PERBANDINGAN PENGGUNAAN MODEL REGRESI COX DAN MODEL REGRESI COX DENGAN *TIME-DEPENDENT VARIABLE* UNTUK MENGATASI *NONPROPORTIONAL HAZARD*

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains dalam Bidang Statistika

Oleh :

ABDULLAH YAZID

0610950001-95

PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2011

LEMBAR PENGESAHAN SKRIPSI

PERBANDINGAN MODEL REGRESI *COX* DAN MODEL REGRESI *COX* DENGAN *TIME-DEPENDENT VARIABLE* UNTUK MENGATASI *NONPROPORTIONAL HAZARD*

Oleh :
ABDULLAH YAZID
0610950001-95

Setelah dipertahankan di depan Majelis Penguji
pada tanggal 28 Juli 2011
dan dinyatakan memenuhi syarat untuk memperoleh gelar
Sarjana Sains dalam Bidang Statistika

Pembimbing I	Pembimbing II
Ir. Heni Kusdarwati, MS.	**Adji Achmad R. F., SSi., M.Sc**
NIP. 19611208 198701 2 001	**NIP. 19810908 200501 1 002**

Mengetahui,
Ketua Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Brawijaya
Malang

Dr. Abdul Rouf Alghofari, MSc
NIP. 19670907 199203 1 001

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama	**:**	**ABDULLAH YAZID**
NIM	**:**	**0610910001-95**
Program Studi	**:**	**Statistika**
Penulis Skripsi berjudul	**:**	**Perbandingan Model Regresi *Cox* dan Model Regresi *Cox* dengan *time-dependent variable* untuk mengatasi *nonproportional hazard***

Dengan ini menyatakan bahwa :

1. **Isi dari skripsi yang saya buat adalah benar-benar karya sendiri dan ridak menjiplak karya orang lain, selain nama-nama yang termaktub di isi dan tertulis di daftar pustaka dalam skripsi ini.**
2. **Apabila di kemudian hari ternyata skripsi yang saya tulis terbukti hasil jiplakan, maka saya akan bersedia menanggung segala resiko yang akan saya terima.**

Demikian pernyataan ini dibuat dengan segala kesadaran.

Malang, 28 Juli 2011
Yang menyatakan,

(Abdullah Yazid)
NIM. 0610950001

PERBANDINGAN PENGGUNAAN MODEL REGRESI COX DAN MODEL REGRESI COX DENGAN *TIME-DEPENDENT VARIABLE* UNTUK MENGATASI *NONPROPORTIONAL HAZARD*

ABSTRAK

Metode yang dapat dipakai untuk memodelkan antara variabel respon yang berupa waktu *survival* dengan satu atau lebih variabel prediktor adalah regresi *Cox*. Model regresi *Cox* juga dikenal dengan istilah *proportional hazard model* karena asumsi proporsional pada fungsi hazardnya. Akibat dari pelanggaran asumsi ini ialah model yang dihasilkan tidak sesuai. Salah satu metode yang dapat digunakan pada kasus *nonproportional hazard* ialah model regresi *Cox* dengan *time-dependent variable*. Model regresi *Cox* ini dapat digunakan untuk mengatasi *nonproportional hazard*. Tujuan penelitian ini ialah membentuk model regresi *Cox* dan model regresi *Cox* dengan *time-dependent variable* untuk mengatasi *nonproportional hazard* serta membandingkan kedua model tersebut menggunakan nilai *cross validation* Q^2 dan AIC. Berdasarkan nilai Q^2, model regresi *Cox* dan model regresi *Cox* dengan *time-dependent variable* memiliki selisih nilai Q^2 sangat kecil (0.0000001). Hal ini mengindikasikan bahwa kemampuan prediksi model regresi *Cox* dan model regresi *Cox* dengan *time-dependent variable* sama baiknya. Tetapi jika dilihat dari nilai AIC dan pemenuhan asumsi *proportional hazard,* model regresi *Cox* dengan *time-dependent variable* telah memenuhi asumsi *proportional hazard* dan memiliki nilai AIC terkecil sehingga dapat dikatakan bahwa model ini lebih baik daripada model regresi *Cox.*

Kata Kunci : Regresi Cox, *time-dependent variable, proportional hazard*

COMPARISON OF COX REGRESSION MODEL AND COX REGRESSION MODEL WITH TIME-DEPENDENT VARIABLE TO OVERCOME NONPROPORTIONAL HAZARD

ABSTRACT

The method can be used for modeling the response variables in the form of survival time with one or more predictor variables is the Cox regression. Cox regression model also known as proportional hazard model because its assumption of proportional hazard function. As a result of violations of this assumption is that the result of model is not appropriate. One of method that can be used in case the hazard is nonproportional is the Cox regression model with time-dependent variable. This model can be used to overcome nonproportional hazard. The purpose of this study is to get the form of cox regression model and cox regression model with time-dependent variable to overcome nonproportional hazard as well as comparing both models using the cross validation Q^2 and AIC. Based on the value of Q^2, Cox regression models and Cox regression models with time-dependent variable has a very small difference in value of Q^2 (0.0000001). This indicates that the predictive ability of of Cox regression models and Cox regression models with time-dependent variable as well. According to AIC and the assumption of proportional hazard that satisfied, cox regression models with time-dependent variable has the proportional hazard and the smallest AIC value, so it can be said that this model is better than the cox regression model.

Keywords : Cox regression, time-dependent variable, proportional hazard

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kehadirat Allah SWT atas segala rahmat dan hidayahNya sehingga dapat menyelesaikan skripsi dengan judul **Perbandingan Model Regresi *Cox* dan Model Regresi *Cox* dengan *time-dependent variable* untuk Mengatasi *nonproportional hazard*.**

Model regresi *Cox* dengan *time-dependent variable* dapat digunakan untuk mengatasi masalah *nonproportional hazard* pada model regresi *Cox*. Untuk itu perlu diketahui kemampuan prediksi dari kedua model dan menentukan model yang paling baik digunakan pada data yang tidak memenuhi asumsi *proportional hazard*.

Penulis mengucapkan terimakasih kepada semua pihak yang telah membantu dalam menyusun skripsi ini, terutama kepada:

1. Ibu Ir. Heni Kusdarwati, MS. (selaku Dosen Pembimbing I) dan Bapak Adji Achmad Rinaldo Fernandes, SSi, M.Sc (selaku dosen pembimbing II) atas bimbingan moral, akhlak dan spiritual yang diberikan selama penulisan skripsi.
2. Ibu Dr. Ir. Ni Wayan Surya Wardhani, MS. selaku dosen penguji.
3. Bapak Dr. Abdul Rouf Alghofari, MSc selaku Ketua Jurusan Matematika Universitas Brawijaya
4. Seluruh Bapak dan Ibu Dosen serta Staf pengajaran Jurusan Matematika Universitas Brawijaya.
5. Bapak, Ibu, kakak-kakak dan Adik-adikku tercinta, dan teman-teman Statistika 2006 atas segala dukungan dan motivasi.
6. Nonny Puspita Anggraini, atas dukungan dan bantuannya.
7. Semua pihak yang mendukung baik secara langsung maupun tidak langsung yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa dalam penulisan skripsi masih banyak kekurangan. Oleh karena itu, saran dan kritik yang membangun sangat diharapkan. Semoga skripsi ini dapat bermanfaat bagi pembaca dan semua pihak yang memerlukannya.

Malang, Juli 2011
Penulis

DAFTAR ISI

BAB V KESIMPULAN DAN SARAN

DAFTAR TABEL

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR LAMPIRAN

Halaman

BAB I
PENDAHULUAN

1.1 Latar Belakang

Analisis *survival* adalah suatu teknik analisis yang digunakan untuk menganalisis data *survival* dari satu atau beberapa kelompok individu (Lee, 1997). Sedangkan data *survival* adalah data tentang jangka waktu terjadinya suatu kejadian mulai dari waktu awal sampai waktu akhir. Durasi waktu antar dua kejadian tersebut didefinisikan sebagai waktu *survival* dan dilambangkan dengan *T*. Metode yang dapat dipakai untuk memodelkan antara variabel respon yang berupa waktu *survival* dengan satu atau lebih variabel prediktor adalah regresi *cox* (Fox, 2002). Regresi *Cox* yang diperkenalkan oleh D. R. Cox pada tahun 1972 dan digunakan untuk menjelaskan pengaruh antara kegagalan individu pada suatu waktu dengan satu atau lebih variabel prediktor dimana variabel prediktor dapat bersifat kontinyu maupun kategorik.

Model regresi *Cox* juga dikenal dengan istilah *proportional hazard model* karena asumsi proporsional pada fungsi hazardnya. *Proportional hazard* merupakan asumsi yang penting yang mendasari regresi *Cox* yang merupakan rasio antara dua level fungsi hazard. Fungsi *Hazard* untuk level satu adalah proporsional terhadap fungsi hazard untuk level dua jika rasio keduanya bernilai konstan dan tidak tergantung waktu. Pada penelitian Kurniawati (2010) disebutkan bahwa jika asumsi ini tidak terpenuhi berarti komponen linier dari model berubah-ubah tergantung waktu dan dikatakan *nonproportional hazard*. Akibatnya, model yang dihasilkan tidak sesuai. Salah satu metode yang dapat digunakan pada kasus *nonproportional hazard* ialah *stratified proportional hazard model*. Kriestianianatha (2007) telah meneliti penggunaan metode ini dalam mengatasi *nonproportional hazard* pada peubah bebas bersifat kategori. Salah satu kesulitan dalam metode ini ialah menentukan banyaknya kelompok (strata) pada peubah bebas bersifat kontinyu. Jika kelompok yang terbentuk terlalu sedikit, maka sisaan akan bias. Sebaliknya jika kelompok terlalu banyak, maka perhitungan tidak efisien. Metode lain yang dapat dipakai pada kasus *nonproportional hazard* ialah model regresi *Cox* dengan *time-dependent variable*.

Time-dependent variable diartikan sebagai variabel yang nilainya berubah-ubah setiap saat (Kleinbaum dan Klein, 2005), misalnya tekanan darah, kadar kolesterol. Ata dan Sozer (2007) mengatakan bahwa jika terdapat *time-dependent variable* dalam model, asumsi *proportioanal hazard* tidak lagi diperlukan dalam model regresi *Cox*.

Model regresi *Cox* dengan *time-dependent variable* merupakan pengembangan dari model regresi *Cox* dan dapat digunakan untuk membentuk model ketika asumsi *proportioanal hazard* tidak terpenuhi.

Shumway (2001) telah meneliti penggunaan model regresi *Cox* dengan *time-dependent variable* dan regresi logistik pada data kebangkrutan Bank. Dari hasil penelitian tersebut, Shumway (2000) mengatakan bahwa regresi *Cox* dengan *time-dependent variable* menghasilkan prediksi yang lebih akurat pada data kebangkrutan Bank. Berdasarkan hal tersebut, maka dianggap perlu untuk membahas penggunaan model regresi *Cox* dengan *time-dependent variable* dibandingkan dengan model regresi *Cox* dibidang ilmu yang lain, yaitu bidang kesehatan, pada data yang tidak memenuhi asumsi *proportional hazard*.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah :

1. Bagaimana membentuk model regresi *Cox*?
2. Bagaimana membentuk model regresi *Cox* dengan *time-dependent variable?*
3. Bagaimana mengatasi *nonproportional hazard* pada regresi *Cox*?
4. Lebih baik mana antara model regresi *Cox* dan model regresi *Cox* dengan *time-dependent variable* berdasarkan nilai *cross validation* Q^2 dan AIC.

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini meliputi :

1. Hanya digunakan satu variabel tergantung waktu *(time-dependent variables)*.
2. Data yang digunakan ialah data yang tidak memenuhi asumsi *proportional hazard*.
3. Tidak dilakukan *repeated measurement* pada individu yang diamati (hanya satu kali pengamatan tiap individu).
4. Untuk pembentukan variabel tergantung waktu, digunakan fungsi waktu dari variabel prediktor dikalikan dengan variabel prediktornya.

1.4 Tujuan

Tujuan dari penelitian ini ialah :

1. Membentuk model regresi *cox*.
2. Membentuk model regresi *cox* dengan *time-dependent variable.*
3. Mengatasi *nonproportional hazard* pada regresi *Cox* dengan penambahan *time-dependent variable* pada model. Uji *proportional hazard* menggunakan grafik plot antara Log {-log[S(t, X)]} terhadap waktu *survival*
4. Menentukan model mana yang paling baik antara regresi *cox* dan regresi *cox* dengan *time-dependent variable* dilihat dari nilai *cross validation* Q^2 dan AIC.

1.5 Manfaat

Setelah dilakukan penelitian, diharapkan dapat memberikan pengetahuan lebih luas tentang penerapan regresi *Cox* dengan *time-dependent variable* untuk mengatasi asumsi *proportional hazard* yang tidak terpenuhi.

BAB II
TINJAUAN PUSTAKA

2.1 Eksplorasi Data

Eksplorasi data merupakan teknik penyajian dan peringkasan data baik secara tabel maupun grafik. Tujuan dari eksplorasi data ialah untuk mengetahui struktur dan pola data sehingga dapat memberikan keyakinan bahwa data tersebut dapat diwakili oleh suatu model. Eksplorasi ini dapat dilakukan dengan menggunakan grafik histogram untuk mengetahui gambaran umum model.

2.2 Analisis *Survival*

Analisis *survival* adalah suatu metode yang berhubungan dengan waktu, mulai dari titik permulaan atau *start point* sampai dengan terjadinya suatu akhir kejadian atau *end point* (Collet, 2003). Analisis *survival* dapat mengetahui peluang kejadian mampu bertahan, melalui pendekatan analisis *Kaplan-Meier*, dan dapat diketahui faktor-faktor yang mempengaruhi ketahanan tersebut dengan pendekatan regresi *Cox*. Data *survival* yang merupakan kumpulan durasi waktu antara titik waktu saat individu direkrut ke dalam sebuah penelitian sampai saat di mana individu tersebut mengalami kejadian yang dispesifikasikan, diperoleh dari catatan waktu yang dicapai suatu objek sampai terjadinya peristiwa tertentu yang disebut *failure event*. Menurut Lee (1997) dalam menentukan waktu *survival*, terdapat tiga hal mendasar yang harus diperhatikan yaitu :

1. Titik permulaan (*time origin* atau *start point*)
2. Titik akhir/kejadian khusus (*end point*)
3. Skala pengukuran.

Perbedaan antara analisis *survival* dengan analisis statistik lainnya adalah adanya pengamatan tersensor. Menurut Miller (1998) data dikatakan tersensor jika pengamatan waktu *survival* hanya sebagian, tidak sampai terjadi *failure event*. Penyebab terjadinya data tersensor antara lain :

1. *Loss to follow up* terjadi bila objek pindah, meninggal atau menolak untuk berpartisipasi.

2. *Drop Out* terjadi bila perlakuan dihentikan karena alasan tertentu.
3. *Termination of Study* terjadi masa penelitian berakhir sementara objek yang diobservasi belum mencapai *failure*.

