PENAKSIRAN PARAMETER MODEL REGRESI LOGISTIK DENGAN METODE *LOCAL MAXIMUM LIKELIHOOD*

S K R I P S I

Oleh:

SUTRIANAH BURHAN

H12108852

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2012

PENAKSIRAN PARAMETER MODEL REGRESI LOGISTIK DENGAN METODE *LOCAL MAXIMUM LIKELIHOOD*

S K R I P S I

Oleh:

SUTRIANAH BURHAN

H 12108852

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2012

PENAKSIRAN PARAMETER MODEL REGRESI LOGISTIK

DENGAN METODE *LOCAL MAXIMUM LIKELIHOOD*

S K R I P S I

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada

Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Hasanuddin

Makassar

Oleh :

SUTRIANAH BURHAN

H12108852

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2012

<u>LEMBAR KEOTENTIKAN</u>

Saya yang bertanda tangan di bawah ini menyatakan dengan sesungguh-sungguhnya bahwa skripsi yang saya buat dengan judul :

"PENAKSIRAN PARAMETER MODEL REGRESI LOGISTIK DENGAN METODE *LOCAL MAXIMUM LIKELIHOOD*"

adalah benar hasil kerja saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 13 Agustus 2012

<u>SUTRIANAH BURHAN</u>
NIM : H 121 08 852

PENAKSIRAN PARAMETER MODEL REGRESI LOGISTIK

DENGAN METODE *LOCAL MAXIMUM LIKELIHOOD*

Disetujui Oleh :

Pembimbing Utama	Pembimbing Pertama
Drs.Alimin Bado, MS.	**A.Kresna Jaya, S.Si, M.Si**
NIP. 19540617 197702 1 001	**NIP. 19620926 198702 2 001**

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

Pada hari ini.Senintanggal 13Agustus 2012, Panitia Ujian Skripsi menerima dengan baik skripsi yang berjudul :

PENAKSIRAN PARAMETER MODEL REGRESI LOGISTIK DENGAN METODE *LOCAL MAXIMUM LIKELIHOOD*

yang diajukan untuk memenuhi salah satu syarat guna memperoleh gelar Sarjana Sains pada Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Makassar,13 Agustus 2012

PANITIA UJIAN SKRIPSI

		Tanda Tangan
1. Ketua	: **Drs. NirwanIlyas,M.Si**	u/n. Ka. Jur. Mat. (..........................)
2. Sekretaris	: **Anisa, S.Si, M.Si**	(..........................)
3. Anggota	: **Dr.LoekyHaryanto,MS.MSc.MAT**	(..........................)
4. Anggota	: **Drs. AliminBado, MS.**	(..........................)
5. Anggota	: **A.KresnaJaya, S.Si, M.Si**	(..........................)

KATA PENGANTAR

بِسْمِ اللهِ الرَّحْمٰنِ الرَّحِيْمِ

Puji dan syukur penulis panjatkan kehadirat **Allah SWT**, atas segala rahmat dan karunia-Nya, nikmat kesehatan dan kekuatan yang diberikan kepada penulis sehingga dapat menyelesaikan penulisan skripsi dengan judul **"Penaksiran Parameter Regresi Logistik dengan Metode *Local Maximum Likelihood*"** sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin Makassar.

Tak lupa penulis haturkan shalawat dan salam kepada Rasulullah **Muhammad SAW**, sebagai satu-satunya suri tauladan dalam menjalankan kehidupan dunia dan akhirat.

