ISBN:1858-1382

J M S K
Jurnal Matematika, Statistika, & Komputasi
http://journal.unhas.ac.id/index.php/jmsk
Vol. 4, No.1, 43-53, Juli 2007

Seleksi Model Multinomial Logit Melalui Akaike's Information Criterion (AIC)

Erna Tri Herdiani*, Amran**

Abstrak

Regresi kualitatif akhir-akhir ini banyak digunakan sebagai jawaban atas beragamnya data set yang terdiri dari data kualitatif dan data kuantitatif. Dalam penelitian ini kami memiliki data set dalam bentuk respon yang terdiri dari data kualitatif terurut berdistribusi multnomial sedangkan peubah prediksi berbentuk data kategorikal. Dengan kedua asumsi tersebut penulis memilih menggunakan Model Multinomial Logit, Alan Agresti (1990). Dari proses tersebut penaksiran parameter dilakukan dengan metode kuadrat terkecil di perumum dan metode maksimum likelihood Gonzales, P.L. (1993), selanjutnya di dalam penelitian ini penulis menggunakan Akaike's Information Criterion (AIC) untuk memilih model terbaik dalam menganalisis data kategorikal..

Kata Kunci : *gelombang permukaan datar, metode elemen batas.*

1. Pendahuluan

Analisis regresi banyak digunakan dalam memprediksi dan menentukan faktor yang berpengaruh terhadap suatu respon. Salah satu asumsi yang harus dipenuhi ialah model error yang diambil haruslah berdistribusi normal. Yang menjadi permasalahan sekarang bagaimana kalau model error ternyata tidak berdistribusi normal dan data berbentuk kategorikal atau kualitatif.

Untuk mengatasi masalah tersebut dalam tulisan ini dibahas Model Linier Diperumum (MLD). MLD adalah suatu model yang digunakan untuk menyelidiki hubungan antara variabel prediktor dengan variabel respon dengan asumsi error tidak harus berdistribusi normal. Salah satu bentuk MLD yang akan dibahas dalam tulisan ini ialah model Multinomial Logit (MML), karena komponen random berbentuk data kategorik dan variabel respon yang juga berbentuk kategorik, dengan jumlah variabel lebih dari dua.

Dalam menyeleksi MML, Tenenhaus dkk. [3] menggunakan nilai p yang kecil, tetapi untuk data penyakit Demam Berdarah Dengue (DBD) di Rumah Sakit Umum Pemerintah (RSUP) Wahidin Sudirohusodo Makassar. Pada DBD ini nilai p memiliki nilai yang sama, lalu penulis memberikan alternatif lain yakni menyeleksi model dengan menggunakan kriteria informasi yang disebut dengan AIC (Akaike's Information Criterion). Kriteria ini diperkenalkan pertama kali oleh Akaike (1973,1974) untuk memilih model terbaik.

2. Model Multinomial Logit

MLD merupakan metode membentuk model yang digunakan untuk menyelidiki hubungan antara variabel prediktor dan variabel respon. Metode ini pertama kali diperkenalkan oleh Nelder dan Wedderburn (1972) untuk menentukan taksiran model linier dengan asumsi

* dan ** *Staf pengajar pada Jurusan Matematika F.MIPA Universitas Hasanuddin Makassar*

model error tidak harus berdistribusi normal, tetapi distribusinya termasuk dalam keluarga eksponensial. Model linier diperumum memiliki tiga komponen, yaitu:

a. Komponen random terdiri dari variabel respon $Y_i,\ i = 1,2,\ldots,n$ saling bebas, dimana n ukuran sampel, mean = μ_i dan peluang densitasnya termasuk keluarga eksponensial.

Definisi 1. Keluarga eksponensial.

Misalkan *Y* = Variabel respon, $Y = (y_1, y_2, y_3, \ldots, y_N)^t$ dimana *N* adalah banyaknya observasi dan *Y* saling bebas.

