Model Log-Linier dan Regresi Logistik

Julio Adisantoso, G16109011/STK

26 Mei 2010

Ringkasan

Regresi log-linier adalah suatu pendekatan pemodelan linier terampat yang dapat digunakan untuk data yang menyebar Poisson. Model log-linier merupakan pengembangan dari analisis tabel silang dua arah atau lebih dimana terdapat hubungan antara dua atau lebih variabel kategori yang dianalisis menggunakan logaritme alami terhadap setiap isi sel dalam tabel.

Hasil analisis menunjukkan bahwa model log-linier yang melibatkan tiga variabel untuk data tabel tiga arah dengan interaksi dua level merupakan model yang sesuai dibanding hanya menggunakan model *additive* sempurna. Model ini juga memberikan hasil dugaan yang relatif sama dengan model regresi logistik nominal dengan menggunakan fungsi hubung logit, baik berdasarkan hasil analisis numerik terhadap data contoh maupun berdasarkan tinjauan matematis.

1 Pendahuluan

Hasil pengukuran suatu variabel sering mempunyai ciri berupa dua atau lebih kemungkinan nilai yang dikenal sebagai variabel kategorik. Variabel kategorik yang tidak memiliki urutan disebut sebagai variabel nominal sedangkan yang memiliki urutan disebut variabel ordinal. Kedua jenis variabel ini, baik nominal maupun ordinal sering disebut juga sebagai variabel multinomial.

Dalam analisis data dimana variabel respon adalah nominal, digunakan suatu metode yang merupakan pengembangan dari regresi logistik dan dikenal sebagai regresi logistik nominal atau *nominal logistic regression*, sedangkan untuk variabel respon ordinal digunakan regresi logistik ordinal atau atau *nominal logistic regression* (McCullagh & Nelder, 1983).

Pada keadaan tertentu, variabel respon yang berupa frekuensi mengikuti sebaran Poisson. Dalam analisis data dimana variabel respon adalah frekuensi dengan sebaran Poisson, digunakan model Poisson dan log-linier (Dobson, 2001). Fakta menunjukkan bahwa sebaran binomial maupun multinomial dapat diturunkan dari sebaran peubah acak Poisson yang saling bebas. Oleh karena terdapat hubungan antara model multinomial untuk proporsi dan model log-linier untuk frekuensi. Makalah ini akan mengkaji perbandingan antara model log-linier dan model multinomial dengan menggunakan data pada buku Dobson (2001) bab 9.9.

2 Model

Ketika variabel respon berupa nilai kategori dengan dua atau lebih kategori, terdapat dua pendekatan yang dapat dilakukan dalam model linier terampat. Pertama adalah menggunakan model regresi logistik dari respon dikotomus dan model regresi nominal atau ordinal untuk lebih dari dua kategori, yang keduanya disebut sebagai regresi logistik multinomial. Pendekatan kedua adalah menggunakan model log-linier untuk respon frekuensi yang mengikuti sebaran Poisson.

2.1 Regresi Logistik Multinomial

Regresi logistik multinomial (nominal dan ordinal) merupakan salah satu pendekatan pemodelan yang dapat digunakan untuk mendeskripsikan hubungan beberapa variabel kovariat X dengan suatu variabel respon multinomial (*polytomous*). Model regresi logistik nominal digunakan ketika tidak ada urutan di antara kategori respon. Satu kategori diantaranya dipilih sebagai **kategori acuan**. Misalnya terdapat J kategori respon dan kategori 1 sebagai acuan, maka model logit untuk kategori selain kategori acuan dapat dituliskan sebagai

$$\text{logit}(\pi_j) = \log\left(\frac{\pi_j}{\pi_1}\right) = \mathbf{x}_j^T\beta_j, \text{ untuk } j = 2, ..., J \tag{1}$$

Terdapat $(J-1)$ persamaan logit digunakan secara simultan untuk menduga parameter β_j sehingga penduga linier dari $\mathbf{x}_j^T\beta_j$ dapat dihitung. Dari persamaan (1) dapat diperoleh

$$\widehat{\pi}_j = \widehat{\pi}_1 \exp\left(\mathbf{x}_j^T\widehat{\beta}_j\right)$$

Karena $\widehat{\pi}_1 + \widehat{\pi}_2 + ... + \widehat{\pi}_J = 1$, maka dapat diperoleh

$$\begin{aligned}\widehat{\pi}_1 &= \frac{1}{1+\sum_{j=2}^J \exp\left(\mathbf{x}_j^T\widehat{\beta}_j\right)} \\ \widehat{\pi}_j &= \frac{\exp\left(\mathbf{x}_j^T\widehat{\beta}_j\right)}{1+\sum_{j=2}^J \exp\left(\mathbf{x}_j^T\widehat{\beta}_j\right)}, \text{ untuk } j = 2, ..., J \end{aligned} \tag{2}$$

Jika ada urutan pada kategori respon (respon ordinal) maka model yang digunakan adalah regresi logistik ordinal. Misalkan z adalah variabel kontinu yang dapat dipotong-potong dengan titik-titik C_1, ..., C_{J-1} untuk mendefinisikan J kategori ordinal yang masing-masing dengan peluang π_1, ..., π_J dimana $\sum_{j=1}^J \pi_j = 1$. Ada beberapa model yang dapat digunakan untuk regresi logistik ordinal ini, antara lain model logit kumulatif, *proportional odds*, *adjacent categories logit*, dan *continuation ratio logit*.

Cumulative odds untuk kategori ke-j adalah

$$\frac{P(z \leq C_j}{P(z > C_j} = \frac{\pi_1 + ... + \pi_j}{\pi_{j+1} + ... + \pi_J}$$

sehingga model kumulatif logit adalah

$$\log\left(\frac{\pi_1 + ... + \pi_j}{\pi_{j+1} + ... + \pi_J}\right) = \mathbf{x}_j^T \beta_j \tag{3}$$

Jika penduga linier $\mathbf{x}_j^T \beta_j$ pada persamaan (3) memiliki *intercept* β_{0j} untuk kategori ke-j tetapi variabel kovariat tidak tergantung pada j, maka digunakan model *proportional odds*, yaitu

$$\log\left(\frac{\pi_1 + ... + \pi_j}{\pi_{j+1} + ... + \pi_J}\right) = \beta_{0j} + \beta_1 x_1 + ... + \beta_{p-1} x_{p-1} \tag{4}$$

Alternatif lainnya dari model kumulatif odd adalah rasio dari peluang sukses untuk kategori yang bersebelahan, yaitu

$$\frac{\pi_1}{\pi_2}, \frac{\pi_2}{\pi_3}, ..., \frac{\pi_{J-1}}{\pi_J}$$

sehingga model *adjacent logit* menjadi

$$\log\left(\frac{\pi_j}{\pi_{j+1}}\right) = \mathbf{x}_j^T \beta_j \tag{5}$$

Model rasio peluang lainnya adalah

$$\frac{\pi_1}{\pi_2}, \frac{\pi_1 + pi_2}{\pi_3}, ..., \frac{\pi_1 + ... + \pi_{J-1}}{\pi_J}$$

atau

$$\frac{\pi_1}{\pi_2 + ... + \pi_J}, \frac{pi_2}{\pi_3 + ... + \pi_J}, ..., \frac{\pi_{J-1}}{\pi_J}$$

sehingga model logit rasio menjadi

$$\log\left(\frac{\pi_j}{\pi_{j+1} + ... + \pi_J}\right) = \mathbf{x}_j^T \beta_j \tag{6}$$

2.2 Regresi Log-Linier

Misalkan Y_1, Y_2, ..., Y_N adalah peubah acak saling bebas dimana Y_i adalah jumlah kejadian dari frekuensi n_i untuk kovariat ke-i. Nilai harapan dari Y_i dapat ditulis sebagai

$$E(Y_i) = \mu_i = n_i \theta_i$$

Pengaruh θ_i dalam variabel penjelas biasanya dimodelkan sebagai

$$\theta_i = e^{\mathbf{x}_i^T\beta}$$

sehingga model linier terampat menjadi

$$E(Y_i) = \mu_i = n_i e^{\mathbf{x}_i^T\beta}, \text{ dimana } Y_i \sim Poisson(\mu_i)$$