2.2.1 Fungsi *Survival*

Fungsi *survival* dapat digunakan untuk menyatakan peluang suatu individu bertahan dari waktu mula-mula sampai suatu waktu t. Waktu *survival* suatu individu, *t,* dapat dianggap sebagai nilai variabel *T* yang bernilai non-negatif (Collet, 1994). Jika *T* melambangkan waktu *survival* dan mempunyai distribusi peluang *f(t)*, maka Fungsi sebaran kumulaif *F(T)* dinyatakan sebagai berikut:

$$F(T) = P(T \le t) = \int_0^t f(u)du \tag{2.1}$$

Fungsi survival, S(t), didefinisikan sebagai peluang suatu objek bertahan setelah waktu ke-*t*, yaitu :

$$S(t) = P(T > t) = 1 - P(T < t) = 1 - F(t) \tag{2.2}$$

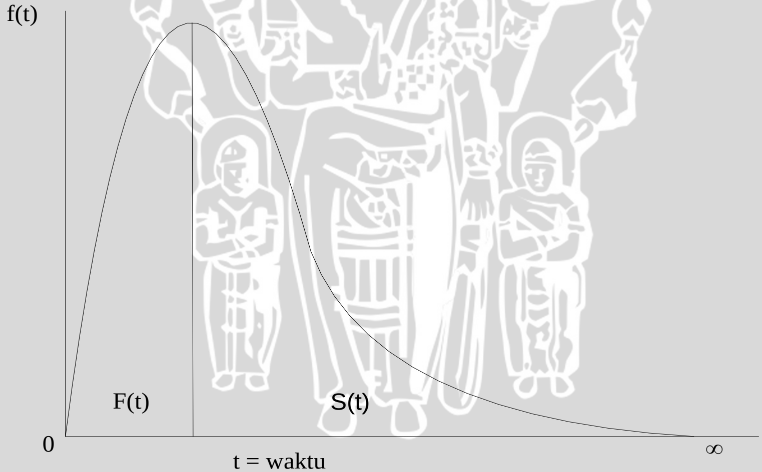
Menurut Kleinbaum dan Klein (2005), *S(t)* adalah fungsi *non-increasing* terhadap waktu *t* dan dinyatakan sebagai :

$$S(t) = \begin{cases} 1 \;\; untuk\; t = 0 \\ 0 \;\; untuk\; t = \infty \end{cases} \tag{2.3}$$

Berdasarkan persamaan (2.3) pada waktu *t* = 0 maka *S(t)* = S(0) = 1 yang diartikan sebagai awal dari pengujian di mana tidak ada satupun objek yang mengalami kejadian yang dispesifikasikan dan peluang hidup dari suatu objek bernilai satu. Pada waktu *t* = ∞ maka *S(t)* = S(∞) = 0, artinya jika periode pengujian meningkat sampai tidak terbatas maka pada akhirnya tidak akan ada suatu objek yang dapat bertahan hidup sehingga peluang hidup dari suatu objek akan mendekati nilai nol. Secara grafik, fungsi *S(t)* diilustrasikan pada Gambar 2.1.

Gambar 2.1. Grafik Fungsi *S(t)* Berdasarkan Nilai *t*

Hubungan antara *f(t)*, *F(T)* dan *S(t)* dapat dilihat dari gambar 2.2.



Gambar 2.2. Grafik Fungsi kepekatan peluang, $f(t)$, fungsi sebaran kumulatif, $F(t)$ dan fungsi *survival*, $S(t)$.

2.2.2 Fungsi *Hazard*

Fungsi *hazard* didefinisikan sebagai peluang suatu individu mengalami kegagalan pada waktu *t*, dengan syarat individu tersebut telah bertahan sampai waktu tersebut. Fungsi ini menyatakan angka laju kegagalan individu yang bertahan sampai waktu ke-*t*.

Misal peluang bahwa variabel random *T* lebih besar atau sama dengan nilai *t*, berada antara *t* dan *t+δt*, dapat ditulis sebagai $P(t \leq T \leq t + \delta t | T \geq t)$ sehingga fungsi *hazard* dapat dinyatakan sebagai berikut :

$$h(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T < t + \delta t | T \geq t)}{\delta t} \right\} \quad (2.4)$$

Berdasarkan persamaan 2.4. *h(t)* mengukur peluang kematian suatu objek pada interval (*t, t+ δt*), berdasarkan syarat bahwa objek tersebut telah bertahan hidup pada waktu *t*. Sebagai contoh jika waktu *survival* dihitung dalam hari, *h(t)* merupakan pendekatan peluang suatu objek yang hidup pada hari *t*, mati pada hari yang lain. Untuk alasan ini, fungsi *hazard* seringkali mudah diinterpretasikan sebagai resiko kematian pada waktu *t*.

Jika dihubungkan dengan fungsi *survival* maka fungsi *hazard* akan lebih berarti jika persamaan dalam (2.4) diubah kedalam persamaan berikut :

$$\frac{P(t \leq T < t + \delta t)}{P(T \geq t)} = \frac{F(t + \delta t) - F(t)}{S(t)} \quad (2.5)$$

Berdasarkan persamaan (2.5), persamaan (2.4) dapat ditulis kembali sebagai berikut :

$$h(t) = \lim_{\delta t \to 0} \left\{ \frac{F(t + \delta t) - F(t)}{\delta t} \right\} \frac{1}{S(t)} = \frac{f(t)}{S(t)} \quad (2.6)$$

karena $f(t) = \frac{-d}{dt} S(t)$ maka

$$h(t) = \left(\frac{-d}{dt} S(t) \right) / S(t) \tag{2.7}$$

dengan mengintegralkan $h(t)$ didapatkan persamaan (2.8)

$$\int_0^t h(t)\ dt = -\int_0^t \frac{\frac{dt}{dt} S(t)}{S(t)} dt = -\log S(t) \tag{2.8}$$

kemudian dieksponensialkan, maka diperoleh persamaan :

$$S(t) = \exp\left(-\int_0^t h(t)\ dt \right) \tag{2.9}$$

karena fungsi kumulatif *hazard* adalah :

$$H(t) = \int_0^t h(t)\ dt \tag{2.10}$$

dari persamaan (2.9) dan (2.10) diperoleh hubungan yaitu :

$$H(t) = -\log S(t) \tag{2.11}$$

2.3 Model Regresi *Cox*

Pemodelan data *survival* merupakan kombinasi variabel prediktor yang mempengaruhi fungsi hazard. Pendekatan untuk data *survival* adalah Model regresi *Cox*. Model regresi *Cox* juga dikenal dengan istilah *proportional hazard model* karena asumsi *proportional* pada fungsi hazardnya. (Cox, 1972).

Apabila resiko kegagalan (*failure*) pada waktu tertentu dari suatu kejadian bergantung pada $\mathbf{x_1}$, $\mathbf{x_2}$, …, $\mathbf{x_p}$ dari p variabel prediktor yang tidak tergantung waktu *(time-independent variable)*, maka nilai variabel tersebut diasumsikan telah tercatat sebagai *time origin*. Apabila fungsi h_0 (t) untuk setiap individu dari semua variabel prediktor anggota vektor **x** bernilai nol, fungsi h_0 (t) dapat disebut sebagai *baseline hazard function*. Secara umum model regresi *Cox* didefinisikan :

$$h(t, X) = h_0(t) \exp\left[\sum_{i=1}^{p} \beta_i X_i \right]$$

$$h(t,X) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p) \qquad (2.12)$$

dimana :

h(t,X)	: fungsi *hazard* pada saat *t*
$h_0(t)$	: *baseline hazard function*
β	: koefisien parameter regresi *Cox*
X	: variabel prediktor
p	: banyaknya variabel prediktor

Menurut Chan (2004), model regresi *Cox* menunjukkan bahwa rasio kematian antar objek di dalam kelompok ditunjukkan oleh $\exp(\beta_j)$ kali rasio kematian antar objek di dalam kendali kelompok secara terus-menerus. Model regresi *Cox* dapat diinterpretasikan sebagai *hazard ratio* atau nisbah peluang kegagalan atau kematian objek pada suatu level dari faktor relatif terhadap peluang kegagalan pada level lainnya dari faktor tersebut. Oleh karena itu dalam penerapannya terkadang tidak memerlukan pendugaan fungsi *baseline hazard*.

2.3.1 Pendugaan Parameter Model Regresi *Cox*

Untuk menentukan model regresi *Cox* diperlukan estimasi koefisien variebel prediktor $x_1, x_2, \ldots, x_p$ yaitu $\beta_1, \beta_2, ..., \beta_p$. Koefisien β dalam model *hazard proportional* dapat ditaksir menggunakan metode *Maximum Partial Likelihood*. Apabila terdapat sebanyak n sampel, diantaranya terdapat r jarak waktu kegagalan *(failure)* dengan waktu yang berbeda, dengan urutan waktu kegagalan $t_{(1)} < t_{(2)} <\ldots < t_{(r)}$ dengan $t_{(j)}$ sebagai urutan waktu kegagalan ke-*j*, maka suatu objek yang mendapat resiko pada waktu $t_{(j)}$ dinyatakan sebagai $R(t_{(j)})$, di mana $R(t_{(j)})$ adalah kelompok objek yang masih hidup dan tidak tersensor oleh waktu $t_{(j)}$. Penjumlahan dari $R(t_{(j)})$ disebut dengan sekumpulan resiko. Menurut *Cox* (1972) fungsi *likelihood* untuk *model hazard proportional* adalah :

$$L(\beta') = \prod_{j=1}^{r} \frac{\exp(\beta' \mathbf{x}_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta' \mathbf{x}_l)} \qquad (2.13)$$

$X_{(j)}$ adalah vektor variabel prediktor dari individu yang meninggal pada saat ke-*j* dengan urutan waktu $t_{(j)}$.

Apabila data terdiri dari *n* pengamatan ditulis sebagai $t_1, t_2, \ldots, t_n$, dengan indikator kejadian (δ_i) maka fungsi *partial likelihood* pada persamaan (2.13) dapat dinyatakan dalam bentuk :

$$L(\beta') = \prod_{i=1}^{n} \left[\frac{\exp(\beta' \mathbf{x_i})}{\sum_{l \in R(t_i)} \exp(\beta' \mathbf{x}_l)} \right]^{\delta_i'} \tag{2.14}$$

dimana :

$$\delta_i = \begin{cases} 1, \text{ jika terjadi event atau kejadian} \\ 0, \text{ tersensor} \end{cases}$$

dan $R(t_{(i)})$ adalah kelompok individu yang beresiko saat t_i, dan δ_i adalah indikator sensoring yang bernilai nol jika t_i, i=1,2,..,n adalah tersensor dan bernilai 1 untuk lainnya. Penyebut merupakan jumlah dari semua peluang kegagalan dari individu yang mungkin pada waktu $t_{(j)}$. Fungsi kesesuaian *log-likelihood* adalah :

$$\log L(\beta') = \sum_{i=1}^{n} \delta_i \left\{ \beta' \mathbf{x_i} - \log \sum_{l \in R(t_{(i)})} \exp(\beta' \mathbf{x}_l) \right\} \tag{2.15}$$

Penaksiran nilai parameter β diperoleh dengan memaksimumkan fungsi *log-likelihood* dengan menggunakan metode iterasi *Newton Raphson*. Penaksiran parameter $\beta_1, \beta_2, \ldots, \beta_p$ diperoleh dari penyelesaian sejumlah *px1* vektor persamaan yang dinyatakan dengan skor koefisien vektor $u(\beta)$. Skor koefisien untuk β_j adalah $u(\beta_j) = \frac{\partial \log L(\beta)}{\partial \beta_j}$, sehingga

$$u(\beta_j) = \sum_{j=1}^{n} \delta_j \left\{ x_j - \left(\frac{\sum_{l \in R(t_i)} x_{jl} \exp(\beta' x_l)}{\sum_{l \in R(t_i)} \exp(\beta' x_l)} \right) \right\}$$

(2.16)

Misal matriks **I**(β) adalah matriks *pxp* yang merupakan turunan kedua dari fungsi *log-likelihood* yang bernilai negatif maka **I**(β) adalah :

$$-\frac{\partial^2 \log L(\beta)}{\partial\beta_j \partial\beta_k} \tag{2.17}$$

dengan $j = 1, 2, \ldots, p$ dan $k = 1, 2, \ldots, p$, maka **I**(β) disebut matriks Hessian atau matriks informasi pengamatan.

Berdasarkan prosedur iterasi *Newton-Raphson*, penaksiran dari parameter β pada *s+1* yang disimbolkan $\hat{\beta}_{s+1}$ adalah :

$$\hat{\beta}_{s+1} = \hat{\beta}_s + I^{-1}\left(\hat{\beta}_s\right) u\left(\beta_s\right) \tag{2.18}$$

dengan

$$I^{-1}\left(\beta_{jk}\right) = \left\{ \left[\frac{\sum_{l \in R(t_i)} x_{jl} x_{kl} \exp(\beta' x_l)}{\sum_{l \in R(t_i)} \exp(\beta' x_l)} \right] - \left[\left(\frac{\sum_{l \in R(t_i)} x_{jl} \exp(\beta' x_l)}{\sum_{l \in R(t_i)} \exp(\beta' x_l)} \right) \left(\frac{\sum_{l \in R(t_i)} x_{kl} \exp(\beta' x_l)}{\sum_{l \in R(t_i)} \exp(\beta' x_l)} \right) \right] \right\} \tag{2.19}$$

di mana :

s : 0, 1, 2, …

$u\left(\hat{\beta}_s\right)$: vektor skor koefisien

$I^{-1}\left(\hat{\beta}_s\right)$: invers matriks informasi yang diamati

Proses iterasi dimulai dengan menentukan nilai awal $\hat{\beta}_0 = 0$ dan proses akan dihentikan jika perubahan pada nilai β_{s+1} dan β_s kurang dari 10^{-6}.

2.3.2 Uji Signifikansi Parameter Model Regresi *Cox*

Pengujian signifikansi parameter model regresi *Cox* meliputi uji simultan dan parsial.

2.3.2.1 Uji Simultan

Uji simultan digunakan untuk memeriksa pengaruh variabel prediktor secara bersama-sama terhadap variabel respon. Uji yang digunakan ialah uji *likelihood ratio*.

Uji hipotesis :

H_0 : $\beta_1 = \beta_2 = \ldots = \beta_p = 0$

H_1 : minimal ada satu $\beta_j \neq 0$, untuk j = 1, 2, …, p.

Statistik Uji :

$$\chi^2_{LR} = 2\left[Ln\, L(\hat{\beta}) - Ln\, L(\beta_0)\right] \tag{2.20}$$

Apabila $\chi^2_{LR} > \chi^2_{p,\alpha}$ atau *p-value* $\leq \alpha$ maka H_0 ditolak yang berarti secara bersama-sama variabel prediktor berpengaruh nyata terhadap variabel respon.

2.3.2.2 Uji Parsial

Uji signifikansi parameter secara parsial digunakan untuk memeriksa pengaruh dari masing-masing variabel prediktor terhadap variabel respon secara individu pada model. Uji ini dilakukan dengan merasio antara penduga koefisien dengan *standard error* penduga koefisien. Rasio antar keduanya disebut dengan statistik *Wald* dan uji yang digunakan disebut uji *Wald*.

Uji hipotesis :

H_0 : $\beta_j = 0$

H_1 : $\beta_j \neq 0$, untuk j = 1, 2, …, p.

