

Pemodelan Pasien Kanker Payudara Menggunakan Regresi Logistik Biner (Studi Kasus : Pasien Kanker Payudara di Rumah Sakit Umum Daerah Arifin Ahmad Pekanbaru)

Rahmadeni¹, Eka Safitri²

^{1,2}Jurusan Matematika Fakultas Sains dan Teknologi UIN Suska Riau, UIN Sultan Syarif Kasim Riau
Jl. HR. Soebrantas No. 155 Simpang Baru, Panam, Pekanbaru, 28293
E-mail: r4dieni@gmail.com

(Received: 29 April 2016; Revised: 20 Juni 2016; Accepted: 20 Juni 2016)

ABSTRAK

Kanker Payudara merupakan salah satu jenis kanker yang tercatat sebagai tingkat tertinggi di dunia. Pasien kanker payudara memiliki dua kemungkinan, yaitu kanker *benign* (jinak) dan *malignant* (ganas). Data yang digunakan dalam penelitian ini adalah umur, berat badan, dan kadar hemoglobin pasien kanker payudara yang diambil dari bagian Rekam Medis Rumah Sakit Umum Daerah (RSUD) Arifin Ahmad Pekanbaru. Adapun parameter β_0 , β_1 , β_2 dan β_3 dari distribusi Logistik Biner dapat diestimasi dengan menggunakan metode *Maximum Likelihood Estimation* (MLE) dan diperoleh $\beta_0 = -3,757$, $\beta_1 = 0,104$, $\beta_2 = 0,053$, dan $\beta_3 = -0,198$.

Kata Kunci: *benign*, kanker payudara, *malignant*, metode *Maximum Likelihood Estimation*.

ABSTRACT

Breast cancer is one type of cancer that is listed as the world's highest level. Breast cancer patients have two possibilities, namely benign dan malignant cancer. The data used in this research were age, body weight, and hemoglobin levels o breast cancer patients who were taken from the Medical Record in the General Hospital Arifin Ahmad Pekanbaru. The parameters β_0 , β_1 , β_2 and β_3 of Binary Logistic Distribution can be estimated using Maximum Likelihood Estimation Method (MLE) and obtained $\beta_0 = -3,757$, $\beta_1 = 0,104$, $\beta_2 = 0,053$, and $\beta_3 = -0,198$.

Keywords: *benign*, breast cancer, *malignant*, *Maximum Likelihood Estimation method*.

Corresponding Author

Rahmadeni,
Program Studi Matematika, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarfi Kasim Riau,
Email: r4dieni@gmail.com

Pendahuluan

Saat ini penyakit tidak menular, termasuk kanker menjadi masalah kesehatan utama baik di dunia maupun di Indonesia. Kanker yang tercatat sebagai tingkat tertinggi di Indonesia adalah kanker yang sering ditemukan pada wanita, yaitu kanker payudara. Kanker payudara sudah menjadi suatu hal yang sangat menakutkan bagi kaum wanita baik di Indonesia maupun di seluruh dunia. Terutama bagi wanita yang berumur antara 30-40 tahun, tingkat

pasien kanker payudara meningkat secara drastis dari tahun ke tahun (Kompas, 2015). Berdasarkan Estimasi Globocan, *International Agency for Research on Cancer* (IARC) Tahun 2012, insiden kanker payudara tercatat sebesar 40,3% per 100.000 perempuan atau kira-kira 48.998 orang.

Dari berbagai survei diatas menunjukkan bahwa angka penderita kanker payudara sangat tinggi dan akan semakin meningkat untuk tahun-tahun berikutnya, sehingga perlu adanya penanggulangan dari masalah ini. Diagnosis pasien kanker payudara mempunyai dua kemungkinan

(dikotomi), yaitu *benign* (jinak) dan *malignant* (ganas). Dalam hal ini, untuk penelitian tidak hanya dapat digunakan dengan regresi linier biasa, tetapi dapat digunakan metode regresi logistik biner. Metode regresi logistik biner dapat digunakan untuk meneliti hubungan antara beberapa variabel penjelas dengan variabel respon dikotomi.