$y_i \sim f(y_i, \theta_i)$ dimana $i = 1,2,\ldots,N$.

jika fungsi peluang y_i dapat dituliskan sebagai $f(y_i, \theta_i) = a(\theta_i)\, b(y_i)\, e^{y_i Q(\theta_i)}$ maka *Y* disebut famili eksponensial.

b. Komponen sistematik terdiri dari variabel prediktor x_i, dengan $x_i = (1, x_{i1}, x_{i2}, \ldots, x_{ik})^t$ dimana *k* menyatakan banyaknya prediktor. Komponen ini menjelaskan prediktor $\{x_i\}$ dalam suatu persamaan:

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} \tag{1}$$

Kombinasi linear ini disebut prediktor linear

c. Link adalah suatu fungsi yang menghubungkan antara komponen sistematik dengan mean dari komponen random. Misalkan $E(Y_i) = h(\mu_i) = \eta_i$ dimana *h* disebut link dengan syarat *h* merupakan fungsi dari μ yang monoton dan differensiabel.

Misalkan *Y* sebagai variabel respon dengan *r* kategori pada *s* populasi dan berdistribusi multinomial. *Y* dapat ditulis sebagai tabel distribusi frekuensi sebagai berikut:

Tabel I. Tabel Frekuensi *N*

Populasi	1	. . .	j	. . .	r	
1	n_{11}	. . .	n_{1j}	. . .	n_{1r}	n_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
s	n_{s1}	. . .	n_{sj}	. . .	n_{sr}	n_s

Dimana n_{ij} menyatakan banyaknya sampel dari populasi ke-*i* yang termasuk kategori ke-*j*, dengan $i = 1,2,\ldots,s$ dan $j = 1,2,\ldots,r$.

Distribusi multinomial termasuk keluarga eksponensial,

$$n_{ij} \sim multinomial\,(n, \pi_{ij})\, dengan\ i = 1,2,\ldots,r\ \ dan\ \ j = 1,2,\ldots,r$$

Fungsi peluang n_{ij} dapat dituliskan sebagai:

$$a(\pi_{ij}) = \sum_{j=1}^{r} \pi_{ij} = 1 \quad , \quad b(n_{ij}) = \frac{n_i!}{\prod_{j=1}^{r} n_{ij}!} \quad , \quad Q(\pi_{ij}) = \ln(\pi_{ij})$$

Selanjutnya, misalkan P adalah tabel frekuensi relatif dengan elemen p_{ij} yang merupakan relatif respon Y kategori j pada populasi i, dimana $p_{ij} = \left(\frac{n_{ij}}{n_i}\right)$

Tabel 2. Tabel Frekuensi Relatif

Populasi	1	...	j	...	r	
1	p_{11}	...	p_{1j}	...	p_{ir}	p_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
s	p_{s1}	...	p_{sj}	...	p_{sr}	p_s

Model dengan asumsi respon multinomial dengan link generalized logit disebut model multinomial respon yang kemudian dikenal dengan Model Multinomial logit. Bentuk umum model multinomial logit adalah :

$$\ln\left(\frac{\pi_{ij}}{\pi_{ir}}\right) = X_{ij}\beta \tag{3}$$

Selanjutnya akan ditaksir β dengan menggunakan metode Kuadrat Terkecil Diperumum dan Metode Maksimum Likelihood (ML)

a. Metode Kuadrat Terkecil Diperumum

Misalkan persamaan umum:

$$F(P) = X\beta + \varepsilon \tag{4}$$

dimana $E(\varepsilon) = 0$ *dan* $V(\varepsilon) = V_F = Var(F(P))$. Agar variansi konstan maka persamaan (4) dikalikan dengan $V_F^{-\frac{1}{2}}$, menjadi:

$$V_F^{-1/2}F(P) = V_F^{-1/2}X\beta + V^{-1/2}\varepsilon \tag{5}$$

Jadi persamaan (5) telah memiliki variansi konstan. Selanjutnya akan ditentukan penaksiran parameter β :

$$V_F^{-1/2}F(P) = V_F^{-1/2}X\beta + V^{-1/2}\varepsilon$$

$$\Rightarrow V_F^{-1/2}\varepsilon = V_F^{-1/2}F(P) - V_F^{-1/2}X\beta$$

$$\varphi = \left(V_F^{-1/2}\varepsilon\right)^t V_F^{-1/2}\varepsilon$$

$$= \left[(F(P) - X\beta)^t\left(V_F^{-1/2}\right)^t\right]\left[V_F^{1/2}F(P) - V_F^{-1/2}X\beta\right]$$