Dengan menggunakan fungsi hubung log, maka diperoleh

$$\log \mu_i = \log n_i + \mathbf{x}_i^T\beta \tag{7}$$

Misalkan data frekuensi dalam tabulasi silang dua dimensi dinotasikan sebagai Y_{jk}, yaitu frekuensi untuk sel ke-(j,k) dimana j=1,...,J dan k=1,...,K sedemikian sehingga $\sum_{j=1}^{J}\sum_{k=1}^{K} Y_{jk} = n$. Jika Y_{jk} adalah peubah acak saling bebas yang mengikuti sebaran Poisson dengan parameter $E(Y_{jk}) = \mu_{jk}$, maka jumlahnya akan mengikuti sebaran Poisson dengan parameter $E(n) = \mu = \sum\sum \mu_{jk}$. Dengan demikian, sebaran peluang bersama dari Y_{jk} merupakan sebaran multinomial, yaitu

$$f(y \mid n) = n! \prod_{j=1}^{J}\prod_{k=1}^{K} \frac{\theta_{jk}^{y_j k}}{y_{jk}!}$$

dimana $\theta_{jk} = \mu_{jk}/\mu$, yang dapat diartikan sebagai peluang pengamatan pada sel ke-(j,k) dari tabel silang dua dimensi. Nilai harapan dari Y_{jk} adalah

$$E(Y_{jk}) = \mu_{jk} = n\theta_{jk}$$

Dengan fungsi hubung log maka diperoleh

$$\log \mu_{jk} = \log n + \log \theta_{jk}$$

Oleh karena itu, dengan fungsi hubung log diperoleh komponen linier

$$\log E(Y_i) = \text{konstanta } + \mathbf{x}_i^T\beta \tag{8}$$

Pengertian model log-linier digunakan untuk menjelaskan semua model linier terampat pada persamaan (8) ini.

Jika tidak ada hubungan antara dua variabel (baris dan kolom pada tabel) atau saling bebas, maka peluang bersama θ_{jk} adalah hasil kali dari peluang marjinalnya, yaitu

$$\theta_{jk} = \theta_{j.}\theta_{.k}, \text{ untuk } j = 1, ..., J \text{ dan } k = 1, ..., K.$$

Uji hipotesis dilakukan dengan membandingkan model aditif

$$\log E(Y_{jk}) = \log n + \log \theta_{j.} + \log \theta_{.k} \tag{9}$$

dengan

$$\log E(Y_{jk}) = \log n + \log \theta_{jk} \tag{10}$$

Hal ini sama dengan sidik ragam (ANOVA) untuk percobaan dua faktor tanpa ulangan sehingga persamaan (10) dapat dianalogikan sebagai

$$\log E(Y_{jk}) = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$$

dan persamaan (9) dapat dianalogikan sebagai

$$\log E(Y_{jk}) = \mu + \alpha_j + \beta_k$$

dan model minimalnya adalah

$$\log E(Y_{jk}) = \mu$$

Tabel 1: Tabel silang dua arah Y dan X

		Y			
		0	1	...	$J-1$
X	0	μ_{00}	μ_{01}	...	$\mu_{0,J-1}$
	1	μ_{10}	μ_{11}	...	$\mu_{1,J-1}$
	.	...	...	...	...
	.	...	...	...	...
	$I-1$	$\mu_{I-1,0}$	$\mu_{I-1,1}$	...	$\mu_{I-1,J-1}$

Sebagai contoh, Tabel 1 menunjukkan tabel silang dua arah Y (ada J kategori) dan X (ada I kategori). Maka model log-linier penuh (*saturated*) dapat dituliskan sebagai

$$\log(\mu_{ij}) = \mu + \alpha_i^X + \beta_j^Y + (\alpha\beta)_{ij}^{XY} \iff \mu_{ij} = e^{\mu+\alpha_i^X+\beta_j^Y+(\alpha\beta)_{ij}^{XY}} \tag{11}$$

dimana μ adalah rata-rata umum, α_i^X adalah pengaruh baris, β_j^Y adalah pengaruh kolom, dan $(\alpha\beta)_{ij}^{XY}$ adalah pengaruh interaksi baris dan kolom. Pada keadaan tertentu, ada beberapa parameter yang membutuhkan kendala, salah satunya adalah kendala sudut (*corner constraints*), yaitu $\alpha_0^X = \beta_0^Y = \alpha_{i0}^{XY} = \beta_{0j}^{XY} = 0$ untuk semua i dan j. Sebagai contoh, untuk $I = J = 2$, maka isi sel tabel 2x2 seperti dicantumkan pada Tabel 2.

Dengan demikian, rasio odd untuk tabel silang pada Tabel 2 adalah

$$\frac{\mu_{00}/\mu_{01}}{\mu_{10}/\mu_{11}} = \frac{\mu_{00}\mu_{11}}{\mu_{11}\mu_{10}} = \frac{e^{\mu}.e^{\mu+\alpha_1^X+\beta_1^Y+(\alpha\beta)_{11}^{XY}}}{e^{\mu+\beta_1^Y}.e^{\mu+\alpha_1^X}} = e^{(\alpha\beta)_{11}^{XY}} \tag{12}$$

Tabel 2: Isi Sel pada tabel silang dua arah Y dan X

		Y	
		0	1
X	0	e^{μ}	$e^{\mu+\beta_1^Y}$
	1	$e^{\mu+\alpha_1^X}$	$e^{\mu+\alpha_1^X+\beta_1^Y+(\alpha\beta)_{11}^{XY}}$

sehingga $(\alpha\beta)_{11}^{XY}$ dari persamaan (12) merupakan log dari rasio odd. Jika $(\alpha\beta)_{11}^{XY} = 0$, maka tidak ada interaksi antara X dan Y sehingga menjadi model *additive* sempurna, yaitu

$$\log(\mu_{ij}) = \mu + \alpha_i^X + \beta_j^Y$$

2.3 Hubungan Analisis Log-Linier dan Logit

Dari sel-sel yang terdapat pada Tabel 2, jika Y merupakan variabel respon dan X merupakan variabel penjelas, maka diperoleh model logit pada $X = 1$ sebagai berikut:

$$\begin{aligned} \text{logit}[P(Y=1))] &= \log\left(\frac{P(Y=1)}{P(Y=0)}\right) \\ &= \log\left(\frac{e^{\mu+\alpha_1^X+\beta_1^Y+(\alpha\beta)_{11}^{XY}}}{e^{\mu+\alpha_1^X}}\right) \\ &= \beta_1^Y + (\alpha\beta)_{11}^{XY} x \qquad (13) \end{aligned}$$

Nilai *intercept* β_1^Y disebut sebagai *baseline log odds*, yaitu nilai *log odds* dari $Y = 1$ dengan syarat $X = 0$. Sedangkan koefisien dari x merupakan beda antara *log odds* $Y = 1$ pada $X = 0$ dan $X = 1$.

Untuk tabel silang tiga arah yang melibatkan tiga variabel kategori, model log-linier dengan interaksi dua level adalah

$$\log(\mu_{ijk}) = \mu + \alpha_i^X + \beta_j^Y + \gamma_k^Z + (\alpha\beta)_{ij}^{XY} + (\alpha\gamma)_{ik}^{XZ} + (\beta\gamma)_{jk}^{YZ}$$

Misalkan Y merupakan variabel respon biner, maka

$$\begin{aligned} \text{logit}[P(Y=1)] &= \log\left(\frac{p(Y=1)}{1-p(Y=1)}\right) \\ &= \log\left(\frac{P(Y=1 \mid X=i, Z=k)}{P(Y=0 \mid X=i, Z=k)}\right) \\ &= \log\left(\frac{\mu_{i1k}}{\mu_{i0k}}\right) \end{aligned}$$

$$
\begin{aligned}
&= \log\left(\frac{e^{\mu+\alpha_i^X+\beta_1^Y+\gamma_k^Z+(\alpha\beta)_{i1}^{XY}+(\alpha\gamma)_{ik}^{XZ}+(\beta\gamma)_{1k}^{YZ}}}{e^{\mu+\alpha_i^X+\beta_0^Y+\gamma_k^Z+(\alpha\beta)_{i0}^{XY}+(\alpha\gamma)_{ik}^{XZ}+(\beta\gamma)_{0k}^{YZ}}}\right) \\
&= (\beta_1^Y-\beta_0^Y)+[(\alpha\beta)_{i1}^{XY}-(\alpha\beta)_{i0}^{XY}]+[(\beta\gamma)_{1k}^{YZ}-(\beta\gamma)_{0k}^{YZ}] \\
&= \mu+\alpha_i^X+\gamma_k^Z \qquad (14)
\end{aligned}
$$

Dengan demikian, untuk tabel silang tiga arah dengan Y sebagai respon biner, terdapat hubungan antara model log-linier dengan model logistik seperti tercantum pada Tabel 3.