Statistik Uji :

$$W = \frac{\hat{\beta}_j^{\,2}}{\left[Se(\hat{\beta}_j)\right]^2} \sim \chi^2_1 \tag{2.21}$$

di mana : $\hat{\beta}_j$: koefisien regresi pada variabel ke-*j*

$\left[Se(\hat{\beta}_j)\right]^2$: ragam koefisien regresi

Apabila $W > \chi_1^2$ atau *p-value* ≤ α maka H_0 ditolak yang berarti bahwa variabel prediktor ke-*j* berpengaruh nyata terhadap variabel respon (Kutner *et al.*, 2005).

2.3.3 Sisaan Model Regresi *Cox*

Pemerikasaan sisaan pada model regresi *Cox* bertujuan untuk memeriksa apakah model sudah sesuai. Sisaan yang sering digunakan dalam pengujian model regresi *Cox* adalah sisaan *Cox-Snell*. Sisaan *Cox-Snell* dapat diartikan sebagai nilai harapan setiap pengamatan. Sisaan *Cox-Snell* untuk individu ke-*i* dapat dirumuskan :

$$r_{Ci} = \exp\left(\hat{\beta}' x_i\right)\hat{H}_0\left(t_i\right) \qquad (2.22)$$

dimana :

r_{Ci} = Sisaan *Cox-Snell* untuk individu ke-*i*

$\hat{H}_0\left(t_i\right)$ = dugaan dari fungsi kumulatif garis dasar *hazard* pada waktu t_i

Menurut Nelson-Aalen dalam Collet (2003) r_{Ci} adalah nilai dari:

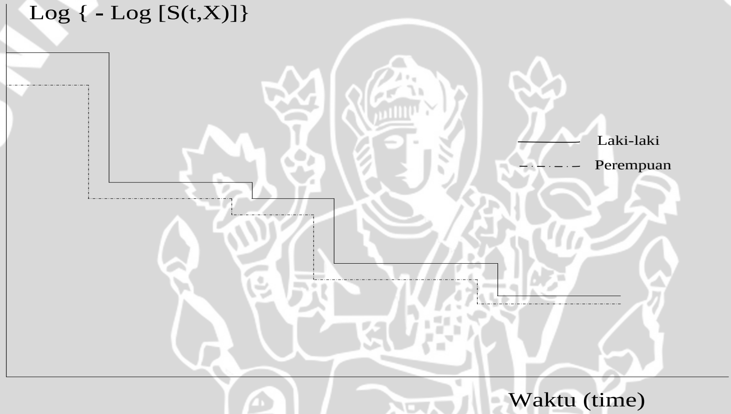
$$\hat{H}_i\left(t_i\right) = -\log \hat{S}_i\left(t_i\right) \qquad (2.23)$$

di mana $\hat{H}_i(t_i)$ dan $\hat{S}_i\left(t_i\right)$ adalah nilai penduga dari kumulatif *hazard* dan fungsi *survivor* dari objek ke-*i* pada waktu t_i

2.3.4 Pengujian Asumsi *Proportional Hazard*

Proportional hazard merupakan suatu asumsi penting yang mendasari regresi *Cox*. *Proportional hazard* merupakan rasio antara dua level fungsi *hazard*. Fungsi hazard untuk level satu adalah *proportional* terhadap fungsi *hazard* untuk level dua jika rasio keduanya bernilai konstan dan tidak tergantung waktu. Asumsi *proportional hazard* berarti *hazard ratio* yang konstan sepanjang waktu. Pelanggaran terhadap asumsi ini mengakibatkan komponen linier dari model berubah-ubah terhadap waktu sehingga model yang dihasilkan tidak sesuai.

Menurut Ata dan Sozer (2007), cara untuk memeriksa asumsi proporsional *hazard* ialah secara visual dengan melihat grafis dari plot antara Log {-log[S(t, X)]} terhadap waktu *survival*. Jika grafik Log {-log[S(t, X)]} terhadap waktu *survival* untuk beberapa kategori dalam suatu variabel prediktor terlihat sejajar atau tidak bersilangan, mengindikasikan bahwa asumsi *proportional hazard* terpenuhi. Sebaliknya jika terlihat bersilangan maka asumsi *proportional hazard* tidak terpenuhi. Untuk lebih jelasnya tentang asumsi ini dapat dilihat pada Gambar 2.3.



Gambar 2.3 Pengujian Asumsi *Proportional Hazard*

Jika asumsi ini tidak terpenuhi maka model yang dihasilkan dikatakan *nonproportional hazard* dan mengakibatkan model yang dihasilkan tidak sesuai. Salah satu metode yang dapat digunakan jika terdapat *nonproportional hazard* ialah model regresi *Cox* dengan *time-dependent variable.*

2.4 Regresi *Cox* dengan *Time-Dependent Variable*

2.4.1 *Time-Dependent Variable*

Time-dependent variable diartikan sebagai variabel yang nilainya berubah-ubah setiap saat. Sebaliknya *time-independent variable* ialah variabel yang nilainya konstan atau tidak tergantung waktu. Pada banyak kasus data variabel prediktor yang diteliti terutama dalam bidang kedokteran seperti tekanan darah, berat badan relatif, dan kadar kolesterol ialah data yang dikumpulkan berdasarkan suatu periode waktu tertentu. Nilai pengamatan dari variabel tersebut dapat berubah-ubah sepanjang waktu sehingga besar kemungkinan *hazard ratio* yang dihasilkan juga berubah-ubah menurut waktu dan dikatakan fungsi *hazard* untuk level satu tergantung waktu terhadap fungsi *hazard* untuk level dua (tidak konstan). Oleh karena itu model regresi *Cox* biasa tidak sesuai lagi untuk digunakan. Untuk mengatasi hal ini digunakan model regresi *Cox* dengan *time-dependent variable*.

2.4.2 Model Regresi *Cox* dengan *Time-Dependent Variable*

Sebagaimana model regresi *Cox* pada persamaan (2.12), model regresi *Cox* dengan *time-dependent variable* terdiri dari *baseline hazard function* $h_0(t)$ dan fungsi eksponen. Perbedaannya terletak pada fungsi eksponennya, pada regresi *Cox* hanya terdapat variabel prediktor X_i yang nilainya tidak tergantung waktu (konstan) atau disebut *time-independent variable*. Pada model regresi *Cox* dengan *time-dependent variable*, selain *time-independent variable* yang dilambangkan dengan X_i, juga terdapat *time-dependent variable* yang dilambangkan dengan $X_j(t)$. Model regresi *Cox* dengan *time-dependent variable* dapat ditulis sebagai berikut (Kleinbaum and Klein, 2005) :

$$h(t, X(t)) = h_0(t) \exp\left[\sum_{i=1}^{p1} \beta_i X_i + \sum_{j=1}^{p2} \gamma_j X(t)_j\right] \quad (2.24)$$

dimana :

$h_0(t)$: fungsi *baseline hazard* pada saat t

β	:	parameter atau koefisien regresi untuk *time-independent variable*.
γ	:	parameter atau koefisien regresi untuk *time-dependent variable*.
p1	:	banyaknya *time-independent variable*
p2	:	banyaknya *time-dependent variable*
$X(t)_j$	:	variabel yang tergantung waktu *(time-dependent variable)*

Ata dan Sozer (2007) mengatakan bahwa jika terdapat *time-dependent variable* dalam model, asumsi *proportioanal hazard* tidak lagi diperlukan dalam model regresi *Cox*. Model regresi *Cox* dengan *time-dependent variable* merupakan pengembangan dari model regresi *Cox* dan dapat digunakan untuk membentuk model ketika asumsi *proportioanal hazard* tidak terpenuhi. Shumway (2000) telah meneliti penggunaan model regresi *Cox* dengan *time-dependent variable* dan regresi logistik pada data kebangkrutan Bank. Dari hasil penelitian tersebut, Shumway (2000) mengatakan bahwa regresi *Cox* dengan *time-dependent variable* menghasilkan prediksi yang lebih akurat pada data kebangkrutan Bank.

2.4.3 Pendugaan Parameter Model Regresi *Cox* dengan *Time-Dependent Variable*

Untuk menentukan model regresi *Cox* dengan *time-dependent variable* diperlukan estimasi koefisien variabel prediktor X_1, $X_2,\ldots,X_{p1}$ yaitu $\beta_1,\beta_2,\ldots, \beta_{p1}$ dan koefisien variabel tergantung waktu $X(t)_{p1+1}$, $X(t)_{p1+2},\ldots$, $X(t)_p$ γ_{p1+1}, $\gamma_{p1+2},\ldots$, γ_p. Sama halnya dengan model regresi *Cox*, Koefisien β dan γ dalam model regresi *Cox* dengan *time-dependent variable* dapat ditaksir menggunakan metode *Maximum Partial Likelihood*. Apabila terdapat sebanyak n sampel, diantaranya terdapat r jarak waktu kegagalan *(failure)* dengan waktu yang berbeda, dengan urutan waktu kegagalan $t_{(1)} < t_{(2)} <\ldots < t_{(r)}$ dengan $t_{(j)}$ sebagai urutan waktu kegagalan ke-*j*, maka suatu objek yang mendapat resiko pada waktu $t_{(j)}$ dinyatakan sebagai $R(t_{(j)})$, di mana $R(t_{(j)})$ adalah kelompok objek yang masih hidup dan tidak tersensor oleh waktu $t_{(j)}$. Penjumlahan dari $R(t_{(j)})$ disebut dengan sekumpulan resiko. Misalkan $W_{(j)}$ adalah vektor variabel prediktor pada model regresi *Cox* dengan *time-dependent variable* dari individu yang meninggal pada saat ke-*j* dengan urutan waktu $t_{(j)}$, dan α ialah koefisien model regresi cox dengan *time-dependent variable*. $W_{(j)}$ dan α didefinisikan sebagai berikut :

$$W = \begin{bmatrix} X_{1,1} & \cdots & X_{1,p1} & X(t)_{1,p1+1} & \cdots & X(t)_{1,p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ X_{n,1} & \cdots & X_{n,p1} & X(t)_{n,p1+1} & \cdots & X(t)_{n,p} \end{bmatrix}$$

$$\alpha = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{p1} \\ \gamma_1 \\ \vdots \\ \gamma_p \end{bmatrix}$$

dimana : p_1 = banyaknya variabel prediktor yang tidak tergantung waktu

Mengacu pada pendugaan parameter model regresi *Cox*, fungsi *likelihood* untuk model regresi *Cox* dengan *time-dependent variable* adalah :

$$L(\alpha') = \prod_{\mathrm{j}=1}^{\mathrm{r}} \frac{\exp(\alpha' \mathbf{W}_{(\mathrm{j})})}{\sum_{l \in \mathrm{R}(\mathrm{t}_{(\mathrm{j})})} \exp(\alpha' \mathbf{W}_{l})} \qquad (2.25)$$

$W_{(j)}$ adalah vektor variabel prediktor dari individu yang meninggal pada saat ke-*j* dengan urutan waktu $t_{(j)}$.

Apabila data terdiri dari *n* pengamatan ditulis sebagai $t_1, t_2, \ldots, t_n$, dengan indikator kejadian (δ_i) maka fungsi *partial likelihood* pada persamaan (2.25) dapat dinyatakan dalam bentuk :

$$L(\alpha^{'}) = \prod_{i=1}^{n}\left[\frac{\exp(\alpha^{'}\mathrm{W}_{(i)})}{\sum_{l\in R(t_i)}\exp(\alpha^{'}\mathrm{W}_{l})}\right]^{\delta_i^{'}} \tag{2.26}$$

dimana :

$$\delta_i = \begin{cases} 1, \text{ jika terjadi event atau kejadian} \\ 0, \text{ tersensor} \end{cases}$$

dan $R(t_{(i)})$ adalah kelompok individu yang beresiko saat t_i, dan δ_i adalah indikator sensoring yang bernilai nol jika t_i, i=1,2,..,n adalah tersensor dan bernilai 1 untuk lainnya. Penyebut merupakan jumlah dari semua peluang kegagalan dari individu yang mungkin pada waktu $t_{(j)}$. Fungsi kesesuaian *log-likelihood* adalah :

$$\log \mathrm{L}(\alpha^{'}) = \sum_{i=1}^{n}\delta_i\left\{\alpha^{'}\mathrm{W}_{(i)} - \log\sum_{l\in R(t_{(i)})}\exp\,(\alpha^{'}\mathrm{W}_l)\right\} \tag{2.27}$$

Penaksiran nilai parameter α diperoleh dengan memaksimumkan fungsi *log-likelihood* persamaan (2.27) dengan menggunakan metode numerik. Fungsi yang maksimum dapat diperoleh dengan menggunakan metode iterasi *Newton Raphson*. Dengan p=p1+p2, penaksiran parameter $\beta_1,\beta_2,\ldots,\beta_{p1}$ dan $\gamma_1, \gamma_2, \ldots, \gamma_{p2}$ diperoleh dari penyelesaian sejumlah *px1* vektor persamaan yang dinyatakan dengan skor koefisien vektor $u(\alpha)$. Skor koefisien untuk α_j adalah $u(\alpha_j) = \dfrac{\partial \log L(\alpha)}{\partial \alpha_j}$, sehingga

$$u(\alpha_j) = \sum_{j=1}^{n}\delta_i\left\{W_j - \left(\frac{\sum_{l\in R(t_i)} W_{jl}\exp(\alpha' W_l)}{\sum_{l\in R(t_i)}\exp(\alpha' W_l)}\right)\right\} \tag{2.28}$$

Misal matriks I(α) adalah matriks *pxp* yang merupakan turunan kedua dari fungsi *log-likelihood* yang bernilai negatif maka I(α) adalah :

$$I(\alpha) = -\frac{\partial^2 \log L(\alpha)}{\partial \alpha_j \partial \alpha_k} \tag{2.29}$$

dengan $j = 1, 2,\ldots, p$ dan $k = 1, 2,\ldots, p$, maka I(α) disebut matriks Hessian atau matriks informasi pengamatan.

Berdasarkan prosedur iterasi *Newton-Raphson*, penaksiran dari parameter α pada $s+1$ yang disimbolkan $\hat{\alpha}_{s+1}$ adalah :

$$\hat{\alpha}_{s+1} = \hat{\alpha}_s + I^{-1}(\hat{\alpha}_s)u(\hat{\alpha}_s) \tag{2.30}$$

dengan

$$I^{-1}(\alpha_{jk}) = \left\{ \left[\frac{\sum_{l \in R(t_i)} W_{jl} W_{kl} \exp(\alpha' W_l)}{\sum_{l \in R(t_i)} \exp(\alpha' W_l)} \right] - \left[\left(\frac{\sum_{l \in R(t_i)} W_{jl} \exp(\alpha' W_l)}{\sum_{l \in R(t_i)} \exp(\alpha' W_l)} \right) \left(\frac{\sum_{l \in R(t_i)} W_{kl} \exp(\alpha' W_l)}{\sum_{l \in R(t_i)} W_l \exp(\alpha' W_l)} \right) \right] \right\} \tag{2.31}$$

di mana :

s : 0, 1, 2, …

$u(\hat{\alpha}_s)$: vektor skor koefisien

$\boldsymbol{I}^{-1}(\hat{\alpha}_s)$: invers matriks informasi yang diamati

Proses iterasi dimulai dengan menentukan nilai awal $\hat{\alpha}_0 = 0$ dan proses akan dihentikan jika perubahan pada nilai α_{s+1} dan α_s kurang dari 10^{-6}.

