.0626079  0.4041736
x2           0.08194769 0.38312738  0.2138915   0.3963666  0.2067472
x3           0.05067898 0.26765422  0.1893450   0.4527248  0.1119421
x4          -0.09978648 0.07422173 -1.3444375   0.1188950 -0.8392823
x5          -0.21654992 0.31906647 -0.6786985   0.2830599 -0.7650321

Estimated Scale Parameter:  1.040032
Number of Iterations:  1

Working Correlation
        [,1] [,2] [,3] [,4] [,5] [,6]       [,32]
 [1,]     1    0    0    0    0    0          0
 [2,]     0    1    0    0    0    0          0
 [3,]     0    0    1    0    0    0          0
 [4,]     0    0    0    1    0    0          0
 [5,]     0    0    0    0    1    0          0
 [6,]     0    0    0    0    0    1          0
   :      :    :    :    :    :    :    :     :
   :      :    :    :    :    :    :    :     :
[32,]     0    0    0    0    0    0    0     1
```

LAMPIRAN 5

FUNGSI GEE DALAM R

Struktur fungsi yang digunakan dalam paket R adalah

```
gee(formula, id,
    data, subset, na.action,
    R = NA, b = NA,
    tol = 0.001, maxiter = 25,
    family = gaussian, corstr = "independence",
    Mv = 1, silent = TRUE, contrasts = NULL,
    scale.fix = FALSE, scale.value = 1, v4.4compat =
    FALSE)
```

Keterangan :

`formula` sebuah expresi rumus seperti model regresi lainnya, dalam bentuk `respon ~ predictors.`

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

`data` sebuah optimalitas data frame yeng menginterpretasikan variabel yang terdapat dalam `formula` bersama – sama dengan `id` dan `n` variabel.

`subset` expresi yang menyatakan bahwa bagian/ subset dari baris pada data yang harus digunakan dalam kecocokan. Hal ini dapat menjadi sebuah sebuah vektor numerik yang mengidentifikasikan jumlah observasi yang akan dimasukkan atau sebuah vektor karakter pada nama – nama baris yang akan dimasukkan. Semua observasi dimasukkan secara default.

`na.action`	sebuah fungsi untuk menyaring data yang hilang (missing data), hanya `na.omit` yang dapat digunakan.
`R`	sebuah matrik kuadrat pada dimensi ukuran cluster maksimum yang berisi penggunaan korelasi khusus.
`b`	sebuah inisial penduga parameter.
`tol`	toleransi yang digunakan dalam algoritma.
`maxiter`	maksimum iterasi
`family`	sebuah obyek keluarga (family), sebuah list fungsi dan expresi untuk pendefinisian fungsi link dan fungsi varian. Keluarga distribusi yang ada adalah `gaussian`, `binomial` `poisson`, `gamma` dan `quasi`. Beberapa link yang tidak terstruktur adalah `1/mu^2` dan `sqrt` serta fungsi varian invers gaussian.
`corstr`	sebuah karakter string khusus pada struktur korelasi, yaitu : `independence`, `fixed`, `stat_M_dep`, `non_stat_M_dep`, `exchangeable`, `AR-M`, dan `unstructured`.
`Mv`	jika `corstr` adalah `stat_M_dep`, `non_stat_M_ dep`, atau `AR-M`, maka `Mv` harus dispesifikasikan.
`silent`	sebuah variabel logis yang mengontrol apakah penduga parameter pada setiap iterasi dicetak.

`contrasts` sebuah daftar yang memberikan perbedaan untuk beberapa atau semua faktor yang muncul dalam model. Elemen pada daftar mempunyai nama yang sama seperti variabel dan harus seperti salah satu matrik contrast (dikhususkan pada beberapa rank-penuh dengan beberapa baris yang ada pada level dalam faktor) atau fungsi lain untuk menghitung seperti sebuah matrik yang memenuhi jumlah level.

`scale.fix` sebuah variabel logis; jika benar (T) skala parameter adalah fixed pada nilai `skala.value`

`scala.value` variabel numerik yang memberikan nilai yang mana skala parameter harus fixed, hanya digunakan jika `scale.fix = True`.

`v4.4compat` variabel logis yang mengharapkan compability pada parameter korelasi dengan versi sebelumnya.