Dalam penyelesaian skripsi ini, penulis telah melewati perjuangan panjang dan pengorbanan yang tidak sedikit. Namun berkat rahmat dan izin-Nya serta dukungan dan bantuan dari berbagai pihak baik moril maupun materil, langsung maupun tidak langsung, sehingga akhirnya skripsi ini dapat terselesaikan. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang tulus dan penghargaan yang tak terhingga kepada Ayahanda dan Ibunda tercinta **Burhan B, BA** dan **Dra. Hj. Subaedah**, yang telah membesarkan dan mendidik penulis dengan penuh kesabaran, limpahan cinta, kasih sayang, dan penuh ketulusan hati serta telah memberikan dukungan dan doa kepada penulis dan menjadi motivasi terbesar bagi penulis untuk segera menyelesaikan studi. Ucapan terimah kasih juga penulis ucapkan kepada Kakekku dan Tanteku tersayang **H.Abd.Salam, Salmiah, Hj. Nuraeni** yang selalu memberikan doa, kasih sayang yang tak terhingga dan menjadi semangat bagi penulis. Serta kepada Saudara-saudaraku tersayang **Sulfianah**, **Suparmanto**, **Sulfikar**, **Sulkiram**, **Sulaeman** yang selalu memberikan perhatian, bantuan serta kasih sayang yang tak terhingga.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. Bapak **Drs. Alimin Bado, MS** selaku dosen pembimbing utama atas kesediaan untuk membimbing dan memberikan saran, kritikan dalam penyusunan skripsi ini.
2. Bapak **Andi Kresna Jaya, S.Si, M.Si** selaku dosen pembimbing pertama atas kesediaan dan kesabaran untuk membimbing dan membagi ilmunya kepada penulis dalam penyusunan skripsi ini.
3. Bapak **Drs. Nirwan Ilyas, M.Si** selaku ketua penguji, Ibu **Anisa, S.Si, M.Si** selaku sekretaris penguji, dan Bapak **Dr. Loeky Haryanto, MS, M.Sc, MAT** selaku anggota tim penguji yang telah meluangkan waktunya untuk memberikan saran dan arahan kepada penulis dalam penyusunan skripsi ini.
4. Bapak **Dr. Loeky Haryanto, MS, M.Sc, MAT** selaku penasehat akademik yang telah meluangkan waktunya untuk memberikan saran dan kritik.
5. Ibu **Dr. Hasmawati, M.Si** selaku ketua jurusan Matematika, Bapak **Dr. Nurdin, M.Si**, selaku sekretaris jurusan Matematika, Bapak **Drs. Saleh AF, M.Si** selaku ketua program studi Statistika, seluruh Dosen Jurusan

Matematika FMIPA Unhas atas ilmu yan telah dibagi kepada penulis, serta staf pegawai jurusan Matematika FMIPA Unhas (Bapak **Nasir** dan Bapak **Sutamin**) terima kasih atas fasilitas dan kemudahan yang diberikan kepada penulis sejak kuliah hingga terselesaikannya penyusunan skripsi ini.

6. My best friend **Armianty, Jumarni, Sutriani, Umairah Amina,** dan **Wirnancy Juliasari** kalian tidak akan terlupakan dalam kehidupan ini. Terima kasih untuk semua canda, tawa, suka, duka, serta pengalaman-pengalaman hidup yang akan terkenang selamanya.
7. Terima kasih buat **Herdianti**, **Dewi Santri, Nuraidah Puput S, Fera Rusanti, Aqifah Noerfitri, Hikmat Ghazali** atas bantuan dan dukungannya selama ini.
8. Seluruh saudara-saudariku **Statistika 08 (Army, Cute, Marni, Ume, Irna, Puput, Qkhy, Fera, Ila, Balqis, Anty, Santri, Sitti, Hilda, Lita, Ilen, Ayu, Rahmah, Hikmat, Santuo, Zerah, Ichal, Adin, Yusran, Qadli, Dading, Muhlis, Ahmi, Agung, Rian, Ute, Cullank, Ammank, Ahdi, Eka, Ury, Adnan, Koko, Asnawi).** Kebersamaan kita telah memberikan warna tersendiri yang akan selalu membekas di hati.
9. Kawan-kawan **Matematika 08** terkhusus **Nurfuaidah S** dan **Nurahmah**, **MIPA 08**, Kanda-kanda, teman-teman, dan adik-adik **Warga Himatika FMIPA Unhas**, **Pengurus BEM FMIPA Unhas Periode 2011/2012** tanpa terkecuali. Terima kasih atas dukungan kalian selama ini.

