PEMERIKSAAN DATA BINOMIAL PADA MODEL LOGISTIK SEDERHANA DENGAN MENGGUNAKAN PROSEDUR *FORWARD SEARCH*

SRI WAHYUNI

0300010531

UNIVERSITAS INDONESIA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

DEPARTEMEN MATEMATIKA

DEPOK

2006

PEMERIKSAAN DATA BINOMIAL PADA MODEL LOGISTIK SEDERHANA DENGAN MENGGUNAKAN PROSEDUR *FORWARD SEARCH*

Skripsi diajukan sebagai salah satu syarat

untuk memperoleh gelar Sarjana Sains

Oleh :

SRI WAHYUNI

0300010531

DEPOK

2006

SKRIPSI	:	PEMERIKSAAN DATA BINOMIAL PADA MODEL LOGISTIK SEDERHANA DENGAN MENGGUNAKAN PROSEDUR *FORWARD SEARCH*
NAMA	:	SRI WAHYUNI
NPM	:	0300010531

SKRIPSI TELAH DIPERIKSA DAN DISETUJUI

DEPOK, 28 JULI 2006

Dra. TITIN SISWANTINING, DEA	Dra. IDA FITHRIANI, M.Si
PEMBIMBING I	PEMBIMBING II

Tanggal lulus Ujian Sidang Sarjana : 28 Juli 2006

Penguji I	:	Dra. SASKYA MARY, M.Si
Penguji II	:	Dra. IDA FITHRIANI, M.Si
Penguji III	:	Dra. SITI NURROHMAH, M.Si

KATA PENGANTAR

Puji syukur kepada Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini.

Selesainya skripsi ini tidak terlepas dari bimbingan, bantuan dan dorongan dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis ingin menyampaikan banyak terima kasih kepada :

1. Ibu, bapak, kakak dan adikku atas cinta, kasih sayang dan doanya yang diberikan kepada penulis selama ini.
2. Ibu Dra. Titin Siswantining, DEA. dan Ibu Dra. Ida Fithriani, M.Si. selaku dosen pembimbing skripsi yang telah sabar memberikan bimbingan, saran dan bantuan selama penyusunan skripsi ini.
3. Ibu Dra. Saskya Mary, M.Si. dan Ibu Dra. Kasiyah, M.Sc. selaku dosen pembimbing akademik yang telah sabar meluangkan waktu untuk angkatan 2000.
4. Seluruh staf pengajar dan karyawan Departemen Matematika FMIPA UI, khususnya Mba Fevi untuk pinjaman skripsinya, Kak Ari dan Bu Yekti untuk waktunya, Mba Rusmi untuk bantuannya selama peminjaman buku perpustakaan dan Mba Santi atas bantuannya untuk mengurus surat-surat yang dibutuhkan oleh penulis.
5. Teman-teman Matematika UI angkatan 2000 : Suharto Anggono (untuk semua waktu, kesabaran dan bantuannya yang diberikan untuk penulis), Lenny (untuk bantuan tulisan dan pinjaman skripsinya),

teman-teman terdekat (Helen, Rere, Nunik, Ayu untuk kebersamaan dan dukungan motivasi selama ini, khususnya Rere untuk segala informasi tentang penaksiran parameter dan Newton Raphson, serta Ayu untuk bantuan dan informasinya sehingga penulis dapat mendownload *software* S-PLUS versi 7.0), Nina dan Rena (atas dorongan semangat dan masukan-masukan yang diberikan untuk penulis), Ruri, Samsul, Jawa, Tile, Yudhi, Zaenal, Agus (teman-teman seperjuangan semester terakhir), Tuti, Trias, Ismi, Sigit, Sofian, Adrian, Pida, Meno, Ummu, Ina, Lisna, Danu, Satyo, Mira, Ones, Ribhan, Baar, Dony, Hendri, Sulist, Alan, Cepot, Ole, Rahmat, Medi, Benny, Gandhi, serta teman-teman yang sudah menempuh hidup baru (Kiki, Ema, Dizul, Doddy, Citra, dan Bho).

6. Mas Ali dan Ani untuk pinjaman laptopnya, Ajat (M'04) untuk bantuannya, serta semua pihak yang tidak bisa disebutkan satu per satu yang secara langsung atau tidak langsung telah memberikan kontribusi dalam penulisan skripsi ini.

Penulis mohon maaf bila masih terdapat banyak kekurangan dalam penulisan skripsi ini. Semoga skripsi ini bermanfaat bagi pembaca.

Penulis

2006

ABSTRAK

Pencocokan suatu model logistik terhadap data binomial memerlukan penaksiran parameter dari model. Proses penaksiran parameter ini menggunakan keseluruhan pengamatan yang ada dalam data. Setelah taksiran parameter dihasilkan dan model tersebut digunakan untuk mencocokkan data, mungkin tidak setiap pengamatan dalam data sesuai dengan model. Hal tersebut dapat disebabkan karena data mengandung outlier, yaitu pengamatan yang nilainya jauh lebih besar atau lebih kecil jika dibandingkan dengan mayoritas pengamatan dalam data. Selain itu, ketidakcocokan model terhadap data dapat pula disebabkan karena kesalahan dalam pemilihan fungsi penghubung yang digunakan dalam model. Pada tugas akhir ini, dibahas suatu prosedur *forward search* untuk memeriksa kesesuaian data binomial pada model logistik sederhana. Plot dari statistik-statistik diagnostik yaitu residual deviansi dan statistik skor selama proses *forward search* digunakan untuk melihat sejauh mana pengaruh dari setiap pengamatan individual dalam data binomial. Statistik skor akan digunakan dalam pengujian kecukupan fungsi penghubung.

Kata kunci : outlier; prosedur *forward search*; residual deviansi; uji kecukupan fungsi penghubung.

ix + 48 hlm.; lamp

Bibliografi : 8 (1990-2003)

DAFTAR ISI

Halaman

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Pencocokan sebuah model terhadap suatu data memerlukan penaksiran parameter dari model. Proses penaksiran parameter dalam model menggunakan seluruh pengamatan yang ada dalam data. Setelah taksiran parameter dari model dihasilkan, dan model tersebut digunakan untuk mencocokkan data, mungkin tidak setiap pengamatan dalam data tersebut sesuai dengan model yang digunakan.

Ada beberapa hal yang menyebabkan ketidakcocokan model terhadap suatu data. Salah satu penyebabnya karena data yang digunakan dalam penaksiran parameter mengandung outlier, yaitu pengamatan yang nilainya jauh lebih besar atau lebih kecil jika dibandingkan dengan mayoritas pengamatan dalam data. Selain itu, ketidakcocokan model terhadap suatu data dapat pula disebabkan karena kesalahan dalam pemilihan fungsi penghubung yang digunakan dalam model. Fungsi penghubung adalah suatu fungsi tertentu yang menghubungkan mean dari variabel random dengan kombinasi linier dari variabel independen.

Metode yang digunakan untuk memeriksa kesesuaian dari sebuah model dalam mencocokkan suatu data disebut dengan metode diagnostik.

Salah satu prosedur diagnostik yang dapat digunakan adalah prosedur *forward search*.

Pemeriksaan sebuah model dalam mencocokkan suatu data juga berlaku untuk data yang variabel randomnya menyatakan banyaknya sukses dari n percobaan yang saling bebas, dan memiliki distribusi binomial. Data yang demikian disebut dengan data binomial. Salah satu model yang dapat dibentuk dari data binomial adalah model logistik. Prosedur *forward search* dapat digunakan dalam pemeriksaan data binomial pada model logistik sederhana.

1.2 Permasalahan

Bagaimana menggunakan prosedur *forward search* untuk memeriksa data binomial pada model logistik sederhana.

1.3 Tujuan Penulisan

Skripsi ini bertujuan untuk menjelaskan pemeriksaan data binomial pada model logistik sederhana dengan menggunakan prosedur *forward search*.

1.4 Pembatasan Masalah

Statistik diagnostik yang digunakan untuk memeriksa data binomial pada model logistik sederhana dibatasi pada residual deviansi dan statistik skor. Statistik skor akan digunakan dalam pengujian kecukupan fungsi penghubung.

1.5 Sistematika Penulisan

Sistematika penulisan skripsi ini dibagi menjadi lima bab. Bab II merupakan landasan teori tentang distribusi binomial, model logistik sederhana untuk data binomial, penaksiran parameter dalam model logistik sederhana, serta statistik diagnostik yaitu residual deviansi dan statistik skor, dimana statistik skor akan digunakan dalam pengujian kecukupan fungsi penghubung. Pada Bab III dibahas tahapan-tahapan pada prosedur *forward search* untuk memeriksa data binomial pada model logistik sederhana melalui statistik diagnostik yaitu residual deviansi dan statistik skor. Contoh aplikasi prosedur *forward search* untuk memeriksa data binomial pada model logistik sederhana akan diberikan pada Bab IV. Sedangkan Bab V berisi kesimpulan.

BAB II

LANDASAN TEORI

Pada bab II dibahas mengenai distribusi binomial, bentuk model logistik sederhana untuk data binomial dan penaksiran parameter pada model logistik sederhana. Selain itu dibahas pula statistik diagnostik yaitu residual deviansi dan statistik skor. Statistik skor akan digunakan dalam pengujian kecukupan fungsi penghubung pada model logistik sederhana.

2.1 Distribusi Binomial

Didefinisikan variabel random biner

$$R = \begin{cases} 1, & \text{jika kejadian sukses} \\ 0, & \text{jika kejadian gagal} \end{cases}$$

dengan probabilitas sukses $\Pr(R=1)=\pi$ dan probabilitas gagal $\Pr(R=0)=1-\pi$. Jika r adalah nilai pengamatan dari variabel random R, maka R berdistribusi bernoulli dengan fungsi probabilitas

$$\Pr(R=r)=\pi^{r}(1-\pi)^{1-r}, \quad r = 0,1$$

dengan mean dari variabel random R adalah $E(R)=\pi$.

Untuk n variabel random $R_1, R_2, ..., R_n$ yang saling bebas, maka $\Pr(R_j=1)=\pi_j$, untuk j = 1, 2, …, n. Misalkan π_j sama untuk setiap j, maka dapat didefinisikan $Y=\sum_{j=1}^{n} R_j$ dengan Y menyatakan banyaknya sukses dalam n percobaan yang saling bebas. Variabel random Y berdistribusi binomial dengan parameter n dan π, yang ditulis dengan $Y \sim B(n,\pi)$. Fungsi probabilitas dari Y diberikan oleh

$$\Pr(Y = y) = \binom{n}{y} \pi^{y} (1-\pi)^{n-y}, \quad y = 0,1,...,n$$

Mean dari variabel random Y adalah $E(Y) = n\pi$.

Secara umum, misalkan variabel random $Y_1, Y_2, ..., Y_N$ saling bebas dan menyatakan banyaknya sukses dalam N subgrup yang berbeda, seperti pada Tabel 2.1. Maka untuk i = 1, 2, …, N, variabel random Y_i berdistribusi binomial dengan parameter n_i dan π_i, yang ditulis dengan $Y_i \sim B(n_i, \pi_i)$, dengan n_i menyatakan ukuran subgrup ke-i dan π_i menyatakan probabilitas sukses pada subgrup ke-i.

Tabel 2.1 Frekuensi sukses dan gagal dalam N subgrup

	Subgrup			
	1	2	…	N
Sukses	y_1	y_2	…	y_N
Gagal	$n_1 - y_1$	$n_2 - y_2$	…	$n_N - y_N$
Total	n_1	n_2	…	n_N

Fungsi probabilitas dari Y_i diberikan oleh

$$f(y_i, \pi_i) = \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i}, \; y_i = 0, 1, ..., n_i$$

Mean dari variabel random Y_i adalah $E(Y_i) = n_i \pi_i$.

