

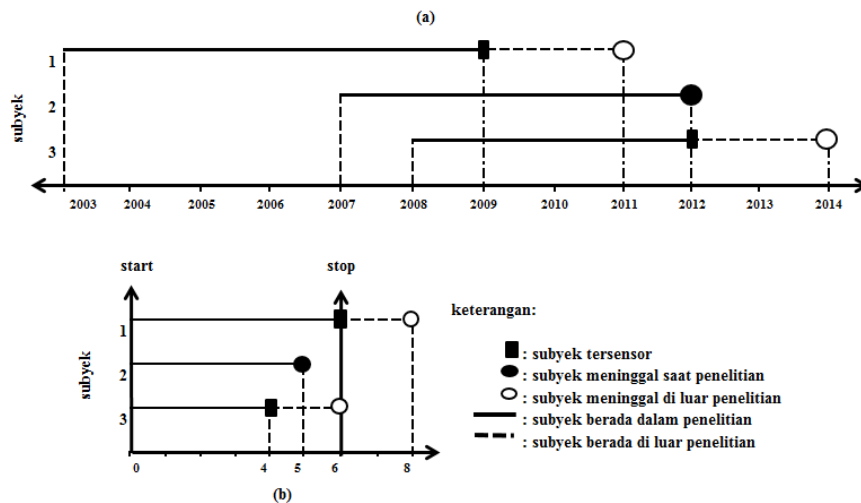
BAB II LANDASAN TEORI

A. Analisis Survival

Analisis survival atau analisis ketahanan hidup adalah analisis data yang berhubungan dengan waktu, mulai dari awal sampai terjadinya suatu peristiwa khusus (Colled, 2003). Jangka waktu dari awal dilakukan pengamatan pada suatu individu (*time origin*) sampai terjadinya suatu peristiwa khusus (*end point* atau *failure event*) disebut dengan waktu survival. Peristiwa khusus (*failure event*) tersebut dapat berupa kegagalan, kematian, kambuh atau sembuhnya dari suatu penyakit, respon dari suatu percobaan, atau peristiwa lain yang dipilih sesuai dengan kepentingan peneliti.

1. Waktu Survival (*Survival Time*)

Waktu survival (*survival time*) dalam analisis survival adalah periode amatan berupa interval waktu antara permulaan pengamatan hingga terjadinya kejadian yang diamati.



Gambar 2.1 *Calendar Time* (a) dan *Survival Time* (b)

Pada gambar 2.1 diatas, *survival time* berbeda dengan *calendar time*, *survival time* diukur dan ditetapkan berdasarkan mulainya peristiwa tertentu. Misalkan peneliti sedang menyelidiki penyakit “A”. Periode amatan dilakukan peneliti selama 6 tahun terhadap masing-masing subjek. Terdapat 3 subjek penelitian yang sudah terserang penyakit tersebut. Subjek (1) mulai diamati pada tahun 2003 dan penelitian berakhir tahun 2009. Subjek (1) meninggal tahun 2011, jadi tahun 2010 dan tahun 2011 subjek ini tidak masuk pada waktu penelitian. Subjek (2) mulai diamati pada tahun 2007 dan meninggal tahun 2012 sebelum penelitian berakhir. Subjek (3) mulai diamati tahun 2008 dan meninggal tahun 2014 sementara peneliti menghentikan penelitian pada tahun 2012, dua tahun lebih awal dari rencana awal periode pengamatan dikarenakan suatu hal.

Pada konsep *calendar time*, ketiga subjek mulai diamati pada waktu yang berbeda, sedangkan pada konsep analisis survival waktu diatur seolah-olah mulai pengamatan pada saat yang bersamaan seperti pada gambar (b). Penentuan *start* dan *stop* dalam analisis survival sangat penting untuk menentukan siapa saja subjek yang berisiko untuk suatu kejadian.

2. Penyensoran

Penyensoran adalah salah satu langkah yang harus dilakukan untuk mengatasi ketidaklengkapan suatu data pengamatan. Data dikatakan tersensor apabila data tidak dapat diamati secara lengkap karena subjek penelitian hilang atau mengundurkan diri atau sampai akhir penelitian subjek tersebut belum

mengalami kejadian tertentu, sedangkan data yang dapat diamati secara lengkap sampai penelitian berakhir disebut data yang tidak tersensor (Lee & Wang, 2003).

Menurut Kleinbaum dan Klein (2005) tiga penyebab data dikatakan tersensor antara lain:

- a. *Loss to follow up*, yaitu subjek menghilang selama masa pengamatan, misal subjek pindah atau menolak untuk diamati.
- b. Subjek tidak mengalami kejadian selama penelitian.
- c. Subjek terpaksa diberhentikan dari pengamatan karena meninggal sebelum pengamatan berakhir atau alasan lain.

Menurut David Collet (2003: 2) dalam analisis survival terdapat 3 tipe penyensoran yaitu:

- a. Sensor kanan (*right censoring*)

Sensor yang terjadi dikarenakan objek pengamatan belum mengalami kejadian hingga akhir periode pengamatan, sedangkan waktu awal dari objek pengamatan dapat diamati secara penuh. Misalkan suatu individu diamati selama lima tahun dari awal pengamatan, kemudian pada tahun ketiga individu tersebut pindah ke negara lain dan tidak dapat diamati lagi (*lost to follow up*). Individu ini memiliki waktu survival dalam penelitian setidaknya dua tahun, sehingga waktu pengamatan individu tersebut dikatakan tersensor kanan.

- b. Sensor kiri (*left censoring*)

Sensor yang terjadi dikarenakan waktu awal dari subjek pengamatan tidak dapat teramati pada awal pengamatan, sementara kegagalan dapat diamati secara penuh sebelum penelitian berakhir. Sebagai contoh, peneliti mengamati pasien penyakit kanker, peneliti dapat mencatat kejadian tepatnya seseorang tersebut positif kanker di tes pertamanya, namun peneliti tidak memiliki catatan tentang waktu tepatnya seseorang tersebut mulai berpenyakit kanker, dengan demikian pasien kanker tersebut tersensor kiri yaitu ketika mengalami kejadian pertama dengan hasil positif kanker.

c. *Sensor interval (interval censoring)*

Sensor interval adalah sensor yang waktu survivalnya berada dalam suatu selang tertentu. Sebagai contohnya, jika catatan medis menunjukkan bahwa pada usia 45 tahun pasien kanker dalam contoh di atas kondisinya sehat dan belum berpenyakit kanker, kemudian pasien melakukan tes pertama saat berumur 50 tahun dan terdiagnosis terkena penyakit kanker, dengan demikian usia saat didiagnosis positif kanker adalah antara 45 dan 50 tahun.