10. Teman-teman seperjuangan selama KP (**Aqifah Noerfitri**) dan KKTS (**Aqifah Noerfitri** dan **Nur Hilal),** terima kasih telah banyak membantu dan memberi dukungan.

Semua pihak yang tak sempat disebutkan satu persatu atas segala bentuk bantuan dan perhatiannya hingga terselesaikannya skripsi ini. Semoga segala bantuan dan partisipasi yang telah diberikan bernilai ibadah dan mendapat pahala yang setimpal di sisi **Allah SWT**. Akhir kata, semoga tulisan ini dapat memberikan manfaat bagi semua pihak yang membutuhkan dan terutama bagi penulis.

Amin Yaa Rabbal Alamin.

Makassar, 13 Agustus 2012

Penulis

ABSTRAK

Regresi logistik termasuk dalam keluarga eksponensial. Model regresi logistik juga termasuk dalam *Generalized Linier Model* (GLM) dengan variabel respon biner yang berdistribusi Bernoulli. Pada tulisan ini, yang dibahas adalah penaksiran parameter model regresi logistik dengan metode *Local Maximum Likelihood*. Pada hasil penaksiran ini tidak memberikan penyelesaian karena parameter-parameternya masih saling terkait satu sama lain. Oleh karena itu, penaksiran parameternya harus dilakukan dengan metode iterasi Newton-Raphson dengan pendekatan deret Taylor linear orde dua. Model regresi logistik diaplikasikan pada data sekunder kejadian kanker. Setelah dilakukan uji kelayakan model regresi dengan uji *goodness of fit.* Model taksiran regresi logistik layak digunakan.

Kata kunci : regresi, GLM, regresi logistik, *Local Maximum Likelihood*, iterasi Newton-Raphson, uji*goodness of fit*.

ABSTRACT

Logistic regression including the exponential family. Logistic regression model was also included in the Generalized Linear Model (GLM) with binary response variables Bernoulli distributed. In this paper, which discussed the estimation parameter logistic regression model with Local Maximum Likelihood method. In the results of this assessment does not provide a solution because its parameters are interlinked with each other. Therefore, the parameter estimation should be done with the Newton-Raphson iteration method with the approach of linear Taylor series second-order. Logistic regression model was applied to the secondary data of cancer incidence. Having tested the feasibility of a regression model with a goodness of fit test. The estimated logistic regression models fit for use.

Key words : Regression, GLM, logistic regression, Local Maximum Likelihood, Newton-Raphson iteration, goodness of fit test

DAFTAR ISI

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam ilmu statistik, model regresi merupakan komponen penting dalam beberapa analisis data dengan menggambarkan hubungan antara variabel respon dan satu atau beberapa variabel penjelas. Variabel respon dapat berupa data yang bersifat kuantitatif akan tetapi sering juga ditemui kasus dengan variabel respon berupa data yang bersifat kualitatif atau kategori. Variabel respon yang bersifat kualitatif yang dimaksud adalah variabel respon yang hanya memiliki dua hasil yang mungkin. Contoh: setuju atau tidak setuju, sukses atau gagal, lulus atau tidak lulus. Yang biasanya menggunakan dua buah bilangan 0 atau 1 untuk menggantikan kategori pada variabel respon sehingga dapat didekati dengan distribusi Bernoulli. Distribusi Bernoulli merupakan salah satu distribusi yang termasuk dalam keluarga eksponensial. Distribusi-distribusi yang termasuk dalam keluarga eksponensial yaitu distribusi Poisson, Gamma, Normal, Binomial.