2.2 Model Logistik Sederhana untuk Data Binomial

Untuk data dengan variabel random berdistribusi binomial, responnya berupa proporsi sukses, yaitu Y/n. Misalkan ingin diketahui hubungan antara proporsi sukses Y/n dengan sebuah variabel independen X, maka hubungan tersebut tidak dapat dijelaskan oleh model regresi sederhana berikut

$$E(Y/n) = \beta_0 + \beta_1 x$$

$$\pi = \beta_0 + \beta_1 x \quad (2.2.1)$$

dengan

π adalah probabilitas sukses,

x adalah nilai dari variabel independen X,

β_0 dan β_1 adalah parameter-parameter yang tidak diketahui.

Ruas kanan dari persamaan (2.2.1) disebut juga prediktor linier, yang dinotasikan dengan η. Jadi

$$\eta = \beta_0 + \beta_1 x \quad (2.2.2)$$

Nilai η terletak pada interval $(-\infty, \infty)$, sedangkan nilai π harus terbatas pada interval (0,1). Oleh karena itu, agar interval kedua ruas pada persamaan (2.2.1) sesuai, diperlukan suatu transformasi dari nilai-nilai π agar berada pada interval $(-\infty,\infty)$. Transformasi ini disebut juga sebagai fungsi penghubung. Salah satu fungsi penghubung yang dapat digunakan untuk model pada persamaan (2.2.1) adalah fungsi penghubung logistik.

Fungsi penghubung logistik dari suatu probabilitas sukses π adalah $\log\left(\frac{\pi}{1-\pi}\right)$, dimana log yang dimaksud dalam hal ini adalah logaritma berbasis e. Bentuk $\log\left(\frac{\pi}{1-\pi}\right)$ dapat ditulis juga dengan $g(\pi)$=logit(π). Fungsi penghubung logistik dari π dapat diinterpretasikan sebagai logaritma dari rasio probabilitas sukses π terhadap probabilitas gagal (1-π). Dengan menggunakan fungsi penghubung logistik (logit), sembarang nilai π dalam interval (0,1) akan bersesuaian dengan nilai logit(π) dalam interval $(-\infty,\infty)$. Bentuk model logistik sederhana diberikan oleh:

$$g(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x \qquad (2.2.3)$$

atau

$$\pi = \frac{\exp(\beta_0 + \beta_1 x)}{1+\exp(\beta_0 + \beta_1 x)} \qquad (2.2.4)$$

dengan

π adalah probabilitas sukses,

x adalah nilai dari variabel independen X,

β_0 dan β_1 adalah parameter-parameter yang tidak diketahui.

Dari (2.2.2) dan (2.2.3) diperoleh

$$g(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \eta \tag{2.2.5}$$

Jadi fungsi penghubung $g(\pi)$ menghubungkan probabilitas sukses π dengan prediktor linier η. Model logistik sederhana digunakan untuk mempelajari hubungan antara probabilitas sukses π dengan sebuah variabel independen X yang diamati.

2.3 Penaksiran Parameter dalam Model Logistik Sederhana

Anggap N buah subgrup dari pengamatan binomial yang saling bebas, yaitu $y_1, y_2, ..., y_N$, sehingga untuk i = 1, 2, …, N, variabel random Y_i berdistribusi binomial dengan parameter n_i dan π_i, ditulis $Y_i \sim B(n_i, \pi_i)$.

Parameter $\boldsymbol{\beta} = \left[\beta_0, \beta_1\right]^T$ dalam model logistik akan ditaksir dengan metode maksimum *likelihood*. Prinsip dari metode maksimum *likelihood* adalah mencari taksiran parameter $\boldsymbol{\beta} = \left[\beta_0, \beta_1\right]^T$, yaitu $\hat{\boldsymbol{\beta}} = \left[\hat{\beta}_0, \hat{\beta}_1\right]^T$, yang memaksimumkan fungsi *likelihood*.

Karena $Y_i \sim B(n_i, \pi_i)$, maka fungsi probabilitas dari Y_i diberikan oleh

$$f(y_i; \pi_i) = \binom{n_i}{y_i} \pi_i^{y_i} (1-\pi_i)^{n_i - y_i} \text{ , } y_i = 0, 1, ..., n_i \tag{2.3.1}$$

Karena subgrupnya saling bebas maka fungsi *likelihood* didapat sebagai hasil perkalian dari masing-masing fungsi probabilitas Y_i pada persamaan (2.3.1) yaitu

$$
\begin{aligned}
L(\boldsymbol{\beta}) &= \prod_{i=1}^{N} f(y_i;\pi_i) \\
&= \prod_{i=1}^{N} \binom{n_i}{y_i} \pi_i^{\,y_i}(1-\pi_i)^{n_i-y_i} \qquad (2.3.2)
\end{aligned}
$$

dengan $\boldsymbol{\beta} = \left[\beta_0, \beta_1\right]^T$ adalah vektor dari parameter yang tidak diketahui.

Untuk mempermudah penghitungan dibentuk logaritma dari fungsi *likelihood* yaitu

$$
\begin{aligned}
\log L(\boldsymbol{\beta}) &= \sum_{i=1}^{N} \log\left\{\binom{n_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{n_i-y_i}\right\} \\
&= \sum_{i=1}^{N}\left\{y_i \log \pi_i + (n_i - y_i)\log(1-\pi_i) + \log\binom{n_i}{y_i}\right\} \\
&= \sum_{i=1}^{N}\left\{y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + n_i \log(1-\pi_i) + \log\binom{n_i}{y_i}\right\}
\end{aligned}
$$

Dari persamaan (2.2.3) dan (2.2.4) diperoleh

$$
\begin{aligned}
\log L(\boldsymbol{\beta}) &= \sum_{\mathrm{i}=1}^{\mathrm{N}}\left\{y_i\left(\beta_0+\beta_1 x_i\right) + n_i \log\left(\frac{1}{1+\exp\left(\beta_0+\beta_1 x_i\right)}\right) + \log\binom{n_i}{y_i}\right\} \\
&= \sum_{\mathrm{i}=1}^{\mathrm{N}}\left\{y_i\left(\beta_0+\beta_1 x_i\right) - n_i \log\left(1+\exp\left(\beta_0+\beta_1 x_i\right)\right) + \log\binom{n_i}{y_i}\right\}
\end{aligned}
$$

Fungsi $\log L(\boldsymbol{\beta})$ di atas disebut fungsi *log-likelihood* untuk model logistik sederhana.

Nilai β_0 dan β_1 diperoleh dengan mendiferensialkan $\log L(\boldsymbol{\beta})$ secara parsial terhadap masing-masing parameter β_0 dan β_1, dan hasilnya disamakan dengan nol.

$$\begin{aligned}\frac{\partial \log L(\boldsymbol{\beta})}{\partial \beta_0} &= \sum_{i=1}^{N} y_i - \frac{n_i \exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)} = \sum_{i=1}^{N} y_i - n_i \pi_i = 0 \\ \frac{\partial \log L(\boldsymbol{\beta})}{\partial \beta_1} &= \sum_{i=1}^{N} x_i y_i - \frac{n_i x_i \exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)} = \sum_{i=1}^{N} x_i y_i - n_i x_i \pi_i = 0\end{aligned} \tag{2.3.3}$$

Nilai β_0 dan β_1 yang diperoleh sebagai hasil penyelesaian persamaan (2.3.3) akan memaksimumkan $\log L(\boldsymbol{\beta})$ apabila turunan parsial kedua dari $\log L(\boldsymbol{\beta})$ memenuhi

$$\Delta = \left\{ \left(\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0^2} \right) \left(\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_1^2} \right) - \left(\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} \right)^2 \right\} > 0$$

dan (2.3.4)

$$\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0^2} < 0 \text{ atau } \frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_1^2} < 0$$

untuk setiap nilai β_0 dan β_1[Soemartojo, 1993].

Turunan parsial kedua dari $\log L(\boldsymbol{\beta})$ adalah sebagai berikut

$$\begin{aligned}\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0^2} &= \frac{\partial}{\partial \beta_0}\left[\sum_{i=1}^{N} y_i - \frac{n_i \exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)} \right] = -\sum_{i=1}^{N} n_i \pi_i (1-\pi_i) \\ \frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_0} &= \frac{\partial}{\partial \beta_1}\left[\sum_{i=1}^{N} y_i - \frac{n_i \exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)} \right] = -\sum_{i=1}^{N} n_i x_i \pi_i (1-\pi_i) \\ \frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} &= \frac{\partial}{\partial \beta_0}\left[\sum_{i=1}^{N} x_i y_i - \frac{n_i x_i \exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)} \right] = -\sum_{i=1}^{N} n_i x_i \pi_i (1-\pi_i) \\ \frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_1^2} &= \frac{\partial}{\partial \beta_1}\left[\sum_{i=1}^{N} x_i y_i - \frac{n_i x_i \exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)} \right] = -\sum_{i=1}^{N} n_i x_i^2 \pi_i (1-\pi_i)\end{aligned}$$

Pada Lampiran 1 telah ditunjukkan bahwa persamaan (2.3.4) dipenuhi untuk setiap nilai β_0 dan β_1. Oleh karena itu, nilai β_0 dan β_1 yang diperoleh sebagai hasil penyelesaian persamaan (2.3.3) disebut dengan taksiran maksimum *likelihood*, dan dinotasikan dengan $\hat{\beta}_0$ dan $\hat{\beta}_1$.

Karena persamaan (2.3.3) tidak linier dalam parameter β_0 dan β_1, maka nilai $\hat{\beta}_0$ dan $\hat{\beta}_1$ dicari dengan menggunakan iterasi numerik. Salah satu prosedur numerik yang dapat digunakan adalah prosedur *Newton Raphson*.

Prosedur iterasi *Newton Raphson* untuk mencari nilai taksiran $\hat{\beta}_0$ dan $\hat{\beta}_1$ pada persamaan (2.3.3) adalah sebagai berikut

1. Pilih taksiran awal $\hat{\boldsymbol{\beta}}^{(0)} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}^{(0)}$
2. Pada setiap iterasi ke-(t+1), dengan $t = 0, 1, \ldots$, hitung taksiran baru yaitu $\hat{\boldsymbol{\beta}}^{(t+1)}$ secara iteratif menggunakan formula

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} - \left[\mathbf{H}(\hat{\boldsymbol{\beta}}^{(t)})\right]^{-1} \mathbf{U}(\hat{\boldsymbol{\beta}}^{(t)}) \qquad (2.3.5)$$

dengan

$\hat{\boldsymbol{\beta}}^{(t)}$ adalah taksiran dari $\boldsymbol{\beta}$ pada iterasi ke-t,

$\hat{\boldsymbol{\beta}}^{(t+1)}$ adalah taksiran dari $\boldsymbol{\beta}$ pada iterasi ke-(t+1),

$\mathbf{U}\left(\hat{\boldsymbol{\beta}}^{(t)}\right)$ adalah vektor turunan parsial pertama dari $\log L(\boldsymbol{\beta})$ dengan elemen $\dfrac{\partial \log L(\boldsymbol{\beta})}{\partial \beta_j}$, dan j = 0, 1, yang dihitung pada $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(t)}$,

$\mathbf{H}\left(\hat{\boldsymbol{\beta}}^{(t)}\right)$ adalah matriks turunan parsial kedua dari $\log L(\boldsymbol{\beta})$ dengan elemen $\dfrac{\partial^2 \log L(\boldsymbol{\beta})}{\partial\beta_k \beta_j}$, dan j,k = 0, 1, yang dihitung pada $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(t)}$.