2.5 Kemampuan Prediksi

Untuk melihat kemampuan prediksi yang dihasilkan model regresi *Cox* maupun model regresi *Cox* dengan *time-dependent variable* digunakan nilai *cross validation* Q^2. Polanski *et al* (2004) mengungkapkan bahwa kemampuan prediksi dengan nilai *cross validation* Q^2 dapat digunakan untuk mengetahui seberapa baik keakuratan prediksi yang dihasilkan dari model yang terbentuk. Nilai Q^2 dapat diperoleh dengan rumus sebagai berikut :

$$Q^2 = 1 - \frac{\sum_i \left(y_i - \hat{y}_i\right)^2}{\sum_i \left(y_i - \bar{y}\right)^2} \quad (2.32)$$

di mana :

Q^2 : nilai *cross validation*

n : banyaknya pengamatan

$\mathbf{y_i}$: variabel respon pengamatan ke-i

$\hat{\mathbf{y}}_{\mathbf{i}}$: nilai prediksi bagi y_i

$\bar{\mathbf{y}}$: rata-rata variabel respon

Besaran nilai *cross validation* Q^2 memiliki rentang $0 \leq Q^2 < 1$, di mana semakin mendekati 1 berarti model yang dihasilkan memiliki prediksi yang semakin akurat.

2.6 *Akaike's Information Criterion* (AIC)

Akaike's Information Criterion (AIC) dikembangkan oleh Hirotsugu Akaike pada tahun 1971 merupakan ukuran kebaikan penduga model statistik dengan mempertimbangkan banyaknya parameter dalam model. AIC tidak digunakan untuk pengujian suatu hipotesis, melainkan sebagai alat ukur perbandingan model dimana model dengan nilai AIC paling kecil dianggap sebagai model yang terbaik (Hu,2007). Secara umum AIC dapat dirumuskan sebagai

$$AIC = 2k - 2\ln(L) \quad (2.33)$$

di mana:

k : banyaknya parameter dalam model

L : *maximum likelihood* model yang diduga

Semakin banyak peubah prediktor yang diuji, maka kebaikan model juga akan meningkat. Oleh karena itu AIC tidak hanya bermanfaat untuk melihat kesesuaian model, tetapi juga sesuai digunakan untuk melihat kebaikan model untuk fungsi peningkatan sejumlah parameter.

BAB III
METODE PENELITIAN

3.1 Data

Data yang digunakan dalam penelitian ini diambil dari www.uhasselt.be. dengan variabel respon berupa waktu *survival* dan variabel prediktor bersifat kontinyu maupun kategorik. Penelitian dilakukan terhadap 1177 pasien yang mengidap penyakit AIDS dengan variabel yang diamati meliputi :

T : waktu *survival* (hari)
S : status kematian (1=mati, 0=sensor)
SEX : sex (0=laki-laki, 1=perempuan)
IVDRUG : Penggunaan obat (0=tidak pernah, 1=pernah)
MACSTAT : Status Mac (1=ya, 0=tidak mac)
MACTIME : Waktu mengidap penyakit Mac (hari)
KARNOF : skor Karnof

Variabel prediktor dalam kasus ini ialah sex, ivdrug, macstat, karnof dengan variabel respon T (waktu survival). Mactime merupakan waktu kapan diketahui status pasien mengidap penyakit mac (bukan waktu survival). Variabel tergantung waktu *(time-dependent variable)* dalam kasus ini ialah hasil perkalian antara macstat dan mactime.

3.2 Metode

Data pada penelitian ini diolah dengan bantuan *software Statistical Product and Service Solution* (SPSS) 15 dan SAS 9.1.

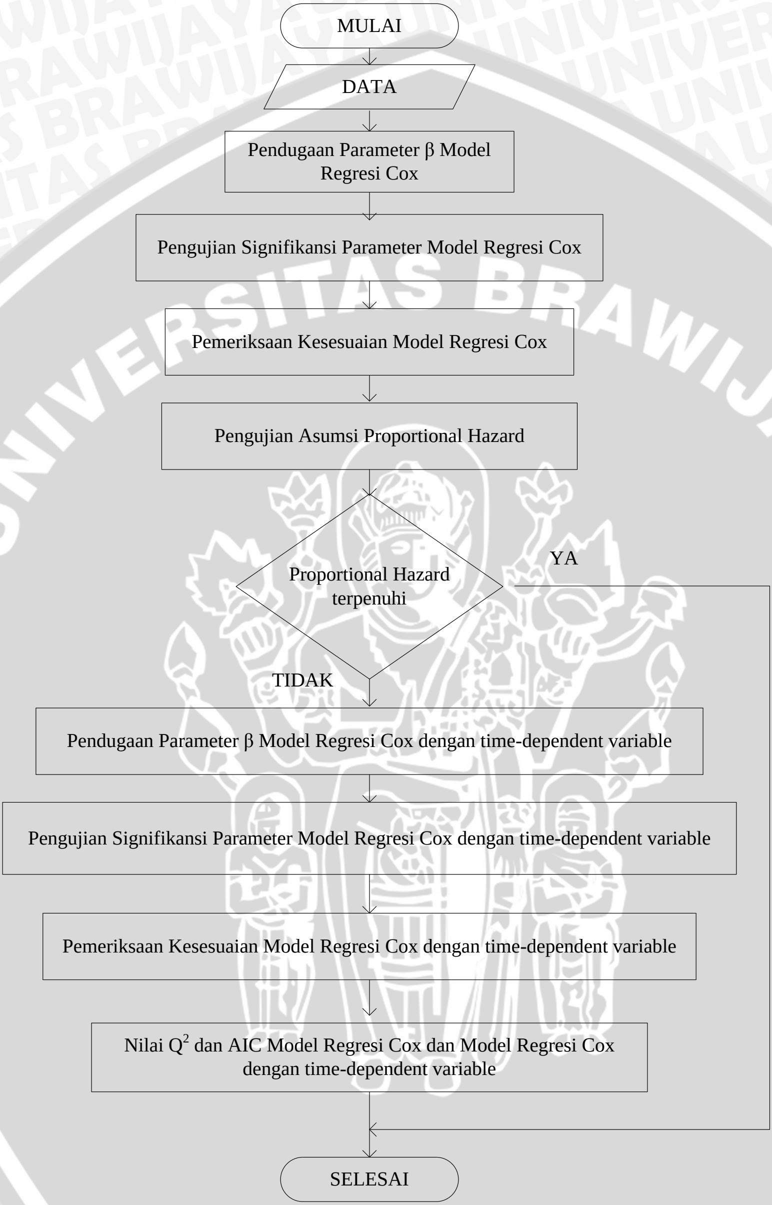
Langkah-langkah analisis pada penelitian ini adalah :

1. Eksplorasi Data.
2. Meregresikan variabel prediktor dengan variabel respon dengan menggunakan Metode Regresi *Cox*. Berikut langkah-langkah untuk melakukan analisis regresi *Cox* :
   a. Pendugaan koefisien β model regresi *Cox*. Koefisien β diduga dengan menggunakan metode maksimum parsial *likelihood* dengan memaksimumkan fungsi *likelihood* pada persamaan (2.15).

b. Pengujian signifikansi parameter β model regresi *Cox* dengan persamaan (2.20) dan (2.21).
c. Pemeriksaan asumsi *proportional hazard* model regresi *Cox* secara grafis dengan melihat plot antara Log {-log[S(t, X)]} terhadap waktu *survival*.
d. Pemeriksaan kesesuaian model dengan melihat plot antara Log Sisaan Cox-Snell dengan Log *cumulative hazard*

3. Meregresikan variabel prediktor yang mengandung *time-dependent variable* dengan variabel responnya. Berikut langkah-langkah untuk melakukan analisis regersi *Cox* dengan *time-dependent variable* :
   a. Pendugaan parameter β dan γ model regresi *Cox* dengan *time-dependent variable*. Koefisien β dan γ diduga dengan menggunakan metode maksimum parsial *likelihood* dengan memaksimumkan fungsi *likelihood* pada persamaan (2.27).
   b. Pengujian signifikansi parameter untuk masing-masing koefisien β dan γ model regresi *Cox* dengan *time-dependent variable* dengan persamaan (2.20) dan (2.21).
   c. Pemeriksaan asumsi *proportional hazard* model regresi *Cox* dengan *time-dependent variable* secara grafis dengan melihat plot antara Log {-log[S(t, X)]} terhadap waktu *survival*.
   d. Pemeriksaan kesesuaian model dengan melihat plot antara Log Sisaan Cox-Snell dengan Log *cumulative hazard.*
4. Membandingkan nilai *cross validation* Q^2 dan AIC model regresi *Cox* dan model regrei *Cox* dengan *time-dependent variable* dengan persamaan (2.32) dan (2.33).

Langkah-langkah tersebut dapat digambarkan dengan diagram alir seperti pada Gambar 3.1.



Gambar 3.1 Diagram Alir Metode Penelitian

BAB IV
HASIL DAN PEMBAHASAN

4.1 Eksplorasi Data

Eksplorasi data bertujuan untuk mengetahui gambaran awal model regresi *cox* dengan variabel respon berupa waktu *survival* dan variabel prediktor meliputi skor karnof, sex, status mac dan ivdrug (penggunaan obat-obatan).

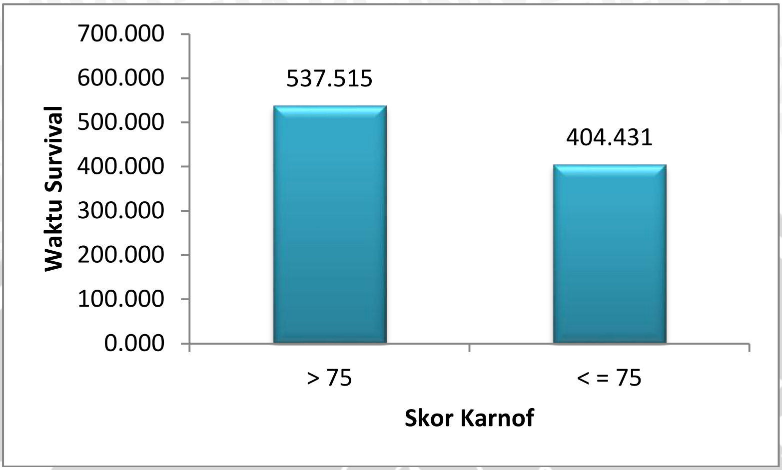
4.1.1 Skor Karnof

Skor Karnof pada variabel prediktor dapat dibagi menjadi dua kelas atau kategori yaitu pasien dengan skor karnof rendah (kurang dari 75) dan pasien dengan skor karnof tinggi (lebih dari 75). Tabel 4.1 menunjukkan banyaknya pasien yang memiliki skor karnof rendah maupun tinggi :

Tabel 4.1 Rata-rata Waktu Survival Berdasarkan Skor Karnof

Pasien	Frekuensi	Rata-rata Waktu Survival
Skor Karnof rendah	102	404.43
Skor Karnof Tinggi	1075	537.51
Total	1177	

Dari Tabel 4.1 dapat dilihat bahwa sebagian besar pasien atau sebesar 1075 pasien memiliki skor karnof tinggi dengan rata-rata waktu *survival* sebesar 537.51 hari. Sedangkan sisanya 102 atau sekitar 514 pasien memiliki skor karnof yang rendah dengan rata-rata waktu *survival* sebesar 404.43 hari. Tabel 4.1 dapat disajikan dalam bentuk grafik sebagai berikut :



Gambar 4.1 Rata-rata Waktu Survival Berdasarkan Skor Karnof

Dari Gambar 4.1 dapat dilihat bahwa pasien yang memiliki skor karnof rendah mempunyai rata-rata waktu *survival* sekitar 537 hari dan pasien yang memiliki skor karnof tinggi mempunyai rata-rata waktu *survival* sekitar 404 hari. Selisih rata-rata waktu *survival* keduanya sebesar 133 hari. Hal ini dapat mengindikasikan bahwa skor karnof mempengaruhi waktu bertahan hidup pasien penderita penyakit *aids*. Hal ini akan dibuktikan pada hasil pembentukan model regresi *Cox*.

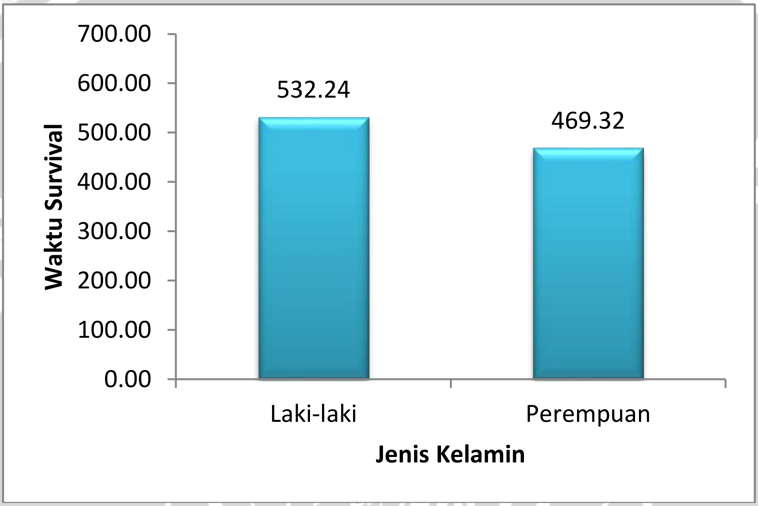
4.1.2 Jenis Kelamin (Sex)

Pasien dikelompokkan kedalam kategori laki-laki dan perempuan. Tabel 4.2 berikut menunjukkan banyaknya pasien berjenis kelamin laki-laki dan perempuan :

Tabel 4.2 Rata-rata Waktu Survival Berdasarkan Jenis Kelamin

Pasien	Frekuensi	Rata-rata Waktu Survival
Laki-laki	1060	532.24
Perempuan	117	469.32
Total	1177	

Berdasarkan Tabel 4.2 dapat dilihat bahwa sebagian besar pasien berjenis kelamin laki-laki atau sekitar 1060 pasien dengan rata-rata waktu survival sebesar 532.24 hari. Sedangkan sisanya 117 pasien berjenis kelamin perempuan dengan rata-rata waktu survival sebesar 469.32 hari. Tabel 4.1 dapat tampilkan dalam bentuk grafik sebagai berikut :



Gambar 4.2 Rata-rata Waktu Survival Berdasarkan Jenis Kelamin

Dari Gambar 4.2 dapat dilihat bahwa pasien laki-laki mempunyai rata-rata waktu *survival* sekitar 532 hari dan pasien perempuan mempunyai rata-rata waktu *survival* sekitar 469 hari. Selisih rata-rata waktu *survival* keduanya sebesar 63 hari. Hal ini dapat mengindikasikan bahwa Jenis Kelamin (Sex) tidak mempengaruhi waktu bertahan hidup pasien penderita penyakit *aids*. Hal ini akan dibuktikan pada hasil pembentukan model regresi *Cox*.