Analisis yang digunakan untuk membantu menyelesaikan masalah tersebut adalah model regresi logistik. Regresi logistik dapat dibedakan menjadi dua, yaitu: Regresi logistik biner dan Regresi logistik multinomial. Regresi logistik biner digunakan ketika hanya ada dua kemugkinan variabel respon sedangkan regresi logistik multinomial digunakan ketika pada variabel respon terdapat lebih dari dua kategori. Analisis regresi logistik biner bertujuan utuk memperoleh hubungan antara X dan Y. Berapapun nilai X bila disubtitusikan ke dalam fungsi logistik hasilnya akan berkisar antara 0 dan 1.

Model regresi logistik dengan pendekatan *Generalized Linier Model* (GLM) dapat digunakan tanpa memenuhi asumsi seperti pada regresi linier klasik. Salah satu metode yang dapat digunakan dalam mengestimasi parameter model logistik biner adalah *Local Maximum Likelihood Estimation*. Penaksiran yang dilakukan menggunakan teknik *Local Maximum Likelihood* yang diperkenalkan oleh Tibshirani (1982) dalam konteks data tersensor dan model Hazard proportional.

Berdasarkan uraian latar belakang diatas, penulis mengambil judul **"Penaksiran Parameter Regresi Logistik dengan Metode *Local Maximum Likelihood*".**

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana cara menaksir parameter pada regresi logistik biner dengan metode *Local Maximum Likelihood* ?
2. Bagaimana penerapan regresi logistik biner dengan metode *Local Maximum Likelihood* pada suatu data penelitian?

1.3 Batasan Masalah

Permasalahan dalam tugas akhir ini dibatasi pada model regresi logistik biner dan metode yang digunakan untuk mengestimasi model regresi logistik biner yaitu metode *Local Maximum Likelihood.* Data yang digunakan adalah data kejadian kanker payudara pada wanita di R.S. Bhayangkara Mappa Oudang, R.S. Pelamonia dan RSUP. Dr. Wahidin Sudirohusodo Makassar tahun 2011

1.4 Tujuan Penulisan

Berdasarkan rumusan masalah, maka tugas akhir ini memiliki tujuan sebagai berikut :

1. Menentukan taksiran parameter pada regresi logistik biner dengan metode *Local Maximum Likelihood.*
2. Menerapkan regresi logistik biner dengan metode *Local Maximum Likelihood* pada suatu data penelitian.

1.5 Manfaat Penulisan

Manfaat yang diperoleh dari penulisan ini adalah sebagai referensi kepada pembaca tentang penaksiran parameter dari model regresi logistik biner dengan menggunakan metode *Local Maximum Likelihood* dan tentang model regresi logistik biner.

BAB II

TINJAUAN PUSTAKA

2.1 Regresi

Persamaan regresi adalah persamaan matematika yang memungkinkan untuk meramalkan nilai-nilai suatu variabel respon dari nilai-nilai satu atau lebih variabel penjelas (Walpole, 1995).

Kuat atau tidaknya hubungan variabel penjelas (X) dan variabel respon (Y) diukur dengan suatu nilai yang disebut dengan koefisien korelasi, sedangkan besarnya pengaruh X terhadap Y, diukur dengan koefisien regresi. Persamaan regresi juga menggambarkan relasi dari variabel-variabel yang ada di dalamnya.

Di dalam pemakaiannya, variabel respon (Y) ternyata juga dipengaruhi oleh faktor lain selain variabel penjelas (X) yang tidak dimasukkan kedalam persamaan tersebut. Oleh karena persamaan dari regresi perlu untuk mengambarkan bentuk dari data dengan tepat, maka dimasukkanlah *error* ε ke dalam persamaan regresi tersebut. Karena *error* itu tidak dapat dihilangkan sama sekali, maka resiko itu akan selalu ada. Resiko hanya bisa diperkecil dengan memperkecil kesalahan (*minimized error*).