$$= F^t(P)\left(V_F^{-1/2}\right)^t V_F^{-1/2}F(P) - 2\beta^t X^t\left(V_F^{-1/2}\right)^t V_F^{-1/2}F(P) + \beta^t X^t\left(V_F^{-1/2}\right)^t V_F^{-1/2}X\beta$$

$$\frac{\partial\varphi}{\partial\beta} = -2X^tV_F^{-1}F(P) + 2XV_F^{-1}X\beta = 0 \Rightarrow X^tV_F^{-1}X\beta = X^tV_F^{-1}F(P)$$

Jadi,

$$\hat{\beta} = \left(X^tV_F^{-1}X\right)^{-1}X^tV_F^{-1}F(P) \tag{6}$$

Jika $\hat{V}_F = S$ maka

$$\hat{\beta} \approx \left(X^t S^{-1} X\right)^{-1} X^t S^{-1} F(P) \tag{7}$$

b. Metode Maksimum Likelihood

Misalkan diketahui model logistik :

$$\ln\left(\frac{\pi_{ij}}{\pi_{ir}}\right) = X_{ij}\beta \tag{8}$$

untuk *i = 1,2,…,s dan j = 1,2,…,r-1* Dimana X_{ij} adalah baris ke-*j* dari matriks X_i .

Fungsi respon Generalized logit:

$$F_h(\pi_i) = \ln\left(\frac{\pi_{ih}}{\pi_{ir}}\right), \text{ dimana } h = 1,2,\ldots,r\text{-}1 \tag{9}$$

syarat yang harus dipenuhi $\sum_{j=1}^{r} \pi_{ij} = 1$, dengan

$$\pi_{ij} = f_{ij}(\beta) = \begin{cases} \dfrac{e^{X_{ij}\beta}}{1 + \sum_{j=1}^{r-1} e^{X_{ij}\beta}} & jika \quad j = 1,2,\ldots,r-1 \\[2ex] \dfrac{1}{1 + \sum_{j=1}^{r-1} e^{X_{ij}\beta}} & jika \quad j = r \end{cases}$$

Kemudian akan ditentukan penaksiran parameter β melalui maksimum likelihood. Fungsi Likelihood untuk variabel multinomial adalah:

$$\varphi(\mathrm{mod}\, el) = P\left(n_{11},\ldots,n_{ij},\ldots,n_{ir}\right)\ldots P\left(n_{s1},\ldots,n_{sj},\ldots,n_{sr}\right)$$

$$= n_i!\frac{\pi_{11}^{11}\ldots\pi_{ij}^{ij}\ldots\pi_{ir}^{n_{ir}}}{n_{i1}!\ldots n_{ij}\ldots n_{ir}} \ldots n_s!\frac{\pi_{s1}^{n_{s2}}\ldots\pi_{sj}^{n_{sj}}\ldots\pi_{sr}^{n_{sr}}}{n_{i1}!\ldots n_{ij}\ldots n_{ir}}$$

$$= \prod_{i=1}^{s} n_i!\frac{f_{i1}^{n_{11}}(\beta)\ldots f_{ij}^{n_{ij}}(\beta)\ldots f_{ir}^{n_{ir}}(\beta)}{n_{i1}!\ldots n_{ij}\ldots n_{ir}}$$

Sekarang akan ditentukan nilai β yang memaksimumkan $\varphi(\mathrm{mod}\, el)$ menggunakan Metode Newton Raphson. Dengan persamaan :

$$\beta^{(t+1)} = \beta^{(t)} - \left(H^{(t)}\right)^{-1}\overline{q}^{(t)}$$

Dimana $\overline{q}' = \begin{pmatrix} \frac{\partial\varphi}{\partial\beta_1} \\ \frac{\partial\varphi}{\partial\beta_2} \\ \vdots \end{pmatrix}$ dan H ≡ Matriks Hessian, diasumsikan tidak singular.