B. Dasar Teori Analisis Survival

Untuk T suatu variabel acak positif dan menunjukkan waktu survival setiap subjek, maka nilai-nilai yang mungkin untuk T yaitu $T \geq 0$. Menurut Lee & Wang (2003: 8), distribusi dari T dapat dinyatakan dalam tiga cara yaitu sebagai berikut:

1. Fungsi Kepadatan Peluang

Fungsi kepadatan peluang atau *PDF(Probability Density Function)* adalah peluang suatu individu mati atau mengalami kejadian sesaat dalam interval waktu t sampai $t + \Delta t$. Fungsi kepadatan peluang $f(t)$ dirumuskan sebagai berikut (Lee & Wang, 2003: 10),

$$f(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{P(t < T < (t + \Delta t))}{\Delta t} \right] = \lim_{\Delta t \rightarrow 0} \left[\frac{F(t + \Delta t) - F(t)}{\Delta t} \right] \quad (2.1)$$

Jika T merupakan variabel acak positif pada interval $[0, \infty)$, maka $F(t)$ merupakan fungsi distribusi kumulatif kontinu dari T . Didefinisikan sebagai peluang suatu individu mengalami kejadian kurang dari sama dengan waktu t , yaitu,

$$F(t) = P(T \leq t) = \int_0^t f(x) dx \quad (2.2)$$

Berdasarkan Persamaan (2.2) diperoleh:

$$f(t) = \frac{d(F(t))}{dt} = F'(t) \quad (2.3)$$

2. Fungsi Survival

Fungsi survival $S(t)$ didefinisikan sebagai peluang suatu individu dapat bertahan hidup dengan waktu survival sampai dengan waktu $t (t > 0)$, yaitu sebagai berikut:

$$S(t) = P(T \geq t) \quad (2.4)$$

Sesuai dengan definisi fungsi distribusi kumulatif $F(t)$ dari T , fungsi survival dapat dinyatakan dengan,

$$S(t) = 1 - P(T \leq t)$$

$$= 1 - F(t)$$

Fungsi survival juga dapat dinyatakan dalam bentuk fungsi kepadatan peluang yaitu,

$$S(t) = P(T \geq t) = \int_t^{\infty} f(t) dt$$

Diperoleh hubungan antara fungsi kepadatan peluang, fungsi distribusi kumulatif dari T, dan fungsi survival yaitu,

$$\begin{aligned} F(t) &= 1 - S(t) \\ &= \frac{d(F(t))}{dt} = \frac{d(1 - S(t))}{dt} \\ &= F'(t) = -S'(t) \\ &= f(t) = -S'(t) \end{aligned} \tag{2.5}$$

diperoleh,

$$f(t) = F'(t) = -S'(t) \tag{2.6}$$

3. Fungsi Hazard

Fungsi *hazard* $h(t)$ didefinisikan sebagai kelajuan suatu individu mengalami kejadian dalam interval waktu dari t sampai $t + \Delta t$ dengan syarat individu tersebut masih bertahan hidup sampai dengan waktu t , dapat dinyatakan dengan persamaan berikut:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \tag{2.7}$$

berdasarkan teori peluang, bahwa peluang kejadian A dengan syarat kejadian B yaitu:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2.8)$$

T merupakan variabel acak, dari Persamaan (2.7) diperoleh:

$$\begin{aligned}
h(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\
&= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, T \geq t)}{P(T \geq t) \cdot \Delta t} \\
&= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{S(t) \cdot \Delta t} \\
&= \frac{1}{S(t)} \cdot \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \\
&= \frac{1}{S(t)} \cdot \lim_{\Delta t \rightarrow 0} \frac{P(T < t + \Delta t) - P(T < t)}{\Delta t} \\
&= \frac{1}{S(t)} \cdot \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \\
&= h(t) = \frac{f(t)}{S(t)} \quad (2.9)
\end{aligned}$$

Persamaan (2.5) disubstitusikan ke Persamaan (2.9), diperoleh:

$$\begin{aligned}
h(t) &= \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)} \\
&= -S'(t) \frac{1}{S(t)} \\
&= -S'(t) \log S(t) \\
&= -S'(t) \frac{d \log S(t)}{dS(t)} \\
&= -\frac{dS(t)}{dt} \cdot \frac{d \log S(t)}{dS(t)} \\
&= -\frac{d}{dt} \log S(t)
\end{aligned}$$

diperoleh persamaan,

$$h(x)dx = \frac{d \log S(x)}{dx} dx$$

diintegrasikan menjadi,

$$\begin{aligned} \int_0^t h(x)dx &= - \int_0^t \frac{d \log S(x)}{dx} dx \\ &= - \int_0^t h(x)dx = \int_0^t \frac{d \log S(x)}{dx} dx \\ &= - \int_0^t h(x)dx = \log S(x) \Big|_0^t \\ &= - \int_0^t h(x)dx = \log S(t) - \log S(0) \end{aligned}$$

Diketahui $S(0)=1$ dan $\log S(0)=0$, oleh karena itu diperoleh:

$$\begin{aligned} - \int_0^t h(x)dx &= \log S(t) \\ S(t) &= \exp \left[- \int_0^t h(x)dx \right] \end{aligned} \tag{2.10}$$

Berdasarkan fungsi *hazard* yang diperoleh dari Persamaan (2.10), menurut Lee & Wang (2003: 16) fungsi kumulatif *hazard*($H(t)$) adalah:

$$H(t) = \int_0^t h(x)dx \tag{2.11}$$

dan didapatkan hubungan dengan fungsi survival, yaitu,

$$S(t) = \exp[-H(t)] \tag{2.12}$$

C. Model Cox Proportional Hazard

Salah satu tujuan model *Cox Proportional Hazard* adalah untuk memodelkan hubungan antara waktu survival dengan variabel-variabel yang diduga mempengaruhi waktu survival. Model *Cox Proportional Hazard* memiliki asumsi bahwa fungsi *hazard* dari individu yang berbeda adalah proporsional, atau rasio fungsi *hazard* dari dua individu yang berlainan adalah konstan (Lee & Wang, 2003).

Risiko kematian individu pada waktu tertentu bergantung pada nilai x_1, x_2, \dots, x_p dari p variabel bebas X_1, X_2, \dots, X_p . Himpunan nilai variabel bebas pada model Cox dipresentasikan oleh x , sehingga $x = (x_1, x_2, \dots, x_p)$. Model Cox dapat dituliskan sebagai berikut (Kleinbaum & Klein, 2005: 98).

$$h(t, x) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad (2.13)$$

dengan,

$h_0(t)$: fungsi dasar *hazard*,

$\beta_1, \beta_2, \dots, \beta_p$: parameter regresi,

x_1, x_2, \dots, x_p : nilai dari variabel bebas X_1, X_2, \dots, X_p

Analisis survival mengenal dua model, yaitu parametrik dan semiparametrik. Model parametrik antara lain model *Weibull* yang berdistribusi *weibull* dan model *Gamma* yang berdistribusi *gamma* (Kleinbaum & Klein, 2005: 357). Model *Cox Proportional Hazard* merupakan model berdistribusi semiparametrik karena model Cox tidak memerlukan informasi tentang distribusi yang mendasari waktu survival dan parameter regresi dapat diestimasi dari model Cox tanpa harus menentukan fungsi *hazard* dasar (Lee & Wang, 2003: 298).