3. Iterasi berlanjut hingga diperoleh $\hat{\boldsymbol{\beta}}^{(t+1)} \approx \hat{\boldsymbol{\beta}}^{(t)}$. Hentikan prosedur iterasi jika norm dari selisih antara $\hat{\boldsymbol{\beta}}^{(t+1)}$ dengan $\hat{\boldsymbol{\beta}}^{(t)}$ sangat kecil (misal $\left\|\hat{\boldsymbol{\beta}}^{(t+1)} - \hat{\boldsymbol{\beta}}^{(t)}\right\| < \varepsilon$, dengan ε adalah suatu bilangan positif yang sangat kecil sekali). Kemudian ambil nilai $\hat{\boldsymbol{\beta}}^{(t+1)}$ sebagai $\hat{\boldsymbol{\beta}}$.

Jika dituliskan dalam bentuk matriks, persamaan (2.3.5) dapat dinyatakan sebagai berikut

$$\begin{aligned}
\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix}^{(t+1)} &= \begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix}^{(t)} - \left(\begin{bmatrix} -\sum_{i=1}^{N} n_i\pi_i(1-\pi_i) & -\sum_{i=1}^{N} n_i x_i\pi_i(1-\pi_i) \\ -\sum_{i=1}^{N} n_i x_i\pi_i(1-\pi_i) & -\sum_{i=1}^{N} n_i x_i^2\pi_i(1-\pi_i)\end{bmatrix}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(t)}}\right)^{-1} \begin{bmatrix}\sum_{i=1}^{N} y_i - n_i\pi_i \\ \sum_{i=1}^{N} x_i(y_i - n_i\pi_i)\end{bmatrix}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(t)}} \\
&= \begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix}^{(t)} - \left(-\begin{bmatrix} \sum_{i=1}^{N} n_i\pi_i(1-\pi_i) & \sum_{i=1}^{N} n_i x_i\pi_i(1-\pi_i) \\ \sum_{i=1}^{N} n_i x_i\pi_i(1-\pi_i) & \sum_{i=1}^{N} n_i x_i^2\pi_i(1-\pi_i)\end{bmatrix}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(t)}}\right)^{-1} \begin{bmatrix}\sum_{i=1}^{N} y_i - n_i\pi_i \\ \sum_{i=1}^{N} x_i(y_i - n_i\pi_i)\end{bmatrix}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(t)}} \\
&= \begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix}^{(t)} + \left(\begin{bmatrix} \sum_{i=1}^{N} n_i\pi_i(1-\pi_i) & \sum_{i=1}^{N} n_i x_i\pi_i(1-\pi_i) \\ \sum_{i=1}^{N} n_i x_i\pi_i(1-\pi_i) & \sum_{i=1}^{N} n_i x_i^2\pi_i(1-\pi_i)\end{bmatrix}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(t)}}\right)^{-1} \begin{bmatrix}\sum_{i=1}^{N} y_i - n_i\pi_i \\ \sum_{i=1}^{N} x_i(y_i - n_i\pi_i)\end{bmatrix}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(t)}}
\end{aligned} \tag{2.3.6}$$

dengan t = 0, 1, 2, …

Iterasi pada persamaan (2.3.6) akan dijalankan menggunakan *software*

S-PLUS versi 7.0.

2.4 Statistik Diagnostik

2.4.1 Residual Deviansi

Ukuran kecocokkan antara sebuah pengamatan dari suatu variabel respon dengan nilai prediksinya dinamakan residual. Ukuran ini memberikan informasi mengenai seberapa baik model cocok dengan setiap pengamatan dalam data.

Salah satu residual yang digunakan dalam model logistik adalah residual deviansi. Residual deviansi dibentuk dari Deviansi yang diperoleh setelah mencocokkan sebuah model logistik terhadap data binomial.
Deviansi (D) didefinisikan oleh :

$$D = 2\left[\log L(\hat{\boldsymbol{\beta}}^{\max}) - \log L(\hat{\boldsymbol{\beta}})\right] \qquad (2.4.1)$$

dengan

$\log L(\hat{\boldsymbol{\beta}}^{\max})$ adalah fungsi *log-likelihood* untuk model *saturated* (model dengan banyaknya parameter sama dengan banyaknya subgrup dalam data binomial), yang dihitung pada taksiran maksimum *likelihood* $\hat{\boldsymbol{\beta}}^{\max}$, dimana $\hat{\boldsymbol{\beta}}^{\max}$ adalah vektor dari taksiran parameter berukuran N,

$\log L(\hat{\boldsymbol{\beta}})$ adalah fungsi *log-likelihood* untuk model *current* (model yang digunakan), yang dihitung pada taksiran maksimum *likelihood* $\hat{\boldsymbol{\beta}}$.

Dari persamaan (2.3.2), fungsi *likelihood* dari parameter $\boldsymbol{\beta}$ adalah

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} \binom{n_i}{y_i} \pi_i^{y_i} (1-\pi_i)^{n_i - y_i}$$

Fungsi *log-likelihood* dari parameter $\boldsymbol{\beta}$ adalah

$$\begin{aligned} \log L(\boldsymbol{\beta}) &= \log \left\{ \prod_{i=1}^{N} \binom{n_i}{y_i} \pi_i^{y_i} (1-\pi_i)^{n_i - y_i} \right\} \\ &= \sum_{i=1}^{N} \left\{ y_i \log \pi_i + (n_i - y_i) \log(1-\pi_i) + \log \binom{n_i}{y_i} \right\} \end{aligned}$$

Untuk model *saturated*, taksiran probabilitas sukses sama dengan nilai proporsi sukses dari data, $\tilde{\pi}_i = \dfrac{y_i}{n_i}$, dengan i = $1, 2, \ldots, N$, sehingga

$$\log L(\hat{\boldsymbol{\beta}}^{\max}) = \sum_{i=1}^{N} \left\{ y_i \log\left(\frac{y_i}{n_i}\right) + (n_i - y_i) \log\left(1 - \frac{y_i}{n_i}\right) + \log \binom{n_i}{y_i} \right\} \qquad (2.4.2)$$

Untuk model *current* , taksiran probabilitas sukses adalah $\hat{\pi}_i$, dengan i = $1, 2, \ldots, N$. Nilai prediksi dari variabel random Y_i adalah $\hat{y}_i = n_i \hat{\pi}_i$ maka $\hat{\pi}_i = \dfrac{\hat{y}_i}{n_i}$, sehingga

$$\log L(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{N} \left\{ y_i \log\left(\frac{\hat{y}_i}{n_i}\right) + (n_i - y_i) \log\left(1 - \frac{\hat{y}_i}{n_i}\right) + \log \binom{n_i}{y_i} \right\} \qquad (2.4.3)$$

Persamaan (2.4.1) menjadi

$$D = \sum_{i=1}^{N} \left\{ \begin{array}{l} 2\left[y_i \log\left(\frac{y_i}{n_i}\right) + (n_i - y_i)\log\left(1 - \frac{y_i}{n_i}\right) + \log\binom{n_i}{y_i} \right] - \\ 2\left[y_i \log\left(\frac{\hat{y}_i}{n_i}\right) + (n_i - y_i)\log\left(1 - \frac{\hat{y}_i}{n_i}\right) + \log\binom{n_i}{y_i} \right] \end{array} \right\} \tag{2.4.4}$$

Model *saturated* adalah model dengan banyaknya parameter sama dengan banyaknya subgrup dalam data binomial. Dalam model *saturated*, taksiran probabilitas sukses sama dengan nilai proporsi sukses dari data yaitu $\tilde{\pi}_i = \frac{y_i}{n_i}$, sehingga model ini mencocokkan data dengan sempurna [Collett, 1991]. Oleh karena itu, nilai *log-likelihood* untuk model *saturated* lebih besar daripada nilai *log-likelihood* untuk model *current*. Hal ini mengakibatkan pengurangan dari nilai *log-likelihood* untuk model *saturated* dengan nilai *log-likelihood* untuk model *current* akan bernilai positif. Sehingga, suku-suku dalam persamaan (2.4.4) akan bernilai positif. Dengan demikian, persamaan (2.4.4) dapat dinyatakan dengan

$$D = \sum_{i=1}^{N} d_i^2 \tag{2.4.5}$$

Dari (2.4.4) dan (2.4.5), diperoleh

$$
\begin{aligned}
d_i^2 &= 2\left[y_i \log\left(\frac{y_i}{n_i}\right) + (n_i - y_i)\log\left(1-\frac{y_i}{n_i}\right) + \log\binom{n_i}{y_i}\right] - \\
&\quad 2\left[y_i \log\left(\frac{\hat{y}_i}{n_i}\right) + (n_i - y_i)\log\left(1-\frac{\hat{y}_i}{n_i}\right) + \log\binom{n_i}{y_i}\right] \\
&= 2\left[y_i \log\left(\frac{y_i / n_i}{\hat{y}_i / n_i}\right) + (n_i - y_i)\log\left(\frac{(n_i - y_i)/n_i}{(n_i - \hat{y}_i)/n_i}\right)\right] \\
&= 2\left[y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i)\log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right)\right] \\
&= 2 y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + 2(n_i - y_i)\log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right)
\end{aligned}
$$

Residual deviansi (d_i) didefinisikan sebagai akar kuadrat bertanda dari d_i^2,

yaitu

$$
d_i = \operatorname{sgn}(y_i - \hat{y}_i)\left\{2y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + 2(n_i - y_i)\log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right)\right\}^{1/2} \quad ; \; i = 1, 2, ..., N
$$

dengan

y_i menyatakan nilai pengamatan dari variabel random Y_i untuk subgrup ke-i,

$\hat{y}_i$ menyatakan nilai prediksi dari variabel random Y_i untuk subgrup ke-i, dimana $\hat{y}_i = n_i \hat{\pi}_i$, dengan n_i menyatakan ukuran subgrup ke-i.

Nilai d_i akan bertanda positif atau negatif berdasarkan pada sgn(y_i - $\hat{y}_i$), jika $y_i \geq \hat{y}_i$ maka d_i akan bernilai positif dan bernilai negatif jika $y_i < \hat{y}_i$.

Residual yang digunakan untuk model logistik adalah residual deviansi karena residual ini lebih stabil ketika probabilitas yang dicocokkan dekat dengan 0 atau 1 [Atkinson & Riani, 2001].

2.4.2 Uji Kecukupan Fungsi Penghubung

Pengujian kecukupan fungsi penghubung atau *goodness of link* untuk data binomial digunakan untuk memeriksa apakah suatu fungsi penghubung yang diasumsikan, $g(\pi)$, memang menggambarkan hubungan yang sesungguhnya antara mean dari respon Y/n, $E(Y/n) = \pi$, dengan prediktor linier, η. Prediktor linier, η, untuk model logistik sederhana adalah

$$\eta = \beta_0 + \beta_1 x$$

Jika $g(\pi)$ memang menggambarkan hubungan yang sesungguhnya antara π dan η, maka berlaku bahwa $g(\pi) = \eta$.