2.2 Generalized Linier Model (GLM)

Generalized Linier Model adalah sebuah pengembangan dari model linier klasik. Model linier klasik mengasumsikan variabel responnya berdistribusi normal. Sementara pada kenyataannya sering ditemukan suatu kondisi dengan distribusi variabel responnya tidak selalu berdistribusi normal. Untuk mengatasi kondisi dengan variabel respon yang ada tidak berdistribusi normal, tetapi masih saling bebas, maka para statistikawan yang dipelopori oleh Nelder telah mengembangkan model linier yang dikenal dengan istilah Generalized Linier Model. Model linier ini menggunakan asumsi bahwa variabel respon memiliki distribusi yang termasuk dalam keluarga eksponensial. Beberapa distribusi yang termasuk dalam distribusi keluarga eksponensial adalah distribusi normal, Poisson, binomial, eksponensial, dan distribusi gamma.

2.3 Distribusi Bernoulli

Sebuah percobaan Bernoulli dilakukan beberapa kali sehingga kemungkinan keberhasilan, misalkan p, tetap sama dari satu percobaan ke percobaan yang lain. Artinya, p menyatakan probabilitas keberhasilan pada setiap percobaan.

Misalkan Y adalah variabel acak yang terkait dengan percobaan Bernoulli dengan mendefinisikan sebagai berikut:

$$Y\ (sukses) = 1 \quad dan \quad Y\ (gagal) = 0$$

Mempunyai dua hasil yang muncul yaitu sukses dan gagal dengan masing-masing peluang p dan $p - 1$, maka sukses dinotasikan dengan 1 dan gagal dengan 0. Maka fungsi kepadatan peluang Y dapat ditulis sebagai berikut:

$$f(y) = p^y(1-p)^{1-y}, \quad y = 0{,}1$$

Bentuk distribusi peluang untuk peubah acak Y dari percobaan Bernoulli disebut distribusi Bernoulli. Nilai yang diharapkan dari Y yaitu:

$$\mu = E(Y) = \sum_{y=0}^{1} yp^y(1-p)^{1-y} = (0)(1-p) + (1)(p) = p$$

Dan variansi dari Y yaitu:

$$\sigma^2 = var\ (Y) = E[(y-p)^2] = \sum_{y=0}^{1}(y-p)^2p^y(1-p)^{1-y}$$

$$= p^2(1-p) + (1-p)^2p$$

$$= p(1-p)$$

Standar deviasi Y adalah $\sigma = \sqrt{p(1-p)}$

2.4 Regresi Logistik

Regresi logistik biner merupakan salah satu pendekatan model matematika yang digunakan untuk menganalisis hubungan beberapa faktor dengan sebuah variabel yang bersifat dikotomus (biner). Pada regresi logistik jika variabel responnya terdiri dari dua kategori misalnya $Y = 1$ menyatakan hasil yang diperoleh "sukses" dan $Y = 0$ menyatakan hasil yang diperoleh "gagal" maka regresi logistik tersebut menggunakan regresi logistik biner.

Menurut Agresti (1990), variabel Y yang demikian lebih tepat dikatakan sebagai variabel indikator dan memenuhi distribusi Bernoulli. Fungsi distribusi peluang untuk Y dengan parameter p_i adalah

$$f(y_i; p_i) = \begin{cases} {p_i}^{y_i}(1 - p_i)^{1-y_i} & y_i = 0{,}1 \\ 0 \; untuk & y_i \; yang \; lain \end{cases} \tag{2.1}$$

dengan $p_i = P(Y_i = 1)$. Dari fungsi distribusi tersebut diperoleh rata-rata:

$$E(Y_i) = 1. P(Y_i = 1) + 0. P(Y_i = 0)$$

$$= P(Y_i = 1)$$

Misalkan probabilitas ini dinotasikan sebagai $p(\boldsymbol{x})$ yang bergantung dengan variabel penjelas $X = (X_1, \cdots, X_k)$ dengan $E(Y) = p$ dan $0 \leq p \leq 1$, sehingga diperoleh