Kenyataannya, data yang diperoleh tidak dapat memberikan informasi distribusi waktu survival, sehingga bentuk $h_0(t)$ dari fungsi *hazard* dasar juga tidak dapat diketahui.

Model semiparametrik lebih sering digunakan karena walaupun bentuk fungsional $h_0(t)$ tidak diketahui, tapi model *Cox Proportional Hazard* ini tetap dapat memberikan informasi berupa *hazard ratio (HR)* yang tidak bergantung pada $h_0(t)$. *Hazard ratio* didefinisikan sebagai rasio dari *hazard rate* satu individu dengan *hazard rate* dari individu lain, hal ini ditunjukkan sebagai berikut.

Misalnya individu A memiliki fungsi *hazard* dasar $h_A(t, X^*)$ dengan $X^* = (X_1^*, X_2^*, \dots, X_m^*)$ dan individu B memiliki fungsi *hazard* dasar $h_B(t, X)$, dengan $X = (X_1, X_2, \dots, X_m)$, maka diperoleh *hazard ratio* sebagai berikut:

$$\begin{aligned}
 HR &= \frac{h_A(t, X^*)}{h_B(t, X^*)} = \frac{h_0(t) \exp[\sum_{p=1}^m \beta_p X_p^*]}{h_0(t) \exp[\sum_{p=1}^m \beta_p X_p]} \\
 &= \exp\left(\sum_{p=1}^m \beta_p X_p^* - \sum_{p=1}^m \beta_p X_p\right) \\
 &= \exp\left[\sum_{p=1}^m \beta_p (X_p^* - X_p)\right] \tag{2.14}
 \end{aligned}$$

Persamaan (2.13) dapat dituliskan dalam bentuk sebagai berikut:

$$\begin{aligned}
 \ln h(t, x) &= \ln([h_0(t)] \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)) \\
 \ln \frac{h(t, x)}{h_0(t)} &= \log(\exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)) \\
 \ln \frac{h(t, x)}{h_0(t)} &= \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \tag{2.15}
 \end{aligned}$$

Pada model Cox, untuk mengestimasi parameter regresi $(\beta_1, \beta_2, \dots, \beta_p)$ dapat dilakukan tanpa mengestimasi fungsi *hazard* dasar. Persamaan (2.15) merupakan model *hazard ratio* yang didefinisikan sebagai *hazard* (kegagalan) dari satu individu dibagi dengan *hazard* individu yang berbeda, dapat dinyatakan sebagai berikut (Kleinbaum & Klein, 2005: 24).

$$\log[HR(x)] = (\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad (2.16)$$

Saat variabel bebas dengan rasio *hazard* kurang dari 1, peningkatan nilai variabel bebas berhubungan dengan menurunnya risiko kematian dan lebih panjangnya waktu bertahan hidup. Ketika rasio *hazard* lebih besar dari 1, peningkatan nilai variabel bebas berhubungan dengan peningkatan risiko kematian dan lebih pendeknya waktu bertahan hidup.

1. Estimasi parameter model *Cox Proportional Hazard*

Estimasi parameter β_j dengan $j= 1, 2, \dots, p$ pada model *Cox Proportional Hazard* dapat dilakukan salah satunya dengan menggunakan metode *Maximum Partial Likelihood Estimation (MPLE)*. Terlebih dahulu akan dijelaskan tentang penyusunan fungsi *partial likelihood*. Fungsi *likelihood* adalah fungsi dari parameter-parameter β yang tidak diketahui nilainya yang menggambarkan peluang bersama dari observasi data (Kleinbaum & Klein, 2005:98).

Fungsi *likelihood* yang digunakan dalam model Cox adalah fungsi *partial likelihood*. Fungsi *partial likelihood* memperhatikan peluang untuk subjek yang mengalami kejadian dan urutan kejadian. Misalkan terdapat k waktu kejadian dan

pada saat t_j terjadi kejadian dengan $j = 1, 2, \dots, k$ dan kejadian yang diperhatikan adalah meninggal, maka fungsi *partial likelihood*-nya adalah L_j dengan $R(t_j)$ adalah himpunan individu yang berisiko pada waktu t_j yang terdiri dari semua individu yang bertahan hidup hingga t_j . Fungsi *partial likelihood* dinyatakan sebagai berikut:

$$L(\beta) = \prod_{j=1}^k L(t_j, \beta) \quad (2.17)$$

Notasi (t_j, β) menunjukkan *likelihood* untuk kejadian saat t_j dengan himpunan risiko $R(t_j)$.

Untuk menggambarkan fungsi *partial likelihood*, diberikan contoh berikut. Akan dilakukan penelitian dengan subjek yang akan diteliti adalah sebanyak 4 subjek yang berpenyakit kanker, variabel yang berpengaruh adalah Variabel Umur (x_1) dan Variabel Terapi (x_2). Subjeknya adalah Berry, Meggy, Rossy, dan Fenny, data diperoleh dari keempat subjek seperti tercantum pada tabel 2.1 berikut:

Tabel 2. 1 Data Subjek Penelitian Penyakit Kanker

Subjek(i)	Waktu Survival		t	Status	x_1	x_2
	Mulai	Berhenti				
Berry	0	9	9	1	1	1
Meggy	9	21	12	1	0	1
Rossy	21	34	13	0	0	0

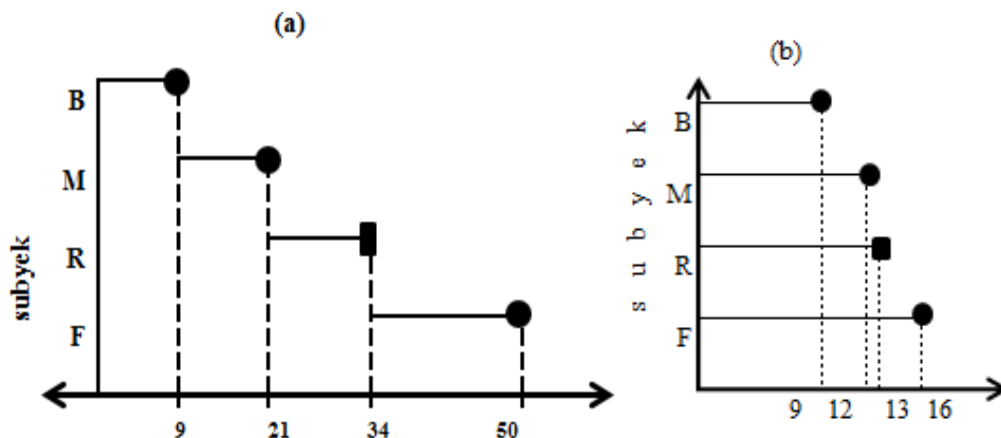
Fenny	34	50	16	1	1	0
-------	----	----	----	---	---	---