Perhitungan nilai β maksimum melalui metode Newton Raphson ini dapat diperoleh dengan menggunakan software ststistik SAS, procedure CATMOD.

3. Akaike's Information Criterion (AIC)

Seleksi model merupakan suatu tahapan penting untuk memutuskan model yang terbaik. Akaike (1973,1974) memperkenalkan suatu kriteria informasi yang disebut dengan AIC (Akaike's Information Criterion).

Misalkan $x_1,..., x_n$ merupakan sampel acak dari sebuah distribusi g yang tidak diketahui dan $\hat{\theta}_k$ adalah penduga maksimum likelihood yang diberikan f_k.

Definisi 2. Jarak Kullback Leibler

Misalkan f adalah distribusi data sample yang diasumsikan dan g adalah distribusi eksperimental data sampel , maka :

$$\begin{aligned} D(f,g) &= E_g \log \frac{g(X)}{f(X)} \\ &= E_g \log g(X) - E_g \log f(X) \end{aligned}$$

dimana $D(f,g) \geq 0$

Pilih model yang memaksimumkan $Q_k = E\left\{\log f_k\left(Z,\hat{\theta}_k\right)\right\}$ (10)

dimana: Z adalah variabel bebas $x_1,..., x_n$. Dan

Log-Likelihood dari θ_k dari model $f_k = (x,\theta_k)$ adalah :

$$\log L(\theta_k) = \sum \log f_k(x_i,\theta_k) \tag{11}$$

Misalkan θ_* adalah penyelesaian dari $\lambda(\theta_k) = 0$, dimana $\lambda(\theta_k) = E\{\log f_k(Z,\theta_k)\}$, θ_* adalah parameter yang di duga oleh $\hat{\theta}$.

Didefinisikan:

$$J_k \equiv E\left(\frac{\partial \log f_k(Z,\theta_k)}{\partial \theta_k}\right)\left(\frac{\partial \log f_k(Z,\theta_k)}{\partial \theta_k{}'}\right)\Bigg|_{\theta_k=\theta_*} \quad \text{dan} \tag{12}$$

$$I_k = -E\left(\frac{\partial^2 \log f_k(Z,\theta_k)}{\partial \theta \partial \theta_k{}'}\right)\Bigg|_{\theta_k=\theta} \tag{13}$$

Sehingga

$$E\left\{n\left(\hat{\theta}_k - \theta_*\right)' I_k \left(\hat{\theta}_k - \theta_*\right)\right\} \approx trace\left(J_k I_k^{-1}\right) \tag{14}$$

Selanjutnya akan ditentukan ruang sampel untuk metode AIC, Ekspansi deret Taylor orde dua dari log likelihood di sekitar $\hat{\theta}_k s$, diperoleh:

$$\log L(\theta_k) = \log L(\hat{\theta}_k) + \frac{\partial \log L(\hat{\theta}_k)}{\partial \theta_k}(\theta_k - \hat{\theta}_k) + \frac{1}{2}(\theta_k - \hat{\theta}_k)' \frac{\partial^2 \log L(\theta_k^*)}{\partial \theta_k \partial \theta_k'}(\theta_k - \hat{\theta}_k)$$

$$= \log L(\hat{\theta}_k) + \frac{\partial \log L(\hat{\theta}_k)}{\partial \theta_k} + \frac{1}{2}(\theta_k - \hat{\theta}_k)' \frac{\partial^2 \log L(\theta_k^*)}{\partial \theta_k \partial \theta_k'}(\theta_k - \hat{\theta}_k)$$

$$= \log L(\hat{\theta}_k) + \frac{n}{2}(\theta_k - \hat{\theta}_k)' \cdot \frac{1}{n}\frac{\partial^2 \log L(\theta_k^*)}{\partial \theta_k \partial \theta_k'}(\theta_k - \hat{\theta}_k)$$

$$\approx \log L(\hat{\theta}_k) - \frac{n}{2}(\theta_k - \hat{\theta}_k)' I_k (\theta_k - \hat{\theta}_k) \tag{15}$$

Untuk $\theta_k = \theta_*$ maka $\log L(\hat{\theta}_*) \approx \log L(\hat{\theta}_k) - \frac{1}{2} trace(J_k I_k^{-1})$ (16)