Anggap fungsi penghubung yang digunakan dalam pencocokan model terhadap data adalah $g(\pi)$, sedangkan fungsi penghubung yang sesungguhnya adalah $g^*(\pi) = \eta$. Misalkan $h(\eta) = g\{g^{*-1}(\eta)\}$, maka

$$g(\pi) = g\{g^{*-1}(\eta)\} = h(\eta) \qquad (2.4.6)$$

Jika fungsi penghubung yang digunakan tepat, maka $h(\eta)$ adalah fungsi linier dalam η, dapat dinyatakan dengan $h(\eta) = \eta$. Sebaliknya jika fungsi penghubung yang digunakan tidak tepat, maka $h(\eta)$ menjadi tidak linier dalam η. Oleh karena itu, perlu diuji apakah $g(\pi)$ adalah sebuah fungsi linier dari η.

Pandang persamaan (2.4.6). Dengan menggunakan ekspansi deret Maclaurin orde kedua dari $h(\eta)$ diperoleh

$$g(\pi) = h(\eta) = h(0) + h^{'}(0)\eta + h^{"}(0)\frac{\eta^2}{2!} \qquad (2.4.7)$$

Misalkan $h(0) = \alpha$, $h^{'}(0) = \delta$ dan $h^{"}(0)/2 = \gamma$ dengan α, δ dan γ adalah skalar. Maka persamaan (2.4.7) menjadi

$$\begin{aligned} g(\pi) &= \alpha + \delta\eta + \gamma\eta^2 \\ &= \alpha + \delta(\beta_0 + \beta_1 x) + \gamma\eta^2 \\ &= \alpha + \delta\beta_0 + \delta\beta_1 x + \gamma\eta^2 \end{aligned}$$

dengan mengganti $\alpha + \delta\beta_0 = \beta_0^{'}$ dan $\delta\beta_1 = \beta_1^{'}$ maka diperoleh

$$\begin{aligned} g(\pi) &= \beta_0^{'} + \beta_1^{'} x + \gamma\eta^2 \\ &= \beta_0^{'} + \beta_1^{'} x + \gamma\eta^2 \end{aligned}$$

$$g(\pi) = \eta^{'} + \gamma\eta^2 \qquad (2.4.8)$$

dimana $\eta^{'} = \beta_0^{'} + \beta_1^{'} x$

Persamaan (2.4.8) kemudian akan digunakan untuk mencocokkan data. Suku kedua pada bagian kanan persamaan (2.4.8) menyatakan kontribusi suku non linier dari η. Sehingga pengujian kecukupan fungsi penghubung ekivalen dengan pengujian terhadap signifikansi γ.

Hipotesis dari pengujian ini adalah

$$H_0 : \gamma = 0$$

$$H_1 : \gamma \neq 0$$

Hipotesis nol dari pengujian di atas menyatakan bahwa fungsi $h(\eta)$ linier dalam η, yang berarti fungsi penghubung yang digunakan, yaitu $g(\pi)$, telah tepat. Sedangkan hipotesis alternatif menyatakan bahwa fungsi $h(\eta)$ tidak linier dalam η, yang berarti fungsi penghubung $g(\pi)$ tidak tepat.

Jadi pengujian kecukupan fungsi penghubung untuk data binomial melibatkan penambahan sebuah suku tambahan yaitu η^2.

Statistik uji yang digunakan adalah statistik skor, yaitu :

$$S = \frac{U_\gamma}{\sqrt{Var(U_\gamma)}} \tag{2.4.9}$$

dengan U_γ adalah turunan parsial pertama dari fungsi *log-likelihood* terhadap parameter γ, untuk model (2.4.8).

Untuk memperoleh statistik skor, terlebih dahulu dihitung kuadrat dari taksiran prediktor linier $\hat{\eta}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, dengan $i = 1, 2, \ldots, N$, menggunakan fungsi penghubung yang diasumsikan, yaitu $g(\pi)$, dengan model (2.2.5), yaitu model tanpa η^2. Parameter dalam model (2.2.5) ditaksir menggunakan metode maksimum *likelihood* seperti pada subbab 2.3.

Fungsi *log-likelihood* untuk model (2.4.8) adalah sebagai berikut

$$\log L(\boldsymbol{\beta}^{'},\gamma) = \sum_{\mathrm{i=1}}^{\mathrm{N}} \log\left\{\binom{n_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{n_i-y_i}\right\}$$

$$= \sum_{\mathrm{i=1}}^{\mathrm{N}}\left\{y_i \log \pi_i + (n_i - y_i)\log(1-\pi_i) + \log\binom{n_i}{y_i}\right\}$$

$$= \sum_{\mathrm{i=1}}^{\mathrm{N}}\left\{y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + n_i \log(1-\pi_i) + \log\binom{n_i}{y_i}\right\}$$

$$= \sum_{\mathrm{i=1}}^{\mathrm{N}}\left\{y_i\left(\beta_0^{'} + \beta_1^{'}x_i + \gamma\hat{\eta}_i^2\right) + n_i \log\left(\frac{1}{1+\exp\left(\beta_0^{'} + \beta_1^{'}x_i + \gamma\hat{\eta}^2\right)}\right) + \log\binom{n_i}{y_i}\right\}$$

$$= \sum_{\mathrm{i=1}}^{\mathrm{N}}\left\{y_i\left(\beta_0^{'} + \beta_1^{'}x_i + \gamma\hat{\eta}_i^2\right) - n_i \log\left(1+\exp\left(\beta_0^{'} + \beta_1^{'}x_i + \gamma\hat{\eta}_i^2\right)\right) + \log\binom{n_i}{y_i}\right\}$$

dimana $\boldsymbol{\beta}^{'} = [\beta_0^{'}, \beta_1^{'}]^T$, $g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right)$ dan $\pi_i = \frac{\exp(\beta_0^{'} + \beta_1^{'}x_i + \gamma\hat{\eta}_i^2)}{1+\exp(\beta_0^{'} + \beta_1^{'}x_i + \gamma\hat{\eta}_i^2)}$

Turunan parsial pertama dari $\log L(\boldsymbol{\beta}^{'},\gamma)$ terhadap parameter γ dinotasikan dengan U_γ yaitu

$$U_\gamma = \frac{\partial \log L(\boldsymbol{\beta}^{'},\gamma)}{\partial\gamma} = \sum_{i=1}^{N}\left\{y_i\hat{\eta}_i^2 - \frac{n_i\hat{\eta}_i^2\exp(\beta_0^{'} + \beta_1^{'}x_i + \gamma\hat{\eta}_i^2)}{1+\exp(\beta_0^{'} + \beta_1^{'}x_i + \gamma\hat{\eta}_i^2)}\right\}$$

$$= \sum_{i=1}^{N}\left\{y_i\hat{\eta}_i^2 - n_i\hat{\eta}_i^2\pi_i\right\}$$

Statistik skor dalam (2.4.9) menjadi

$$S = \frac{U_\gamma}{\sqrt{Var\left(U_\gamma\right)}}$$

$$= \frac{\sum_{i=1}^{N}\left\{y_i\hat{\eta}_i^2 - n_i\hat{\eta}_i^2\pi_i\right\}}{\sqrt{Var\left(\sum_{i=1}^{N}\left\{y_i\hat{\eta}_i^2 - n_i\hat{\eta}_i^2\pi_i\right\}\right)}}$$

$$= \frac{\sum_{i=1}^{N}\hat{\eta}_i^2\left(y_i - n_i\pi_i\right)}{\sqrt{Var\left(\sum_{i=1}^{N}\hat{\eta}_i^2\left(y_i - n_i\pi_i\right)\right)}}$$

dengan

U_γ adalah turunan parsial pertama dari fungsi $\log L(\boldsymbol{\beta}^{'},\gamma)$ terhadap parameter γ, yang dihitung pada $\gamma = 0$. Semua parameter $\beta_0^{'}$ dan $\beta_1^{'}$ diganti dengan nilai taksirannya, yaitu $\hat{\beta}_0^{'}$ dan $\hat{\beta}_1^{'}$.

$Var(U_\gamma)$ adalah variansi dari U_γ.

Di bawah H_o benar, statistik skor mendekati distribusi Normal(0,1). Dengan tingkat signifikansi α, H_o ditolak jika $|S| > z_{\alpha/2}$. Nilai $z_{\alpha/2}$ diperoleh dari tabel distribusi Normal(0,1). Jika H_o tidak ditolak artinya fungsi $h(\eta)$ linier dalam η, yang berarti fungsi penghubung pada model (2.2.5) telah tepat.

BAB III

PEMERIKSAAN DATA BINOMIAL PADA MODEL LOGISTIK SEDERHANA DENGAN MENGGUNAKAN PROSEDUR *FORWARD SEARCH*

Pada bab III dibahas mengenai tahapan-tahapan pada prosedur *forward search* untuk memeriksa data binomial pada model logistik sederhana. Melalui tahapan-tahapan pada prosedur *forward search* ini, akan diperiksa statistik-statistik diagnostik yang telah dibahas pada subbab 2.4 yaitu residual deviansi dan statistik skor.

3.1 Prinsip Umum *Forward Search*

Forward search adalah suatu metode yang digunakan untuk mendeteksi keberadaan outlier dalam data. Seperti yang telah dinyatakan sebelumnya bahwa outlier merupakan pengamatan yang nilainya jauh lebih besar atau lebih kecil dibandingkan mayoritas pengamatan dalam data.

Prosedur *forward search* diawali dengan pemilihan sebuah *initial subset* dari data, kemudian pada tahapan selanjutnya bekerja secara *forward*, yaitu menambahkan satu per satu pengamatan ke dalam *subset* yang digunakan untuk mencocokkan model terhadap data, dan memonitor statistik diagnostik setiap penambahan satu pengamatan ke dalam *subset*.

Ide dasar dari prosedur *forward* adalah mengurutkan pengamatan-pengamatan berdasarkan kedekatannya terhadap model. Ukuran kedekatan antara pengamatan dalam data binomial terhadap model logistik diketahui berdasarkan nilai residual deviansinya.

Hasil dari langkah *forward search* terlihat pada plot dari statistik-statistik diagnostik. Plot ini memperlihatkan pengaruh dari pengamatan-pengamatan individual terhadap residual deviansi dan statistik skor. Melalui plot ini akan dimonitor nilai-nilai dari statistik diagnostik pada setiap tahapan prosedur.

3.2 Tahapan-tahapan pada Prosedur *Forward Search*

Prosedur *forward search* terdiri dari tiga tahap yaitu pemilihan sebuah *initial subset*, penambahan satu per satu pengamatan ke dalam *subset* yang digunakan dalam pencocokan model, serta pemonitoran statistik-statistik diagnostik setiap penambahan satu pengamatan ke dalam *subset* selama proses *forward search.*

3.2.1 Pemilihan Sebuah *Initial Subset*

Misalkan N menyatakan banyaknya subgrup dalam data binomial dan p menyatakan banyaknya parameter yang tidak diketahui dalam model logistik, dengan $N > p$. Dalam data binomial, pengamatannya bersesuaian dengan subgrup.