Pada Tabel 2.1 di atas, kolom i adalah kolom untuk individu, t adalah waktu survival (dalam bulan). Pada kolom status, angka 1 menunjukkan subjek mengalami kejadian dan 0 untuk subjek yang tersensor, x_1 dan x_2 adalah Variabel Umur dan terapi yang dikategorikan sebagai berikut:

$$\text{kategori umur}(x_1): \begin{cases} 0, x_1 \leq 25 \text{ tahun} \\ 1, x_1 > 25 \text{ tahun} \end{cases}$$

$$\text{kategori terapi}(x_2): \begin{cases} 1, \text{mendapat terapi} \\ 0, \text{tidak mendapat terapi} \end{cases}$$

Tabel 2.1 di atas dapat diilustrasikan dengan gambar 2.2 sebagai berikut:



keterangan:

- : subyek tersensor
- : subyek meninggal pada saat penelitian
- : subyek berada dalam penelitian

Gambar 2. 2Calendar Time (a) dan Survival Time (b)

Pada data dapat diketahui bahwa Berry mengalami kejadian saat $t=9$, Menggy mengalami kejadian pada waktu $t=12$, karena Rossy menghilang saat

penelitian di $t=13$, maka Rossy tersensor pada waktu $t=13$ dan Fenny mengalami kejadian pada waktu $t=16$. Meggy dan Fenny berusia kurang dari sama dengan 25 tahun, sedang umur Berry dan Rossy lebih dari 25 tahun. Berry dan Meggy melakukan terapi, sedangkan Rossy dan Fenny tidak melakukan terapi. Sesuai dengan Persamaan (2.13), maka persamaan model *Cox Proportional Hazard* pada kasus di atas adalah sebagai berikut:

$$h(t) = h_0(t) \exp(\beta_1 \text{umur} + \beta_2 \text{terapi}) \quad (2.18)$$

Pada Tabel 2.1 terlihat bahwa Berry mengalami kekambuhan pada bulan ke 9. Pada bulan ke-9 tersebut semua subjek berisiko. Meggy mengalami kekambuhan pada bulan ke 12, pada waktu ini yang masuk dalam himpunan risiko adalah Meggy, Rossy, dan Fenny, dengan begitu pada L_2 penyebutnya adalah penjumlahan dari tiga fungsi *hazard* yang masuk pada risiko $t=12$. Rossy hilang saat penelitian berlangsung, sehingga Rossy dikatakan tersensor jadi Rossy tidak berisiko dan Fenny mengalami kekambuhan saat $t=16$ sehingga masuk ke risiko pada waktu $t=16$. Dapat dituliskan fungsi *hazard* pada masing-masing individu adalah,

Tabel 2. 2 Fungsi Hazard

ID	Fungsi <i>hazard</i>
Berry	$h_0(t) \exp(\beta_1 + \beta_2)$
Meggy	$h_0(t) \exp(\beta_2)$
Rossy	$h_0(t) \exp(0)$

Fenny	$h_0(t) \exp(\beta_1)$
-------	------------------------

Berdasarkan informasi tersebut, sesuai dengan Persamaan (2.13)

fungsi *partial likelihood* kasus di atas adalah sebagai berikut:

$$L_1 = \left[\frac{h_0(t) \exp(\beta_1 + \beta_2)}{h_0(t) \exp(\beta_1 + \beta_2) + h_0(t) \exp(\beta_2) + h_0(t) \exp(0) + h_0(t) \exp(\beta_1)} \right]$$

$$L_2 = \left[\frac{h_0(t) \exp(\beta_2 x_2)}{h_0(t) \exp(\beta_2 x_2) + h_0(t) \exp(\beta_1 x_1) + h_0(t) \exp(0)} \right]$$

$$L_3 = \left[\frac{h_0(t) \exp(\beta_1 x_1)}{h_0(t) \exp(\beta_1 x_1)} \right]$$

Fungsi *partial likelihood* dapat dikatakan sebagai perkalian dari *likelihood-likelihood* yang kejadiannya teramati, sehingga *partial likelihood* dari contoh di atas adalah sebagai berikut:

$$L = L_1 \times L_2 \times L_3$$

$$L = \left[\frac{h_0(t) \exp(\beta_1 + \beta_2)}{h_0(t) \exp(\beta_1 + \beta_2) + h_0(t) \exp(\beta_2) + h_0(t) \exp(0) + h_0(t) \exp(\beta_1)} \right] \\ \times \left[\frac{h_0(t) \exp(\beta_2 x_2)}{h_0(t) \exp(\beta_2 x_2) + h_0(t) \exp(\beta_1 x_1) + h_0(t) \exp(0)} \right] \\ \times \left[\frac{h_0(t) \exp(\beta_1 x_1)}{h_0(t) \exp(\beta_1 x_1)} \right] \quad (2.19)$$

Pada *Cox Proportional Hazard*, fungsi *baseline hazard* ($h_0(t)$) dapat dikeluarkan dari model, sehingga *partial likelihood*-nya adalah:

$$L = \left[\frac{\exp(\beta_1 + \beta_2)}{\exp(\beta_1 + \beta_2) + \exp(\beta_2) + \exp(0) + \exp(\beta_1)} \right] \\ \times \left[\frac{\exp(\beta_2 x_2)}{\exp(\beta_2) + \exp(\beta_1) + \exp(0)} \right] \times \left[\frac{\exp(\beta_1)}{\exp(\beta_1)} \right] \quad (2.20)$$

Misalkan terdapat data untuk n individu dan masing-masing mempunyai vektor kovariat $X_i = [X_{i1}, \dots, X_{ip}]'$. Dari n individu tersebut, misalkan terdapat r individu yang mengalami kejadian, maka terdapat $n - r$ individu yang tersensor. Jika diurutkan, urutannya menjadi $t_1 < t_2 < \dots < t_i < \dots < t_r$, dengan t_i merupakan urutan kejadian ke- i . Diasumsikan hanya terdapat satu individu yang mengalami kematian pada tiap waktu kegagalan. Apabila vektor variabel bebas dari individu yang mati pada waktu t_i dinotasikan dengan x_i , maka peluangnya menjadi sebagai berikut:

$$P(\text{individu dengan variabel } x_i \text{ mati saat } t_i | \text{satu kematian saat } t_i). \quad (2.21)$$