Tabel 3: Model Log-Linier dan Logistik pada tabel tiga arah

Simbol Log-Linier	Model Logistik
(Y, XZ)	μ
(XY, XZ)	$\mu+\alpha_i^X$
(YZ, XZ)	$\mu+\gamma_k^Z$
(XY, YZ, XZ)	$\mu+\alpha_i^X+\gamma_k^Z$
(XYZ)	$\mu+\alpha_i^X+\gamma_k^Z+(\alpha\gamma)_{ik}^{XZ}$

2.4 Uji Kebaikan Model

Untuk mengukur tentang kesesuaian model regresi logistik maupun log-linier, ada beberapa ukuran statistik yang dapat dijadikan kriteria, di antaranya adalah Pearson Chi-square residual, Deviance, Uji Rasio likelihood, dan uji lainnya misalkan AIC. Nilai Pearson chi-squares residual dapat dihitung melalui persamaan:

$$r_i = \frac{o_i - e_i}{\sqrt{e_i}}$$

dimana o_i dan e_i adalah nilai observasi dan nilai dugaan (harapan) untuk setiap i=1,2,...,n. Nilai $\sum_{i=1}^{n} r_i$ mengikuti sebaran χ^2 dengan derajat bebas $n-p$, p adalah banyaknya parameter.
Kriteria lainnya untuk mengukur kesesuaian model adalah deviance yang didefinisikan sebagai nilai maksimum dari fungsi log-likelihood untuk model fit $l(\mathbf{b})$ dan untuk model maksimum $l(\mathbf{b}_{max})$,

$$D = 2[l(\mathbf{b}_{max}) - l(\mathbf{b})]$$

Untuk model yang baik, nilai deviance juga mempunyai kedekatan dengan sebaran χ^2 dengan derajat bebas np. Kondisi lain jika nilai antara Pearson Chi-square dan deviance relative sama dengan derajat bebas np, maka model yang dihasilkan kemungkinan mempunyai tingkat kesesuaian yang cukup.

Kriteria ketiga adalah uji rasio likelihood yang dapat diimplementasikan untuk menduga kesesuaian dari pendugaan parameter regresi dengan menggunakan MLE

(*Maximum Likelihood Estimation*). Uji ini untuk melihat kontribusi variabel penjelas terhadap variabel respon di dalam model. Dengan demikian, uji hipotesis yang dilakukan adalah

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = ... = \beta_k = 0 \\ H_1 &: \text{sedikitnya ada satu } \beta \neq 0 \end{aligned}$$

Statistik uji yang digunakan adalah $G = 2(l_1 - l_0)$ dimana l_1 adalah likelihood tanpa variabel penjelas dan l_0 adalah likelihood dengan variabel penjelas. Nilai G mengikuti sebaran χ^2 dengan derajat bebas k.

Ukuran lain yang dapat mengukur kebaikan model adalah AIC (*Akaike Information Criteria*) dan SC (*Schwartz Criteria*). AIC didefinisikan sebagai $G^2 - 2d$, d adalah derajat bebas. Nilai AIC yang semakin kecil mengindikasikan model yang baik.

3 Analisis Data

3.1 Bahan dan Metode

Analisis data dilakukan terhadap soal nomor 9.5 pada buku Dobson (2001), yaitu hasil survei tentang tingkat kepuasan kondisi rumah di Copenhagen. Responden dipilih di daerah hunian rumah yang disewa pada tahun 1960-1968. Tingkat kepuasan diukur berdasarkan derajat kontak mereka dengan penghuni lainnya. Data dikelompokkan berdasarkan tipe rumah seperti yang dicantumkan pada Tabel 4.

Tabel 4: Data yang dianalisis

	Tingkat Kepuasan					
Derajat	Rendah		Sedang		Tinggi	
Kontak	Rendah	Tinggi	Rendah	Tinggi	Rendah	Tinggi
Tower block	65	34	54	47	100	100
Apartment	130	141	76	116	111	191
House	67	130	48	105	62	104

Tingkat kepuasan terdiri atas tiga level, yaitu rendah, sedang, dan tinggi; derajat kontak terdiri atas dua level yaitu rendah dan tinggi; sedangkan tipe rumah terdiri atas tiga kategori yaitu *tower block*, *apartment*. dan *house*. Data dianalisis menggunakan program SAS v9.1 sebagai berikut:

1. Pertama data dianalisis menggunakan model regresi logistik nominal untuk melihat hubungan antara tingkat kepuasan dengan dua variabel lainnya, yaitu derajat kontak dan tipe rumah.

2. Data juga dianalisis menggunakan model regresi logistik ordinal logit kumulatif untuk selanjutnya dibandingkan dengan model regresi logistik nominal.

3. Hasil tahap (1) dan (2) dibandingkan untuk mendapatkan model terbaik.
4. Selanjutnya data dianalisis untuk melihat hubungan antara tingkat kepuasan (diperlakukan sebagai variabel kategori nominal) dengan derajat kontak reponden terhadap penghuni lainnya, terpisah untuk setiap tipe rumah, menggunakan model log-linier.
5. Analisis pada tahap (4) dilanjutkan dengan melakukan secara simultan untuk semua tipe rumah.
6. Hasil analisis tahap (4) dan (5) dibandingkan dengan model yang diperoleh pada tahap (3).

Model regresi logistik nominal dengan fungsi hubung logit yang dilakukan dapat dituliskan sebagai

$$\log\left(\frac{\pi_j}{\pi_1}\right) = \beta_{0j} + \beta_{1j}T_2 + \beta_{2j}T_3 + \beta_{3j}K_2, \text{ untuk } j = 2,3$$

dimana

$$\begin{aligned}
S_2 &= \begin{cases} 1, & \text{tingkat kepuasan sedang} \\ 0, & \text{tingkat kepuasan lainnya} \end{cases} \\
S_3 &= \begin{cases} 1, & \text{tingkat kepuasan tinggi} \\ 0, & \text{tingkat kepuasan lainnya} \end{cases} \\
T_2 &= \begin{cases} 1, & \text{tipe apartment} \\ 0, & \text{tipe lainnya} \end{cases} \\
T_3 &= \begin{cases} 1, & \text{tipe house} \\ 0, & \text{tipe lainnya} \end{cases} \\
K_2 &= \begin{cases} 1, & \text{derajat kontak tinggi} \\ 0, & \text{derajat kontak lainnya} \end{cases}
\end{aligned}$$

Model regresi logistik ordinal logit kumulatif yang digunakan dalam analisis data adalah

$$\begin{aligned}
\log\left(\frac{\pi_1}{\pi_2+\pi_3}\right) &= \beta_{01} + \beta_{11}T_2 + \beta_{12}T_3 + \beta_2 K_2 \\
\log\left(\frac{\pi_1+\pi_2}{\pi_3}\right) &= \beta_{02} + \beta_{11}T_2 + \beta_{12}T_3 + \beta_2 K_2
\end{aligned}$$

Model regresi log-linier yang digunakan dalam analisis data terdiri dari beberapa tahap. Pertama adalah model *additive* regresi log-linier untuk variabel tingkat kepuasan dan derajat kontak untuk setiap tipe rumah. Dengan demikian terdapat tiga model *additive* yang masing-masing berbentuk

$$\log(\mu_{jk}) = \mu + \beta 1_j^{S_2} + \beta 2_j^{S_3} + \gamma_k^{K_2}$$

atau

$$\mu_{jk} = e^{\mu+\beta 1_j^{S_2}+\beta 2_j^{S_3}+\gamma_k^{K_2}}$$

dimana μ adalah rataan umum, $\beta 1_j^{S_2}$ adalah pengaruh tingkat kepuasan sedang, $\beta 2_j^{S_3}$ adalah pengaruh tingkat kepuasan tinggi, dan $\gamma_k^{K_2}$ adalah pengaruh derajat kontak tinggi. Selanjutnya dianalisis untuk model *saturated* sebagai berikut:

$$\log(\mu_{jk}) = \mu + \beta 1_j^{S_2} + \beta 2_j^{S_3} + \gamma_k^{K_2} + (\beta_1\gamma)_{jk}^{S_2K_2} + (\beta_2\gamma)_{jk}^{S_3K_2}$$