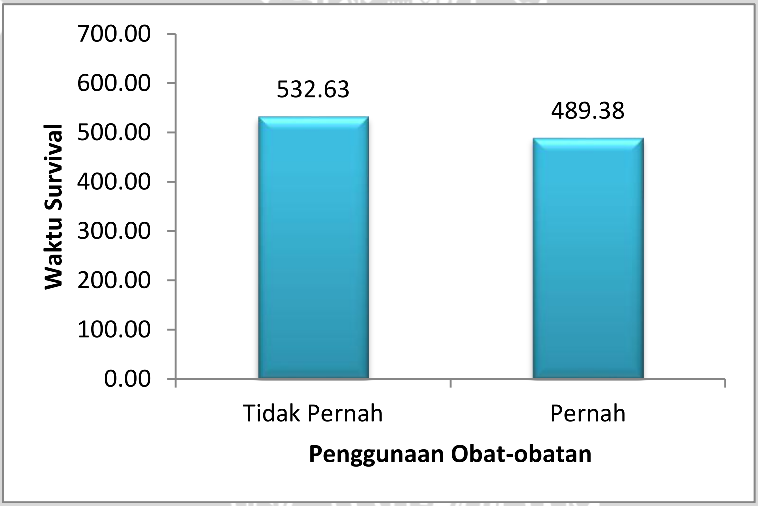
4.1.3 Penggunaan Obat (ivdrug)

Penggunaan obat oleh pasien digolongkan menjadi dua, yaitu pasien yang pernah menggunakan obat-obatan dan pasien yang belum/tidak pernah menggunakan obat-obatan. Tabel 4.3 menyajikan frekuensi pasien yang menggunakan obat-obatan dan pasien yang tidak menggunakan obat-obatan :

Tabel 4.3 Rata-rata Waktu Survival Berdasarkan ivdrug

Pasien	Frekuensi	Rata-rata Waktu Survival
Tidak Pernah	996	532.63
Pernah	181	489.38
Total	1177	

Berdasarkan Tabel 4.3 dapat dilihat bahwa sebagian besar pasien tidak pernah menggunakan obat-obatan atau sekitar 996 pasien dengan rata-rata waktu survival sebesar 532.63 hari. Sedangkan sisanya 181 pasien pernah menggunakan obat-obatan dengan rata-rata waktu survival sebesar 489.38. Tabel 4.3 dapat disajikan dalam bentuk grafik sebagai berikut :



Gambar 4.3 Rata-rata Waktu Survival Berdasarkan Penggunaan Obat-obatan

Dari Gambar 4.3 dapat dilihat bahwa pasien yang tidak pernah menggunakan obat-obatan mempunyai rata-rata waktu *survival* sekitar 532 hari dan pasien yang menggunakan obat-obatan mempunyai rata-rata waktu *survival* sekitar 489 hari. Selisih rata-rata waktu *survival* keduanya sebesar 43 hari. Hal ini dapat mengindikasikan bahwa penggunaan obat-obatan (ivdrug) tidak mempengaruhi waktu bertahan hidup pasien penderita penyakit *aids*. Hal ini dapat dibuktikan pada hasil pembentukan model regresi *Cox*.

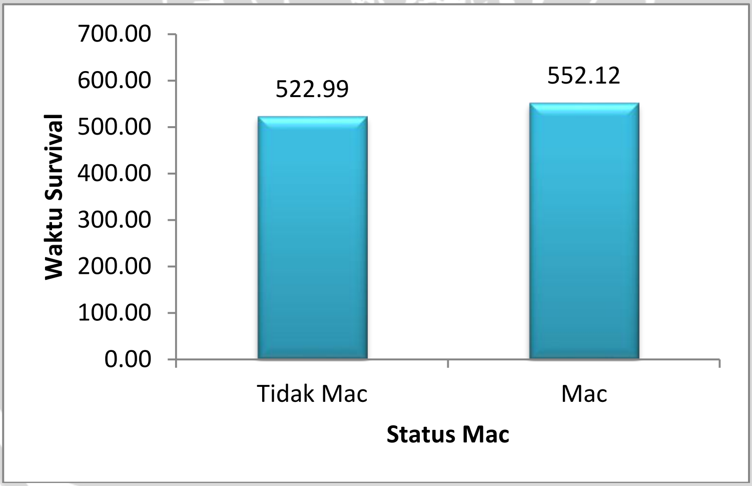
4.1.4 Status Mac

Variabel ini berupa nilai frekuensi yang menunjukkan pasien yang terkena penyakit mac atau tidak terkena penyakit mac. Tabel 4.1 berikut menunjukkan banyaknya pasien yang terkena penyakit mac maupun pasien yang tidak terkena penyakit mac :

Tabel 4.4 Rata-rata Waktu Survival Berdasarkan Status Mac

Pasien	Frekuensi	Rata-rata Waktu Survival
Tidak MAC	1056	522.99
MAC	121	552.12
Total	1177	

Berdasarkan Tabel 4.4 dapat dilihat bahwa sebagian besar pasien tidak menderita penyakit mac atau sekitar 1056 pasien dengan rata-rata waktu survival sebesar 522.99. Sedangkan sisanya 121 pasien menderita penyakit mac dengan rata-rata waktu survival sebesar 552.12.



Gambar 4.4 Rata-rata Waktu Survival Berdasarkan Status Mac

Dari Gambar 4.4 dapat dilihat bahwa pasien yang tidak terkena penyakit mac mempunyai rata-rata waktu *survival* sekitar 523 hari dan pasien yang terkena penyakit mac mempunyai rata-rata waktu *survival* sekitar 552 hari. Selisih rata-rata waktu *survival* keduanya sebesar 29 hari. Hal ini mengindikasikan bahwa status mac (macstat) tidak mempengaruhi waktu bertahan hidup pasien penderita penyakit *aids*. Hal ini akan dibuktikan pada hasil pembentukan model regresi *Cox*.

4.2 Model Regresi *Cox*

4.2.1 Pendugaan Parameter β Model Regresi *Cox*

Pendugaan koefisien β dilakukan dengan memaksimumkan *log-likelihood* pada persamaan (2.15) dengan metode iterasi *Newton Raphson*. Nilai β model regresi *cox* dapat dilihat pada Tabel 4.5.

Tabel 4.5 Nilai Koefisien β Model Regresi *Cox*

Variabel	Koefisien β
macstat	-0.0285
sex	-0.2370
skor karnof	-0.0453
ivdrug	-0.0951

Dari hasil yang diperoleh pada Tabel. 4.5 Model persamaan regresi *cox* yang terbentuk ialah :

$$h(t,X) = \exp\left(-0.0285\ \text{MacStat} - 0.2370\ \text{Sex} - 0.0453\ \text{SkorKarnof} - 0.095\ \text{ivdrug}\right)$$

dengan :

t = waktu *survival*

X = himpunan variabel prediktor yang terdiri dari :

1. macstat : status mac
2. sex : jenis kelamin
3. skor karnof : tingkat kesehatan pasien
4. ivdrug : penggunaan obat-obatan

4.2.2 Pengujian Signifikansi Parameter β Model Regresi *Cox*

4.2.2.1 Uji Simultan

Uji secara simultan model regresi *Cox* dengan menggunakan uji *likelihood ratio*. Nilai χ^2_{LR} model regresi *Cox* ialah sebesar 84.703 dengan p-value sebesar 0.000. Karena *p-value* < 0.05 maka H_0 ditolak dan dapat dikatakan bahwa secara bersama-sama variabel macstat, sex, skor karnof, dan ivdrug memberikan pengaruh yang signifikan terhadap peluang kematian pasien.

4.2.2.2 Uji Parsial

Uji secara parsial menggunakan uji *Wald*. Tolak H_0 jika $W > \chi^2_{1,0.05} = 3.841$ atau *p-value* ≤ 0.05. Pengujian signifikansi disajikan pada tabel 4.6 sebagai berikut :

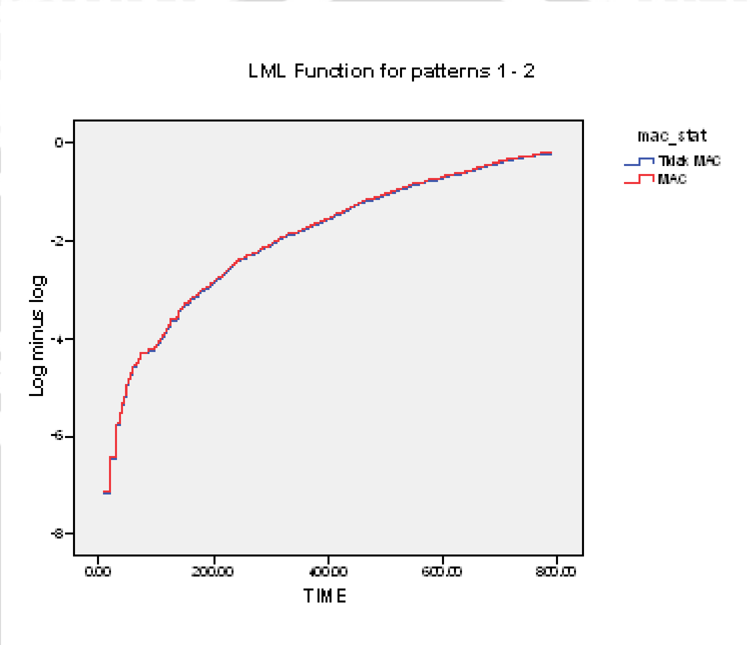
Tabel 4.6 Pengujian Signifikansi β Model Regresi *Cox*

Variabel	Koefisien β	Wald	Sig.	Exp(B)
macstat	-0.0285	0.0445	0.8330	0.9719
sex	-0.2370	2.6607	0.1029	0.7890
skor karnof	-0.0453	80.3390	0.0000	0.9557
ivdrug	-0.0951	0.6031	0.4374	0.9092

Dari Tabel 4.6 dapat diketahui bahwa variabel yang memberikan pengaruh signifikan terhadap peluang kegagalan pasien ialah skor karnof karena memiliki p-value = 0.000 < 0.05. Besar nilai pengaruh variabel skor karnof ialah sebesar 0.9557 dengan koefisien parameter bertanda negatif, artinya setiap meningkatnya kesehatan pasien akan menurunkan resiko kematian pasien tersebut sebesar 0.9557 kali. Variabel yang lain yaitu macstat, sex, dan ivdrug memiliki p-value > 0.05 sehingga ketiga variabel tersebut tidak memberikan pengaruh signifikan terhadap peluang kematian pasien.

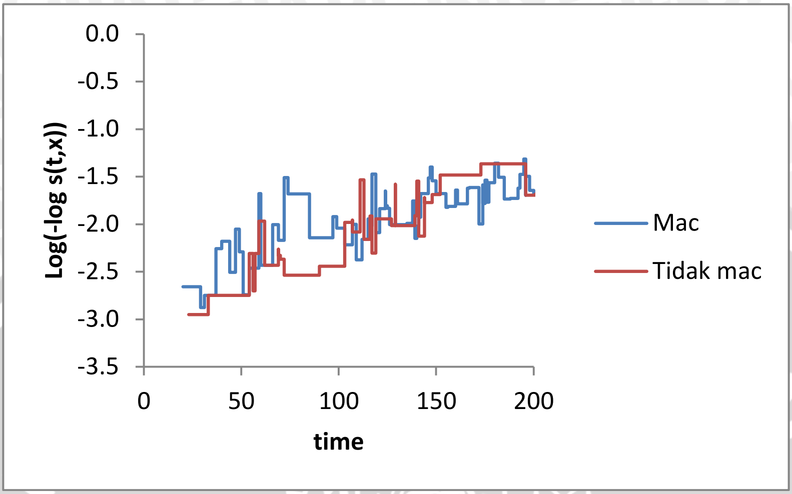
4.2.3 Pengujian Asumsi *Proportional Hazard* Model Regresi *Cox*

Pemeriksaan asumsi *proportional hazard* dilakukan secara grafis dengan melihat plot antara Log {-log[S(t, X)]} terhadap waktu *survival*. Jika plot antara Log {-log[S(t, X)]} terhadap waktu *survival* untuk beberapa kategori dalam variabel prediktor terlihat sejajar atau terdapat perbedaan yang konstan, maka asumsi *proportional hazard* dikatakan terpenuhi. Hasil uji asumsi *proportional hazard* pada variabel macstat, sex, dan ivdrug dapat dilihat pada Gambar 4.5, Gambar 4.6, dan Gambar 4.8



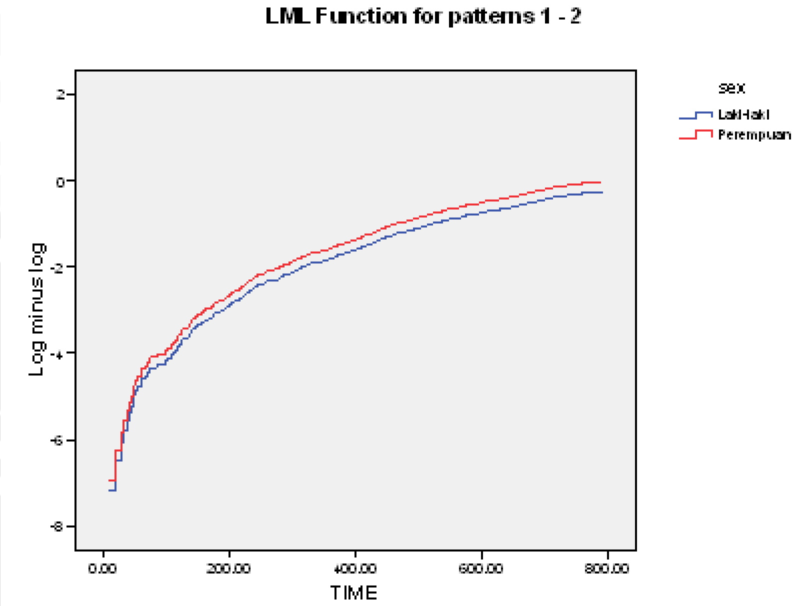
Gambar 4.5 Grafik Log {-log[S(t, X)]} terhadap waktu *survival* untuk variabel macstat

Agar terlihat dengan jelas, waktu survival pada Gambar 4.5 dapat dipotong pada waktu t > 200 dan di ambil pada waktu t=30 sampai t=200 sehingga Gambar 4.5 menjadi :



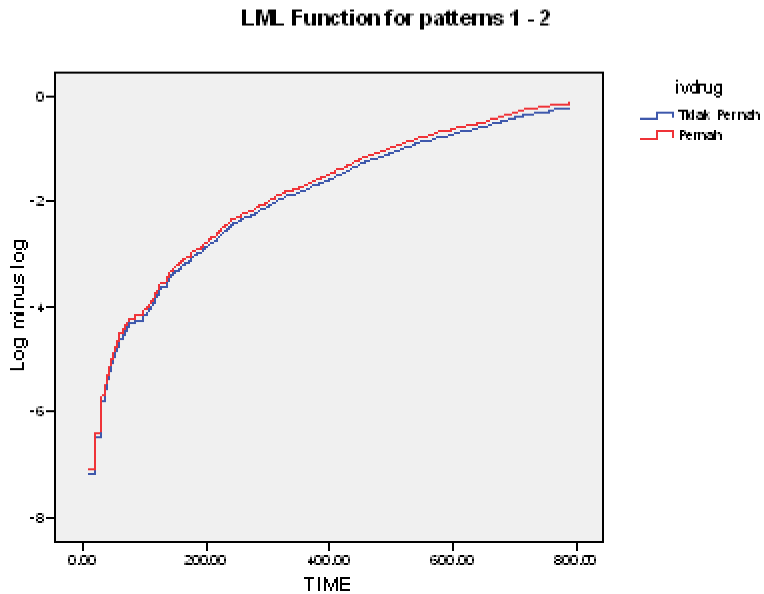
Gambar 4.6 Grafik Log {-log[S(t, X)]} terhadap waktu *survival* untuk variabel macstat

Dari Gambar 4.6 pada variabel macstat dapat dilihat antara kategori pasien yang terjangkit penyakit mac atau tidak terjangkit penyakit mac saling bersilangan (tidak sejajar) di beberapa selang waktu t yaitu sekitar t=60, t=110, t=150 sehingga asumsi *proportional hazard* pada varaibel ini tidak terpenuhi.