Jika kejadian A adalah individu dengan variabel x_i mati saat t_i dan kejadian B adalah satu kematian saat t_i , maka peluang Persamaan (2.21) menjadi sebagai berikut:

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(\text{individu dengan variabel } x_i \text{ mati saat } t_i)}{P(\text{satu kematian saat } t_i)} \end{aligned} \quad (2.22)$$

Diasumsikan bahwa waktu kematian independen satu sama lain, maka penyebut Persamaan (2.22) diatas merupakan penjumlahan dari peluang kematian semua individu yang beresiko mati pada waktu t_i . Apabila individu-individu tersebut dinotasikan dengan l dan $R(t_i)$ merupakan himpunan dari individu-individu yang beresiko pada waktu t_i , maka Persamaan (2.22) menjadi sebagai berikut:

$$\frac{P(\text{individu dengan variabel } x_i \text{ mati saat } t_i)}{\sum_{l \in R(t_i)} P(\text{individu } l \text{ mati saat } t_i)}. \quad (2.23)$$

Peluang individu yang mati saat t_i dapat diganti dengan peluang individu yang mati pada interval $(t_i, t_i + \Delta t)$, dan kemudian baik penyebut maupun pembilang pada persamaan (2.23) tersebut dibagi dengan Δt , sehingga didapatkan persamaan sebagai berikut:

$$\frac{P[\text{individu dengan variabel } x_i \text{ mati pada } (t_i, t_i + \Delta t)]/\Delta t}{\sum_{l \in R(t_i)} P[\text{individu } l \text{ mati pada } (t_i, t_i + \Delta t)]/\Delta t} \quad (2.24)$$

Nilai limit dari Persamaan (2.24) dimana $\Delta t \rightarrow 0$ merupakan rasio peluang dari Persamaan (2.23), sehingga,

$$\frac{\text{Fungsi } hazard \text{ untuk individu dengan variabel } x_i \text{ yang mati pada } t_i}{\sum_{l \in R(t_i)} [\text{Fungsi } hazard \text{ untuk individu } l \text{ yang mati pada } t_i]}.$$

Apabila individu ke- j mati pada saat t_i , fungsi *hazard* pada pembilang persamaan tersebut dapat ditulis $h_j(t_i)$, sedangkan penyebut persamaan tersebut merupakan penjumlahan dari fungsi *hazard* untuk individu yang beresiko pada waktu t_i yang dapat ditulis $h_l(t_i)$. Peluang bersyarat untuk Persamaan (2.22) adalah sebagai berikut:

$$\frac{h_j(t_i)}{\sum_{l \in R(t_i)} h_l(t_i)}.$$

Pada *Cox likelihood*, fungsi *baseline hazard* dapat dikeluarkan dari model, maka persamaan diatas menjadi sebagai berikut:

$$\frac{\exp(\sum_{j=1}^k \beta_j X_{j(i)})}{\sum_{l \in R(t_i)} \exp(\sum_{j=1}^k \beta_j X_{jl})}$$

Setiap kegagalan (*failure*) menyumbang sebuah faktor, oleh karena itu fungsi *partial likelihood*-nya adalah sebagai berikut:

$$L(\beta) = \prod_{i=1}^r \frac{\exp(\sum_{j=1}^k \beta_j X_{j(i)})}{\sum_{l \in R(t_i)} \exp(\sum_{j=1}^k \beta_j X_{jl})} \quad (2.25)$$

diperoleh fungsi *log partial likelihood* yaitu sebagai berikut:

$$\ln L(\beta) = \ln \prod_{i=1}^r \frac{\exp(\sum_{j=1}^p \beta_j x_{j(i)})}{\sum_{l \in R(t_i)} \exp(\sum_{j=1}^p \beta_j x_{jl})} \quad (2.26)$$

$$\begin{aligned} &= \sum_{i=1}^k \left[\ln \left(\exp \left(\sum_{j=1}^p \beta_j x_{j(i)} \right) \right) - \ln \left(\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^p \beta_j x_{jl} \right) \right) \right] \\ &= \sum_{i=1}^k \left[\left(\sum_{j=1}^p \beta_j x_{j(i)} \right) - \left(\ln \sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^p \beta_j x_{jl} \right) \right) \right] \end{aligned} \quad (2.27)$$

Turunan pertama dari $\ln L(\beta)$ terhadap β_j yaitu,

$$\begin{aligned} \frac{\partial \ln L(\beta)}{\partial \beta_j} &= \frac{\partial \left(\sum_{i=1}^k \left[\left(\sum_{j=1}^p \beta_j x_{j(i)} \right) - \ln \left(\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^p \beta_j x_{jl} \right) \right) \right] \right)}{\partial \beta_j} \\ &\Leftrightarrow \frac{\partial \ln L(\beta)}{\partial \beta_j} = \sum_{i=1}^k \left[\sum_{j=1}^p x_{j(i)} - \frac{\sum_{l \in R(t_i)} x_{jl} \exp \left(\sum_{j=1}^p \beta_j x_{jl} \right)}{\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^p \beta_j x_{jl} \right)} \right] \end{aligned} \quad (2.28)$$

Pendugaan β_j dapat diperoleh dengan memaksimumkan turunan pertama fungsi *log partial likelihood* yaitu dengan menyelesaikan persamaan berikut:

$$\frac{\partial \ln L(\beta)}{\partial \beta_j} = 0$$

diperoleh,

$$\sum_{i=1}^k \left[\sum_{j=1}^p x_{j(i)} - \frac{\sum_{l \in R(t_i)} x_{jl} \exp\left(\sum_{j=1}^p \beta_j x_{jl}\right)}{\sum_{l \in R(t_i)} \exp\left(\sum_{j=1}^p \beta_j x_{jl}\right)} \right] = 0 \quad (2.29)$$