Model terakhir yang digunakan untuk analisis data adalah model yang melibatkan semua variabel, dimulai dengan model *additive* sempurna, yaitu:

$$\log(\mu_{ijk}) = \mu + \alpha 1_j^{T_2} + \alpha 2_j^{T_3} + \beta 1_j^{S_2} + \beta 2_j^{S_3} + \gamma_k^{K_2}$$

dimana $\alpha 1_j^{T_2}$ adalah pengaruh tipe rumah *apartment* dan $\alpha 2_j^{T_3}$ adalah pengaruh tipe rumah *house*. Model selanjutnya adalah model *saturated* dengan interaksi yang dibatasi hanya untuk dua level, yaitu

$$\begin{aligned}\log(\mu_{ijk}) &= \mu + \alpha 1_j^{T_2} + \alpha 2_j^{T_3} + \beta 1_j^{S_2} + \beta 2_j^{S_3} + \gamma_k^{K_2} + \\ &\quad (\alpha_1\beta_1)_{ij}^{T_2S_2} + (\alpha_1\beta_2)_{ij}^{T_2S_3} + (\alpha_1\gamma)_{ik}^{T_2K_2} + \\ &\quad (\beta_1\gamma)_{jk}^{S_2K_2} + (\beta_1\gamma)_{jk}^{S_3K_2} \qquad (15)\end{aligned}$$

dan model *saturated* untuk semua interaksi, yaitu:

$$\begin{aligned}\log(\mu_{ijk}) &= \mu + \alpha 1_j^{T_2} + \alpha 2_j^{T_3} + \beta 1_j^{S_2} + \beta 2_j^{S_3} + \gamma_k^{K_2} + \\ &\quad (\alpha_1\beta_1)_{ij}^{T_2S_2} + (\alpha_1\beta_2)_{ij}^{T_2S_3} + (\alpha_1\gamma)_{ik}^{T_2K_2} + \\ &\quad (\beta_1\gamma)_{jk}^{S_2K_2} + (\beta_1\gamma)_{jk}^{S_3K_2} + \\ &\quad (\alpha_1\beta_1\gamma)_{ijk}^{T_2S_2K_2} + (\alpha_1\beta_2\gamma)_{ijk}^{T_2S_3K_2} + \\ &\quad (\alpha_2\beta_1\gamma)_{ijk}^{T_3S_2K_2} + (\alpha_2\beta_2\gamma)_{ijk}^{T_3S_3K_2} \qquad (16)\end{aligned}$$

3.2 Hasil Analisis

Data disusun dan dibaca dengan prosedur SAS sebagai berikut:

```
data tugas3;
input count T2 T3 S2 S3 K2 Y $ @@@@@@@;
datalines;
 65 0 0 0 0 0 0  130 1 0 0 0 0 0   67 0 1 0 0 0 0
 34 0 0 0 0 1 0  141 1 0 0 0 1 0  130 0 1 0 0 1 0
 54 0 0 1 0 0 1   76 1 0 1 0 0 1   48 0 1 1 0 0 1
 47 0 0 1 0 1 1  116 1 0 1 0 1 1  105 0 1 1 0 1 1
100 0 0 0 1 0 2  111 1 0 0 1 0 2   62 0 1 0 1 0 2
100 0 0 0 1 1 2  191 1 0 0 1 1 2  104 0 1 0 1 1 2
;
```

3.2.1 Plot Data

Hasil plot proporsi data antara tingkat kepuasan responden (S) dengan derajat kontak (K) menunjukkan hasil yang berbeda pada setiap tipe rumah (Gambar 1). Pada tipe TOWER HOUSE tingkat kepuasan responden lebih tinggi pada derajat kontak rendah, dan sebaliknya, pada tipe APARTMENT tingkat kepuasan responden lebih rendah pada derajat kontak rendah. Pada kedua tipe rumah ini terlihat ada perpotongan garis yang menunjukkan adanya korelasi antara tingkat kepuasan dan derajat kontak. Sedangkan pada tipe HOUSE, kedua garis terlihat sejajar dan tingkat kepuasan responden lebih tinggi pada derajat kontak rendah. Perbedaan ketiga kurva pada Gambar 1 tersebut menunjukkan bahwa tipe rumah memberi kontribusi yang berbeda, atau terdapat interaksi antara tipe rumah, derajat kontak, dan tingkat kepuasan responden.

3.2.2 Model Regresi Logistik

Berdasarkan tabel data sebelumnya, selanjutnya digunakan PROC LOGISTIC untuk menduga model nominal dengan fungsi hubung logit dan model ordinal sebagai berikut:

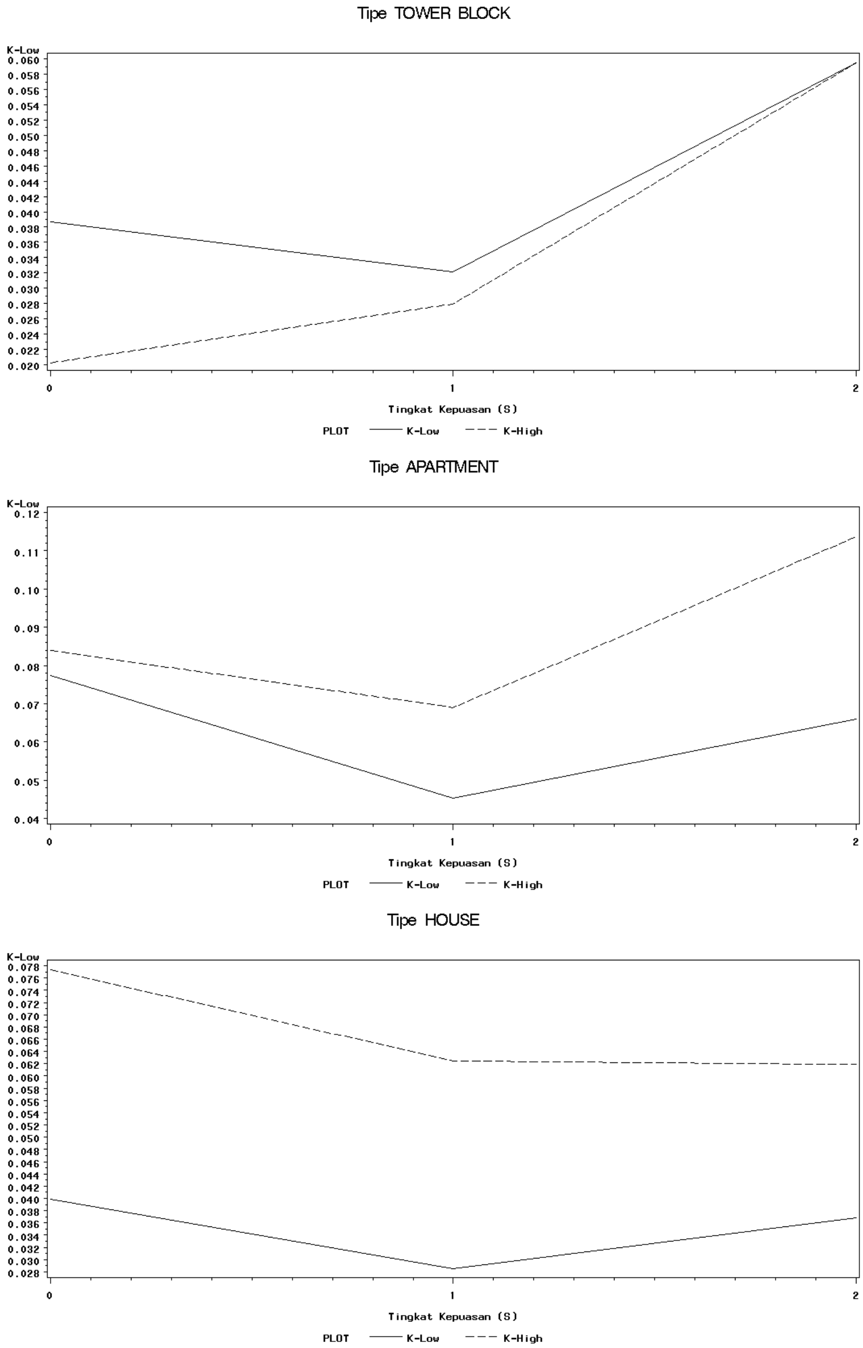
```
proc logistic data=tugas3;
   title 'Hasil analisis model logistik nominal';
   weight count;
   model Y (REFERENCE="0")= T2 T3 K2/link=glogit scale=none aggregate;
   output out = hasil PRED=PREDICTED PREDPROBS=I C;
   run;
proc print data=hasil;
   run;

proc logistic data=tugas3;
   title 'Hasil analisis model logistik ordinal';
   weight count;
   model Y (REFERENCE="0")= T2 T3 K2/link=clogit scale=none aggregate;
   output out = hasil PRED=PREDICTED PREDPROBS=I C;
   run;
proc print data=hasil;
   run;
```