Gambar 4.7 Grafik Log {-log[S(t, X)]} terhadap waktu *survival* untuk variabel sex

Dari Gambar 4.7 pada variabel sex dapat dilihat adanya perbedaan antara kategori pasien laki-laki atau perempuan pada semua selang waktu *t* (tidak bersilangan), sehingga asumsi *proportional hazard* pada varaibel sex ini terpenuhi.



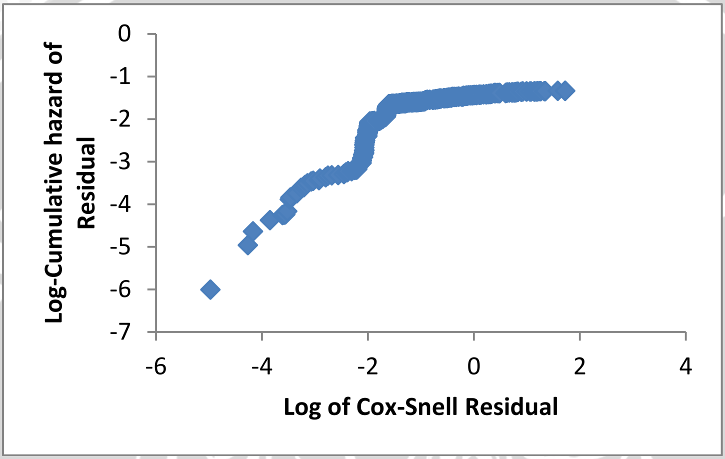
Gambar 4.8 Grafik Log {-log[S(t, X)]} terhadap waktu *survival* untuk variabel ivdrug

Dari Gambar 4.8 pada variabel ivdrug dapat dilihat adanya perbedaan antara pasien yang pernah menggunakan obat-obatan dan tidak pernah menggunakan obat-obatan pada semua selang waktu *t* (tidak bersilangan), sehingga asumsi *proportional* hazard pada varaibel ivdrug ini terpenuhi.

Dari ketiga variabel prediktor yaitu macstat, sex , dan ivdrug. Hanya variabel macstat yang tidak memenuhi asumsi *proportional hazard* dan dikatakan *nonproportional hazard*. Hal ini menunjukkan bahwa *hazard ratio* yang dihasilkan tidak konstan dan bergantung waktu dan mengakibatkan model yang dihasilkan tidak sesuai..

4.2.4 Pemeriksaan Kesesuaian Model

Langkah selanjutnya ialah pemeriksaan kesesuaian model. Pemeriksaan kesesuaian model dilakukan menggunakan sisaan Cox-Snell dengan melihat plot antara Log dari sisaan Cox-Snell terhadap Log *Cumulative Hazard* seperti pada Gambar 4.9



Gambar 4.9 Plot Log Sisaan Cox-Snell terhadap Log *Cumulative Hazard*

Dari Gambar 4.9 dapat dilihat bahwa plot antara Log Sisaan Cox-Snell terhadap Log *Cumulative Hazard* tidak membentuk garis lurus. Hal ini mengindikasikan bahwa model regresi *Cox* yang dihasilkan tidak sesuai. Ketidaksesuaian model dapat disebabkan karena asumsi *proportional hazard* tidak terpenuhi.

4.3 Model Regresi *Cox* dengan *Time-Dependent Variable*

4.3.1 Pendugaan Parameter β dan γ Model Regresi Cox dengan *Time-Dependent Variable*

Pendugaan koefisien β dan γ dilakukan dengan memaksimumkan *log-likelihood* pada persamaan (2.30) dengan metode iterasi *Newton Raphson*. Nilai β model regresi *cox* dengan *time-dependent variable* dapat dilihat pada Tabel 4.7.

Tabel 4.7 Nilai β dan γ Model Regresi *Cox* dengan *time-dependent variable*

Variabel	Koefisien β dan γ
macstat	-1.1628
sex	-0.2454
skor karnov	-0.0450
ivdrug	-0.0854
macstat*mactime	-0.0030

Dari hasil yang diperoleh pada Tabel 4.7. Model persamaan regresi *cox* dengan *time-dependent variable* yang terbentuk ialah :

$$h(t, X(t)) = \exp\big(-1.163 \text{ macstat } - 0.245 \text{ sex} - 0.045 \text{ skor karnof} \\ - 0.095 \text{ ivdrug} - 0.003 \text{ macstat} * \text{mactime}\big)$$

dengan t merupakan waktu *survival* dan X ialah himpunan variabel prediktor yang diteliti. β merupakan koefisien dari variabel macstat, sex, skor karnof, dan ivdrug. Sedangkan γ ialah koefisien dari variabel tergantung waktu (macstat*mactime).

4.3.2 Pengujian Signifikansi Parameter β Model Regresi Cox dengan *Time-Dependent Variable*

4.3.2.1 Ui Simultan

Uji secara simultan model regresi *Cox* dengan menggunakan uji *likelihood ratio*. Nilai χ^2_{LR} model regresi *Cox* dengan *time-dependent variable* ialah sebesar 105.490 dengan *p-value* sebesar 0.000. Karena p-value < 0.05 maka H_0 ditolak dan dapat dikatakan bahwa secara bersama-sama variabel macstat, sex, skor karnof, macstat*mactime, dan ivdrug memberikan pengaruh yang signifikan terhadap peluang kematian pasien.

4.3.2.2 Uji Parsial

Uji Parsial menggunakan uji *Wald*. Tolak H_0 jika $W > \chi^2_{1,0.05} = 3.841$ atau *p-value* ≤ 0.05. Pengujian signifikansi disajikan pada tabel 4.8 sebagai berikut :

Tabel 4.8 Pengujian Signifikansi β Model Regresi *Cox* dengan *time-dependent variable*

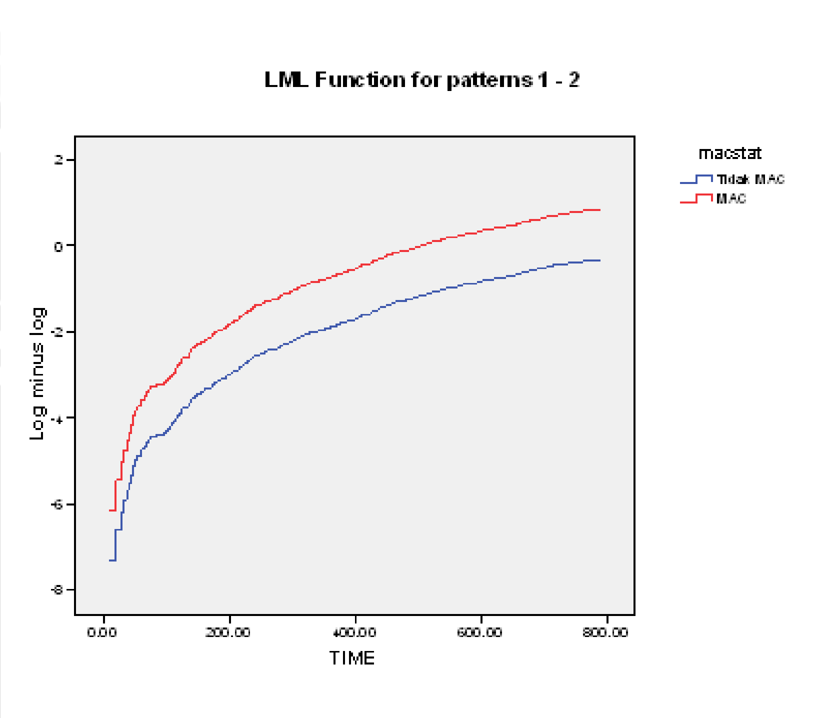
Variabel	Koefisien β	Wald	Sig.	Exp(B)
macstat	-1.1628	18.0736	0.0000	0.3126
sex	-0.2454	2.8477	0.0915	0.7824
skor karnov	-0.0450	78.1937	0.0000	0.9560
ivdrug	-0.0854	0.4854	0.4860	0.9181
macstat*mactime	-0.0030	17.3739	0.0000	0.9970

Dari Tabel 4.8 dapat dilihat bahwa hanya variabel sex dan ivdrug yang tidak memberikan pengaruh signifikan terhadap peluang kegagalan pasien karena memiliki *p-value* > 0.005. Sedangkan variabel macstat, skor karnof, dan macstat*mactime yang digunakan sebagai *time-dependent variable* menunjukkan pengaruh signifikan terhadap peluang kegagalan pasien karena semuanya memiliki nilai *p-value* yang sangat kecil (0.000) < 0.05.

Pada variabel macstat, besar nilai pengaruh pasien yang terkena penyakit mac (status=1) relatif terhadap pasien yang tidak terkena penyakit (status=0) adalah 0.3126 atau dapat dikatakan bahwa dengan menggunakan analisis regresi *cox* dengan *time-dependent variable,* pasien yang terkena penyakit mac mempunyai peluang meninggal sebesar 0.3126 kali lebih tinggi daripada pasien yang tidak terkena penyakit mac. Pada variabel skor karnof, besar pengaruh terhadap resiko kematian ialah sebesar 0.9560 dengan koefisien parameter bertanda negatif, artinya setiap meningkatnya kesehatan pasien akan menurunkan resiko kematian pasien tersebut sebesar 0.9560 kali

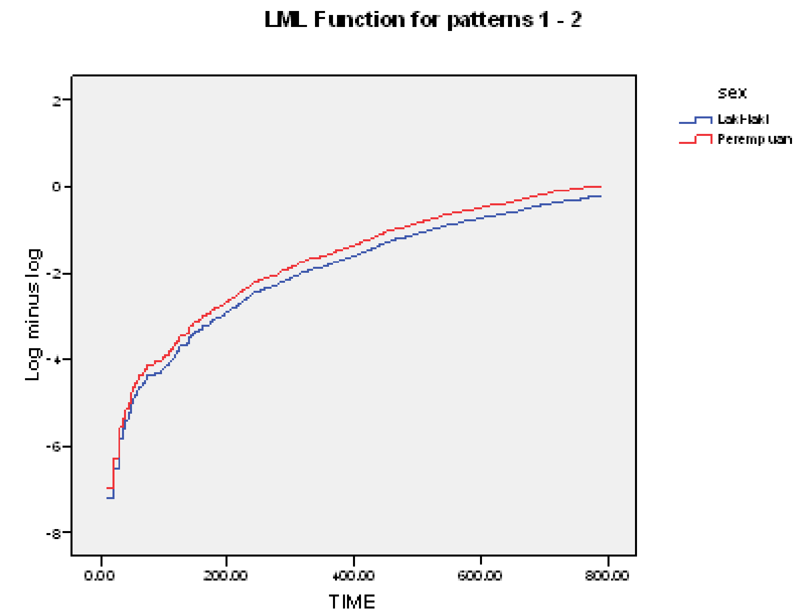
4.3.3 Pengujian Asumsi *Proportional Hazard* Model Regresi *Cox* dengan *Time-Dependent Variable*

Hasil uji asumsi *proportional hazard* setelah penambahan *time-dependent variable* dapat dilihat pada Gambar 4.10, 4.11, dan 4.12.



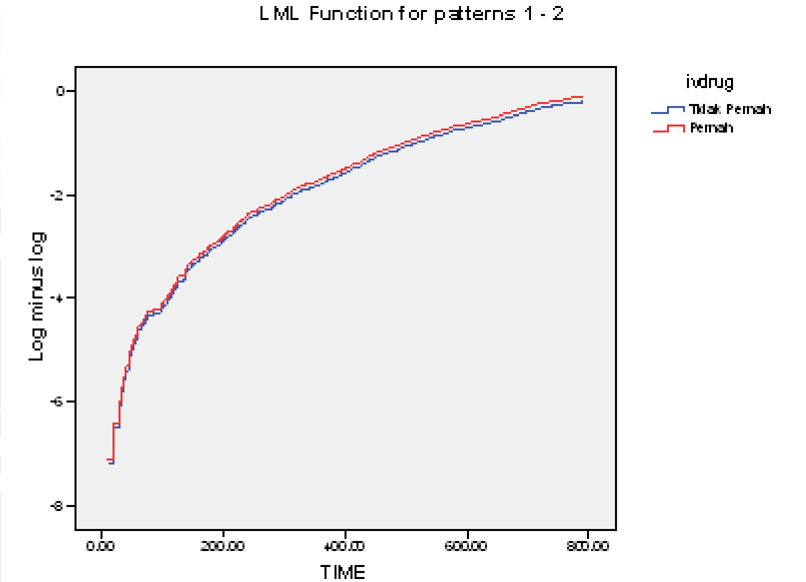
Gambar 4.10 Grafik Log {-log[S(t, X)]} terhadap waktu *survival* untuk variabel macstat

Dari Gambar 4.10 pada variabel macstat dapat dilihat dengan jelas adanya perbedaan antara kategori pasien yang terjangkit penyakit mac atau tidak terjangkit penyakit mac pada semua selang waktu *t* (tidak bersilangan), sehingga asumsi *proportional hazard* pada varaibel macstat ini terpenuhi dan model yang dihasilkan layak.



Gambar 4.11 Grafik Log {-log[S(t, X)]} terhadap waktu *survival* untuk variabel sex

Dari Gambar 4.11 pada variabel sex dapat dilihat adanya perbedaan antara kategori pasien laki-laki atau perempuan pada semua selang waktu *t* (tidak bersilangan), sehingga asumsi *proportional hazard* pada varaibel sex ini terpenuhi dan model yang dihasilkan layak.

