Model Discrete Choice dan Regresi Logistik

Oleh :
Jaka Nugraha[1], Suryo Guritno[2], Sri Haryatmi Kartiko[2]
1. Jurusan Statistika, FMIPA UII. Email: jnugraha@fmipa.uii.ac.id
2. Jurusan Matematika FMIPA UGM

Abstrak

Setiap manusia, lembaga, perusahaan akan dihadapkan pada pilihan-pilihan dan harus memilih satu dari semua alternatif yang tersedia. Banyak hal/faktor yang digunakan untuk mengambil keputusan. Keputusan/pilihan diambil didasarkan pada harapan bahwa pilihan tersebut memberikan manfaat terbesar diantar pilihan yang ada. Dalam tulisan ini dibahas pemodelan untuk data respon polikotomis menggunakan dua pendekatan distribusi, yaitu distribusi nilai ekstrim dan distribusi logistik. Dari dua distribusi tersebut diperoleh model probabilitas yang sama.

Kata kunci : Model discrete Choice, Logit, Link

Discrete Choice Models and Logistic Regression

Abstract

Everyone, institution, and company will faced to several alternatives and must chose one of them. A lot of factor used to make a decision. The decision/choice is based on expectation that the choices will give the biggest advantage among the others. This paper will dicribe modeling for polychotomous response data by using two distributions approximation : extreme value and logistic distributions. Both of distribution give the same model of probability.

Key Words : Discrete Choice Models, Logit, link.

1. Pendahuluan

Setiap manusia, lembaga, perusahaan akan dihadapkan pada pilihan-pilihan. Kita harus memilih satu dari semua alternatif yang tersedia, mulai dari satu alternatif saja (tidak ada alternatif lain), dua alternatif sampai dengan banyak alternatif. Banyak hal/faktor yang digunakan untuk mengambil keputusan. Pertama, faktor internal yaitu sifat sifat atau karakteristik yang melekat pada diri pembuat keputusan. Kedua, faktor eksternal yaitu keadaan/ faktor yang berasal dari luar/lingkungan pembuat keputusan. Kedua faktor tersebut dapat bersifat kualitatif maupun kuantitatif.

Dipresentasikan dalam Seminar Nasional MIPA 2006 dengan tema **"Penelitian, Pendidikan, dan Penerapan MIPA serta Peranannya dalam Peningkatan Keprofesionalan Pendidik dan Tenaga Kependidikan"** yang diselenggarakan oleh Fakultas MIPA UNY, Yogyakarta pada tanggal 1 Agustus 2006

Sementara itu keputusan/pilihan dilakukan berdasarkan atas asas manfaat dan mudhorot (resiko) bagi pembuat keputusan. Suatu pilihan dilakukan jika pilihan itu paling menguntungkan dibandingkan dengan semua alternatif pilihan yang lain, atau bisa juga karena pilihan itu paling kecil resikonya dibanding dengan alternatif yang lain.

Disamping itu, kadang kita tertarik untuk menduga nilai respon yang dikaitkan dengan sekumpulan kovariate dengan menggunakan fungsi penghubung tertentu. Ketika respon kontinyu, umumnya dipilih fungsi penghubung identitas, sedangkan untuk respon biner dipilih fungsi penghubung *logit*. Terdapat perbedaan antara analisis regresi logistik ganda dengan analisis regresi ganda. Analisis regresi ganda digunakan pada distribusi normal, sedangkan analisis regresi logistik digunakan untuk data yang berdistribusi binomial.

Regresi logistik pada respon dikotomis tersebut biasa dinamakan regresi logistik biner. Sering kali respon mempunyai lebih dari dua alternatif jawaban misalkan tentang warna kesukaan (merah, kuning, hijau, dsb) atau contoh yang lain tentang penilaian (sangat baik, baik, cukup, buruk, sangat buruk). Pada data polikotomis tersebut dapat digunakan analisis regresi multinomial.

Dalam makalah ini akan dibahas pemodelan untuk data polikotomis menggunakan analisis regresi multinomial berdasarkan link logit dan model *discrete choice* yang didasarkan atas utilitis (manfaat) bagi pembuat keputusan (responden).