Model dugaan regresi logistik nominal yang dihasilkan adalah

$$\begin{aligned} \log\left(\frac{\widehat{\pi}_2}{\widehat{\pi}_1}\right) &= -0.1071 - 0.4069T_2 - 0.3371T_3 + 0.2959K_2 \\ \log\left(\frac{\widehat{\pi}_3}{\widehat{\pi}_1}\right) &= 0.5605 - 0.6413T_2 - 0.9453_T3 + 0.3282K_2 \end{aligned} \tag{17}$$

sedangkan dugaan regresi logistik ordinal adalah

$$\begin{aligned} \log\left(\frac{\widehat{\pi}_1}{\widehat{\pi}_2+\widehat{\pi}_3}\right) &= -0.9973 + 0.5009T_2 + 0.7362T_3 - 0.2525K_2 \\ \log\left(\frac{\widehat{\pi}_1+\widehat{\pi}_2}{\widehat{\pi}_3}\right) &= 0.1152 + 0.5009T_2 + 0.7362T_3 - 0.2525K_2 \end{aligned} \tag{18}$$



Gambar 1: Plot data tingkat kepuasan dan derajat kontak setiap tipe rumah

Berdasarkan nilai deviance dan Pearson untuk model nominal dan ordinal pada Tabel 6 menunjukkan bahwa kedua model dapat digunakan untuk menduga parameter. Hal ini juga dapat dilihat dari AIC, SC, dan -2 Log L pada Tabel 7. Namun demikian, model nominal lebih baik dibanding dengan model ordinal, dimana deviance untuk model nominal sebesar 6.8930 lebih baik dibanding untuk

Tabel 5: Penduga maksimum likelihood model logistik

Parameter	Y	DF	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq
Model 1: Nominal						
Intercept	1	1	-0.1071	0.1524	0.4943	0.4820
Intercept	2	1	0.5605	0.1329	17.7811	<.0001
T2	1	1	-0.4069	0.1713	5.6415	0.0175
T2	2	1	-0.6413	0.1501	18.2623	<.0001
T3	1	1	-0.3371	0.1803	3.4947	0.0616
T3	2	1	-0.9453	0.1645	33.0304	<.0001
K2	1	1	0.2959	0.1301	5.1742	0.0229
K2	2	1	0.3282	0.1182	7.7114	0.0055
Model 2: Ordinal						
Intercept 0		1	-0.9973	0.1072	86.5461	<.0001
Intercept 1		1	0.1152	0.1044	1.2175	0.2699
T2		1	0.5009	0.1166	18.4667	<.0001
T3		1	0.7362	0.1267	33.7810	<.0001
K2		1	-0.2525	0.0929	7.3936	0.0065

Tabel 6: Statistik deviance dan pearson untuk model logistik

Kriteria	Model Nominal		Model Ordinal	
	Nilai	Pr>ChiSq	Nilai	Pr>ChiSq
Deviance	6.8930	0.1416	11.6991	0.1109
Pearson	6.9323	0.1395	11.6419	0.1130

model ordinal yang nilainya lebih tinggi, yaitu 11.6991.

3.2.3 Model Regresi Log-Linier

Untuk setiap data tipe rumah diolah dengan program SAS sebagai berikut

```
proc genmod data=tugas3a order=internal;
   title 'Model additive per tipe rumah';
   model count = S2 S3 K2
      /link=log dist=poisson lrci type3 pred;
   run;
proc genmod data=tugas3a order=internal;
   title 'Model saturated per tipe rumah';
   model count = S2 S3 K2 S2*K2 S3*K2
      /link=log dist=poisson lrci type3 pred;
   run;
```

Model *additive* dugaan regresi log-linier yang dihasilkan untuk setiap tipe rumah adalah

$$\log(\mu_{0jk}) = 3.9927 + 0.0200^{S_2} + 0.7032^{S_3} - 0.1906^{K_2}$$

Tabel 7: Statistik fit model logistik

Kriteria	Model Nominal		Model Ordinal	
	Minimum	Fit	Minimum	Fit
AIC	3652.878	3621.480	3652.878	3620.286
SC	3654.658	3628.603	3663.732	3647.422
-2 Log L	3648.878	3605.480	3648.878	3610.286

Tabel 8: Penduga parameter dan deviance model untuk setiap tipe rumah

Parameter	Tower Block		Apartment		House	
	Additive	Saturated	Additive	Saturated	Additive	Saturated
Intercept	*3.9927	*4.1744	*4.7211	*4.8675	*4.2132	*4.2047
S_2	0.0200	-0.1854	*-0.3446	*-0.5368	*-0.2528	-0.3335
S_3	*0.7032	*0.4308	0.1083	-0.1580	-0.1712	-0.0776
K_2	-0.1906	*-0.6480	*0.3459	0.0812	0.6498	*0.6628
$S_2 * K_2$		0.5092		0.3416		0.1199
$S_3 * K_2$		*0.6480		*0.4615		-0.1456
Deviance	6.7424		7.7448		1.2736	
Pearson	6.6422		7.7670		1.2741	

* berbeda nyata pada taraf 5%

$$\begin{aligned} \log(\mu_{1jk}) &= 4.7211 - 0.3446^{S_2} + 0.1083^{S_3} + 0.3459^{K_2} \\ \log(\mu_{2jk}) &= 4.2132 - 0.2528^{S_2} - 0.1712^{S_3} + 0.6498^{K_2} \end{aligned} \tag{19}$$

untuk j=0,1(S_2),2(S_3) dan k=0,1(K_2). Sedangkan model *saturated*-nya adalah

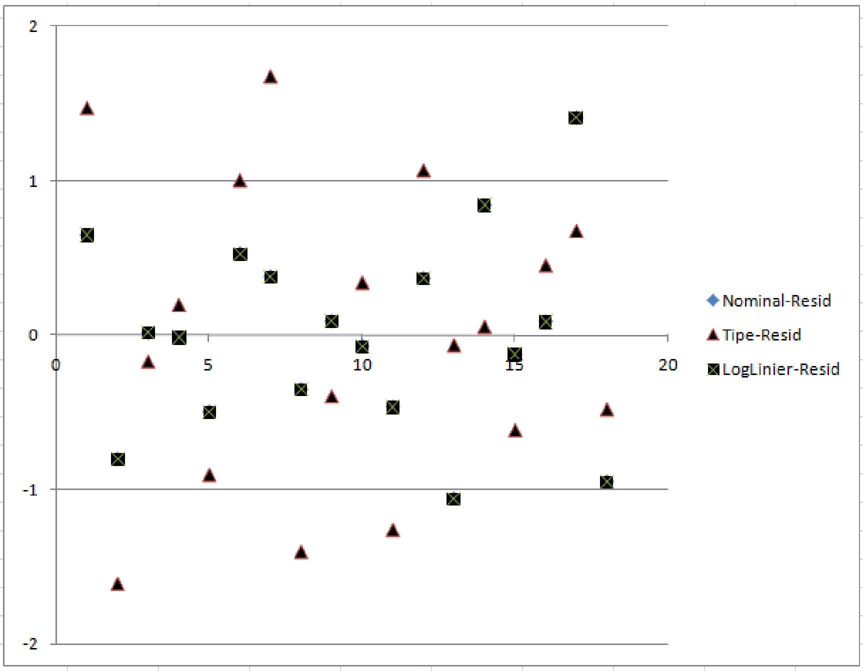
$$\begin{aligned} \log(\mu_{0jk}) &= 4.1744 - 0.1854^{S_2} + 0.4308^{S_3} - 0.6480^{K_2} + 0.5092^{S_2K_2} + 0.6480^{S_3K_2} \\ \log(\mu_{1jk}) &= 4.8675 - 0.5368^{S_2} - 0.1580^{S_3} + 0.0812^{K_2} + 0.3416^{S_2K_2} + 0.4615^{S_3K_2} \\ \log(\mu_{2jk}) &= 4.2047 - 0.3335^{S_2} - 0.0776^{S_3} + 0.6628^{K_2} + 0.1199^{S_2K_2} - 0.1456^{S_3K_2} \end{aligned} \tag{20}$$

Tipe rumah memberikan pengaruh yang berbeda ke dalam setiap model. Hal ini ditunjukkan oleh hasil uji penduga parameter yang berbeda pada Tabel 8, dimana $S2$ tidak nyata pada tipe *Tower Block* tetapi nyata pada tipe lainnya. Sebaliknya, $S3$ nyata pada tipe *Tower Block* tetapi tidak nyata pada tipe lainnya. Sedangkan *intercept* berbeda nyata pada setiap model, dan tidak terlihat ada interaksi antara $S2$ dengan $K2$. Nilai deviance untuk tipe *house* sebesar 1.2736 pada model *saturated* lebih kecil dari nilai $\chi^2_{(2,0.5)}$=5.99, sehingga model ini paling sesuai dibanding model pada dua tipe rumah lainnya. Hal ini juga tercermin dari nilai Pearson chi-square pada Tabel 8.