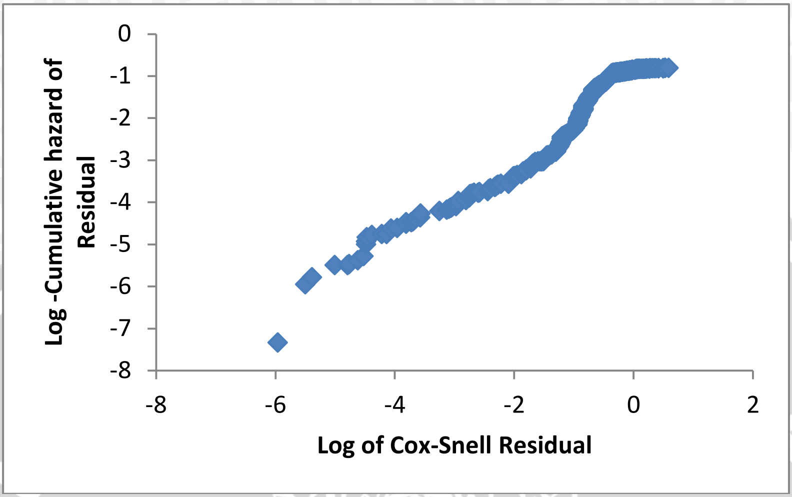


Gambar 4.12 Grafik Log {-log[S(t, X)]} terhadap waktu *survival* untuk variabel ivdrug

Dari Gambar 4.12 pada variabel ivdrug dapat dilihat adanya perbedaan antara pasien yang pernah menggunakan obat-obatan dan pasien yang tidak pernah menggunakan obat-obatan pada semua selang waktu *t* (tidak bersilangan), sehingga asumsi *proportional hazard* pada varaibel ivdrug ini terpenuhi dan model yang dihasilkan sesuai (layak).

4.3.4 Pemeriksaan Kesesuaian Model

Langkah selanjutnya ialah pemeriksaan kesesuaian model. Pemeriksaan kesesuaian model dilakukan menggunakan sisaan Cox-Snell dengan melihat plot antara Log dari sisaan Cox-Snell terhadap Log *Cumulative Hazard* seperti pada Gambar 4.13



Gambar 4.13 Plot Log Sisaan Cox-Snell terhadap Log *Cumulative Hazard*

Dari Gambar 4.13 dapat dilihat bahwa plot antara Log Sisaan Cox-Snell terhadap Log *Cumulative Hazard* membentuk garis lurus. Hal ini mengindikasikan bahwa model regresi *Cox* dengan *time-dependent variable* yang dihasilkan sesuai.

4.4 Nilai Q^2 dan AIC Model Regresi *Cox* dan Model Regresi *Cox* dengan *Time-Dependent Variable*

Nilai *Cross Validation* Q^2 dan AIC untuk model regresi *Cox* dan model regresi *Cox* dengan *time-dependent variable* dapat disajikan sebagai berikut :

Tabel 4.9 Nilai Q^2 Model Regresi *Cox* dan Model Regresi *Cox* dengan *time-dependent variable*

Model	**Q^2**	**AIC**
Model Regresi Cox	0.9999929	6714.609
Mode Regresi Cox dengan *time-dependent variable*	0.9999928	6698.306

Berdasarkan Tabel 4.9, dapat diketahui nilai Q^2 model regresi *Cox* sebesar 0.9999929 dan model regresi *Cox* dengan *time-dependent variable* sebesar 0.9999928 dengan nilai AIC untuk model regresi *Cox* sebesar 6714.609 dan untuk model regresi *Cox* dengan *time-dependent variable* sebesar 6698.306.

Berdasarkan nilai Q^2, model regresi *Cox* dan model regresi *Cox* dengan *time-dependent variable* memiliki selisih nilai Q^2 sangat kecil (0.0000001). Hal ini mengindikasikan bahwa kemampuan model regresi *Cox* dan model regresi *Cox* dengan *time-dependent variable* sama baiknya. Tetapi jika dilihat dari nilai AIC, model regresi *Cox* dengan *time-dependent variable* memiliki nilai AIC terkecil sehingga dapat dikatakan bahwa model ini lebih baik daripada model regresi *Cox*.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari penelitian ini dapat ditarik kesimpulan sebagai berikut :

1. Model regresi *Cox* yang terbentuk pada data pasien penderita penyakit AIDS ialah :
$$h(t,X) = \exp\left(-0.0285\ \text{MacStat} - 0.2370\ \text{Sex} - 0.0453\ \text{SkorKarnof}\right.$$
$$\left. - 0.095\ \text{ivdrug}\right)$$
2. Model regresi *Cox* dengan *time-dependent variable* yang terbentuk pada data pasien penderita penyakit *aids* ialah :
$$h(t,X(t)) = \exp\left(-1.163\ \text{macstat} - 0.245\ \text{sex} - 0.045\ \text{skor karnof}\right.$$
$$\left. - 0.095\ \text{ivdrug} - 0.003\ \text{macstat} * \text{mactime}\right)$$
3. Adanya *nonproportional hazard* pada model regresi *Cox* dapat diatasi dengan penambahan *time-dependent variable*. Hal ini terlihat dari grafik Log {-log[S(t, X)]} terhadap waktu *survival* pada variabel macstat untuk model regresi *Cox* dengan *time-dependent variable* yang menunjukkan kesejajaran antara dua kategori pasien sehingga asumsi *proportional hazard* terpenuhi.
4. Berdasarkan nilai Q^2, baik model regresi *Cox* maupun model regresi *Cox* dengan *time-dependent variable* mampu memberikan tingkat kemampuan prediksi yang sama karena memiliki selisih yang sangat kecil (0.0000001). Hal ini mengindikasikan bahwa kedua model sama baiknya. Tetapi jika dilihat dari nilai AIC, model regresi *Cox* dengan *time-dependent variable* lebih baik digunakan daripada model regresi *Cox* pada data yang tidak memenuhi asumsi *proportional hazard*.

5.2 Saran

Pada penelitian selanjutnya dapat dikembangkan :

1. Membentuk model regresi *Cox* dengan *time-dependent variable* jika terdapat lebih dari satu variabel yang tidak memenuhi asumsi *proportional hazard*.

2. Pembentukan variabel tergantung waktu dapat menggunakan fungsi dari waktu *survival* yaitu hasil interaksi antara variabel prediktor dengan waktu survival (X*t) atau dengan log waktu survival (X*log t).

DAFTAR PUSTAKA

Anderson, P. K. 1982. **Testing Goodness of Fit of Cox Regression and Life Model**. Biometrics, 38, 67-77.

Ata, Nihal and Sozer, M. Tekin. 2007. **Cox Regression Model with Nonproportional Hazard Applied to Lung Cancer Survival Data.** Hacetttepe Journal of Mathematics and Statistics Volume 36(2), 157-167.

Chan, YH. 2004. **Bostatistics 203 : Survival Analysis**. Singapore Med J Vol. 45 (6) : 249

Collet, D. 2003. **Modelling Survival Data in Medical Research Second Edition.** Chapman and Hall. London.

Cox, D.R. 1972. **Regression Model and Life Table (With Discussion)**. Journal of the Royal Statistical Society, B, 74, 187 – 220.

Fisher, Lloyd D. and Lin D. Y. 1999. **Time-Dependent Covariates in The Cox Proportional Hazards Regression Model.** Department of Biostatistics, University of Washington, Seaattle, Washington.

Fox, J. 2002. **Cox Proportional-Hazard Regression for Survival Data**.

Hosmer, DW., Lemeshow, S. 1999. **Applied Survival Analysis: Regression Modelling of Time to Event Data**. John Wiley and Son. Canada.

Kriestianianatha, Reshany. 2007. **Penggunaan model stratified proportional hazard untuk mengatasi non proportional**

hazard pada peubah bebas bersifat kategori model regresi cox. Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Brawiajaya. Tidak dipublikasikan.

Kleinbaum, D.G. and Klein, M. 2005. **Survival Analysis : A Self-Learning Text**. Second Edition. Springer-Verlag. New York.

Klein, P. John and Moeschberger, Melvin L. 2003. **Survival analysis Techniques for Censored and Truncated Data Second Edition.** Springer-Verlag. New York.

Kutner, M. H., Machtseim, and J. Neter. 2005. **Applied Linier Regression Models**. Fourth Edition. Mc Graw Hill. New York.

Lee, E.T. 1997. **Statistical Methods for Survival Data Analysis**. Belmont, CA : Wadsworth.

Miller, R. G. 1998. **Survival Analysis**. John Willey and Sons, New York.

Kurniawati, N. 2010. **Regresi Kuadrat Terkecil Parsial Cox pada Data *Survival***. Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Brawiajaya. Tidak dipublikasikan.

Polanski, J; A. Bak; R. Gieleciak and T. Magdzdiarz. 2004. **Self-Organizing Neural Network for Modelling Robust 3D and 4D QSAR: Application to Dihydofolate Reductase Inhibitors**

Shumway, T. 2001. **Forecasting bankruptcy more accurately : A simple hazard model**, Journal of Business, 74 (1), pp. 101-124.

Lampiran 1. Data Pasien Pengidap Penyakit AIDS (www.uhasselt.be)

Pasien	status	Time (hari)	macstat	mactime (hari)	sex	karnof (%)	ivdrug
1	0	623	1	560	0	90	0
2	0	651	0	651	0	90	0
3	1	464	0	26	0	100	0
4	0	622	0	622	0	80	0
5	0	643	0	643	0	90	0
6	1	216	0	171	0	90	0
7	1	425	1	174	0	90	0
8	1	449	1	449	0	90	0
9	0	811	1	377	0	80	0
10	1	74	0	58	0	60	1
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
1170	0	716	0	716	0	90	0
1171	0	675	0	675	0	90	0
1172	0	620	0	620	0	90	0
1173	0	581	0	544	0	90	0
1174	1	622	0	592	0	80	0
1175	0	629	0	578	0	80	0
1176	1	99	0	53	0	80	0
1177	0	629	0	627	0	80	0

Lampiran 2. Eksplorasi Data Penelitian

Statistics

		STATUS	mac_stat	sex	ivdrug
N	Valid	1177	1177	1177	1177
	Missing	0	0	0	0

mac_stat

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Tidak MAC	1056	89.7	89.7	89.7
	MAC	121	10.3	10.3	100.0
	Total	1177	100.0	100.0	

sex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Laki-laki	1060	90.1	90.1	90.1
	Perempuan	117	9.9	9.9	100.0
	Total	1177	100.0	100.0	

ivdrug

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Tidak Pernah	996	84.6	84.6	84.6
	Pernah	181	15.4	15.4	100.0
	Total	1177	100.0	100.0	

Lampiran 3. Model Regresi *Cox*

Block 0: Beginning Block

mnibus Tests of Model Coefficients

-2 Log Likelihood
6786.503

Block 1: Method = Enter

Omnibus Tests of Model Coefficients[a,b]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
6706.609	84.703	4	.000	79.894	4	.000	79.894	4	.000

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 6786.503

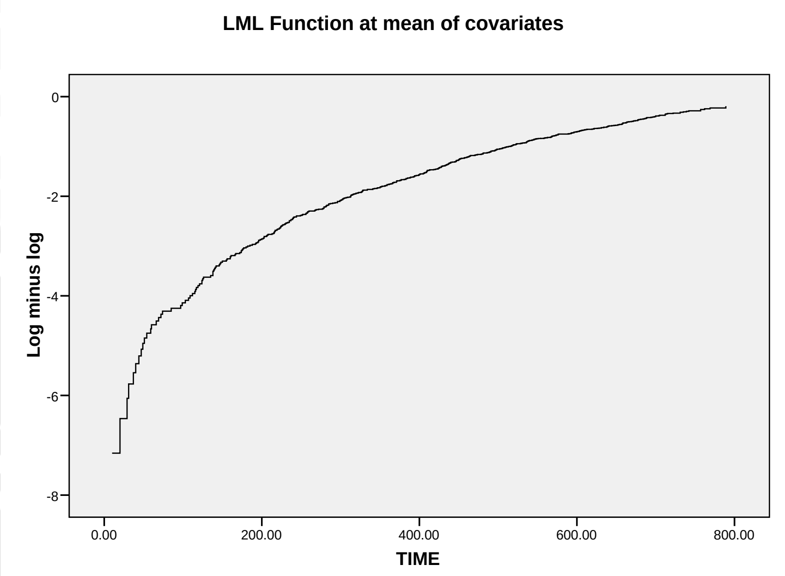
b. Beginning Block Number 1. Method = Enter

Variables in the Equation

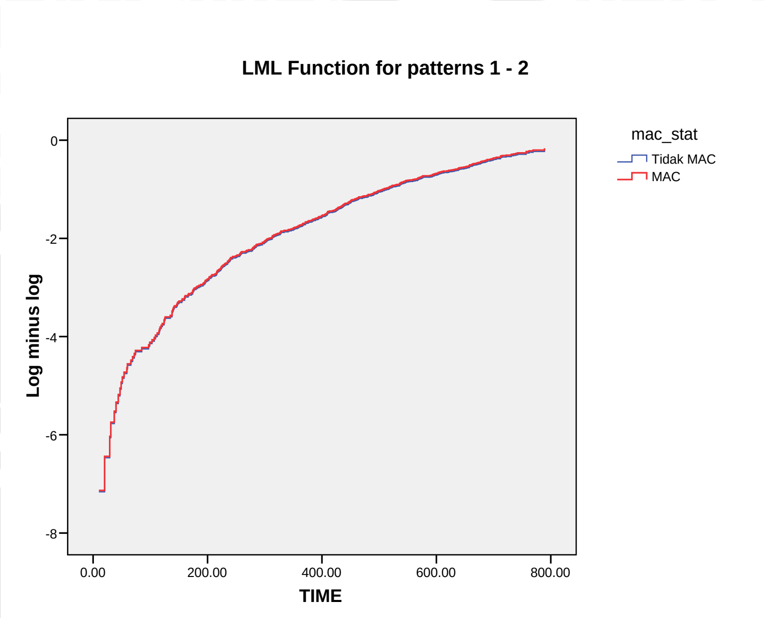
	B	SE	Wald	df	Sig.	Exp(B)
mac_stat	-.029	.135	.044	1	.833	.972
sex	-.237	.145	2.661	1	.103	.789
skor_karnov	-.045	.005	80.339	1	.000	.956
ivdrug	-.095	.123	.603	1	.437	.909

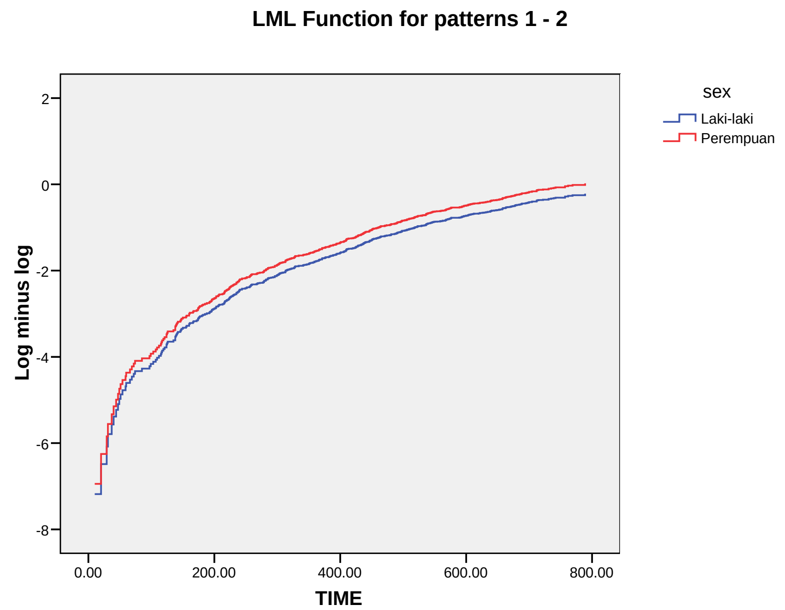
Covariate Means and Pattern Values

	Mean	Pattern	
		1	2
mac_stat	.897	1.000	.000
sex	.902	.902	.902
skor_karnov	86.911	86.911	86.911
ivdrug	.847	.847	.847