$$E(Y^2) = 1^2 p(x) + 0^2[1 - p(x)] = p(x)$$

dan varians dari Y adalah

$$Var\ (Y) = E(Y^2) - [E(Y)]^2 = p(x)[1 - p(x)]$$

Secara umum model probabilitas regresi logistik dengan satu variabel penjelas (x) dapat diformulasikan sebagai berikut:

$$E(y|x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \tag{2.2}$$

dimana $E(y|x)$ merupakan penjumlahan dari $p(x)$. Fungsi $p(x)$ merupakan fungsi nonlinier sehingga perlu dilakukan transformasi logit untuk memperoleh fungsi yang linier agar dapat dilihat hubungan antara variabel respon (y) dengan variabel penjelas (x). Bentuk logit dari $p(x)$ dinyatakan sebagai $g(x)$, yaitu:

$$g(x) = ln\left(\frac{p(x)}{1-p(x)}\right) \tag{2.3}$$

Persamaan (2.2) dan persamaan (2.3) disubtitusikan sehingga diperoleh

$$g(x) = ln\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x \tag{2.4}$$

dengan demikian fungsi respon logit $g(x) = \beta_0 + \beta_1 x$ telah linier dalam parameter-parameternya.

Untuk regresi logistik dengan beberapa variabel penjelas, $\beta_0 + \beta_1 x$ pada persamaan (2.2) diganti dengan $\beta_0 + \sum_{j=1}^{k} \beta_j x_j$, dengan k adalah jumlah variabel penjelas. Selanjutnya model regresi logistik dapat dituliskan dalam bentuk:

$$E(y|x) = \frac{\exp(\beta_0 + \sum_{j=1}^{k} \beta_j x_j)}{1+\exp(\beta_0 + \sum_{j=1}^{k} \beta_j x_j)} \tag{2.5}$$

dengan demikian fungsi respon logit $g(x) = \beta_0 + \sum_{j=1}^{k} \beta_j x_j$

2.5 Maximum Likelihood Estimation

Metode *Maximum Likelihood Estimation* digunakan untuk mengestimasi parameter-parameter dalam regresi logistik dan pada dasarnya metode *Maximum Likelihood Estimation* memberikan nilai estimasi β dengan memaksimumkan fungsi *likelihood*-nya. Secara matematis fungsi padat peluang $f(y_i;p_i)$ dapat dinyatakan:

$$f(y_i;p_i) = {p_i}^{y_i}(1-p_i)^{1-y_i}$$

Karena setiap pengamatan diasumsikan saling bebas maka fungsi *likelihood*-nya merupakan perkalian antara masing-masing fungsi padat peluang $f(y_i;p_i)$, yaitu:

$$L(p_i|y_1,y_2,\cdots,y_n) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i}$$

Nilai parameter β dapat diperoleh dengan memaksimumkan fungsi *likelihood*-nya. Hal tersebut dilakukan dengan metode turunan pertama fungsi *likelihood*-nya terhadap setiap parameternya yang disamakan dengan nol. Terkadang sulit menemukan turunan dari fungsi *likelihood*-nya sehingga yang dilakukan adalah menemukan nilai maksimum dari logaritma natural fungsi *likelihood* tersebut atau fungsi *log-likelihood*.

Fungsi *log-likelihood* adalah bentuk logaritma dari fungsi *likelihood*, yang dituliskan dalam bentuk:

$$\ln L(p_i|y_1,y_2,\cdots,y_n) = \sum_{i=1}^{n} \ln({p_i}^{y_i}(1-p_i)^{1-y_i})$$

Berikut langkah-langkah dalam menentukan penduga parameter dengan metode *Maximum Likelihood Estimation:*

1. Menentukan fungsi *likelihood*

$$L(p_i|y_1, y_2, \cdots, y_n) = \prod_{i=1}^{n} p_i^{y_i}(1 - p_i)^{1-y_i}$$

2. Menentukan fungsi *log-likelihood*

$$\ln L(p_i|y_1, y_2, \cdots, y_n) = \sum_{i=1}^{n} \ln p_i{}^{y_i}(1 - p_i)^{1-y_i}$$