2. Model Multinomial

Model Regresi Logistik digunakan untuk menggambarkan hubungan antara variabel independen/prediktor dengan variabel dependen/respon yang berupa data biner/dikotomi [1]. Variabel responnya (Y), biasanya digunakan nilai 0 (gagal) dan 1 (sukses). Fungsi distribusi yang digunakan adalah distribusi logistik dengan notasi p= $\pi(X) = E(Y|X)$ untuk menyatakan mean bersyarat dari Y jika diberikan vektor kovariate $X_i = (x_1,x_2, \ldots x_p)^t$. Model regresi logistik dapat dinyatakan dalam bentuk

$$\pi(X_i) = \frac{\exp(X_i^t\beta)}{1+\exp(X_i^t\beta)} \tag{1}$$

dengan $\beta = (\beta_1, \ldots, \beta_p)^T$ adalah vektor parameter dan $0 \leq \pi(X) \leq 1$ dan i=1,2,...n.

Untuk menentukan model regresi, harga β ditaksir lebih dahulu dengan menggunakan Metode Kemungkinan Maksimum dan dilanjutkan dengan pengujian terhadap model (uji kecocokan model dan uji koefisien regresi) [2].

Pengembangan atas regresi logistik biner adalah regresi logistik multinomial, yaitu variabel dependen Y bernilai diskrit (1, 2, ...J). Jenis data untuk variabel Y dapat berbentuk data nominal maupun ordinal.

Variabel random Y_i bernilai diskrit dengan indeks 1, 2, …J.

$$\pi_{ij} = P\{Y_i = j\} \tag{2}$$

merupakan probabilitas responden ke-i memilih kategori ke-j untuk i=1,2...n dan j=1,2,...J. Diasumsikan antar pilihan adalah saling asing sehingga $\sum_{j=1}^{J} \pi_{ij} = 1$ untuk setiap i. Sehingga hanya dipunyai J-1 parameter. Dapat disusun distribusi multinomial

$$P\{Y_{i1}=y_{i1}, \ldots, Y_{iJ}=y_{iJ}\} = \binom{n_i}{y_{i1},\ldots,y_{iJ}} \pi_{i1}^{yi1} \ldots \pi_{iJ}^{yiJ} \tag{3}$$

Untuk menyusun model logit, dilakukan dengan cara salah satu kategori (biasanya kategori terakhir) dijadikan baseline, sehingga model logitnya adalah [3]

$$\eta_{ij} = \log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = X_{ij}^{t}\beta = \beta^{t} X_{ij} \tag{4}$$

untuk j=1,2,….,(J-1).

Teorema 1.

Model probabilitas responden/individu ke i memilih alternatif j adalah

$\pi_{ij} = \dfrac{\exp(\eta_{ij})}{\sum_{k=1}^{J} \exp(\eta_{ik})}$ untuk suatu i dan suatu j dimana i=1,2,…n dan j=1,2…,J

Bukti :

$$\eta_{ij} = \log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = X_{ij}^{t}\beta$$

$$\Leftrightarrow \frac{\pi_{ij}}{\pi_{iJ}} = \exp(\eta_{ij})$$

$$\Leftrightarrow \pi_{ij} = \pi_{iJ} \exp(\eta_{ij})$$

Karena $\sum_{j=1}^{J}\pi_{ij} = 1$ dan $\eta_{iJ} = 0$ maka

$$\sum_{j=1}^{J}\pi_{ij} = \sum_{j=1}^{J}\pi_{iJ}\exp(\eta_{ij}) = 1$$

$$\Leftrightarrow \sum_{j=1}^{J}\pi_{ij} = \pi_{iJ}\sum_{j=1}^{J}\exp(\eta_{ij}) = 1$$

$$\Leftrightarrow \pi_{iJ} = \frac{1}{\sum_{k=1}^{J}\exp(\eta_{ik})} \text{ dan } \pi_{ij} = \frac{\exp(\eta_{ij})}{\sum_{k=1}^{J}\exp(\eta_{ik})} \quad \textbf{(terbukti)}$$

Untuk menentukan model regresi, harga β ditaksir lebih dahulu dengan menggunakan Metode Kemungkinan Maksimum.

Teorema 2.