Penelitian tentang kanker payudara berdasarkan faktor resiko dengan menggunakan regresi logistik pernah dilakukan oleh Purwantaka. Ketepatan klasifikasi yang didapatkan dari model regresi logistik pada kasus ini hanya sebesar 37%. Selain itu telah dilakukan juga beberapa penelitian tentang diagnosis kanker payudara berbasis *Support Vector Machine* (SVM) yang memberikan ketepatan klasifikasi diatas 95%. Dan kemudian juga dilakukan analisis perbandingan antara model regresi logistik dan SVM dengan data mamografi pada pasien kanker payudara yang menghasilkan ketepatan klasifikasi pada model regresi logistik sebesar 88,72%, sedangkan dengan SVM didapat ketepatan klasifikasi sebesar 94,34% (Novianti, 2012).

Landasan Teori

Kanker Payudara

Kanker payudara adalah suatu penyakit dimana terjadi pertumbuhan berlebihan atau perkembangan tidak terkontrol dari sel-sel jaringan payudara. payudara wanita terdiri dari lobulus (kelenjar susu), duktus (saluran susu), lemak dan jaringan ikat, pembuluh darah dan *limfe*. Sebagian besar kanker payudara bermula pada sel-sel yang melapisi duktus (kanker duktal), beberapa bermula di lobulus (kanker lobular), serta sebagian kecil bermula di jaringan lain (Novianti, 2012). Tanda awal dari kanker payudara biasanya dengan terdapatnya benjolan pada payudara yang pada mulanya hanya kecil kemudian semakin lama semakin membesar, dan ketika ditekan tidak terasa nyeri.

Beberapa variabel yang penulis ambil dari data pasien kanker payudara pada bagian Rekam Medik di RSUD Arifin Ahmad Pekanbaru adalah sebagai berikut :

a. Umur

Faktor utama kanker payudara adalah bertambahnya umur. Wanita dengan umur 50-69 tahun adalah ketegori umur yang paling beresiko terkena kanker payudara, terutama yang mengalami menopause terlambat.

b. Berat badan

Wanita yang memiliki berat badan berlebih atau biasa disebut obesitas juga memiliki resiko terkena kanker payudara, karena kelebihan lemak

ditubuh meningkatkan hormon estrogen dan insulin yang menjadi penyebab umum kanker payudara.

c. Kadar Hemoglobin

Hemoglobin (Hb) adalah protein kompleks yang ada dalam sel darah merah (eritrosit) mengandung zat besi dan berwarna merah. Kadar hemoglobin yang normal untuk wanita dalah 11,4 sampai 15,1 g/dl, sedangkan untuk pria adalah 3,4 sampai 17,7 g/dl. Jika kadar hemoglobin seseorang yang mengidap kanker payudara terus menurun atau rendah, maka akan berdampak kepada keganasan kanker.

Regresi Logistik Biner

Regresi logistik biner merupakan suatu metode yang digunakan untuk melihat suatu hubungan antara beberapa variabel penjelas dengan suatu variabel respon dikotomi. Variabel respon biasanya disimbolkan dengan y dan variabel penjelas/prediktor disimbolkan dengan x , dimana variabel respon berskala biner atau dikotomus. Dikatakan biner atau dikotomus adalah karena memiliki dua kategori yaitu 0 dan 1, sangat berbeda dengan regresi linier biasa yang variabel responnya bisa bernilai < 0 atau > 1 .

Variabel respon (y) mengikuti distribusi bernouli dengan fungsi probabilitas sebagai berikut:

$$f(y_i) = \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}, \quad (1)$$

Kemudian Bentuk umum dari model peluang regresi logistik biner adalah sebagai berikut :

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \quad (2)$$

Maksimum Likelihood

Maksimum *likelihood* adalah metode pendugaan yang digunakan untuk memaksimumkan fungsi *likelihood*. Sedangkan fungsi *likelihood* itu sendiri adalah fungsi densitas bersama (*joint density function*) dari n variabel random X_1, X_2, \dots, X_n dan dinyatakan dalam bentuk $f(x_1, x_2, \dots, x_n | \theta)$. Jika x_1, x_2, \dots, x_n tetap, maka fungsi *likelihood* adalah fungsi dari parameter θ dan dinotasikan dengan $L(\theta)$ dan *likelihood*nya dapat ditulis sebagai berikut:

$$L(\theta) = f(x_1 | \theta) f(x_2 | \theta) \dots f(x_n | \theta) = \prod_{i=1}^n f(x_i | \theta) \quad (3)$$