3. Memaksimumkan fungsi *log-likelihood*

   Memaksimumkan fungsi *log-likelihood* untuk memperoleh nilai $\hat{\beta}$ dapat dilakukan dengan langkah-langkah berikut:

   a) Nilai $\hat{\beta}$ diperoleh dari turunan pertama dengan disamadengankan nol

   $$\left.\frac{\partial l}{\partial \beta}\right|_{\beta=\hat{\beta}} = 0$$

   b) Nilai $\hat{\beta}$ dikatakan memaksimumkan $L(\beta)$ jika

   $$\left.\frac{\partial^2 l}{\partial^2 \beta}\right|_{\beta=\hat{\beta}} < 0$$

4. Menyelesaikan fungsi *log-likelihood* yang diperoleh pada langkah 2 atau 3 dan mendapatkan $\hat{\beta}$ sebagai estimator *Maximum Likelihood Estimation*.

2.6 Metode Newton-Raphson

Apabila langkah menaksir parameter menggunakan metode *Maximum Likelihood Estimation (MLE)* menghasilkan fungsi *log-likelihood* yang nonlinier, maka penyelesaian fungsi tersebut untuk memperoleh nilai parameter taksirannya menggunakan metode Newton-Raphson (Jong & Heller, 2008:69). Metode ini merupakan metode perhitungan yang iteratif, sehingga akan lebih mudah dikerjakan dengan bantuan komputer.

Metode Newton-Raphson didasarkan pada deret Taylor sebagai berikut:

$$f(x_{t+1}) = f(x_t) + f'(x_t)(x_{t+1} - x_t) + \frac{f''(x_t)}{2!}(x_{t+1} - x_t)^2 + \cdots + \frac{f^n(x_t)}{n!}(x_{t+1} - x_t)^n$$

Fungsi *log-likelihood* dengan parameter θ dapat diselesaikan sehingga diperoleh nilai $\hat{\theta}$ taksiran dengan menggunakan metode Newton-Raphson. Rumus penaksiran parameter $\hat{\theta}$ pada iterasi ke-$(t+1)$ dalam proses iterasi $(t = 0,1,2,\cdots)$ adalah sebagai berikut:

$$\hat{\theta}_{(t+1)} = \hat{\theta}_{(t)} - \boldsymbol{D}\left(\theta_{(t)}\right)^{-\mathbf{1}} \boldsymbol{d}(\theta_{(t)}) \quad (2.7)$$

dengan

$\hat{\theta}_{(t+1)}$: parameter taksiran θ pada iterasi ke-$(t+1)$

$\hat{\theta}_{(t)}$: parameter taksiran θ pada iterasi ke-(t)

$\boldsymbol{d}(\theta_{(t)})$: turunan pertama fungsi *likelihood*, sehingga entri dari $\boldsymbol{d}(\theta_{(t)})$ adalah $\frac{\partial}{\partial \theta_{(t)}} l|_{\theta}$

$\boldsymbol{D}(\theta_{(t)})$: turunan kedua fungsi *likelihood,* sehingga entri dari $\boldsymbol{d}(\theta_{(t)})$ adalah $\frac{\partial^2}{\partial(\theta_{(t)})^2} l|_\theta$

Formulasi (2.7) berlaku bukan hanya θ bernilai skalar (titik) tapi juga berlaku bila θ berbentuk vektor nilai $\theta = (\beta_1, \beta_2, \cdots, \beta_k)^\theta$

2.7 Uji *Goodness of Fit*

Uji *Goodness of Fit* adalah suatu alat statistik yang digunakan untuk memeriksa kelayakan respon dari model yang layak digunakan dibandingkan dengan data yang diamati. Kelayakan dikatakan baik jika ada kesesuaian antara data yang dicocokkan dengan data yang diamati. Metode yang digunakan untuk *goodness of fit* data kategori adalah *Deviance*. Metode *Deviance* didasarkan pada kriteria rasio likelihood untuk membandingkan model *current* (model tanpa variabel penjelas) dengan model penuh (model dengan variabel penjelas).