Misalkan suatu sampel terdiri n observasi dari pasangan observasi (X_i,y_i), i=1,…,n dan $Y_i=(y_{i1},y_{i2},...,y_{iJ})$ dengan y_{ij} =0,1 untuk j=1,2,...J. Model regresi logistik

$$\pi_{ij} = \frac{\exp(\eta_{ij})}{\sum_{k=1}^{J}\exp(\eta_{ik})}$$

penaksir $\beta = (\beta_1,\ldots,\beta_p)^t$ menggunakan metode kemungkinan maksimum adalah penyelesaian dari persamaan kemungkinan :

$$\sum_{i=1}^{n}[y_{ij} - \pi_{ij}] = 0 \text{ dan } \sum_{i=1}^{n}\sum_{j=1}^{J}X_{ij}[y_{ij} - \pi_{ij}] = 0$$

Bukti :

Fungsi likelihood dari sampel random berukuran n (n responden) adalah

$$L(\beta) = \prod_{i=1}^{n}\pi_{i1}^{y_{i1}}\ldots\pi_{iJ}^{y_{iJ}}$$

dengan y_{ij} = 1 jika responden i memilih j dan y_{ij}=0 jika responden i memilih selain j.

Log dari fungsi likelihood tersebut adalah

$$LL(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{J}y_{ij}\ln(\pi_{ij})$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{J} y_{ij}\left(X_{ij}^{t}\beta - \ln\left(\sum_{j=1}^{J}\exp(X_{ij}^{t}\beta)\right)\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{J} y_{ij} X_{ij}^{t}\beta - \sum_{i=1}^{n}\left(\ln\left(\sum_{j=1}^{J}\exp(X_{ij}^{t}\beta)\right)\sum_{j=1}^{J} y_{ij}\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{J} y_{ij} X_{ij}^{t}\beta - \sum_{i=1}^{n}\ln\left(\sum_{j=1}^{J}\exp(X_{ij}^{t}\beta)\right)$$

Derivatif LL(β) terhadap β adalah

$$\frac{\partial LL(\beta)}{\partial\beta} = \sum_{i=1}^{n}\sum_{j=1}^{J} X_{ij} y_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{J} X_{ij}\left(\frac{\exp(X_{ij}^{t}\beta)}{\sum_{j=1}^{J}\exp(X_{ij}^{t}\beta)}\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{J} X_{ij}\left[y_{ij} - \pi_{ij}\right] = 0 \quad \textbf{(terbukti)}$$

Penyelesaian dari persamaan penaksir dari teorema 2 dapat digunakan iterasi dengan menggunakan metode newton raphson dan metode scoring.

3. Model *Discrete Choice* (Pemilihan Diskrit)

Model pemilihan diskrit menggambarkan pembuat keputusan memilih diantara alternatif yang tersedia. Pembuat keputusan dapat berupa orang, rumah tangga, perusahaan atau unit pembuat keputusan yang lain. Himpunan semua pilihan/alternatif disebut *Choice set*. Model pemilihan diskrit digunakan untuk menguji pilihan "yang mana", sedangkan model regresi dipakai untuk menguji "berapa banyak". Walaupun demikian seringkali model pemilihan diskrit juga dapat dipakai untuk menguji "berapa banyak". Model pemilihan diskrit biasanya diturunkan dibawah asumsi manfaat maksimum oleh pembuat keputusan [4].

Seorang pembuat keputusan dinotasikan dengan i, yang berhadapan dengan pilihan sebanyak J anternatif. Pembuat keputusan mempunyai tingkat utiliti (keuntungan) untuk setiap alternatif. Misalkan U_{ij} untuk j=1,…,J adalah utiliti pembuat keputusan (responden) i jika memilih alternatif j. Nilai U_{ij} yang sesungguhya tidak diketahui oleh pengamat (peneliti). Tentunya pembuat keputusan memilih alternatif yang mempunyai utiliti terbesar, sehingga memilih alternatif k jika dan hanya jika $U_{ik} > U_{ij}\ \forall j \neq k$.