Persamaan (2.29) tersebut dapat diselesaikan secara numerik yaitu menggunakan metode Newton Raphson. Selanjutnya akan dicari turunan kedua dari $\ln L(\beta)$, yaitu,

$$\begin{aligned} \frac{\partial^2 \ln L(\beta)}{\partial^2 \beta_j} &= \frac{\partial}{\partial \beta_j} \left(\frac{\partial \ln L(\beta)}{\partial \beta_j} \right) \\ &= \frac{\partial}{\partial \beta_j} \left[\sum_{i=1}^k \left[\sum_{j=1}^p x_{j(i)} - \frac{\sum_{l \in R(t_i)} x_{jl} \exp\left(\sum_{j=1}^p \beta_j x_{jl}\right) \sum_{j=1}^p x_{jl}}{\sum_{l \in R(t_i)} \exp\left(\sum_{j=1}^p \beta_j x_{jl}\right)} \right] \right] \\ &= \sum_{i=1}^k \left[\frac{\left(\sum_{l \in R(t_i)} \exp\left(\sum_{j=1}^p \beta_j x_{jl}\right) \sum_{j=1}^p x_{jl} \right)^2}{\left(\sum_{l \in R(t_i)} \exp\left(\sum_{j=1}^p \beta_j x_{jl}\right) \right)^2} \right. \\ &\quad \left. - \frac{\sum_{l \in R(t_j)} \exp\left(\sum_{j=1}^p \beta_j x_{jl}\right) \left(\sum_{j=1}^p x_{jl}\right)^2}{\sum_{l \in R(t_i)} \exp\left(\sum_{j=1}^p \beta_j x_{jl}\right)} \right] \\ &= - \sum_{i=1}^k \left[\frac{\sum_{l \in R(t_i)} \exp\left(\sum_{j=1}^p \beta_j x_{jl}\right) \left(\sum_{j=1}^p x_{jl}\right)^2}{\sum_{l \in R(t_i)} \exp\left(\sum_{j=1}^p \beta_j x_{jl}\right)} \right. \\ &\quad \left. - \frac{\left(\sum_{l \in R(t_i)} \left(\sum_{j=1}^p \beta_j x_{jl}\right) \sum_{j=1}^p x_{jl} \right)^2}{\sum_{l \in R(t_i)} \exp\left(\sum_{j=1}^p \beta_j x_{jl}\right)} \right] \end{aligned} \quad (2.30)$$

Nilai negatif turunan kedua dari *log partial likelihood* yaitu sebagai berikut:

$$-\frac{\partial^2 \ln L(\beta)}{\partial^2 \beta_j}$$

$$= - \left[- \sum_{i=1}^k \left[\frac{\sum_{l \in R(t_i)} \exp(\sum_{j=1}^p \beta_j x_{jl}) (\sum_{j=1}^p x_{jl})^2}{\sum_{l \in R(t_i)} \exp(\sum_{j=1}^p \beta_j x_{jl})} - \frac{(\sum_{l \in R(t_i)} (\sum_{j=1}^p \beta_j x_{jl}) \sum_{j=1}^p x_{jl})^2}{\sum_{l \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{jl})} \right] \right] \quad (2.31)$$

1. Prosedur Newton Raphson untuk Penaksiran Parameter Model Cox Proportional Hazard

Prosedur Newton Raphson berperan dalam memaksimalkan fungsi *partial likelihood* dalam penaksiran parameter model Cox. Untuk memaksimalkan fungsi *partial likelihood* dalam penaksiran parameter model Cox Proportional Hazard dapat menggunakan prosedur Newton Raphson. Misalkan $L(\beta)$ merupakan fungsi *partial likelihood* dimensi vektor $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$. Misalkan $\mathbf{U}(\beta)$ merupakan vektor berukuran p dari turunan parsial pertama $\ln L_p(\beta)$ seperti Persamaan (2.28) berikut:

$$\mathbf{U}(\beta) = [U_1(\beta), U_2(\beta), \dots, U_p(\beta)]^T \quad (2.32)$$

$$\text{dengan } U_j(\beta) = \frac{\partial \ln L_p(\beta)}{\partial \beta_j}, j = 1, 2, \dots, p$$

Misalkan $\mathbf{I}(\beta)$ adalah matriks Hessian berukuran $p \times p$ dari turunan *partial* kedua $\ln L_p(\beta)$ yaitu,

$$I(\beta) = (I_{ij}(\beta)) \text{ dengan } i, j = 1, 2, \dots, p \quad (2.33)$$

$$\text{dengan } I_{ij}(\beta) = \frac{\partial^2 \ln L(\beta)}{\partial \beta_i \partial \beta_j}$$

$$\mathbf{I}(\boldsymbol{\beta}) = \begin{bmatrix} \frac{\partial^2 \ln L(\boldsymbol{\beta})}{(\partial \beta_1)^2} & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_1 \beta_2} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_1 \beta_p} \\ \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_2 \beta_1} & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_2 \beta_2} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_2 \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_p \beta_1} & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial \beta_p \beta_2} & \cdots & \frac{\partial^2 \ln L(\boldsymbol{\beta})}{(\partial \beta_p)^2} \end{bmatrix}$$

Algoritma pada metode Newton Raphson yaitu sebagai berikut:

$$\hat{\boldsymbol{\beta}}_{c+1} = \hat{\boldsymbol{\beta}}_c - \mathbf{I}(\hat{\boldsymbol{\beta}}_c)^{-1} \mathbf{U}(\hat{\boldsymbol{\beta}}_c) \quad (2.34)$$

dengan $c=0,1,2,\dots$ dan $\mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}_c)$ merupakan invers dari $\mathbf{I}(\hat{\boldsymbol{\beta}}_c)$.

Langkah iterasi dengan metode Newton Raphson sebagai berikut:

1. Menetapkan nilai awal $\hat{\boldsymbol{\beta}}_0 = \mathbf{0}$
2. Menghitung $\hat{\boldsymbol{\beta}}_1 = \hat{\boldsymbol{\beta}}_0 - \mathbf{I}(\hat{\boldsymbol{\beta}}_0)^{-1} \mathbf{U}(\hat{\boldsymbol{\beta}}_0)$
3. Iterasi dilakukan sampai memperoleh nilai yang konvergen : $\hat{\boldsymbol{\beta}}_{c+1} \cong \hat{\boldsymbol{\beta}}_c$

Varians dari $\hat{\beta}_j$ yaitu sebagai berikut (Hosmer Lemeshow,1999:72):

$$V\hat{a}r(\hat{\boldsymbol{\beta}}) = \mathbf{I}(\hat{\boldsymbol{\beta}})^{-1} \quad (2.35)$$

Pada estimasi standar deviasi, dinotasikan dengan $S\hat{E}(\hat{\boldsymbol{\beta}})$ yaitu akar kuadrat positif dari varian $\hat{\beta}_j$ sebagai berikut (Hosmer Lemeshow,1999:72):

$$S\hat{E}(\hat{\boldsymbol{\beta}}) = \sqrt{V\hat{a}r(\hat{\boldsymbol{\beta}})} = \mathbf{I}(\hat{\boldsymbol{\beta}})^{-1} \quad (2.36)$$