Hasil analisis log-linier setiap tipe rumah menunjukkan hasil yang berbeda, yang sesuai dengan hasil plot data pada Gambar 1. Hal ini menunjukkan bahwa tipe

Tabel 9: Penduga parameter dan deviance model semua variabel

Parameter	Additive		2 Level		3 Level	
	Penduga	Pr>ChiSq	Penduga	Pr>ChiSq	Penduga	Pr>ChiSq
Intercept	4.0470	<.0001	4.0943	<.0001	4.1744	<.0001
T_2	0.6484	<.0001	0.7402	<.0001	0.6931	<.0001
T_3	0.2546	0.0001	0.2395	0.0910	0.0303	0.8618
S_2	-0.2400	0.0001	-0.1073	0.4816	-0.1854	0.3140
S_3	0.1639	0.0041	0.5608	<.0001	0.4308	0.0069
K_2	0.3058	<.0001	-0.4306	0.0009	-0.6480	0.0022
$T_2 * S_2$			-0.4068	0.0176	-0.3514	0.1332
$T_3 * S_2$			-0.3371	0.0616	-0.1481	0.5747
$S_2 * K_2$			0.2960	0.0229	0.5092	0.0800
$T_2 * S_3$			-0.6416	<.0001	-0.5888	0.0041
$T_3 * S_3$			-0.9456	<.0001	-0.5083	0.0324
$S_3 * K_2$			0.3282	0.0055	0.6480	0.0109
$T_2 * K_2$			0.5744	<.0001	0.7293	0.0028
$T_3 * K_2$			0.8906	<.0001	1.3109	<.0001
$T_2 * S_2 * K_2$					-0.1676	0.6302
$T_3 * S_2 * K_2$					-0.3893	0.2939
$T_2 * S_3 * K_2$					-0.1865	0.5426
$T_3 * S_3 * K_2$					-0.7936	0.0183
Deviance	89.3481		6.8930			
Pearson	85.3473		6.9323			



Gambar 2: Plot residual model nominal dan log-linier

Tabel 10: Penduga parameter dan deviance model semua variabel

Obs	Count	Nominal		Per Tipe		2 Level	
		Pred	Resid	Pred	Resid	Pred	Resid
1	65	59.9994	0.6456	54.2025	1.4666	59.9947	0.6462
2	34	39.0109	-0.8023	44.7975	-1.6132	39.0053	-0.8014
3	54	53.9047	0.0130	55.2975	-0.1745	53.8930	0.0146
4	47	47.1161	-0.0169	45.7025	0.1919	47.1070	-0.0156
5	100	105.0959	-0.4971	109.5000	-0.9079	105.1123	-0.4986
6	100	94.8748	0.5262	90.5000	0.9986	94.8877	0.5248
7	130	125.7698	0.3772	112.2967	1.6706	125.7714	0.3771
8	141	145.2282	-0.3509	158.7033	-1.4053	145.2286	-0.3509
9	76	75.2241	0.0895	79.5608	-0.3992	75.2223	0.0897
10	116	116.7757	-0.0718	112.4392	0.3358	116.7777	-0.0720
11	111	116.0062	-0.4648	125.1425	-1.2642	116.0063	-0.4648
12	191	185.9917	0.3672	176.8575	1.0634	185.9937	0.3671
13	67	76.2321	-1.0574	67.5757	-0.0700	76.2338	-1.0576
14	130	120.7654	0.8403	129.4244	0.0506	120.7662	0.8403
15	48	48.8856	-0.1267	52.4828	-0.6188	48.8847	-0.1265
16	105	104.1137	0.0869	100.5176	0.4471	104.1153	0.0867
17	62	51.8822	1.4047	56.9421	0.6703	51.8814	1.4048
18	104	114.1210	-0.9474	109.0583	-0.4844	114.1186	-0.9472
Pearson			6.9328		15.6832		6.9324

rumah memberikan kontribusi yang berbeda terhadap model, sehingga perlu dilakukan analisis model log-linier yang melibatkan ketiga variabel secara simultan. Oleh karena itu, model selanjutnya yang dianalisis adalah model yang melibatkan semua variabel yang hasilnya dituangkan pada Tabel 9.

Dengan menambah interaksi dua level ke dalam model, diperoleh dugaan yang lebih baik dibanding model *additive* sempurna. Nilai deviance sebesar 6.8930 dan Pearson sebesar 6.9323 jauh lebih rendah dibanding pada model *additive* sempurna yaitu berturut-turut 89.3481 dan 85.3473 (Tabel 9). Uji terhadap setiap parameter juga memperlihatkan hasil yang lebih baik setelah menambahkan interaksi dua level ke dalam model. Model log-linier dengan interaksi sampai dengan dua level dari setiap variabel ternyata menghasilkan dugaan dan *Pearson residual* yang sama dengan model logistik nominal (Tabel 6 dan 10; Gambar 2). Nilai deviance model log-linier dengan interaksi dua level pada Tabel 9 juga sama dengan nilai deviance untuk model logistik nominal pada Tabel 6. Dengan demikian, model yang sudah diturunkan pada persamaan (14) dan hubungan antara model log-linier dan model logit pada Tabel 3 dapat ditunjukkan dalam analisis ini.

Walaupun model log-linier dengan model regresi nominal menunjukkan kesamaan, tetapi intrepretasi terhadap model log-linier lebih kompleks karena melibatkan banyak variabel. Dari hasil penelitiannya, Jeansonne (1975) juga menyatakan bahwa untuk tabel dua arah, frekuensi yang diharapkan pada suatu sel harus lebih besar dari satu, dan tidak lebih dari 20% yang kurang dari lima. Jika terjadi maka ada beberapa cara harus dilakukan, antara lain menggabungkan beberapa variabel atau menambahkan suatu nilai yang sama ke setiap sel.

4 Kesimpulan

Berdasarkan analisis yang telah dilakukan telah diperoleh beberapa kesimpulan, yaitu model log-linier yang melibatkan tiga variabel untuk data tabel tiga arah dengan interaksi dua level merupakan model yang sesuai dibanding hanya menggunakan model *additive* sempurna. Model ini juga memberikan hasil yang sama dengan model regresi logistik nominal dengan menggunakan fungsi hubung logit, baik berdasarkan hasil analisis numerik terhadap data contoh maupun berdasarkan tinjauan matematis. Walaupun demikian, intrepretasi hasil analisis menggunakan model log-linier lebih kompleks dibanding model regresi logistik karena melibatkan banyak variabel, dan modelnya juga lebih kompleks.

5 Daftar Pustaka

Agresti, A. 2007. ***An Introduction to Categorical Data Analysis***. 2^{nd} Ed. John Wiley and Sons, Inc.

Dobson, A.J. 2001. ***An Introduction to Generalized Linear Models***. Chapman Hall/CRC Texts in Statistical Science Series.

Jeansonne, A. 1975. ***Loglinear Models***. http://www.statsoftinc.com/textbook/stloglin.html.

McCullagh,P. and Nelder,J.A. 1983. ***Generalized Linear Models***. 2^{nd} Ed. Chapman and Hall.

Ragavan, A.J. 2008. ***How to use SAS to fit Multiple Logistic Regression Models***. SAS Global Forum 2008.

Saparita, R. 1999. **Model Regresi Logistik untuk Respon Kualitatif**. Buletin IPT No.5 Vol.IV.

Zeileis, A; C.Kleiber and S.Jackman. 2008. ***Regression Models for Count Data in R***. Department of Statistics and Mathematics, Wirtschaftsuniversitat Wien, Austria.