Metode Newton Raphson

Metode iterasi Newton-Raphson, yaitu suatu proses iterasi yang digunakan untuk menyelesaikan persamaan-persamaan non-linier. Proses iterasi adalah suatu teknik penghampiran yang berulang-ulang dimana setiap pengulangan disebut iterasi. Jika hampiran tersebut menghasilkan suatu pemecahan yang sangat dekat dengan pemecahan persamaan yang tidak linier tersebut, maka iterasi telah mengalami proses konvergen.

Uji Hipotesis

Uji hipotesis statistik dilakukan untuk menentukan apakah variabel prediktor dalam model signifikan atau berpengaruh nyata terhadap variabel respon. Uji hipotesis dilakukan dengan cara sebagai berikut :

a. Uji Serentak (simultan)

Uji serentak atau disebut juga uji model Chi-square digunakan untuk menguji parameter hasil estimasi secara bersama. Langkah-langkah pengujian adalah sebagai berikut:

$$1. H_0 = \beta_1 = \beta_2 = \dots = \beta_k = 0 \quad (\text{secara}$$

simultan variabel bebas tidak berpengaruh terhadap variabel terikat)

paling tidak ada satu $\beta_i \neq 0$, untuk

$i = 1, 2, \dots, k$ (secara simultan variabel bebas

berpengaruh terhadap variabel terikat)

2. Statistik uji yang digunakan adalah uji G atau *Likelihood ratio test* :

$$G^2 = -2 \ln \left[\frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^k \pi(x_i)^{y_i} (1-\pi(x_i))^{1-y_i}} \right] \quad (4)$$

Keterangan :

n_1 = banyaknya observasi yang berkategori 1

n_0 = banyaknya observasi yang berkategori 0

n = banyaknya observasi ($n_1 + n_0$)

b. Uji Parsial

Dalam uji parsial ini, akan dilakukan pengujian terhadap setiap variabel bebas secara individual sehingga dapat dilihat apakah suatu variabel bebas layak untuk masuk kedalam model atau tidak. Langkah-langkahnya adalah sebagai berikut :

$$1. H_0: \beta_i = 0 \text{ (tidak ada pengaruh variabel}$$

bebas ke-i terhadap variabel terikat.

$$H_0: \beta_i \neq 0 \text{ (ada pengaruh variabel bebas ke-i}$$

terhadap variabel terikat.

2. Statistik uji yang digunakan adalah uji Wald :

$$w = \frac{\beta_i}{SE(\beta_i)} \quad (5)$$

Keterangan:

w = nilai statistik uji wald

β_i = estimasi parameter ke-i

$SE(\beta_i)$ = standar error estimasi ke-i

Metodologi Penelitian

1. Data

Data yang digunakan adalah data pasien penyakit kanker payudara Tahun 2014 pada bagian Rekam Medis di Rumah Sakit Umum Daerah (RSUD) Arifin Ahmad Pekanbaru. Kemudian menentukan variabel-variabel yang akan digunakan, yaitu :

a. Variabel terikat adalah status pasien dengan dua kategori yaitu kanker payudara jinak dengan kode 0 dan kanker payudara ganas dengan kode 1.

b. Variabel bebas adalah umur (x_1), berat badan (x_2), kadar hemoglobin (x_3).

2. Metodologi Penelitian

Metodologi penelitian dalam penelitian ini adalah sebagai berikut :

1. Mengestimasi parameter dari regresi yang digunakan dengan metode Maximum Likelihood Estimation (MLE) dan iterasi menggunakan metode Newton Raphson dengan langkah-langkah sebagai berikut :

a. Memasukkan nilai dugaan awal yang sebelumnya diperoleh dari program SPSS 16.

b. Mencari turunan parsial terhadap parameter dan matriks jacobi.

c. Mendapatkan invers dari matriks jacobi

d. Selanjutnya melakukan iterasi dengan cara sebagai berikut:

$$x_1^{k+1} = x_1^k - (b_{11}^k f_1^k + b_{12}^k f_2^k + \dots + b_{1p}^k f_p^k)$$

$$x_2^{k+1} = x_2^k - (b_{21}^k f_1^k + b_{22}^k f_2^k + \dots + b_{2p}^k f_p^k) \quad (6)$$

$$\begin{aligned} & \vdots \\ & x_p^{k+1} = x_p^k - (b_{p1}^k f_1^k + b_{p2}^k f_2^k + \dots + b_{pp}^k f_p^k) \end{aligned} \quad (7)$$

$$(8)$$

Proses hampiran iterasi ini dapat diteruskan hingga setiap parameter memberikan selisih yang cukup kecil terhadap parameter sebelumnya atau mencapai konvergen.