Standar deviasi dapat digunakan untuk mencari selang kepercayaan $\hat{\boldsymbol{\beta}}_j$ yaitu $((1 - \alpha)100\%)$ selang kepercayaan untuk $\hat{\boldsymbol{\beta}}_j$ ((Hosmer, Lemeshow, & May, 2008: 72) sebagai berikut:

$$\hat{\boldsymbol{\beta}}_j \pm z_{1-\frac{\alpha}{2}} \mathbf{SE}(\hat{\boldsymbol{\beta}}) \quad (2.37)$$

2. Pengujian Parameter

Menurut David W, Hosmer dan Standley Lemeshow (2008), terdapat tiga cara untuk menguji signifikansi parameter yaitu dengan uji *partial likelihoodratio*, uji *Wald*, dan uji *Score*. Pengujian signifikansi parameter bertujuan untuk memeriksa apakah variabel bebas memiliki pengaruh dalam model.

a. Uji *Partial likelihood Ratio*

Uji *partial likelihood ratio* digunakan untuk menguji hipotesis bahwa satu atau beberapa parameter regresi β_j adalah nol. Berikut langkah-langkah uji *partial likelihoodratio*:

1. Hipotesis

$$H_0: \forall \beta_j = 0, \text{ dengan } j = 1, 2, \dots, p$$

$$H_1: \exists \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, p$$

2. Taraf signifikansi : α

3. Statistik uji:

$$G = -2[\ln L(0) - \ln L(\hat{\beta}_j)] \quad (2.38)$$

dengan,

$L(0) = \log$ *partial likelihood* model tanpa variabel bebas (model *null*)

$L(\hat{\beta}_j) = \log$ *partial likelihood* dari model yang terdiri dari p variabel bebas

4. Kriteria keputusan

H_0 ditolak jika $G \geq \chi^2_{(\alpha; db=p)}$ atau $p\text{-value} \leq \alpha$, dengan p adalah banyaknya variabel bebas.

5. Perhitungan

Perhitungan menggunakan bantuan *software* R 3.1.2, didapatkan nilai *log partial likelihood* dan nilai *p-value* dari uji *partial likelihood ratio*.

6. Kesimpulan

Jika H_0 ditolak, maka $\beta_j \neq 0$, mengindikasikan bahwa variabel bebas berpengaruh terhadap waktu survival.

b. Uji Wald

Sama seperti uji *partial likelihood ratio*, uji *Wald* juga menggunakan distribusi *chi-square* dengan derajat bebas p , perbedaannya adalah uji *likelihood ratio* memberikan hasil yang lebih bagus dan akurat dibandingkan dengan uji *Wald* Kleinbaum dan Klein (2005: 90). Analisis pada uji *Wald* adalah sebagai berikut:

1. Hipotesis:

$$H_0: \forall \beta_j = 0, \text{ dengan } j = 1, 2, \dots, p$$

$$H_1: \exists \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, p$$

2. Taraf signifikansi : α

3. Statistik uji:

$$z^2 = \left(\frac{\hat{\beta}_j}{SE\hat{\beta}_j} \right)^2 \quad (2.39)$$

4. Kriteria keputusan

H_0 ditolak jika $z^2 \geq \chi^2_{(\alpha; db=p)}$ atau $p\text{-value} \leq \alpha$, dengan p adalah banyaknya variabel bebas.

5. Perhitungan

Perhitungan menggunakan bantuan *software* R 3.1.2, didapatkan nilai dari z^2 nilai $p\text{-value}$ dari uji *Wald*.

6. Kesimpulan

Jika H_0 ditolak maka $\beta_j \neq 0$, mengindikasikan bahwa variabel bebas berpengaruh terhadap waktu survival.

c. Uji Score

Uji lain untuk menguji signifikansi parameter yaitu uji *score*. Statistik ujinya adalah menggunakan rasio dari turunan *log partial likelihood* pada Persamaan (2.27), dengan akar kuadrat dari Persamaan (2.33) semuanya dievaluasi terhadap $\beta_j = 0$. Uji *score* sama dengan uji lainnya yaitu menggunakan distribusi *chi-square* dengan derajat bebas p (Lee & Wang, 2003: 225). Berikut langkah-langkah uji *score*:

1. Hipotesis:

$$H_0: \forall \beta_j = 0, \text{ dengan } j = 1, 2, \dots, p$$

$$H_1: \exists \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, p$$

2. Taraf signifikansi : α

3. Statistik uji:

$$z^* = \left(\frac{\partial L \beta_j / \partial \beta_j}{\sqrt{I(\beta_j)}} \Bigg|_{\beta_j = 0} \right)^2 \quad (2.40)$$

4. Kriteria keputusan

H_0 ditolak jika $z^* \geq \chi^2_{(\alpha; db=p)}$ atau $p\text{-value} \leq \alpha$, dengan p adalah banyaknya variabel bebas.

5. Perhitungan

Perhitungan menggunakan bantuan *software* R 3.1.2, didapatkan nilai dari z^* nilai $p\text{-value}$ dari uji *Score*.

6. Kesimpulan

Jika H_0 ditolak, maka $\beta_j \neq 0$, mengindikasikan bahwa variabel bebas berpengaruh terhadap waktu survival.

3. Pemilihan Model Cox Terbaik

Langkah pertama dalam pemilihan model Cox terbaik adalah pemilihan variabel yang masuk atau keluar dari model. Menurut Lee & Wang (2003), pemilihan variabel yang masuk atau keluar dari model dapat dilakukan dengan tiga cara yaitu seleksi *forward*, seleksi *backward* dan prosedur *stepwise*. Seleksi *forward* atau seleksi maju yaitu dengan menambahkan variabel satu demi satu dalam setiap langkahnya. Menurut David W. Hosmer dan Stanley Lemeshow (2008: 416), taraf signifikansi yang digunakan dalam seleksi *forward* disarankan antara 20%-25% untuk memungkinkan lebih banyak variabel yang masuk dalam

model. Pada masing-masing tahapan, akan diputuskan variabel mana yang merupakan prediktor terbaik untuk dimasukkan ke dalam model. Prosedur seleksi *stepwise* merupakan kombinasi dari dua proses yaitu seleksi *forward* dan seleksi *backward*.

Skripsi ini menggunakan seleksi *backward* untuk langkah pemilihan model terbaik. Prosedur Seleksi *backward* adalah dengan proses eliminasi variabel yang masuk kedalam model, dimulai dengan mengeluarkan atau menghapus satu persatu menurut kriteria signifikansi. Pemeriksaan masing-masing variabel digunakan uji *Wald*.