6 Lampiran

6.1 Output SAS untuk model log-linier tipe TOWER BLOCK

The GENMOD Procedure

Model Information

Data Set	WORK.TUGAS3A
Distribution	Poisson
Link Function	Log
Dependent Variable	count

Number of Observations Read	6
Number of Observations Used	6

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2	6.7424	3.3712
Scaled Deviance	2	6.7424	3.3712
Pearson Chi-Square	2	6.6422	3.3211
Scaled Pearson X2	2	6.6422	3.3211
Log Likelihood		1305.2564	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	3.9927	0.1103	3.7702	4.2029	1310.23	<.0001
S2	1	0.0200	0.1414	-0.2575	0.2978	0.02	0.8875
S3	1	0.7032	0.1229	0.4653	0.9476	32.75	<.0001
K2	1	-0.1906	0.1005	-0.3882	0.0059	3.60	0.0578
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
S2	1	0.02	0.8875
S3	1	34.80	<.0001
K2	1	3.62	0.0572

Observation Statistics

Observation	count	Pred	Xbeta	Std	HessWgt
1	65	54.202503	3.9927271	0.1103051	54.202503
2	34	44.797502	3.8021524	0.1145682	44.797502
3	54	55.2975	4.0127277	0.1093947	55.2975
4	47	45.7025	3.822153	0.1136919	45.7025
5	100	109.5	4.6959245	0.0840608	109.5
6	100	90.5	4.5053499	0.0895816	90.5

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	0	0.0000	.
Scaled Deviance	0	0.0000	.
Pearson Chi-Square	0	0.0000	.
Scaled Pearson X2	0	0.0000	.
Log Likelihood		1308.6275	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	4.1744	0.1240	3.9210	4.4080	1132.66	<.0001
S2	1	-0.1854	0.1841	-0.5493	0.1744	1.01	0.3140
S3	1	0.4308	0.1593	0.1213	0.7472	7.31	0.0069
K2	1	-0.6480	0.2117	-1.0734	-0.2407	9.37	0.0022
S2*K2	1	0.5092	0.2908	-0.0580	1.0843	3.07	0.0800
S3*K2	1	0.6480	0.2546	0.1543	1.1542	6.48	0.0109
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
S2	1	1.02	0.3129
S3	1	7.48	0.0062
K2	1	9.87	0.0017
S2*K2	1	3.09	0.0786
S3*K2	1	6.65	0.0099

Observation Statistics

Observation	count	Pred	Xbeta	Std	HessWgt
1	65	65	4.1743873	0.1240347	65
2	34	34	3.5263605	0.1714986	34
3	54	54	3.988984	0.1360828	54
4	47	47	3.8501476	0.145865	47
5	100	100	4.6051702	0.1	100
6	100	100	4.6051702	0.1	100

6.2 Output SAS untuk model log-linier tipe APARTMENT

Model Information

Data Set	WORK.TUGAS3A
Distribution	Poisson
Link Function	Log
Dependent Variable	count

Number of Observations Read	6
Number of Observations Used	6

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2	7.7448	3.8724
Scaled Deviance	2	7.7448	3.8724
Pearson Chi-Square	2	7.7670	3.8835
Scaled Pearson X2	2	7.7670	3.8835
Log Likelihood		2968.1765	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	4.7211	0.0744	4.5727	4.8645	4025.20	<.0001
S2	1	-0.3446	0.0943	-0.5306	-0.1606	13.35	0.0003
S3	1	0.1083	0.0837	-0.0555	0.2726	1.68	0.1955
K2	1	0.3459	0.0734	0.2026	0.4904	22.21	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
S2	1	13.55	0.0002
S3	1	1.68	0.1952
K2	1	22.54	<.0001

Observation Statistics

Observation	count	Pred	Xbeta	Std	HessWgt
1	130	112.29673	4.7211448	0.0744139	112.29673
2	141	158.70327	5.0670362	0.0679337	158.70327
3	76	79.560784	4.3765213	0.0839983	79.560784
4	116	112.43922	4.7224128	0.0783153	112.43922
5	111	125.14248	4.829453	0.0718237	125.14248
6	191	176.85752	5.1753444	0.0650862	176.85752

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	0	0.0000	.
Scaled Deviance	0	0.0000	.
Pearson Chi-Square	0	0.0000	.
Scaled Pearson X2	0	0.0000	.
Log Likelihood		2972.0489	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	4.8675	0.0877	4.6906	5.0346	3080.08	<.0001
S2	1	-0.5368	0.1444	-0.8238	-0.2568	13.82	0.0002
S3	1	-0.1580	0.1292	-0.4125	0.0948	1.49	0.2215
K2	1	0.0812	0.1216	-0.1570	0.3202	0.45	0.5041
S2*K2	1	0.3416	0.1912	-0.0319	0.7183	3.19	0.0740
S3*K2	1	0.4615	0.1704	0.1284	0.7967	7.34	0.0068
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
S2	1	14.32	0.0002
S3	1	1.50	0.2208
K2	1	0.45	0.5039
S2*K2	1	3.21	0.0731
S3*K2	1	7.38	0.0066

Observation Statistics

Observation	count	Pred	Xbeta	Std	HessWgt
1	130	130	4.8675345	0.0877058	130
2	141	141	4.9487599	0.0842152	141
3	76	76	4.3307333	0.1147079	76
4	116	116	4.7535902	0.0928477	116
5	111	111	4.7095302	0.0949158	111
6	191	191	5.2522734	0.0723575	191

6.3 Output SAS untuk model log-linier tipe HOUSE

Model Information

Data Set	WORK.TUGAS3A
Distribution	Poisson
Link Function	Log
Dependent Variable	count

Number of Observations Read	6
Number of Observations Used	6

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2	1.2736	0.6368
Scaled Deviance	2	1.2736	0.6368
Pearson Chi-Square	2	1.2741	0.6371
Scaled Pearson X2	2	1.2741	0.6371
Log Likelihood		1811.2396	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	4.2132	0.0937	4.0255	4.3931	2020.00	<.0001
S2	1	-0.2528	0.1078	-0.4651	-0.0423	5.50	0.0190
S3	1	-0.1712	0.1054	-0.3785	0.0349	2.64	0.1041
K2	1	0.6498	0.0927	0.4697	0.8335	49.11	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
S2	1	5.55	0.0185
S3	1	2.65	0.1035
K2	1	51.73	<.0001

Observation Statistics

Observation	count	Pred	Xbeta	Std	HessWgt
1	67	67.575717	4.2132487	0.0937436	67.575717
2	130	129.42435	4.8630965	0.0780257	129.42435
3	48	52.482767	3.9604849	0.1012307	52.482767
4	105	100.51758	4.6103327	0.0868781	100.51758
5	62	56.942087	4.0420347	0.0986702	56.942087
6	104	109.05829	4.6918825	0.0838806	109.05829

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	0	0.0000	.
Scaled Deviance	0	0.0000	.
Pearson Chi-Square	0	0.0000	.
Scaled Pearson X2	0	0.0000	.
Log Likelihood		1811.8764	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	4.2047	0.1222	3.9553	4.4350	1184.52	<.0001
S2	1	-0.3335	0.1891	-0.7090	0.0344	3.11	0.0778
S3	1	-0.0776	0.1762	-0.4246	0.2679	0.19	0.6599
K2	1	0.6628	0.1504	0.3722	0.9628	19.43	<.0001
S2*K2	1	0.1199	0.2302	-0.3297	0.5739	0.27	0.6024
S3*K2	1	-0.1456	0.2199	-0.5771	0.2860	0.44	0.5080
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
S2	1	3.15	0.0758
S3	1	0.19	0.6597

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
K2	1	20.51	<.0001
S2*K2	1	0.27	0.6019
S3*K2	1	0.44	0.5080

Observation Statistics

Observation	count	Pred	Xbeta	Std	HessWgt
1	67	67	4.2046926	0.1221694	67
2	130	130	4.8675345	0.0877058	130
3	48	48	3.871201	0.1443376	48
4	105	105	4.6539604	0.09759	105
5	62	62	4.1271344	0.1270001	62
6	104	104	4.6443909	0.0980581	104

6.4 Output SAS untuk model lengkap

Model Information

Data Set	WORK.TUGAS3A
Distribution	Poisson
Link Function	Log
Dependent Variable	count

Number of Observations Read	18
Number of Observations Used	18

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	12	89.3481	7.4457
Scaled Deviance	12	89.3481	7.4457
Pearson Chi-Square	12	85.3473	7.1123
Scaled Pearson X2	12	85.3473	7.1123
Log Likelihood		6047.8787	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	4.0470	0.0669	3.9144	4.1766	3658.81	<.0001
T2	1	0.6484	0.0617	0.5282	0.7701	110.43	<.0001
T3	1	0.2546	0.0666	0.1244	0.3856	14.61	0.0001
S2	1	-0.2400	0.0633	-0.3644	-0.1163	14.38	0.0001
S3	1	0.1639	0.0571	0.0522	0.2760	8.24	0.0041
K2	1	0.3058	0.0494	0.2092	0.4027	38.38	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
T2	1	116.30	<.0001
T3	1	14.73	0.0001