Sehingga

$$E\{\log L(\theta_*)\} \approx E\{\log L(\hat{\theta}_k)\} - \frac{1}{2} trace(J_k I_k^{-1})$$

$$n\lambda(\theta_*) \approx E\{\log L(\hat{\theta}_k)\} - \frac{1}{2} trace(J_k I_k^{-1}) \tag{17}$$

Sedangkan Ruang Model untuk Metode AIC, melalui Ekspansi deret Taylor orde 2 dari $\lambda(\theta_k)$ di sekitar

$$\theta_* \ \lambda(\theta_k) = \lambda(\theta_*) + \frac{\partial \lambda(\theta_*)}{\partial \theta_k}(\theta_k - \theta_*) + \frac{1}{2}(\theta_k - \theta_*)' \frac{\partial^2 \lambda(\theta_k^*)}{\partial \theta_k \partial \theta_k'}(\theta_k - \theta_*)$$

$$= \lambda(\theta_*) + \left(\frac{\partial \lambda(\theta_*)}{\partial \theta_k}(\theta_k - \theta_*) + \frac{1}{2}(\theta_k - \theta_*)' \frac{\partial^2 \lambda(\theta_k^*)}{\partial \theta_k \partial \theta_k'}\right)(\theta_k - \theta_*)$$

$$\approx \lambda(\theta_*) + \frac{1}{2}(\theta_k - \theta_*)' E \frac{\partial^2 \lambda(\theta_k^*)}{\partial \theta_k \partial \theta_k'}(\theta_k - \theta_*)$$

$$= \lambda(\theta_*) - \frac{1}{2}(\theta_k - \theta_*)' I_k (\theta_k - \theta_*) \tag{18}$$

Untuk $\theta_k = \hat{\theta}_k$, diperoleh :

$$Q_k = E\lambda(\hat{\theta}_k) \approx E\left[\lambda(\theta_*) - \frac{1}{2}(\hat{\theta}_k - \theta_*)' I_k (\hat{\theta}_k - \theta_*)\right]$$

$$\approx E[\lambda(\theta_*)] - \frac{1}{2} E\left[\left(\hat{\theta}_k - \theta_*\right)' I_k \left(\hat{\theta}_k - \theta_*\right)\right] \approx \lambda(\theta_*) - \frac{1}{2n} E\left[n\left(\hat{\theta}_k - \theta_*\right)' I_k \left(\hat{\theta}_k - \theta_*\right)\right]$$

$$\approx \lambda(\theta_*) - \frac{1}{2n} trace\left(I_k J_k^{-1}\right) \qquad (19)$$

Dari persamaan (16) dan (19), diperoleh:

$$nQ_k \approx n\lambda\theta_* - \frac{1}{2} trace\left(J_k I_k^{-1}\right) \approx E\left\{\log L\left(\hat{\theta}_k\right)\right\} - \frac{1}{2} trace\left(J_k I_k^{-1}\right)$$

$$\approx E\left\{\log L\left(\hat{\theta}_k\right)\right\} - trace\left(J_k I_k^{-1}\right) \qquad (20)$$

Rumus AIC didasari oleh anggapan $J_k \approx I_k$, jadi $trace\left(J_k I_k^{-1}\right) \approx p$, dimana *p* adalah banyaknya parameter dari *k*-model, sehingga

$$AIC(k) = -2 \log L\left(\hat{\theta}_k\right) + 2p \qquad (21)$$

Model tersebut selanjutnya akan digunakan untuk memilih model multinomial logit yang terbaik.