Langkah-langkah untuk memilih sebuah *initial subset* adalah sebagai berikut :

1. Daftarkan semua *subset-subset* yang berisi p subgrup yang mungkin. Banyaknya *subset* yang mungkin adalah sebanyak $\binom{N}{p} = k$ *subset*. Misalkan S_j^p adalah *subset-subset* yang mungkin, dengan $j = 1, 2, \ldots, k$.
2. Cocokkan model logistik sederhana terhadap tiap *subset* S_j^p yang diperoleh dari langkah 1. Parameter dalam model logistik untuk tiap *subset*, $\boldsymbol{\beta}^{(j)} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}^{(j)}$, ditaksir dengan metode maksimum *likelihood* seperti yang telah dibahas pada subbab 2.3.
3. Hitung kuadrat dari residual deviansi, $d_{i,S_j^p}^2$, untuk model yang dihasilkan dari tiap *subset* S_j^p. Kuadrat dari residual deviansi dinyatakan dengan

$$d_i^2 = \left[2y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + 2(n_i - y_i)\log\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right)\right] \quad (3.1)$$

dengan $i = 1, 2, \ldots, N$,

d_i^2 menyatakan kuadrat dari residual deviansi untuk subgrup ke-i,

y_i menyatakan nilai dari variabel random Y_i untuk subgrup ke-i,

$\hat{y}_i = n_i\hat{\pi}_i$ menyatakan nilai prediksi dari variabel random Y_i untuk subgrup ke-i, dengan n_i menyatakan ukuran subgrup ke-i.

4. Untuk tiap *subset* S_j^p, urutkan nilai-nilai d^2_{i,S_j^p} yang diperoleh pada langkah 3, dari kecil ke besar. Selanjutnya untuk tiap *subset* S_j^p tentukan median dari d^2_{i,S_j^p}. Yang dimaksud dengan median adalah nilai kuadrat dari residual deviansi yang ke-med, dengan med didefinisikan oleh

$$med = p + \left\lfloor \frac{N - p}{2} \right\rfloor$$

dengan $\lfloor (N - p)/2 \rfloor$ menyatakan bilangan bulat terbesar yang lebih kecil atau sama dengan $(N - p)/2$.

5. *Subset* yang memiliki median paling kecil dari langkah 4, dipilih sebagai *initial subset*, dan dinotasikan dengan S_*^p. Dengan kata lain, *initial subset*, S_*^p, dipilih sedemikian sehingga memenuhi

$$d^2_{[med],S_*^p} = \min_j (d^2_{[med],S_j^p}) \tag{3.2}$$

Kriteria (3.2) memberikan suatu kecocokkan yang *robust* terhadap data [Atkinson & Riani, 2001].

3.2.2 Penambahan Setiap Pengamatan ke dalam *Subset*

Secara umum, misalkan diberikan sebuah *subset* S_*^m yang berisi m subgrup, $m \geq p$. Pada umumnya satu unit subgrup baru masuk ke dalam *subset* S_*^m pada pergerakan *forward search* dari *subset* yang berukuran m, yaitu S_*^m, ke *subset* yang berukuran $m+1$, yaitu S_*^{m+1}. Namun, kadang-kadang dua atau lebih unit subgrup masuk ke dalam *subset* S_*^m diikuti dengan keluarnya satu atau lebih unit subgrup dari *subset*.

Langkah-langkah dalam tahap ini adalah sebagai berikut :

1. Cocokkan model logistik sederhana terhadap *subset* S_*^m. Parameter dalam model logistik ditaksir menggunakan metode maksimum *likelihood* seperti yang telah dijelaskan pada subbab 2.3. Misalkan $\hat{\boldsymbol{\beta}}^{(m)} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}^{(m)}$ adalah taksiran maksimum *likelihood* untuk *subset* S_*^m.
2. Hitung N kuadrat dari residual deviansi untuk *subset* S_*^m, yaitu d^2_{i,S_*^m}, $i =$ $1, 2, \ldots, N$, menggunakan persamaan (3.1).
3. Urutkan nilai kuadrat dari residual deviansi, d^2_{i,S_*^m}, dari kecil ke besar. Bentuk *subset* baru berisi $m+1$ subgrup dengan d^2_{i,S_*^m} terkecil. *Subset* tersebut dinotasikan dengan S_*^{m+1}.
4. Langkah di atas diulang-ulang hingga N subgrup masuk ke dalam *subset*. Misalkan $\hat{\boldsymbol{\beta}}_p, \ldots, \hat{\boldsymbol{\beta}}_m, \hat{\boldsymbol{\beta}}_{m+1}, \ldots, \hat{\boldsymbol{\beta}}_N$ berturut-turut menyatakan taksiran parameter model logistik sederhana untuk *subset* $S_*^p, \ldots, S_*^m, S_*^{m+1}, \ldots, S_*^N$.

3.2.3 Pemonitoran Statistik Diagnostik Selama Proses *Forward search*

Setiap kali dilakukan penaksiran parameter model logistik dari *subset* yang berukuran m, nilai dari statistik-statistik diagnostik dapat dihitung. Nilai dari statistik-statistik diagnostik tersebut akan diplot terhadap ukuran *subset* m.

Melalui serangkaian masuknya unit-unit subgrup ke dalam *subset* yang digunakan selama proses *forward search* pada tahap 3.2.2, akan diperoleh plot-plot sederhana yang memonitor perubahan pada statistik-statistik diagnostik. Dari plot-plot tersebut, terlihat pengaruh masuknya suatu subgrup ke dalam *subset* terhadap nilai statistik diagnostik. Statistik-statistik diagnostik yang akan diperiksa selama proses *forward search* adalah residual deviansi dan statistik skor. Statistik skor akan digunakan dalam pengujian kecukupan fungsi penghubung.

3.3 Plot Statistik-Statistik Diagnostik

3.3.1 Plot Residual Deviansi

Plot residual deviansi terhadap ukuran *subset* m, memonitor perubahan nilai-nilai residual deviansi untuk tiap subgrup pada masing-masing ukuran *subset*. Nilai-nilai tersebut dapat dihitung setelah parameter dalam model ditaksir dari masing-masing ukuran *subset* m, dengan $m = p$, $p+1,\ldots, N$. Nilai residual deviansi yang besar di antara subgrup yang tidak berada dalam *subset* mengindikasikan adanya outlier. Adanya outlier juga ditandai dengan perubahan yang tidak mulus atau perubahan yang tajam pada nilai residual deviansi dari suatu subgrup.

Dari plot residual deviansi yang dihasilkan, akan terlihat urutan masuknya tiap-tiap subgrup ke dalam *subset*, dengan kriteria nilai residual deviansi terkecil dari model yang dihasilkan pada setiap ukuran *subset* m. Sehingga dari plot residual deviansi juga dapat diidentifikasi subgrup-subgrup yang masuk ke dalam *subset* pada tahap akhir, yaitu subgrup yang memiliki nilai residual deviansi yang besar atau kecil.

3.3.2 Plot Pengujian Kecukupan Fungsi Penghubung

Plot lain yang dapat diperlihatkan oleh hasil prosedur *forward search* adalah plot yang memonitor statistik skor dalam pengujian kecukupan fungsi penghubung logistik dalam model. Nilai dari statistik skor dihitung untuk masing-masing ukuran *subset* m, dengan $m = p+1, p+2, \ldots, N$. Dalam penghitungan statistik skor, fungsi *log-likelihood* pada ukuran *subset* m dihitung untuk sebanyak m subgrup dalam *subset*. Jika masuknya suatu subgrup ke dalam *subset* menyebabkan nilai statistik skor menjadi keluar dari batas kepercayaan, subgrup tersebut dicurigai berbeda dengan yang lain, sehingga dapat disebut dengan outlier.

BAB IV

APLIKASI PROSEDUR *FORWARD SEARCH* UNTUK MEMERIKSA DATA BINOMIAL PADA MODEL LOGISTIK SEDERHANA

4.1 Data

Tabel 4.1 di bawah ini, memuat data yang terdiri dari delapan subgrup, dimana setiap subgrup berisi sekitar 60 ekor serangga. Setiap serangga pada masing-masing subgrup diberikan insektisida dengan dosis yang berbeda. Banyaknya serangga yang mati dicatat untuk setiap subgrup. Dalam hal ini, serangga yang mati merupakan kejadian sukses. Misalkan Y_i adalah variabel random yang menyatakan banyaknya serangga yang mati pada tiap subgrup setelah diberikan insektisida dengan dosis tertentu, dan X_i adalah variabel independen yang menyatakan logaritma (berbasis 10) dari dosis insektisida yang diberikan pada tiap subgrup.

Tabel 4.1 Data kematian serangga

Subgrup i	Logaritma dosis x_i	Banyaknya serangga n_i	Banyaknya serangga yang mati y_i
1	1,6907	59	6
2	1,7242	60	13
3	1,7552	62	18
4	1,7842	56	28

5	1,8113	63	52
6	1,8369	59	53
7	1,8610	62	61
8	1,8839	60	60

Sumber : Dobson, A.J. “An Introduction to Generalized Linear Models”. 1990.

4.2 Tujuan Analisis Data

Analisis data bertujuan untuk memeriksa kesesuaian data binomial pada Tabel 4.1 terhadap model logistik sederhana dengan menggunakan prosedur *forward search*. Kesesuaian data binomial terhadap model logistik sederhana dapat terlihat pada plot statistik-statistik diagnostik yaitu residual deviansi dan statistik skor, dimana statistik skor digunakan dalam pengujian kecukupan fungsi penghubung logistik dalam model.

4.3 Analisis Data

4.3.1 Prosedur Analisis Data

Pemeriksaan data binomial pada Tabel 4.1 terhadap model logistik sederhana menggunakan prosedur *forward search* dilakukan dengan mengaplikasikan fungsi *library forward* pada *software* S-PLUS versi 7.0. Penentuan taksiran parameter dalam model logistik menggunakan metode maksimum *likelihood*.

Berdasarkan output dari *software* S-PLUS versi 7.0, diperoleh urutan masuknya subgrup ke dalam *subset*, dari ukuran *subset* m = 4, m = 5, m = 6,

m = 7 hingga ukuran *subset* m = 8 berturut-turut adalah subgrup ke-8, ke-6, ke-4, ke-1 dan ke-2. Outputnya dapat dilihat pada Lampiran 3.a. Subgrup yang masuk pada tiap *subset* adalah subgrup yang memiliki kuadrat dari residual deviansi terkecil.

4.3.2 Taksiran Model Logistik Sederhana untuk Ukuran *Subset m*

Taksiran model logistik sederhana untuk data pada Tabel 4.1 adalah :

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x$$

atau

$$\hat{\pi} = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x)}{1+\exp(\hat{\beta}_0 + \hat{\beta}_1 x)} \tag{4.1}$$

dengan

$\hat{\pi}$ adalah taksiran probabilitas kematian serangga,

x menyatakan logaritma dari dosis insektisida,

$\hat{\beta}_0$ dan $\hat{\beta}_1$ menyatakan taksiran dari parameter β_0 dan β_1.