Peneliti tidak mengetahui nilai utiliti untuk pembuat keputusan terhadap setiap alternatif. Peneliti hanya mengamati atribut yang ada untuk masing-masing alternatifnya, yang dinotasikan dengan x_{kj} $\forall j$ dan atribut pembuat keputusan yang dinotasikan dengan s_i. Secara fungsi dapat dinotasikan sebagai $V_{ij}= V(x_{ij},s_i)$ $\forall j$ yang biasa dinamakan *representative utility*. Karena nilai utiliti yang sesungguhnya tidak diketahui peneliti maka

$$V_{ij} \neq U_{ij} \text{ dan } U_{ij} = V_{ij} + \varepsilon_{ij}$$

$\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ})$ adalah variabel random yang mempunyai densitas $f(\varepsilon_i)$.

Probabilitas pembuat keputusan i memilih alternatif k dapat dinyatakan sebagai [4]

$$\begin{aligned} P_{ik} &= \Pr(U_{ik} > U_{ij} \ \forall j \neq k) \\ &= \Pr(\varepsilon_{ij} - \varepsilon_{ik} < V_{ik} - V_{ij} \ \forall j \neq k) \\ &= \int_{\varepsilon} I(\varepsilon_{ij} - \varepsilon_{ik} < V_{ik} - V_{ij} \forall j \neq k) f(\varepsilon_i) d\varepsilon_i \end{aligned} \tag{5}$$

I(.) adalah fungsi indikator, yang bernilai 1 jika pernyataan dalam kurung benar dan bernilai 0 jika pernyataan salah. Selanjutnya dapat dipilih atau ditentukan densitas $f(\varepsilon_{ik})$ yang sesuai/tepat, misalnya distribusi nilai ekstrim dan biasa disebut dengan model logit.

Model Logit diturunkan dengan asumsi bahwa ε_{ik} berdistribusi nilai ekstrim (extreme value) yang saling independen untuk semua i. Fungsi densitas extreme value (Gumbel) adalah

$$f(\varepsilon_{ij}) = \varepsilon^{-\varepsilon_{ij}} \varepsilon^{-\varepsilon^{-\varepsilon_{ij}}} \tag{6}$$

dan distribusi kumulatifnya adalah

$$F(\varepsilon_{nj}) = \varepsilon^{-\varepsilon^{-\varepsilon_{nj}}} \tag{7}$$

Variansi dari distribusi ini adalah $\pi^2/6$.

Teorema 3.

Jika $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ})$ adalah variabel random yang mempunyai densitas *extreme value* maka probabilitas pembuat keputusan i memilih alternatif k adalah

$$P_{ik} = \frac{\exp(V_{ik})}{\sum_j \exp(V_{ij})} \text{ untuk k=1,2,...J}$$

Bukti :

Probabilitas pembuat keputusan i memilih alternatif k yang dinyatakan sbb :

$$P_{ik} = \Pr(\varepsilon_{ij} - \varepsilon_{ik} < V_{ik} - V_{ij}\ \forall j\neq k)$$

$$= \Pr(\varepsilon_{ij} < \varepsilon_{ik} + V_{ik} - V_{ij}\ \forall j\neq k)$$

Jika ε_{ik} diketahui dan saling independen, maka

$$P_{ik}|\varepsilon_{ik} = \prod_{j\neq k} \exp(-\exp(-(\varepsilon_{ik} + V_{ik} - V_{ij})))$$

Oleh karena nilai ε_{ik} tidak diketahui maka P_{ik} merupakan integral $P_{ik}|\varepsilon_{ik}$ atas seluruh nilai ε_{ik} terbobot densitasnya, yaitu

$$P_{ik} = \int \prod_{j\neq k} \left(\exp(-\exp(-(\varepsilon_{ik} + V_{ik} - V_{ij})))\right)[\exp(-\varepsilon_{ik})]\exp(-\exp(-\varepsilon_{ik}))d\varepsilon_{ik}$$

Dengan mengambil s=ε_{ik} dan oleh karenaV_{ik}-V_{ik}=0 maka P_{ik} dapat dinyatakan sebagai

$$P_{ik} = \int_{s=-\infty}^{\infty} \prod_{j} \left(\exp(-\exp(-(s + V_{ik} - V_{ij})))\right)[\exp(-s)]ds$$

$$= \int_{s=-\infty}^{\infty} \exp\left(-\sum_j \exp(-(s + V_{ik} - V_{ij}))\right)\exp(-s))ds$$

$$= \int_{s=-\infty}^{\infty} \exp\left(-e^{-s}\sum_j \exp(-(V_{ii} - V_{ij}))\right)e^{-s}ds$$