4. Studi Kasus

Data yang ditampilkan pada Tabel 3. diperoleh sebagai hasil penelitian yang telah dilakukan di Rumah Sakit Dr.Wahidin Sudiro Husodo pada ruang Rekam Medis. Penelitian, tanggal 30 Juni s/d 30 Juli 2004. Data yang diambil berupa data sekunder, yaitu:

- Data pribadi pasien, Terdiri dari Nama, Pekerjaan, Pendidikan, Usia, dan Pendapatan.
- Riwayat penyakit, Terdiri kondisi pasien per-hari selama dirawat menurut stadiumnya atau tingkat keparahan penyakit demam berdarah yang terdiri dari:
    - Stadium 1 : Demam mendadak selam 2-7 hari dengan suhu tubuh yang sangat tinggi (lebih dari 38^0 C), timbul gejala klinis yang lain seperti badan pegal, mual, mata pedas, bintik-bintik merah pada badan serta manifestasi pendarahan teringan atau uji turniket.
    - Stadium 2 : Ditemukan pendarahan kulit dan manifestasi pendarahan lain seperti keluarnya darah dari hidung.
    - Stadium 3 : Pre shock, pendarahan terjadi pada pasien demam berdarah pada saat buang air besar.
    - Stadium 4 : Shock, pendarahan pada saat buang air besar dan kecil serta terdapat DSS (Dengue Shock Syndrome) dengan nadi dan tekanan darah yang tak terukur.
      (Sumber : Kapita Selekta Kedokteran)

Dari tabel tersebut, respon Y menyatakan tingkat penyakit demam berdarah yang terdiri dari 5 kategori, kategori 0 tidak menderita penyakit demam berdarah, kategori 1,2,3 dan 4 berturut – turut menunjukkan tingkat keparahan penyakit tersebut (dalam stadium) dan kategori 5 apabila pasien tidak dapat tertolong lagi atau meninggal dunia. Dan prediktor, terdiri dari:

1. X_1 = Usia penderita, terdiri dari 0 – 12 Tahun, 12 – 25 Tahun dan 25 Tahun atas
2. X_2 = Jenis Kelamin, terdiri dari Perempuan dan Laki-laki
3. X_3 = Kelas Perawatan, terdiri dari Kelas A, B, I, II dan III
4. X_4 = Pendidikan, terdiri dari Berpendidikan dan Tidak berpendidikan
5. X_5 = Waktu Perawatan, terdiri dari Hari pertama dan hari terakhir pengamatan

Tabel 3. Tabel kategorik untuk kutipan data Penyakit Demam Berdarah

Usia	Jns.kel	Kelas	Pend.	Waktu	Tingkat Penyakit / Stadium						Total
X1	X2	X3	X4	X5	0	1	2	3	4	5	
1	1	1	1	1	0	2	1	1	0	0	4
1	1	1	1	2	2	2	0	0	0	0	4
1	1	1	2	1	0	2	1	0	0	0	3
1	1	1	2	2	2	1	0	0	0	0	3
1	1	2	1	1	0	3	0	0	0	0	3
1	1	2	1	2	3	0	0	0	0	0	3
1	1	2	2	1	0	9	1	0	0	0	10
1	1	2	2	2	10	0	0	0	0	0	10
1	1	3	1	1	0	0	1	0	0	0	1
1	1	3	1	2	1	0	0	0	0	0	1
1	1	3	2	1	0	0	0	0	0	0	0
1	1	3	2	2	0	0	0	0	0	0	0
1	1	4	1	1	0	6	1	0	0	0	7
1	1	4	1	2	7	0	0	0	0	0	7
1	1	4	2	1	0	8	2	2	0	0	12
1	1	4	2	2	9	2	0	0	0	1	12
1	1	5	1	1	0	11	5	2	0	0	18
1	1	5	1	2	15	3	0	0	0	0	18
1	1	5	2	1	0	30	7	3	5	0	45
1	1	5	2	2	30	13	0	0	1	1	45
1	2	1	1	1	0	0	0	0	0	0	0
1	2	1	1	2	0	0	0	0	0	0	0
1	2	1	2	1	0	2	0	0	0	0	2
1	2	1	2	2	1	1	0	0	0	0	2
1	2	2	1	1	0	4	1	1	0	0	6
1	2	2	1	2	6	0	0	0	0	0	6
1	2	2	2	1	0	11	0	0	0	0	11
1	2	2	2	2	8	2	0	0	1	0	11
1	2	3	1	1	0	2	0	0	0	0	2
1	2	3	1	2	2	0	0	0	0	0	2
1	2	3	2	1	0	0	0	0	0	0	0
1	2	3	2	2	0	0	0	0	0	0	0
1	2	4	1	1	0	7	0	1	1	0	9
1	2	4	1	2	5	4	0	0	0	0	9
1	2	4	2	1	0	8	0	3	1	0	12
1	2	4	2	2	8	3	0	0	0	1	12
1	2	5	1	1	0	9	0	1	0	0	10
1	2	5	1	2	8	2	0	0	0	0	10
1	2	5	2	1	0	28	4	8	3	0	43
1	2	5	2	2	29	9	0	0	0	5	43
2	1	1	1	1	0	2	1	0	0	0	3
2	1	1	1	2	1	2	0	0	0	0	3