2. Pengujian hipotesis terhadap variabel bebas yaitu umur (x_1), berat badan (x_2), dan kadar hemoglobin (x_3) secara serentak (simultan) serta secara individual (parsial) seperti yang sudah dijelaskan pada landasan teori.

Hasil dan Pembahasan

Metode Maximum Likelihood Estimation (MLE)

Metode *Maximum Likelihood Estimation* (MLE) digunakan untuk mengestimasi parameter distribusi logistik dengan cara mengalikan fungsi densitas peluangnya. Distribusi logistik mempunyai satu parameter yaitu parameter bentuk (β) dengan

fungsi densitasnya dinyatakan sebagai berikut :
 $f(y_i) = \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}$, dimana nilai $y_i = 0 ; 1$.

Dan akan didapat fungsi log-likelihood seperti berikut :

$$= \sum_{i=1}^n y_i \log \pi(x_i) - \sum_{i=1}^n (1 - y_i) \log \pi(x_i) \quad (9)$$

Kemudian untuk memaksimumkan fungsi log likelihood di atas, yang harus dilakukan adalah dengan cara mencari turunan parsial pada persamaan diatas terhadap 4 parameter yaitu $\beta_0, \beta_1, \beta_2$ dan β_3 .

Selanjutnya, akan terbentuk empat persamaan yang homogen yaitu

$$\frac{\partial L(\beta)}{\partial \beta_0} = g_1, \frac{\partial L(\beta)}{\partial \beta_1} = g_2, \frac{\partial L(\beta)}{\partial \beta_2} = g_3, \frac{\partial L(\beta)}{\partial \beta_3} = g_4, \quad (10)$$

Selanjutnya dari persamaan-persamaan yang didapat, akan dicari matriks jacobian seperti penjelasan berikutnya. Sebelum proses iterasi dilakukan, terlebih dahulu akan didapatkan matriks

jacobian yang diperoleh dengan mencari turunan parsialnya dari persamaan yang homogen.

Kemudian, Penulis menyelesaikan proses iterasi menggunakan *software* atau program *maple 13*. Proses hampiran iterasi telah selesai pada iterasi kelima dengan nilai parameter yaitu $\beta_0 = -1,840$, $\beta_1 = -0,133$, $\beta_2 = -0,031$, dan $\beta_3 = 0,529$. Selanjutnya untuk menentukan nilai parameter yang mendekati nilai sebenarnya dilakukan dengan bantuan program SPSS 16, dapat kita lihat pada tabel output SPSS 16 berikut :

Tabel 1. Nilai parameter

	B
Step Umur	.104
1 ^a Berat_badan	.053
Kadar_Hb	-.198
Constant	-3.757

Uji Hipotesis

Uji hipotesis statistik dilakukan untuk menentukan apakah variabel prediktor dalam model signifikan atau berpengaruh nyata terhadap variabel respon. Uji hipotesis dilakukan dengan cara sebagai berikut :

- c. Uji Serentak (simultan)
 Uji serentak atau disebut juga uji model Chi-square digunakan untuk menguji parameter hasil estimasi secara bersama. Langkah-langkah pengujian adalah sebagai berikut:

1. Hipotesis
 $H_0 = \beta_1 = \beta_2 = \dots = \beta_k = 0$ (secara simultan variabel bebas tidak berpengaruh terhadap variabel terikat)
 $H_1 = \beta_1 \neq 0$, untuk $i = 1, 2, \dots, k$ (secara simultan variabel bebas

berpengaruh terhadap variabel terikat) maka,
 H_0 = Tidak adanya pengaruh umur, berat badan, dan kadar hemoglobin terhadap seseorang mengidap penyakit kanker payudara ganas atau jinak.
 H_1 = Adanya pengaruh umur, berat badan, dan kadar hemoglobin terhadap seseorang mengidap penyakit kanker payudara ganas atau jinak.