Misalkan t = exp(-s) sehingga dt = -exp(-s)ds

$$P_{ik} = \int_{\infty}^{0} \exp\left(-t\sum_j \exp(-(V_{ik} - V_{ij}))\right)(-dt)$$

$$= \int_{0}^{\infty} \exp\left(-t\sum_j \exp(-(V_{ik} - V_{ij}))\right)dt$$

$$= \left.\frac{\exp\left(-t\sum_j \exp(-(V_{ik} - V_{ij}))\right)}{-t\sum_j \exp(-(V_{ik} - V_{ij}))}\right|_0^{\infty}$$

$$= \frac{1}{\sum_j \exp(-(V_{ik} - V_{ij}))} = \frac{\exp(V_{ik})}{\sum_j \exp(V_{ij}))} \qquad \textbf{(terbukti)}$$

Formula untuk P_{ik} tersebut dinamakan probabilitas logit. Jika V_{ij} merupakan fungsi linear dari x_{ij} maka dapat dinyatakan menjadi

$$P_{ik} = \frac{\exp(\beta^t x_{ii})}{\sum_{j=1}^{J} \exp(\beta^t x_{ij})} \qquad (8)$$

Untuk sebarang dua alternatif k dan r, rasio probabilitas logitnya dapat dinyatakan sebagai

$$\frac{P_{ik}}{P_{ir}} = \frac{e^{Vik}}{\sum_{j=1}^{J} e^{Vij}} \Bigg/ \frac{e^{Vir}}{\sum_{j=1}^{J} e^{Vij}} = \exp(V_{ik} - V_{ir}) \qquad (9)$$

Rasio ini tidak tergantung pada alternatif lain selain i dan k. Sifat ini dinamakan *independence from irrelevant alternatives* (IIA).

Selanjutnya estimasi parameter β dapat dilakukan dengan prosedur maksimum likelihood. Misalkan N sampel dari individu yang membuat keputusan, probabilitas individu n memilih sebuah alternatif dapat dinyatakan sebagai

$$\prod_j (P_{ij})^{y_{ij}}$$

Dengan y_{ij} = 1 jika individu i memilih j dan nol jika memilih yang lainnya. Dengan mengasumsikan bahwa setiap keputusan antar individu saling independen maka probabilitas masing-masing individu dalam sampel memilih sebuah alternatif adalah

$$L(\beta) = \prod_{i=1}^{n} \prod_j (P_{ij})^{y_{ij}} \qquad (10)$$

Dengan β merupakan vektor parameter dalam model. Fungsi Log likelihood nya menjadi

$$LL(\beta) = \sum_{i=1}^{n} \sum_j y_{ij} \ln(P_{ij}) \qquad (11)$$

Penaksir β adalah nilai β yang memaksimumkan fungsi LL(β).

Teorema 4.

Penaksir β dengan menggunakan prosedur maksimum likelihood adalah penyelesaian dari persamaan

$$\sum_{i=1}^{n} \sum_{j=1}^{J} (y_{ij} - P_{ij}) x_{ij} = 0$$

Bukti:

$$LL(\beta) = \sum_{i=1}^{n} \sum_{j} y_{ij} \ln(P_{ij})$$

$$= \sum_{i=1}^{n} \sum_{j} y_{ij} \ln\left(\frac{\exp(\beta^t x_{ij})}{\sum_{j} \exp(\beta^t x_{ij})} \right)$$

$$= \sum_{i=1}^{n} \sum_{i} y_{ij} (\beta^t x_{ij}) - \sum_{i=1}^{n} \sum_{j} y_{ij} \ln\left(\sum_{j} \exp(\beta^t x_{ij}) \right)$$