Hasil Pengolahan data penyakit DBD menggunakan Metode Model Linier Diperumum dengan bantuan Software SAS diperoleh:

1. Tabel Analisis Variansi

Tabel 4. Tabel Analisis Variansi

Sumber	dk	Metode Kuadrat Terkecil Diperumum		Metode Maksimum Likelihood	
		Chi-Square	**Prob.**	**Chi-Square**	**Prob.**
Konstanta	5	308.80	0.0000	302.89	0.0000
Usia	10	28.78	0.0014	27.25	0.0024
Jenis Kelamin	5	4.83	0.4374	4.13	0.5308
Kelas Perawatan	20	70.84	0.0000	82.39	0.0000
Pendidikan	5	4.69	0.4555	4.61	0.4652
Waktu Perawatan	5	194.91	0.0000	311.01	0.0000
	550	**20.4.78**	**1.0000**	**232.59**	**1.0000**

Keterangan: Variabel Jenis Kelamin dan Pendidikan tidak berarti dengan mengambil tingkat keberartian α = 5%. Sedangkan predictor lainnya berarti, karena prob < α .

2. Menentukan MML data penyakit DBD untuk menaksir peluang sel ke (i,j)

a. Metode Kuadrat Terkecil Diperumum

Untuk j = 1, dimana j = 1,2,3,4,5

$$E(Y) = \ln\left(\frac{\pi_{ij}}{\pi_{i5}}\right) = 0.96 + \begin{pmatrix}1 & 2 & 3\end{pmatrix}\begin{pmatrix}0.00 \\ 0.41 \\ 0.00-0.41\end{pmatrix} + \begin{pmatrix}1 & 2\end{pmatrix}\begin{pmatrix}0.15 \\ -0.15\end{pmatrix} +$$

$$\begin{pmatrix}1 & 2 & 3 & 4 & 5\end{pmatrix}\begin{pmatrix}-0.54 \\ 0.28 \\ -0.75 \\ 0.20 \\ 0.54-0.28+0.75-0.20\end{pmatrix} +$$

$$\begin{pmatrix}1 & 2\end{pmatrix}\begin{pmatrix}-0.18 \\ 0.18\end{pmatrix} + \begin{pmatrix}1 & 2\end{pmatrix}\begin{pmatrix}-0.60 \\ 0.60\end{pmatrix}$$

b. Metode Maksimum Likelihood

Untuk j = 1, dimana j = 1,2,3,4,5

$$E(Y) = \ln\left(\frac{\pi_{ij}}{\pi_{i5}}\right) = 0.79 + \begin{pmatrix}1 & 2 & 3\end{pmatrix}\begin{pmatrix}0.02 \\ 0.38 \\ -0.02+0.38\end{pmatrix} + \begin{pmatrix}1 & 2\end{pmatrix}\begin{pmatrix}0.13 \\ -0.13\end{pmatrix} +$$

$$(1\ 2\ 3\ 4\ 5)\begin{pmatrix} -0.57 \\ 0.35 \\ -0.83 \\ 0.21 \\ 0.57-0.35+0.83-0.21 \end{pmatrix} + (1\ \ 2)\begin{pmatrix} 0.18 \\ -0.18 \end{pmatrix} + (1\ \ 2)\begin{pmatrix} -0.76 \\ 0.76 \end{pmatrix}$$

Model di atas diperoleh berdasarkan tabel taksiran parameter data penyakit DBD. Dari model tersebut juga dapat diperoleh peluang taksiran sel, dimana setiap sel menunjukkan peluang setiap populasi dengan kategori tertentu menderita penyakit DBD pada setiap stadiumnya.