2. Uji serentak menggunakan uji Chi-Kuadrat (χ^2), berdasarkan pengujian yang dilakukan menggunakan program SPSS 16 maka didapat nilai χ^2_{hitung} adalah sebagai berikut :

Tabel 2. Hasil uji Chi-Kuadrat

	Chi-square	Df	Sig
Step 1	13.434	3	.004
Step	13.434	3	.004
Block	13.434	3	.004

Maka, $\chi^2_{hitung} = 13,434 > \chi^2_{tabel} = 7,815$

Tolak $H_0 \rightarrow$

Artinya, diketahui bahwa adanya pengaruh dari umur, berat badan dan kadar hemoglobin terhadap seseorang mengidap penyakit kanker payudara ganas atau jinak.

3. Nilai determinan (R^2) digunakan untuk melihat seberapa besar pengaruh dari variabel penjelas terhadap variabel respon. Menggunakan program SPSS 16 didapat nilai dari R^2 adalah sebagai berikut:

Tabel 3. Nilai determinan

Observed	Predicted		
	Pasien		Percentage correct
	0	1	
Step 0 pasien 0			
1	0	19	0
Overall	0	69	100.0
Percentage			78.4

Berdasarkan tabel, dapat dilihat bahwa nilai dari R^2 adalah 78,4. Artinya terdapat pengaruh sebesar 78,4% dari umur, berat badan, dan kadar hemoglobin terhadap seseorang yang mengidap penyakit kanker payudara ganas atau jinak dan 21,6 % dipengaruhi oleh faktor-faktor lain.

d. Uji Parsial

Dalam uji parsial ini, akan dilakukan pengujian terhadap setiap variabel bebas secara individual sehingga dapat dilihat apakah suatu variabel bebas layak untuk masuk kedalam model atau tidak. Langkah-langkahnya adalah sebagai berikut :

1. Hipotesis
2. $H_0: \beta_i = 0$ (tidak ada pengaruh variabel

bebas ke-i terhadap variabel terikat.

$H_0: \beta_i \neq 0$ (ada pengaruh variabel bebas ke-i

terhadap variabel terikat.

maka,

$H_0 =$ Tidak adanya pengaruh umur, berat badan, dan kadar hemoglobin terhadap seseorang mengidap penyakit kanker payudara ganas atau jinak.

$H_1 =$ Adanya pengaruh umur, berat badan, dan kadar hemoglobin terhadap seseorang mengidap penyakit kanker payudara ganas atau jinak.

3. Uji serentak menggunakan uji Wald dengan nilai $\alpha = 0,05$, berdasarkan pengujian yang dilakukan menggunakan program SPSS 16 dan nilai signifikan masing-masing variabel adalah :

Sig 1 = 0.003 < 0,05 terima H_0

Sig 2 = 0,339 > 0,05 tolak H_0

Sig 3 = 0,299 > 0,05 tolak H_0

Dilihat dari hasil uji diatas, dapat ditarik kesimpulan sebagai berikut :

a. Untuk variabel x_1 (umur), tidak ada

pengaruh yang signifikan terhadap seseorang mengidap penyakit kanker payudara ganas atau jinak.

b. Untuk variabel x_2 (berat badan), adanya

pengaruh yang signifikan terhadap seseorang mengidap penyakit kanker payudara ganas atau jinak.

c. Untuk variabel x_3 (kadar hemoglobin),

adanya pengaruh yang signifikan terhadap seseorang mengidap penyakit kanker payudara ganas atau jinak.