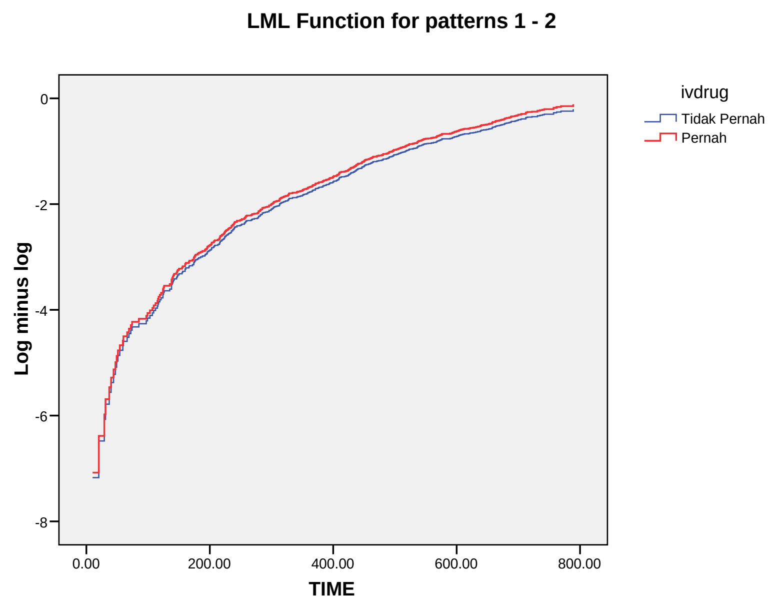

LML Function at mean of covariates
Log minus log
0
-2
-4
-6
-8
0.00
200.00
400.00
600.00
800.00
TIME

Lampiran 4. Uji Asumsi *Proportional Hazard* Model Regresi *Cox*





LML Function for patterns 1 - 2
Log minus log
TIME
sex
Laki-laki
Perempuan
2
0
-2
-4
-6
-8
0.00
200.00
400.00
600.00
800.00



LML Function for patterns 1 - 2
Log minus log
0
-2
-4
-6
-8
0.00
200.00
400.00
600.00
800.00
TIME
ivdrug
Tidak Pernah
Pernah

Lampiran 5. Model Regresi *Cox* dengan *Time-Dependent Variable*

Case Processing Summary

		N	Percent
Cases available in analysis	Event[a]	514	43.7%
	Censored	658	55.9%
	Total	1172	99.6%
Cases dropped	Cases with missing values	0	.0%
	Cases with negative time	0	.0%
	Censored cases before the earliest event in a stratum	5	.4%
	Total	5	.4%
Total		1177	100.0%

a. Dependent Variable: TIME

Categorical Variable Codings[c,d,e]

		Frequency	(1)[a]
mac_stat[b]	.00=Tidak MAC	1056	1
	1.00=MAC	121	0
sex[b]	.00=Laki-laki	1060	1
	1.00=Perempuan	117	0
ivdrug[b]	.00=Tidak Pernah	996	1
	1.00=Pernah	181	0

a. The (0,1) variable has been recoded, so its coefficients will not be the same as for indicator (0,1) coding.

b. Indicator Parameter Coding

c. Category variable: mac_stat

d. Category variable: sex

e. Category variable: ivdrug

Block 0: Beginning Block

mnibus Tests of Model Coefficient:

-2 Log Likelihood
6786.503

Block 1: Method = Enter

Omnibus Tests of Model Coefficients[a,b]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
6688.306	105.490	5	.000	98.197	5	.000	98.197	5	.000

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 6786.503

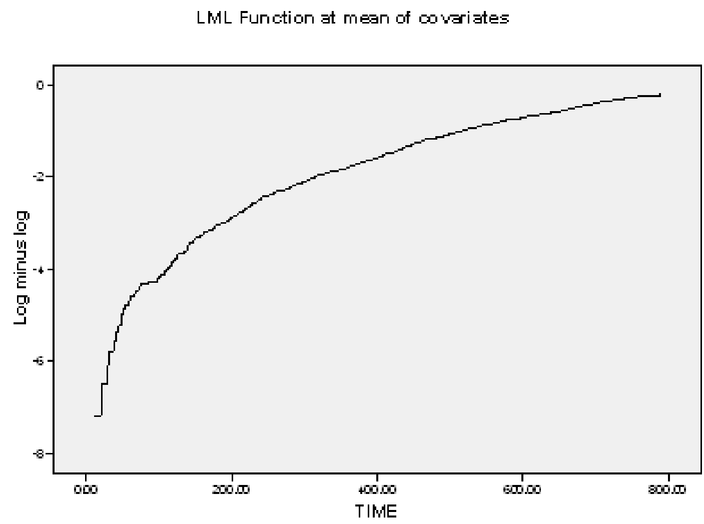
b. Beginning Block Number 1. Method = Enter

Variables in the Equation

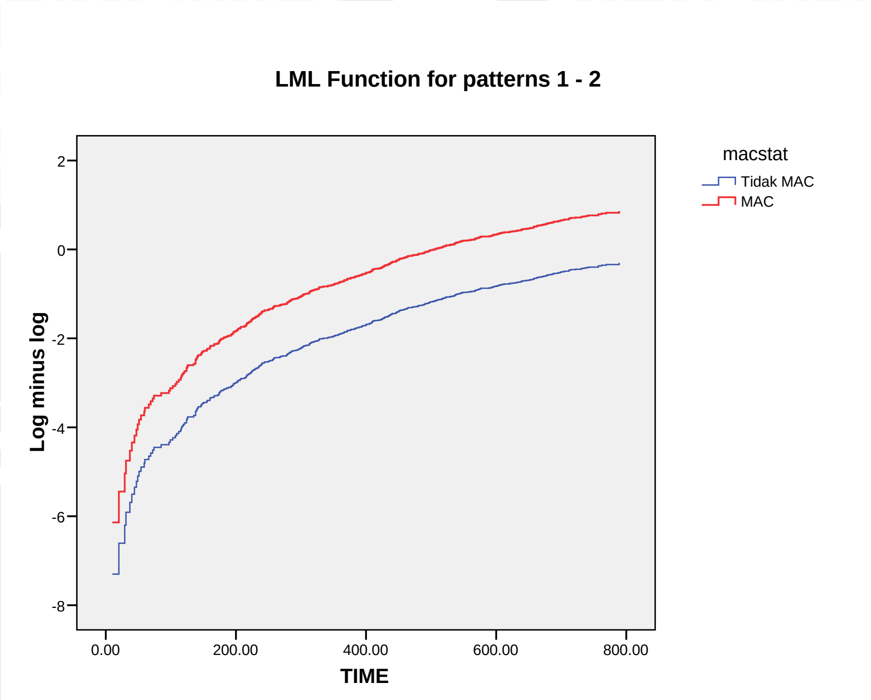
	B	SE	Wald	df	Sig.	Exp(B)
mac_stat	-1.163	.274	18.074	1	.000	.313
sex	-.245	.145	2.848	1	.092	.782
skor_karnov	-.045	.005	78.194	1	.000	.956
ivdrug	-.085	.123	.485	1	.486	.918
macstatXmactime	-.003	.001	17.374	1	.000	.997

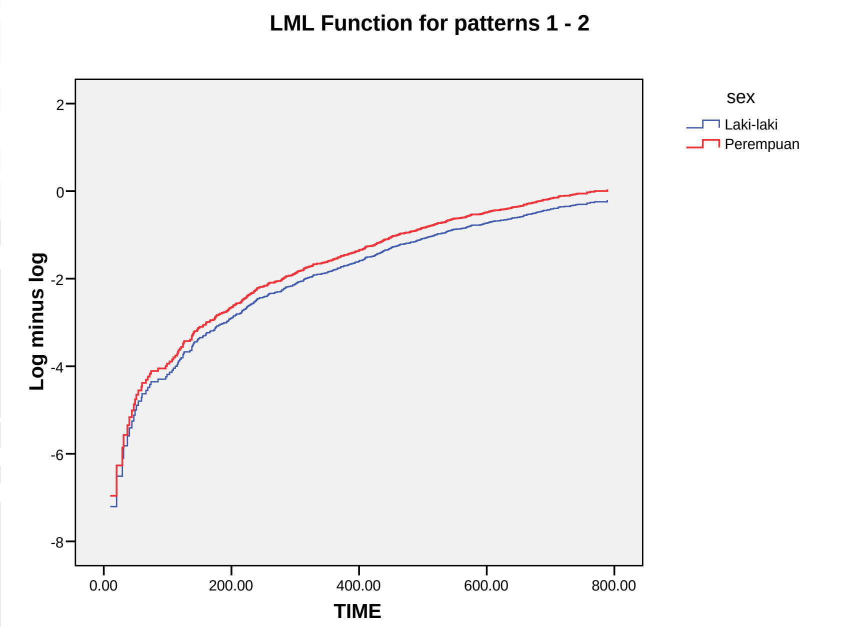
Covariate Means

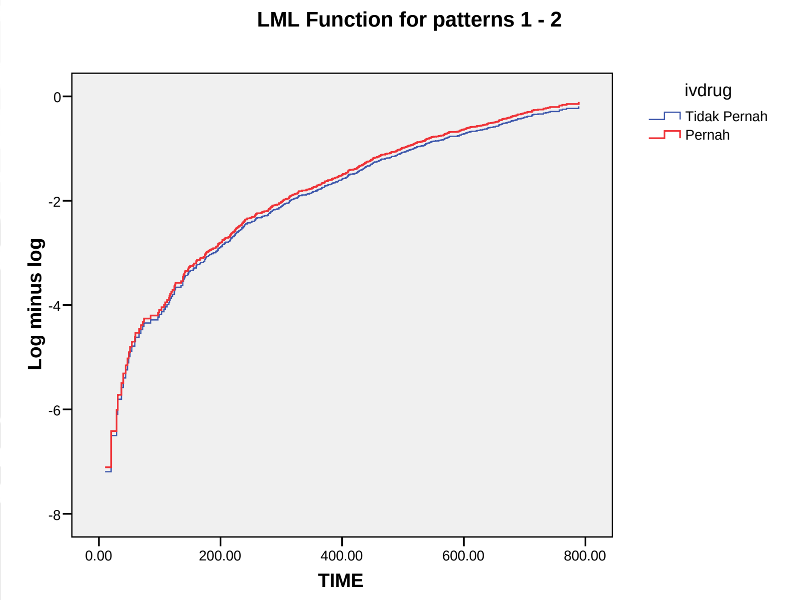
	Mean
mac_stat	.897
sex	.902
skor_karnov	86.911
ivdrug	.847
macstatXmactime	38.717


LML Function at mean of covariates
Log minus log
0
-2
-4
-6
-8
0.00
200.00
400.00
600.00
800.00
TIME

Lampiran 6. Uji Asumsi *Proportional Hazard* Model Regresi Cox dengan *Time-Dependent Variable*




LML Function for patterns 1 - 2
sex
Laki-laki
Perempuan
Log minus log
2
0
-2
-4
-6
-8
0.00
200.00
400.00
600.00
800.00
TIME


LML Function for patterns 1 - 2
Log minus log
0
-2
-4
-6
-8
0.00
200.00
400.00
600.00
800.00
TIME
ivdrug
Tidak Pernah
Pernah

Lampiran 7. Nilai *Akaike's Information Criterion* (AIC)

a. Model Regresi Cox

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	6786.503	6706.609
AIC	6786.503	6714.609
SBC	6786.503	6731.578

b. Model Regresi Cox dengan *time-dependent variable*

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	6786.503	6688.306
AIC	6786.503	6698.306
SBC	6786.503	6719.517

Lampiran 8. Prosedur Iterasi *Newton-Rhapson*

Fungsi *log-likelihood* seperti persamaan (2.15) ialah :

$$\log \mathrm{L}(\beta^{'}) = \sum_{i=1}^{n} \delta_{\mathrm{i}} \left\{ \beta^{'}\mathbf{x_i} - \log \sum_{l \in \mathrm{R}(\mathrm{t}_{(\mathrm{i})})} \exp\,(\beta^{'}\mathbf{x}_{l}) \right\} \qquad (2.15)$$

Karena fungsi *log-likelihood* pada persamaan (2.15) merupakan bentuk model nonliner, maka pendugaan parameter β dapat diselesaikan dengan memaksimumkan fungsi *log-likelihood* tersebut dengan prosedur iterasi *newton-rhapson* dengan rumus:

$$\hat{\beta}_{s+1} = \hat{\beta}_s + I^{-1}\left(\hat{\beta}_s\right) u\left(\beta_s\right)$$

Dimana :

s = 0, 1, 2, 3, …

dengan $u\left(\beta_j\right) = \dfrac{\partial \log L(\beta)}{\partial \beta_j}$, merupakan turunan pertama fungsi *log-likelihood* pada persamaan (2.15) dan disebut skor koefisien vektor $u(\beta)$. Sehingga diperoleh :

$$u\left(\beta_j\right) = \frac{\partial \log L(\beta)}{\partial \beta_j} = 0$$

$$= \sum_{j=1}^{n} \delta_j \left\{ x_j - \frac{\sum_{l \in \mathrm{R}(\mathrm{t}_{(\mathrm{i})})} \mathbf{x}_{jl}\ \exp\,(\beta' \mathbf{x}_l)}{\sum_{l \in \mathrm{R}(\mathrm{t}_{(\mathrm{i})})} \exp\,(\beta' \mathbf{x}_l)\ \ \ln \mathrm{e}} \right\}$$

$$= \sum_{j=1}^{n} \delta_j \left\{ x_j - \left(\frac{\sum_{l \in R(t_i)} x_{jl}\ \exp\left(\beta' x_l\right)}{\sum_{l \in R(t_i)} \exp\left(\beta' x_l\right)} \right) \right\}$$

Sedangkan $\mathbf{I}(\beta) = -\dfrac{\partial^2 \log L(\beta)}{\partial \beta_j \partial \beta_k}$ adalah matriks *pxp* yang merupakan turunan kedua dari fungsi *log-likelihood* yang bernilai negatif. Invers dari $\mathbf{I}(\beta)$ diperoleh :

$$I^{-1}\left(\beta_{jk}\right)=\left\{\left[\frac{\sum\limits_{l\in R(t_i)} x_{jl}\, x_{kl}\, \exp\left(\beta' x_l\right)}{\sum\limits_{l\in R(t_i)} \exp\left(\beta' x_l\right)}\right]-\left[\left(\frac{\sum\limits_{l\in R(t_i)} x_{jl}\, \exp\left(\beta' x_l\right)}{\sum\limits_{l\in R(t_i)} \exp\left(\beta' x_l\right)}\right)\left(\frac{\sum\limits_{l\in R(t_i)} x_{kl}\, \exp\left(\beta' x_l\right)}{\sum\limits_{l\in R(t_i)} \exp\left(\beta' x_l\right)}\right)\right]\right\}$$

dengan $j = 1, 2, \ldots, p$ dan $k = 1, 2, \ldots, p$,

Proses iterasi dimulai dengan menentukan nilai awal $\hat{\beta}_0 = 0$ dan proses akan dihentikan jika perubahan pada nilai β_{s+1} dan β_s kurang dari 10^{-6}.