Output dari taksiran parameter model logistik, yaitu $\hat{\boldsymbol{\beta}}^{(m)} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}^{(m)}$,

untuk tiap ukuran *subset* m dengan m = 4, 5, 6, 7 dan 8, terdapat pada Lampiran 3.b. Dengan memasukkan nilai taksiran parameter ke dalam model logistik pada persamaan (4.1), diperoleh model sebagai berikut :

Model logistik untuk *subset* berukuran m = 4 adalah

$$\hat{\pi} = \frac{\exp(-82,84 + 46,66x)}{1 + \exp(-82,84 + 46,66x)}$$

Model logistik untuk *subset* berukuran m = 5 adalah

$$\hat{\pi} = \frac{\exp(-76,78 + 43,22x)}{1 + \exp(-76,78 + 43,22x)}$$

Model logistik untuk *subset* berukuran m = 6 adalah

$$\hat{\pi} = \frac{\exp(-78,28 + 44x)}{1 + \exp(-78,28 + 44x)}$$

Model logistik untuk *subset* berukuran m = 7 adalah

$$\hat{\pi} = \frac{\exp(-65,14 + 36,71x)}{1 + \exp(-65,14 + 36,71x)}$$

Model logistik untuk *subset* berukuran m = 8 adalah

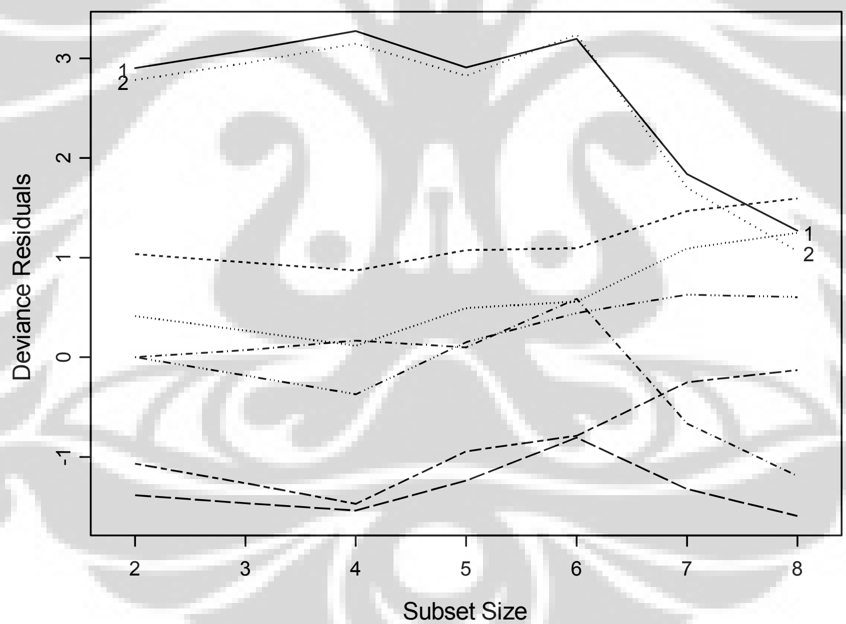
$$\hat{\pi} = \frac{\exp(-60,77 + 34,3x)}{1 + \exp(-60,77 + 34,3x)}$$

Dari Lampiran 3.b, diketahui bahwa perubahan nilai taksiran parameter dalam subset yang berukuran m = 4, 5 dan 6 cukup stabil. Namun, apabila nilai taksiran parameter dalam subset yang berukuran m = 4, 5 dan 6 dibandingkan dengan nilai taksiran parameter untuk subset yang berukuran m = 7 dan 8 mengalami perubahan yang cukup tajam. Hal ini berarti masuknya subgrup ke-1 dan ke-2 berturut-turut pada subset berukuran m = 7 dan m = 8, sangat mempengaruhi taksiran dari parameter dalam model logistik sederhana.

4.3.3 Interpretasi Plot Statistik-statistik Diagnostik

1. Plot Residual Deviansi

Plot nilai-nilai residual deviansi untuk 8 subgrup pada setiap ukuran *subset* m dapat dilihat pada Gambar 4.1. Output dari nilai-nilai residual deviansi tiap subgrup terhadap ukuran *subset* m dapat dilihat pada Lampiran 3.c.



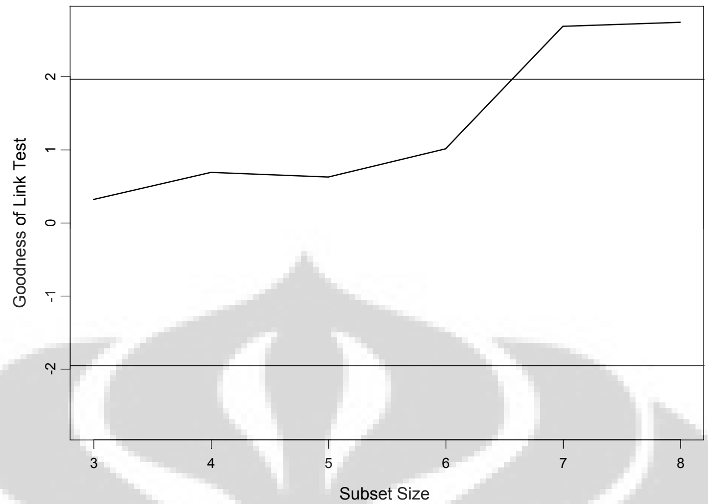
Gambar 4.1 Plot residual deviansi

Gambar 4.1 di atas memonitor nilai-nilai residual deviansi tiap subgrup terhadap ukuran *subset* m. Nilai residual deviansi untuk satu subgrup dihubungkan oleh garis yang sama.

Dari Gambar 4.1 terlihat bahwa subgrup ke-1 dan subgrup ke-2 adalah subgrup yang masuk ke dalam *subset* pada tahap akhir dari proses *forward search*. Selama kedua subgrup ini berada di luar *subset*, nilai residual deviansi untuk kedua subgrup ini jauh lebih besar jika dibandingkan dengan nilai residual deviansi untuk subgrup-subgrup lainnya. Setelah kedua subgrup ini masuk ke dalam *subset*, yaitu pada *subset* yang berukuran m = 7 dan m = 8, nilai residual deviansi untuk subgrup lainnya cenderung menjadi menjauhi nol. Dengan demikian subgrup ke-1 dan subgrup ke-2 dicurigai sebagai outlier.

2. Plot Pengujian Kecukupan Fungsi Penghubung (*Goodness of Link*)

Plot dari pengujian *goodness of link* pada setiap ukuran *subset* m dapat dilihat pada Gambar 4.2. Statistik yang digunakan dalam pengujian ini adalah statistik skor. Nilai-nilai dari statistik skor pada tiap ukuran *subset* m dapat dilihat pada Lampiran 3.d.



Gambar 4.2 Plot pengujian kecukupan fungsi penghubung

Gambar 4.2 di atas memonitor perubahan nilai-nilai statistik skor dalam pengujian kecukupan fungsi penghubung terhadap ukuran *subset* m. Nilai statistik skor dihitung untuk masing-masing ukuran *subset* m. Dari Gambar 4.2 terlihat bahwa nilai statistik skor untuk *subset* yang berukuran m = 7 dan m = 8 keluar dari batas kepercayaan 95% (tingkat signifikansi yang digunakan adalah $\alpha = 0,05$). Jika dilihat dari output nilai statistik skor pada Lampiran 3.d, nilai statistik skor untuk *subset* m = 7 dan m = 8 berturut-turut adalah 2,6935 dan 2,7486. Kedua nilai ini lebih besar dari nilai kritis yang diperoleh dari distribusi Normal(0,1), yaitu $z_{\alpha/2}$ = 1,96. Sehingga dari pengujian kecukupan fungsi penghubung diperoleh penolakan hipotesis nol, yang berarti fungsi penghubung logistik kurang tepat untuk kedua *subset* ini.

Dari Lampiran 3.a, diketahui bahwa *subset* yang berukuran m = 7 dan m = 8 diperoleh berturut-turut setelah masuknya subgrup ke-1 dan ke-2. Dengan demikian berdasarkan pengujian kecukupan fungsi penghubung, fungsi penghubung logistik kurang cocok untuk subgrup ke-1 dan ke-2 dari data pada Tabel 4.1.

Adanya outlier dari data binomial pada Tabel 4.1, dapat disebabkan karena kurangnya suku kuadratik dari variabel independen, atau mungkin juga memang fungsi penghubung yang digunakan dalam model kurang tepat untuk data. Hal ini terlihat dari plot pengujian kecukupan fungsi penghubung yang mengindikasikan ketidakcocokan fungsi penghubung logistik untuk dua *subset* terakhir. Oleh karena itu, untuk mengatasi hal ini, dapat dicoba dengan menambahkan suku kuadratik dari variabel independen atau menggunakan fungsi penghubung lain pada model.

BAB V

KESIMPULAN

Kesimpulan yang dapat diambil dari penulisan skripsi ini adalah sebagai berikut.

1. Ketidakcocokan model logistik sederhana terhadap data binomial dapat disebabkan oleh beberapa hal, diantaranya karena data mengandung outlier dan kesalahan dalam pemilihan fungsi penghubung yang digunakan dalam model.
2. Prosedur diagnostik yang dapat digunakan untuk memeriksa kesesuaian data binomial pada model logistik sederhana adalah prosedur *forward search*.
3. Prosedur *forward search* terdiri dari tiga tahapan, yaitu pemilihan sebuah *initial subset*, penambahan satu per satu subgrup ke dalam *subset* dan pemonitoran statistik diagnostik setiap penambahan satu subgrup ke dalam *subset*.
4. Output dari prosedur *forward search* berupa plot-plot grafik yang memonitor perubahan nilai-nilai statistik diagnostik. Dari plot-plot ini dapat dilihat pengaruh setiap subgrup terhadap residual deviansi dan pengujian kecukupan fungsi penghubung. Dari plot residual deviansi dapat dideteksi outlier dalam data binomial, sedangkan dari plot pengujian kecukupan fungsi penghubung logistik dapat diketahui pengaruh setiap subgrup dalam data binomial terhadap statistik skor.

5. Dari penerapan prosedur *forward search* dalam pemeriksaan data kematian serangga, didapatkan bahwa subgrup ke-1 dan subgrup ke-2 kurang cocok pada model logistik sederhana. Dari plot residual deviansi, terlihat bahwa subgrup ke-1 dan subgrup ke-2 dicurigai sebagai outlier. Sedangkan dari plot pengujian kecukupan fungsi penghubung, terlihat bahwa nilai statistik skor untuk kedua subgrup ini keluar dari batas kepercayaan, sehingga fungsi penghubung logistik kurang cocok untuk kedua subgrup ini.

DAFTAR PUSTAKA

Atkinson, A.C & M. Riani. 2000. *Robust diagnostic regression analysis*. Springer-Verlag, New York.

Atkinson, A.C & M. Riani. 2001. Regression diagnostics for binomial data from the forward search. *Journal of the Royal Statistical Society, Series D* **50**: 63-78.

Collett, D. 1991. *Modelling binary data*: *first edition*. Chapman and Hall, London.

Dobson, A.J. 1990. *An introduction to generalized linear models: first edition*. Chapman and Hall, London.

Green, W.H. 1997. *Econometric analysis: third edition*. Prentice-Hall, Inc., New Jersey.

Hosmer, D.W & S. Lemeshow. 2000. *Applied logistic regression: second edition*. John Wiley and Sons, New York.

Soemartojo, N. 1993. *Kalkulus: edisi ketiga*. Erlangga.

Winkelmann, R. 2003. *Econometric analysis of count data.* Springer-Verlag, New York.

LAMPIRAN 1

Menunjukkan $\Delta = \left\{ \left(\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0^2} \right) \left(\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_1^2} \right) - \left(\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} \right)^2 \right\} > 0$

dan $\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0^2} < 0$ atau $\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_1^2} < 0$

dipenuhi untuk setiap nilai β_0 dan β_1

Turunan parsial kedua dari $\log L(\boldsymbol{\beta})$ adalah

$$\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0^{\ 2}} = -\sum_{i=1}^{N} \frac{n_i \exp(\beta_0 + \beta_1 x_i)}{\left(1 + \exp(\beta_0 + \beta_1 x_i)\right)^2} < 0 \text{ , } \forall \ \beta_0, \beta_1$$

$$\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_1^{\ 2}} = -\sum_{i=1}^{N} \frac{n_i x_i^2 \exp(\beta_0 + \beta_1 x_i)}{\left(1 + \exp(\beta_0 + \beta_1 x_i)\right)^2} < 0 \text{ , } \forall \ \beta_0, \beta_1$$

Jadi $\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0^{\ 2}} < 0$ dan $\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_1^{\ 2}} < 0$, untuk setiap nilai β_0 dan β_1.