Selanjutnya berdasarkan hasil pengujian menggunakan SPSS 16 yang dapat kita lihat pada Tabel 4.4, di dapat nilai dari model Regresi Logistiknya sebagai berikut:

$$\hat{y} = -3,757 + 0,104 x_1 + 0,053 x_2 - 0,198 x_3$$

Dan model umum dari regresi logistik nya adalah :

$$\pi(x_i) = \frac{1}{1 + e^{-(-3,757 + 0,104 x_1 + 0,053 x_2 - 0,198 x_3)}} \cdot (11)$$

Kemudian jika dilihat nilai $\text{Exp}(\beta)$, Maka dapat diinterpretasikan sebagai berikut:

1. Untuk setiap pertambahan Umur 1 tahun mengakibatkan kenaikan resiko seseorang mengidap penyakit kanker payudara sebesar 11 %.
2. Untuk setiap pertambahan berat badan 1 kg mengakibatkan kenaikan resiko seseorang mengidap penyakit kanker payudara sebesar 54 %.
3. Untuk setiap penurunan kadar hemoglobin 1 g/dl mengakibatkan kenaikan resiko seseorang mengidap penyakit kanker payudara sebesar 8,21 %.

Kesimpulan

Parameter β_0 , β_1 , β_2 dan β_3 dari distribusi logistik biner dapat diestimasi dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Pencarian nilai parameter ini dilakukan dengan memaksimalkan fungsi *log-likelihood* dan dilakukan iterasi menggunakan metode Newton Raphson yang dilakukan dengan menggunakan *software* atau program *maple 13*.

Kemudian untuk menentukan nilai parameter yang mendekati nilai sebenarnya dilakukan dengan bantuan program SPSS 16, diperoleh model regresi logistik untuk pasien kanker payudara adalah

$$\pi(x_i) = \frac{1}{1 + e^{-(-3,757 + 0,104 x_1 + 0,053 x_2 - 0,198 x_3)}} \quad (12)$$

Saran

Penelitian ini secara teoritis membahas tentang mendapatkan model Regresi Logistik Biner dari pasien pengidap kanker payudara. Dan sebelumnya dilakukan estimasi parameter menggunakan metode *Maximum Likelihood Estimation* (MLE), kemudian proses iterasi dengan metode Newton Raphson. Oleh sebab itu, disarankan untuk penelitian lebih lanjut baik bagi penulis sendiri ataupun pembaca lain yang berminat dapat untuk menentukan metode atau teknik statistik lain dan dapat mengembangkannya menjadi lebih baik dari penelitian sebelumnya, seperti menggunakan Regresi Logistik Polikotomus dan lain-lain.

Daftar Pustaka

- [1] Fa'rifah, Riska Y dan Purhadi. 2012. Analisis *Survival* Faktor-Faktor yang Mempengaruhi Laju Kesembuhan Pasien Penderita Demam Berdarah Dengue (DBD) di RSUD Haji Surabaya dengan Regresi *Cox*. Jurnal Sains dan Seni ITS, Vol.1, No. 1.
- [2] Hedyani, dkk. 2013. Pemilihan Model Regresi Logistik Terbaik Menggunakan *Akaike's Information Criterion* (Studi Kasus : Penyakit Jantung Koroner di Rumah Sakit Umum Daerah Abdul Wahab Sjahranie Samarinda). Journal Science East Borneo, Vol.1, No.1.
- [3] Nadhifah, Laily. dkk. 2012. Analisis Faktor-Faktor yang Mempengaruhi Bayi Berat Lahir Rendah dengan Model Regresi Logistik Biner Menggunakan Metode Bayes (Studi Kasus di rumah Sakit Umum Daerah Kota Semarang). Universitas Diponegoro : Semarang.
- [4] Novianti, Fourina Ayu dan Santi Wulan P. 2012. Analisis Diagnosis Pasien Kanker

Payudara Menggunakan Regresi Logistik dan *Support Vector Machine* (SVM) Berdasarkan Hasil Mamografi. Jurnal Sains dan Seni ITS, Vol.1, No. 1.

- [5] Nurlaila, Dwi. dkk. 2013. Perbandingan Metode *Maximum Likelihood Estimation* (MLE) dan Metode Bayes dalam Pendugaan Parameter Distribusi Eksponensial. Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster), Vol. 2, No.1.
- [6] Sembiring, R.K. 2003. Analisis Regresi Edisi Kedua. Penerbit ITB : Bandung.
- [7] Walpole, Ronald E. 1992. Pengantar Statistika Edisi Ketiga. PT Gramedia Pustaka Utama : Jakarta.
- [8] Yendra, Rado. Dkk. 2010. Analisis Survival dan Program R. Yayasan Pusaka Riau : Pekanbaru.