Langkah-langkah pengujian dengan uji *goodness of fit* adalah:

1. Merumuskan hipotesis

   H_0: Model layak digunakan

   H_1: Model tidak layak digunakan

2. Memilih taraf signifikansi: α, yaitu $\alpha = 0{,}05$

3. Menentukan statistik uji

   Statistik uji *deviance* didefinisikan oleh

$$Deviance = -2\sum_{i=1}^{n}\left\{y_i \ln\left(\frac{\hat{p}_i}{y_i}\right) + (1 - y_i)\ln\left(\frac{1-\hat{p}_i}{1-y_i}\right)\right\} \quad (2.8)$$

4. Kriteria keputusan:

   H_0 diterima jika $deviance \leq X^2_{(\alpha,n-k)}$

   n: banyaknya pengamatan,

   k: banyaknya parameter

5. Kesimpulan : Jika H_0 diterima, maka dapat disimpulkan bahwa model layak digunakan.

2.8 Kanker Payudara

Kanker payudara (*Carcinoma mammae*) adalah tumor ganas yang menyerang jaringan payudara, jaringan payudara tersebut terdiri dari kelenjar susu, saluran kelenjar dan jaringan penunjang payudara seperti lemak dan saraf. Kanker payudara sebagai jenis kanker yang paling banyak ditakuti oleh wanita karena kanker payudara banyak menyerang wanita.

Penyebab pasti kanker payudara tidak diketahui. Meskipun demikian, riset mengidentifikasi sejumlah faktor yang dapat meningkatkan risiko pada individu tertentu adalah sebagai berikut:

1. Usia

   Risiko utama kanker payudara adalah bertambahnya umur. Wanita di atas 30 tahun berisiko terkena kanker payudara.

2. Tingkat Pendidikan

   Tingkat pendidikan adalah salah satu faktor risiko kanker payudara karena semakin tinggi pendidikan seseorang semakin cepat melakukan pencegahan kanker payudara dan tahu cara deteksi dini.

3. Jenis Pekerjaan

   Salah satu faktor risiko kanker payudara adalah jenis pekerjaan karena beberapa pekerjaan tertentu menyebabkan wanita terpajang faktor risiko kanker payudara seperti stress dan bahan kimia atau sinar radioaktif.

4. Usia saat menstruasi pertama

   Wanita dengan usia saat menstruasi pertama sebelum 12 tahun, lebih berisiko tinggi terkena kanker payudara. Mendapatkan menstruasi pertama pada usia muda dapat menyebabkan terjadinya pertukaran hormon dan peningkatkan pertukaran zat hormon. Perempuan mengalami perubahan hormonal terus-menerus. Saat itulah ada kemungkinan perubahan sel dalam payudara sehingga terjadi mutasi sel.

5. Usia saat menopause

   Wanita yang berisiko terkena kanker payudara adalah wanita yang mengalami menopause terlambat. Semakin cepat menopause mengurangi risiko kanker. Setiap terlambat setahun umur menopause risiko kanker naik dengan 3%. Dengan demikian jika menopause pada usia 55 tahun risikonya naik 35% dibanding menopause 45 tahun.

Tidak ada langkah yang sangat efektif dalam hal pencegahan kanker payudara dan pada tahap awal kanker payudara tidak menimbulkan gejala apapun. Untuk itu dianjurkan untuk melakukan pemeriksaan sebagai berikut:

1. Pemeriksaan payudara sendiri
2. Pemeriksaan payudara oleh tenaga medis (dokter atau bidan)
3. Mamografi

4. Cara lainnya adalah dengan operasi kecil untuk mengambil contoh jaringan dari benjolan itu.