Karena $\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} = \frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_0} = -\sum_{i=1}^{N} \frac{n_i x_i \exp(\beta_0 + \beta_1 x_i)}{\left(1 + \exp(\beta_0 + \beta_1 x_i)\right)^2}$, maka

$$\Delta = \left\{ \left(\frac{\partial^2 \log L(\beta)}{\partial \beta_0^2} \right) \left(\frac{\partial^2 \log L(\beta)}{\partial \beta_1^2} \right) - \left(\frac{\partial^2 \log L(\beta)}{\partial \beta_0 \partial \beta_1} \right)^2 \right\}$$

$$= \left\{ \left(-\sum_{i=1}^{N} \frac{n_i \exp(\beta_0 + \beta_1 x_i)}{\left(1 + \exp(\beta_0 + \beta_1 x_i)\right)^2} \right) \left(-\sum_{i=1}^{N} \frac{n_i x_i^2 \exp(\beta_0 + \beta_1 x_i)}{\left(1 + \exp(\beta_0 + \beta_1 x_i)\right)^2} \right) - \left(-\sum_{i=1}^{N} \frac{n_i x_i \exp(\beta_0 + \beta_1 x_i)}{\left(1 + \exp(\beta_0 + \beta_1 x_i)\right)^2} \right)^2 \right\}$$

$$= \left\{ \left(\sum_{i=1}^{N} \frac{n_i \exp(\beta_0 + \beta_1 x_i)}{\left(1 + \exp(\beta_0 + \beta_1 x_i)\right)^2} \right) \left(\sum_{i=1}^{N} \frac{n_i x_i^2 \exp(\beta_0 + \beta_1 x_i)}{\left(1 + \exp(\beta_0 + \beta_1 x_i)\right)^2} \right) - \left(\sum_{i=1}^{N} \frac{n_i x_i \exp(\beta_0 + \beta_1 x_i)}{\left(1 + \exp(\beta_0 + \beta_1 x_i)\right)^2} \right)^2 \right\}$$

dengan mengganti $\pi_i = \dfrac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$ diperoleh

$$\Delta = \left\{ \left(\sum_{i=1}^{N} n_i \pi_i (1 - \pi_i) \right) \left(\sum_{i=1}^{N} n_i x_i^2 \pi_i (1 - \pi_i) \right) - \left(\sum_{i=1}^{N} n_i x_i \pi_i (1 - \pi_i) \right)^2 \right\}$$

$$= \left\{ \begin{array}{l} \left[\begin{array}{l} \left(n_1 \pi_1 (1 - \pi_1) + n_2 \pi_2 (1 - \pi_2) + \ldots + n_N \pi_N (1 - \pi_N) \right) \\ \left(n_1 x_1^2 \pi_1 (1 - \pi_1) + n_2 x_2^2 \pi_2 (1 - \pi_2) + \ldots + n_N x_N^2 \pi_N (1 - \pi_N) \right) \end{array} \right] - \\ \left[\begin{array}{l} \left(n_1 x_1 \pi_1 (1 - \pi_1) + n_2 x_2 \pi_2 (1 - \pi_2) + \ldots + n_N x_N \pi_N (1 - \pi_N) \right) \\ \left(n_1 x_1 \pi_1 (1 - \pi_1) + n_2 x_2 \pi_2 (1 - \pi_2) + \ldots + n_N x_N \pi_N (1 - \pi_N) \right) \end{array} \right] \end{array} \right\}$$

$$= \left\{ \begin{array}{l} \left[\begin{array}{l} \left(n_1^2 x_1^2 \pi_1^2 (1 - \pi_1)^2 + n_1 n_2 x_2^2 \pi_1 \pi_2 (1 - \pi_1)(1 - \pi_2) + \ldots + n_1 n_N x_N^2 \pi_1 \pi_N (1 - \pi_1)(1 - \pi_N) \right) + \\ (n_1 n_2 x_1^2 \pi_1 \pi_2 (1 - \pi_1)(1 - \pi_2) + n_2^2 x_2^2 \pi_2^2 (1 - \pi_2)^2 + \ldots + n_2 n_N x_N^2 \pi_2 \pi_N (1 - \pi_2)(1 - \pi_N)) + \\ \ldots + \left(n_1 n_N x_1^2 \pi_1 \pi_N (1 - \pi_1)(1 - \pi_N) + n_2 n_N x_2^2 \pi_2 \pi_N (1 - \pi_2)(1 - \pi_N) + \ldots + n_N^2 x_N^2 \pi_N^2 (1 - \pi_N)^2 \right) \end{array} \right] \\ - \left[\begin{array}{l} \left(n_1^2 x_1^2 \pi_1^2 (1 - \pi_1)^2 + n_1 n_2 x_1 x_2 \pi_1 \pi_2 (1 - \pi_1)(1 - \pi_2) + \ldots + n_1 n_N x_1 x_N \pi_1 \pi_N (1 - \pi_1)(1 - \pi_N) \right) + \\ (n_1 n_2 x_1 x_2 \pi_1 \pi_2 (1 - \pi_1)(1 - \pi_2) + n_2^2 x_2^2 \pi_2^2 (1 - \pi_2)^2 + \ldots + n_2 n_N x_2 x_N \pi_2 \pi_N (1 - \pi_2)(1 - \pi_N)) + \\ \ldots + n_1 n_N x_1 x_N \pi_1 \pi_N (1 - \pi_1)(1 - \pi_N) + n_2 n_N x_2 x_N \pi_2 \pi_N (1 - \pi_2)(1 - \pi_N) + \ldots + n_N^2 x_N^2 x_2 \pi_N^2 \\ (1 - \pi_N)^2 \end{array} \right] \end{array} \right\}$$

$$= \left\{ \begin{array}{l} n_1 n_2 \pi_1 \pi_2 (1 - \pi_1)(1 - \pi_2)(x_1 - x_2)^2 + \ldots + n_1 n_N \pi_1 \pi_N (1 - \pi_1) \\ (1 - \pi_N)(x_1 - x_N)^2 + \ldots + n_2 n_N \pi_2 \pi_N (1 - \pi_2)(1 - \pi_N)(x_2 - x_N)^2 + \ldots \end{array} \right\} \qquad (*)$$

Karena $0 < \pi < 1$, maka (*) akan bernilai positif untuk setiap nilai β_0 dan β_1, sehingga $\Delta > 0$ untuk setiap nilai β_0 dan β_1.

Jadi persamaan (2.3.4) dipenuhi untuk setiap nilai β_0 dan β_1.

LAMPIRAN 2

Prosedur Iterasi *Newton Raphson*

Diketahui $\log L(\boldsymbol{\beta})$ adalah fungsi *log-likelihood* dari parameter $\boldsymbol{\beta}$, $\boldsymbol{\beta} = [\beta_0, \beta_1]^T$. Misalkan $\mathbf{U}(\boldsymbol{\beta})$ adalah vektor turunan parsial pertama dari $\log L(\boldsymbol{\beta})$ dengan elemen $\frac{\partial \log L(\boldsymbol{\beta})}{\partial \beta_j}$, untuk $j = 0, 1$. Misalkan $\mathbf{H}(\boldsymbol{\beta})$ adalah matriks turunan parsial kedua dari $\log L(\boldsymbol{\beta})$ dengan elemen $\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_j}$, untuk $j, k = 0, 1$. Matriks $\mathbf{H}(\boldsymbol{\beta})$ disebut matriks Hessian, dan ditulis dengan

$$\mathbf{H}(\boldsymbol{\beta}) = \begin{bmatrix} \dfrac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0^{\,2}} & \dfrac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_0} \\ \dfrac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} & \dfrac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_1^{\,2}} \end{bmatrix}$$

Dengan menggunakan prosedur iterasi *Newton Raphson* akan dicari nilai taksiran dari $\boldsymbol{\beta}$, yaitu $\hat{\boldsymbol{\beta}} = [\hat{\beta}_0, \hat{\beta}_1]^{\mathrm{T}}$ yang merupakan penyelesaian dari persamaan $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$. Jika diberikan nilai taksiran awal dari $\boldsymbol{\beta}$ yaitu $\hat{\boldsymbol{\beta}}^{(0)}$, maka dapat diperoleh pendekatan deret Taylor orde pertama dari vektor $\mathbf{U}(\boldsymbol{\beta})$ di sekitar $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(0)}$ [Green, 1997], yaitu :

$$\mathbf{U}(\boldsymbol{\beta}) \approx \mathbf{U}(\hat{\boldsymbol{\beta}}^{(0)}) + \left.\frac{\partial \mathbf{U}(\boldsymbol{\beta})}{\partial \beta_0}\right|_{\beta=\hat{\beta}^{(0)}} \left(\beta_0 - \hat{\beta}_0^{(0)}\right) + \left.\frac{\partial \mathbf{U}(\boldsymbol{\beta})}{\partial \beta_1}\right|_{\beta=\hat{\beta}^{(0)}} \left(\beta_1 - \hat{\beta}_1^{(0)}\right)$$

$$= \mathbf{U}(\hat{\boldsymbol{\beta}}^{(0)}) + \begin{bmatrix} \left.\frac{\partial \mathbf{U}(\boldsymbol{\beta})}{\partial \beta_0}\right|_{\beta=\hat{\beta}^{(0)}} & \left.\frac{\partial \mathbf{U}(\boldsymbol{\beta})}{\partial \beta_1}\right|_{\beta=\hat{\beta}^{(0)}} \end{bmatrix} \begin{bmatrix} \beta_0 - \hat{\beta}_0^{(0)} \\ \beta_1 - \hat{\beta}_1^{(0)} \end{bmatrix}$$

$$= \mathbf{U}(\hat{\boldsymbol{\beta}}^{(0)}) + \begin{bmatrix} \frac{\partial \mathbf{U}(\boldsymbol{\beta})}{\partial \beta_0} & \frac{\partial \mathbf{U}(\boldsymbol{\beta})}{\partial \beta_1} \end{bmatrix}_{\beta=\hat{\beta}^{(0)}} \begin{bmatrix} \beta_0 - \hat{\beta}_0^{(0)} \\ \beta_1 - \hat{\beta}_1^{(0)} \end{bmatrix}$$

$$= \mathbf{U}(\hat{\boldsymbol{\beta}}^{(0)}) + \begin{bmatrix} \frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0{}^2} & \frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_0} \\ \frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_1{}^2} \end{bmatrix}_{\beta=\hat{\beta}^{(0)}} \begin{bmatrix} \beta_0 - \hat{\beta}_0^{(0)} \\ \beta_1 - \hat{\beta}_1^{(0)} \end{bmatrix}$$

$$= \mathbf{U}(\hat{\boldsymbol{\beta}}^{(0)}) + \mathbf{H}(\hat{\boldsymbol{\beta}}^{(0)})\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}^{(0)}\right)$$

dengan

$\mathbf{U}(\hat{\boldsymbol{\beta}}^{(0)})$ adalah vektor turunan parsial pertama dari $\log L(\boldsymbol{\beta})$ dengan

elemen $\frac{\partial \log L(\beta)}{\partial \beta_j}$, untuk $j = 0, 1$, yang dihitung pada $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(0)}$,

$\mathbf{H}(\hat{\boldsymbol{\beta}}^{(0)})$ adalah matriks Hessian dengan elemen $\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_k \beta_j}$, untuk $j, k =$

$0, 1$, yang dihitung pada $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(0)}$.