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
S2	1	14.49	0.0001
S3	1	8.27	0.0040
K2	1	38.83	<.0001

Observation Statistics

Observation	count	Pred	Xbeta	Std	HessWgt
1	65	57.22668	4.0470202	0.066906	57.22668
2	34	77.693461	4.3527711	0.0640855	77.693461
3	54	45.014255	3.8069792	0.0703911	45.014255
4	47	61.113335	4.1127301	0.0677158	61.113335
5	100	67.420454	4.2109484	0.0648826	67.420454
6	100	91.532977	4.5166993	0.0619701	91.532977
7	130	109.44572	4.6954287	0.0573028	109.44572
8	141	148.58832	5.0011796	0.0539828	148.58832
9	76	86.089519	4.4553877	0.0613359	86.089519
10	116	116.87892	4.7611385	0.0582463	116.87892
11	111	128.94125	4.8593569	0.0549268	128.94125
12	191	175.05632	5.1651078	0.0514537	175.05632
13	67	73.822215	4.3016597	0.0625652	73.822215
14	130	100.22429	4.6074106	0.0595394	100.22429
15	48	58.068229	4.0616187	0.0662789	58.068229
16	105	78.835985	4.3673696	0.0634305	78.835985
17	62	86.972146	4.4655879	0.0603966	86.972146
18	104	118.07722	4.7713388	0.0572563	118.07722

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4	6.8930	1.7233
Scaled Deviance	4	6.8930	1.7233
Pearson Chi-Square	4	6.9323	1.7331
Scaled Pearson X2	4	6.9323	1.7331
Log Likelihood		6089.1063	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	4.0943	0.1127	3.8666	4.3086	1320.42	<.0001
T2	1	0.7402	0.1302	0.4882	0.9990	32.34	<.0001
T3	1	0.2395	0.1417	-0.0368	0.5194	2.86	0.0910
S2	1	-0.1073	0.1524	-0.4066	0.1916	0.50	0.4816
S3	1	0.5608	0.1329	0.3027	0.8243	17.80	<.0001
K2	1	-0.4306	0.1293	-0.6851	-0.1780	11.09	0.0009
T2*S2	1	-0.4068	0.1713	-0.7433	-0.0712	5.64	0.0176
T3*S2	1	-0.3371	0.1804	-0.6914	0.0161	3.49	0.0616
S2*K2	1	0.2960	0.1301	0.0415	0.5517	5.18	0.0229
T2*S3	1	-0.6416	0.1501	-0.9382	-0.3495	18.28	<.0001
T3*S3	1	-0.9456	0.1645	-1.2705	-0.6253	33.05	<.0001
S3*K2	1	0.3282	0.1182	0.0969	0.5604	7.71	0.0055
T2*K2	1	0.5744	0.1256	0.3289	0.8213	20.93	<.0001
T3*K2	1	0.8906	0.1387	0.6198	1.1639	41.21	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
T2	1	34.41	<.0001
T3	1	2.88	0.0895
S2	1	0.50	0.4815
S3	1	18.45	<.0001
K2	1	11.20	0.0008
T2*S2	1	5.64	0.0175
T3*S2	1	3.50	0.0614
S2*K2	1	5.20	0.0226
T2*S3	1	18.79	<.0001
T3*S3	1	34.14	<.0001
S3*K2	1	7.74	0.0054
T2*K2	1	21.12	<.0001
T3*K2	1	42.12	<.0001

Observation Statistics

Observation	count	Pred	Xbeta	Std	HessWgt
1	65	59.994744	4.094257	0.1126728	59.994744
2	34	39.005256	3.6636964	0.1274296	39.005256
3	54	53.892982	3.9870003	0.1171145	53.892982
4	47	47.107018	3.852422	0.1220412	47.107018
5	100	105.11227	4.6550291	0.0887615	105.11227
6	100	94.887726	4.5526944	0.0923709	94.887726
7	130	125.77143	4.8344662	0.0809484	125.77143
8	141	145.22857	4.9783088	0.0764002	145.22857
9	76	75.222278	4.3204474	0.099682	75.222278
10	116	116.77772	4.7602723	0.0846769	116.77772
11	111	116.00629	4.7536444	0.0830601	116.00629
12	191	185.99371	5.2257129	0.0686071	185.99371
13	67	76.233823	4.3338052	0.0992134	76.233823
14	130	120.76618	4.7938563	0.0835209	120.76618
15	48	48.88474	3.8894653	0.1167408	48.88474
16	105	104.11526	4.6454986	0.0899973	104.11526
17	62	51.881437	3.9489611	0.1119801	51.881437
18	104	114.11856	4.7372379	0.0858531	114.11856

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	0	0.0000	.
Scaled Deviance	0	0.0000	.
Pearson Chi-Square	0	0.0000	.
Scaled Pearson X2	0	0.0000	.
Log Likelihood		6092.5528	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	4.1744	0.1240	3.9210	4.4080	1132.66	<.0001
T2	1	0.6931	0.1519	0.3998	0.9964	20.82	<.0001
T3	1	0.0303	0.1741	-0.3114	0.3726	0.03	0.8618
S2	1	-0.1854	0.1841	-0.5493	0.1744	1.01	0.3140
S3	1	0.4308	0.1593	0.1213	0.7472	7.31	0.0069
K2	1	-0.6480	0.2117	-1.0734	-0.2407	9.37	0.0022
T2*S2	1	-0.3514	0.2340	-0.8107	0.1078	2.26	0.1332
T3*S2	1	-0.1481	0.2639	-0.6671	0.3690	0.31	0.5747
S2*K2	1	0.5092	0.2908	-0.0580	1.0843	3.07	0.0800
T2*S3	1	-0.5888	0.2051	-0.9937	-0.1887	8.24	0.0041
T3*S3	1	-0.5083	0.2376	-0.9765	-0.0441	4.58	0.0324
S3*K2	1	0.6480	0.2546	0.1543	1.1542	6.48	0.0109
T2*K2	1	0.7293	0.2441	0.2566	1.2156	8.93	0.0028
T3*K2	1	1.3109	0.2596	0.8086	1.8283	25.49	<.0001
T2*S2*K2	1	-0.1676	0.3481	-0.8528	0.5131	0.23	0.6302
T3*S2*K2	1	-0.3893	0.3709	-1.1185	0.3368	1.10	0.2939
T2*S3*K2	1	-0.1865	0.3063	-0.7915	0.4105	0.37	0.5426

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
T3*S3*K2	1	-0.7936	0.3364	-1.4571	-0.1372	5.57	0.0183
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
T2	1	22.09	<.0001
T3	1	0.03	0.8618
S2	1	1.02	0.3129
S3	1	7.48	0.0062
K2	1	9.87	0.0017
T2*S2	1	2.25	0.1334
T3*S2	1	0.32	0.5746
S2*K2	1	3.09	0.0786
T2*S3	1	8.35	0.0039
T3*S3	1	4.61	0.0318
S3*K2	1	6.65	0.0099
T2*K2	1	9.24	0.0024
T3*K2	1	26.91	<.0001
T2*S2*K2	1	0.23	0.6300
T3*S2*K2	1	1.10	0.2935
T2*S3*K2	1	0.37	0.5418
T3*S3*K2	1	5.62	0.0177

Observation Statistics

Observation	count	Pred	Xbeta	Std	HessWgt
1	65	65	4.1743873	0.1240347	65
2	34	34	3.5263605	0.1714986	34
3	54	54	3.988984	0.1360828	54
4	47	47	3.8501476	0.145865	47
5	100	100	4.6051702	0.1	100
6	100	100	4.6051702	0.1	100
7	130	130	4.8675345	0.0877058	130
8	141	141	4.9487599	0.0842152	141
9	76	76	4.3307333	0.1147079	76
10	116	116	4.7535902	0.0928477	116
11	111	111	4.7095302	0.0949158	111
12	191	191	5.2522734	0.0723575	191
13	67	67	4.2046926	0.1221694	67
14	130	130	4.8675345	0.0877058	130
15	48	48	3.871201	0.1443376	48
16	105	105	4.6539604	0.09759	105
17	62	62	4.1271344	0.1270001	62
18	104	104	4.6443909	0.0980581	104