Derevatif LL(β) terhadap β adalah

$$\frac{\partial LL(\beta)}{\partial \beta} = \sum_{i=1}^{n} \sum_{j} y_{ij} x_{ij} - \sum_{i=1}^{n} \sum_{j} y_{ij} \sum_{j} P_{ij} x_{ij}$$

$$= \sum_{i=1}^{n} \sum_{j} y_{ij} x_{ij} - \sum_{i=1}^{n} \left(\sum_{j} P_{ij} x_{ij} \right) \sum_{j} y_{ij}$$

$$= \sum_{i=1}^{n} \sum_{j} y_{ij} x_{ij} - \sum_{i=1}^{n} \sum_{j} P_{ij} x_{ij}$$

$$= \sum_{i=1}^{n} \sum_{j} (y_{ij} - P_{ij}) x_{ij} = 0 \quad \textbf{(terbukti)}$$

Penyelesaian persamaan penaksir tersebut dapat diselesaikan secara iterasai dengan metode newton raphson maupun metode scoring. Selanjutnya untuk melakukan menguji kesesuaian model dengan data (uji goodness of Fit) dapat digunakan *likelihood ratio index*, yang diefinisikan sebagai

$$\rho = 1 - \frac{LL(\hat{\beta})}{LL(0)} \tag{12}$$

dengan LL(0) adalah nilai log likelihood untuk β=0.

4. Contoh Kasus

Diambil data dari survei Demografi dan Kesehatan di Elsavador [3]. Survey terhadap 3165 wanita yang masih bersuami yang diklasifikasikan ke dalam kelompok usia (interval 5 tahun) dan penggunaan alat kontrasepsi yang diklasifikasikan dalam “steril”, “lainnya” dan “none”. Data disajikan dalam tabel berikut

Tabel 1. data survey Demografi dan Kesehatan di Elsavador.

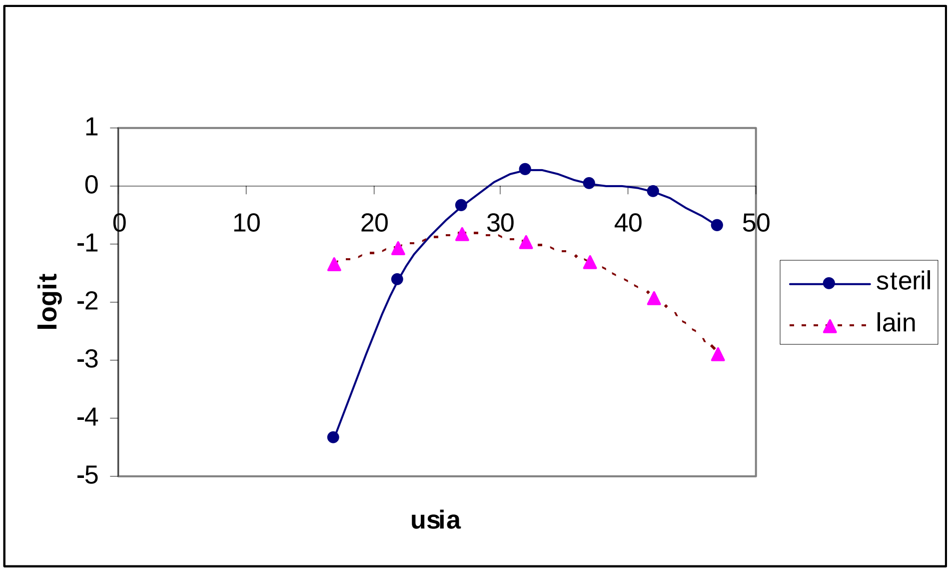
	kontrasepsi		
usia	Steril	lainnya	none

15-19	3	61	232
20-24	80	137	400
25-29	216	131	301
30-34	268	76	203
35-39	197	50	188
40-44	150	24	164
45-49	91	10	183

Dari data pada tabel 1, jika dilakukan uji independensi diperoleh nilai $\chi^2 = 521{,}1$ dengan derajad bebas 12. Jelas bahwa terdapat hubungan atau pengaruh usia terhadap penggunaan kontarasepsi. Selanjutnya jika dilakukan analisis dengan model regresi multinomial maka dapat digunakan model logit dengan fungsi logit

$$\log(\frac{\pi_{ij}}{\pi_{i3}}) = \beta_0 + \beta_1 Usia + \beta_2 (Usia)^2$$

Kategori ke-3 (none) digipakai sebagai base-line. Nilai untuk variabel usia diambil nilai tengah untuk masing-masing kelompok usia. Pengaruh usia terhadap logit diduga berbentuk polinomial orde dua, sebab dari gambar 1, plot usia terhadap nilai logit berbentuk kurva.