Karena ingin dicari penyelesaian dari $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$, maka

$$\mathbf{0} = \mathbf{U}(\boldsymbol{\beta}) = \mathbf{U}(\hat{\boldsymbol{\beta}}^{(0)}) + \mathbf{H}(\hat{\boldsymbol{\beta}}^{(0)})\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}^{(0)}\right)$$

$$\mathbf{H}(\hat{\boldsymbol{\beta}}^{(0)})\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{(0)}\right) = -\mathbf{U}(\hat{\boldsymbol{\beta}}^{(0)})$$

Dengan diasumsikan bahwa matriks $\mathbf{H}(\hat{\boldsymbol{\beta}}^{(0)})$ tidak singular (dari Lampiran 1 telah ditunjukkan bahwa determinan matriks $\mathbf{H}(\boldsymbol{\beta})$ dari fungsi *log-likelihood* binomial, yang dinotasikan dengan Δ tidak nol), maka diperoleh

$$\left[\mathbf{H}(\hat{\boldsymbol{\beta}}^{(0)})\right]^{-1}\mathbf{H}(\hat{\boldsymbol{\beta}}^{(0)})\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}^{(0)}\right)=-\left[\mathbf{H}(\hat{\boldsymbol{\beta}}^{(0)})\right]^{-1}\mathbf{U}(\hat{\boldsymbol{\beta}}^{(0)})$$

$$\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}^{(0)}\right)=-\left[\mathbf{H}(\hat{\boldsymbol{\beta}}^{(0)})\right]^{-1}\mathbf{U}(\hat{\boldsymbol{\beta}}^{(0)})$$

$$\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(0)}-\left[\mathbf{H}(\hat{\boldsymbol{\beta}}^{(0)})\right]^{-1}\mathbf{U}(\hat{\boldsymbol{\beta}}^{(0)}) \qquad \text{(L2.1)}$$

Nilai $\boldsymbol{\beta}$ yang diperoleh dari persamaan (L2.1) dinotasikan dengan $\hat{\boldsymbol{\beta}}^{(1)}$, sehingga persamaan (L2.1) menjadi :

$$\hat{\boldsymbol{\beta}}^{(1)}=\hat{\boldsymbol{\beta}}^{(0)}-\left[\mathbf{H}(\hat{\boldsymbol{\beta}}^{(0)})\right]^{-1}\mathbf{U}(\hat{\boldsymbol{\beta}}^{(0)}) \qquad \text{(L2.2)}$$

Dengan langkah yang sama seperti untuk mendapatkan $\hat{\boldsymbol{\beta}}^{(1)}$ pada persamaan (L2.2), dapat diperoleh taksiran $\hat{\boldsymbol{\beta}}^{(2)}$, $\hat{\boldsymbol{\beta}}^{(3)}$ dan seterusnya.

Secara umum, nilai $\hat{\boldsymbol{\beta}}$ pada iterasi ke-(t+1) adalah sebagai berikut :

$$\hat{\boldsymbol{\beta}}^{(t+1)}=\hat{\boldsymbol{\beta}}^{(t)}-\left[\mathbf{H}(\hat{\boldsymbol{\beta}}^{(t)})\right]^{-1}\mathbf{U}(\hat{\boldsymbol{\beta}}^{(t)})$$

dengan $t = 0, 1, \ldots$

$\hat{\boldsymbol{\beta}}^{(t)}$ adalah taksiran dari $\boldsymbol{\beta}$ pada iterasi ke-t,

$\hat{\boldsymbol{\beta}}^{(t+1)}$ adalah taksiran dari $\boldsymbol{\beta}$ pada iterasi ke-(t+1),

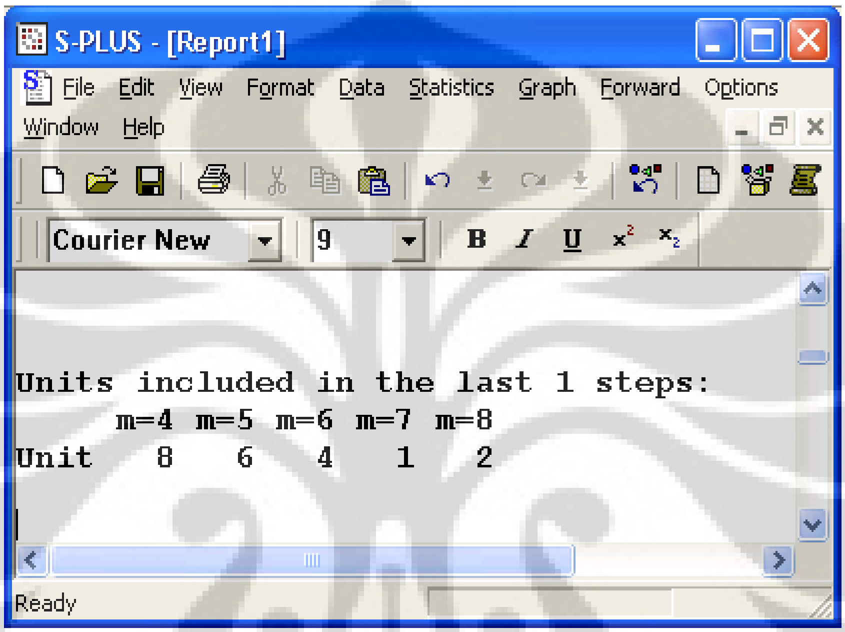
$\mathbf{U}\left(\hat{\boldsymbol{\beta}}^{(t)}\right)$ adalah vektor turunan parsial pertama dari $\log L(\boldsymbol{\beta})$ dengan elemen $\dfrac{\partial \log L(\boldsymbol{\beta})}{\partial \beta_j}$, untuk $j = 0, 1$, yang dihitung pada $\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(t)}$,

$\mathbf{H}\left(\hat{\boldsymbol{\beta}}^{(t)}\right)$ adalah matriks Hessian dengan elemen $\dfrac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_k \beta_j}$, untuk $j, k =$ $0, 1$, yang dihitung pada $\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(t)}$.

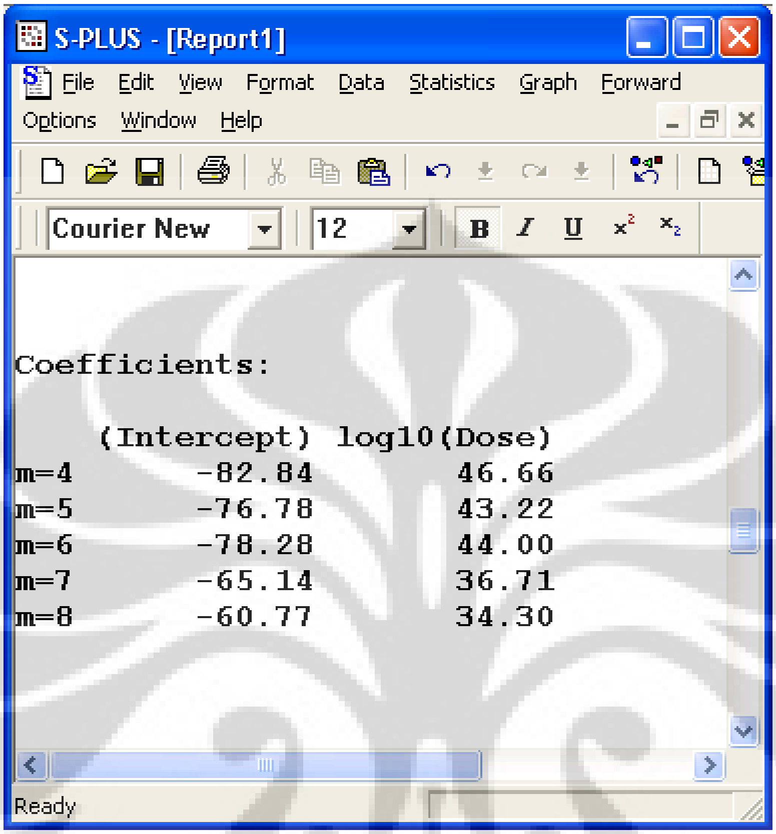
LAMPIRAN 3

Output *Software* S-PLUS Versi 7.0

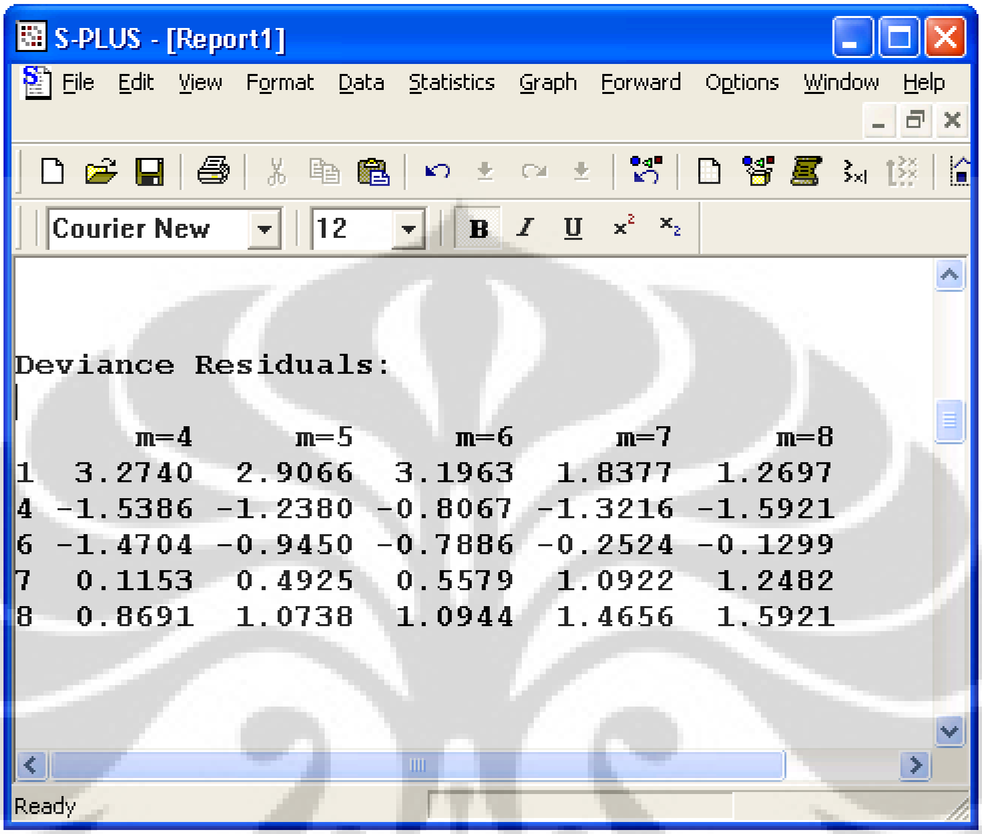
3.a Output Urutan Masuknya Subgrup ke dalam Ukuran *Subset* m



3.b Output dari Taksiran Parameter untuk Ukuran *Subset* m



3.c Output dari Nilai-nilai Residual Deviansi untuk Ukuran *Subset m*



Deviance Residuals:

	m=4	m=5	m=6	m=7	m=8
1	3.2740	2.9066	3.1963	1.8377	1.2697
4	-1.5386	-1.2380	-0.8067	-1.3216	-1.5921
6	-1.4704	-0.9450	-0.7886	-0.2524	-0.1299
7	0.1153	0.4925	0.5579	1.0922	1.2482
8	0.8691	1.0738	1.0944	1.4656	1.5921

3.d Output dari Nilai Statistik Skor dalam Pengujian Kecukupan Fungsi Penghubung untuk Ukuran *Subset* m