3. Menentukan Nilai AIC

Nilai AIC ditentukan dengan menggunakan rumus:

$$AIC(k) = -2\log L(\hat{\theta}_k) + 2p$$

dimana: $L(\hat{\theta}_k)$ adalah Maksimum likelihood dari model multinomial logit

p adalah jumlah parameter yang digunakan dalam model.

Selanjutnya dapat diperolah nilai AIC dari masing-masing model

Tabel 5. Nilai AIC untuk setiap Metode

Akaike's Information Criterion AIC)	
Metode Kuadrat Terkecil Diperumum	Metode Maksimum Likelihood
391,10772	**5146,3225**

Dari Tabel 4. tampak bahwa nilai AIC untuk metode Kuadrat Terkecil Diperumum lebih kecil sehingga dapat dikatakan bahwa model yang paling cocok digunakan untuk data penyakit demam berdarah dengue adalah model multinomial logit dengan menggunakan Metode Kuadarat Terkecil Diperumum / GLS.

5. Kesimpulan

Berdasarkan hasil pengolahan data, diperoleh kesimpulan sebagai berikut:

1. Akaike Information Criterion (AIC) merupakan suatu kriteria informasi yang digunakan untuk memilih model terbaik. Pada data penyakit Demam Berdarah Dengue (DBD) di Rumah Sakit Dr.Wahidin Sudiro Husodo, perhitungan nilai AIC untuk model multinomial logit dengan menggunakan metode kuadrat terkecil diperumum = 391,108 dan metode maksimum likelihood = 5146,323.

   Dari nilai tersebut model yang paling baik digunakan untuk data penyakit DBD adalah model multinomial logit dengan menggunakan metode kuadrat terkecil / GLS.

2. Dari model multinomial logit dengan menggunakan metode kuadrat terkecil diperumum dapat diperoleh peluang taksiran sel, dimana setiap sel menunjukkan peluang setiap populasi dengan kategori tertentu menderita penyakit DBD pada setiap stadiumnya. Secara umum dapat disimpulkan bahwa:
   - Peluang penderita Demam Berdarah Dengue (DBD) berada di stadium 1 lebih besar dibandingkan dengan stadium lainnya yaitu dengan rata-rata 32.16%.
   - Peluang penderita DBD sembuh setelah mendapatkan perawatan di Rumah Sakit Dr. Wahidin Sudiro Husodo yaitu rata-rata sebesar 42,02 % sedangkan peluang penderita

DBD meninggal rata-rata sebesar 8,86 %, hal ini berarti tindakan medis yang dilakukan pihak Rumah Sakit Dr. Wahidin Sudiro Husodo untuk pasien DBD sudah cukup memuaskan.

Saran

Bagi pembaca yang ingin membandingkan hasil yang diperoleh penulis, dapat menggunakan kriteria informasi yang lain untuk menentukan model yang terbaik misalnya Schwartz's Bayesian Criterion (SBC) atau Quasi-likelihood Independence Criterion (QIC) dan dapat membandingkan mana kriteria yang cocok untuk model yang ada serta apa kelebihan dan kekurangannya.

Daftar Pustaka

[1]. A. Agresti, 1990, "*Categorical Data Analysis",* John Wiley & Sons, New York

[2]. A. Agresti, 1996, "*An Introduction to Categorical Data Analysis",* John Wiley & Sons, New York.

[3]. T.W. Anderson, 1984, "*Introduction to Multivariate Statistical Analysis",* (2nded). John Wiley & Sons, New York

[4]. Erna Tri Herdiani, 2000, "*Model Linier dan Model Geometri Data Kategorikal"*, Thesis.

[5]. Y. Parawitan, 2001, "*In All Likelihood, Statistical Modelling and Inferences Using Likelihood",* Clarendon Press-Oxford.

[6]. T. Tenenhaus, Y. Le Roux, C.O. Guimart, 1993, "*Modele'Line'aire Ge'ne'ralise' et Analyse des Correspondances*", Rev. Statistique Applique, XLI(2), 59-86.