Gambar 1. Plot usia terhadap logit antara “steril” vs “none” dan “lain” vs “none”.

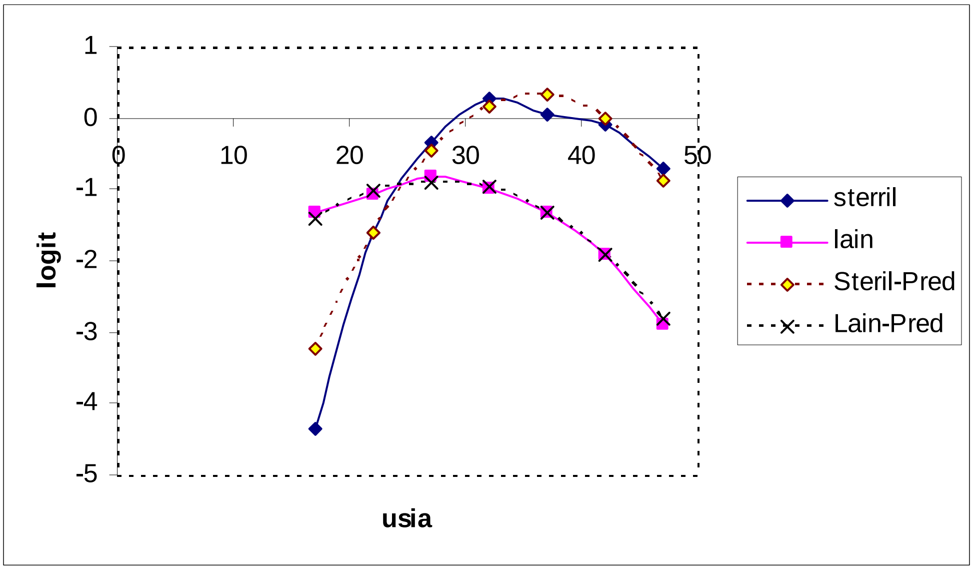
Beberapa paket program statistika telah menyediakan fasilitas analisis untuk model regresi multinomial, misalnya program SPSS dan MINITAB. Dalam hal ini digunakan program SPSS 11.5, diporoleh output sebagai berikut :

Tabel 2. hasil estimasi parameter model multinomial.

RESPON (a)		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
1.00	Intercept	-12.266	.735	278.468	1	.000			
	Usia	.700	.045	243.551	1	.000	2.014	1.844	2.199
	Usia^2	-.010	.001	218.250	1	.000	.990	.989	.992
2.00	Intercept	-4.419	.671	43.398	1	.000			
	Usia	.259	.046	31.340	1	.000	1.296	1.184	1.419
	Usia^2	-.005	.001	39.233	1	.000	.995	.994	.997

Dari tabel 2 dapat disimpulakan bahwa pengaruh usia terhadap logit berbentuk linear dan kuadratis. Nilai Chi-Square untuk statistik Pearson adalah 18,869 dengan derajad bebas 8 dan untuk statistik Deviance besarnya 20,475 dengann derajad bebas 8. Jadi dari nilai statistik ini berarti model layak digunakan.

Perbandingan logit aktual dan logit prediksi disajikan dalam gambar 2. Secara umum model yang diperoleh dapat mempredikasikan dengan baik.



Gambar 2. Plot usia terhadap logit pada nilai aktual dan nial prediksi

4. Kesimpulan

Analisis regresi untuk respon nominal dengan menggunakan model logistic maupun model discrete choice yang didasarkan pada distribusi nilai ekstrim menghasilkan fungsi probabilitas yang sama. Penaksir parameter didapatkan dengan cara menyelesaikan persamaan penaksir secara iterasi.

5. Daftar pustaka

[1] W. Hosmer & L. Stanley, 1989, ***Applied Logistic Regression***, John Wiley & Sons

[2] Agresti, Alan, 1990, ***Categorical Data Analysis***, John wiley & Son

[3] Rodriguez, 2001, ***Generalized Linear statistical Models***, http://data.princeton.edu/wwws509

[4] E. Train, Kenneth, 2003, ***Discrete Choice Methods with Simulation***, Cambridge University Press.