Langkah-langkah pengujian yang dilakukandalam prosedur seleksi *backward* adalah sebagai berikut:

1. Hipotesis:

$H_0: \forall \beta_j = 0$, dengan $j = 1, 2, \dots, p$ (variabel tidak berpengaruh terhadap model)

$H_1: \exists \beta_j \neq 0$, dengan $j = 1, 2, \dots, p$ (variabel berpengaruh terhadap model)

2. Taraf signifikansi : α

3. Statistik uji:

$$z^2 = \left(\frac{\hat{\beta}_j}{SE\hat{\beta}_j} \right)^2$$

4. Kriteria keputusan

H_0 ditolak jika $z^2 \geq \chi^2_{(\alpha; db=p)}$ atau $p\text{-value} \leq \alpha$, dengan p adalah banyaknya variabel bebas.

5. Perhitungan:

Perhitungan menggunakan bantuan *software* R 3.1.2, didapatkan nilai z^2 dan nilai *p-value* masing-masing variabel.

6. Kesimpulan:

Jika H_0 ditolak maka $\beta_j \neq 0$, yang artinya variabel bebas tersebut berpengaruh terhadap model. Sehingga, variabel tersebut tidak perlu dihapus dari model.

Setelah diperoleh variabel yang masuk, kemudian dilanjutkan uji *likelihood ratio* untuk mengetahui apakah ada interaksi antar variabel tersebut, yaitu dengan membandingkan model Cox tanpa interaksi terhadap model Cox dengan penambahan variabel interaksi. Langkah-langkah pemilihan variabel interaksi yang masuk dalam model dapat dilakukan dengan seleksi *forward*, seleksi *backward* maupun prosedur *stepwise* dengan langkah-langkah sama seperti yang telah dijelaskan dalam pemilihan variabel yang masuk kedalam model untuk mendapatkan model Cox terbaik.

D. Pengujian Asumsi Proporsional Hazard

Asumsi terpenting yang harus dipenuhi dalam regresi Cox yaitu asumsi Proporsional *Hazard* yang berarti bahwa rasio fungsi *hazard* konstan dari waktu

ke waktu atau ekuivalen dengan pernyataan bahwa rasio fungsi *hazard* suatu individu terhadap fungsi *hazard* individu yang lain adalah proporsional. Menurut Kleinbaum dan Klein (2005), jika rasio fungsi *hazard* tidak konstan dan asumsi proporsional *hazard* tidak dipenuhi, maka modelCox tidak valid. Ada 2 cara untuk mengecek asumsi *Proportional Hazard* yaitu dengan pendekatan grafik menggunakan plot *log minus log survival* dan dengan menggunakan residual *Schoenfeld* (Lee&Wang, 2003: 326-330), yaitu sebagai berikut:

1. Pengujian Asumsi Proporsional *Hazard* Dengan Grafik *Log Minus Log Survival*

Pendekatan grafik yang digunakan yaitu dengan plot *log minus log survival*. Menurut model regresi Cox, fungsi *hazard* untuk kegagalan individu ke-*i* setiap waktu *t* dapat dituliskan seperti pada Persamaan (2.13) yaitu sebagai berikut:

$$h(t, X) = h_0(t) \exp\left(\sum_{i=1}^p \beta_i X_i\right) \quad (2.41)$$

Notasi X_i dengan $i=1, 2, \dots, p$ menunjukkan nilai dari sebanyak p variabel bebas untuk individu tersebut, β merupakan parameter, dan $h_0(t)$ merupakan fungsi *hazard* dasar, apabila kedua sisi diintegrasikan dari nol hingga t , maka diperoleh sebagai berikut:

$$\int_0^t h(t, X) dt = \exp\left(\sum_{j=1}^p \beta_j X_j\right) \int_0^t h_0(t) dt \quad (2.42)$$

Menggunakan Persamaan (2.11), dapat diperoleh,

$$H(t, X) = \exp\left(\sum_{j=1}^p \beta_j X_j\right) H_0(t) \quad (2.43)$$

Selanjutnya dilakukan logaritma pada Persamaan (2.43) pada kedua sisi sebagai berikut:

$$\log H(t) = \log \exp\left(\sum_{j=1}^p \beta_j X_j\right) + \log H_0(t) \quad (2.44)$$

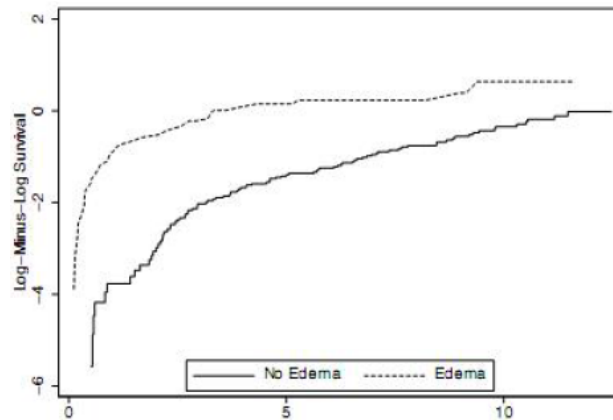
Persamaan di atas ekuivalen dengan persamaan berikut:

$$\log[-\log(S(t))] = \sum_{j=1}^p \beta_j X_j + \log[-\log(S_0(t))] \quad (2.45)$$

Persamaan (2.45) menunjukkan bahwa fungsi *log minus log survival* tidak bergantung terhadap waktu. Ini berarti bahwa fungsi *log minus log survival* pada Persamaan (2.15) berlaku jika digambarkan melawan waktu survival dan kurva akan berbentuk paralel. Pada plot *log minus log survival*, data dikelompokkan sesuai dengan tingkat atau kategori pada masing masing variabel bebas, jika variabel kontinu, maka perlu dikelompokkan menjadi variabel kategori. Jika pada plot *log minus log survival* menunjukkan kurva yang paralel, maka asumsi proporsional *hazard* tidak terpenuhi.

Kelemahan plot *log minus log survival* adalah bersifat subjektif, paralel atau tidaknya kurva sangat bergantung pada cara peneliti menilai (Collet, 2003:142). Sebagai contoh menurut Vittinghoff, Glidden, Shiboski, & McCulloch (2005: 235) misalkan pada kasus pengaruh edema pada penderita *Primary Biliary Cirrhosis (PBC)*. *PBC* adalah penyakit kerusakan saluran-saluran kecil

empedu di hati yang menyebabkan empedu menumpuk di hati. Edema merupakan penyakit dimana seseorang mengalami peningkatan volume cairan pada kaki. Terdapat pasien penderita *PBC* yang mengalami edema dan tidak mengalami edema. Berikut adalah gambar plot *log minus log survival* untuk variabel edema.



Gambar 2. 3 Plot *Log MinusLog Survival* Pada Variabel Edema

Pada Gambar 2.3 terlihat bahwa plot *log minus log survival* pada pasien dengan edema dan tidak mengalami edema mendekati paralel, sehingga mengindikasikan bahwa asumsi proporsional *hazard* pada variabel edema terpenuhi.

2. Pengujian Asumsi Proporsional *Hazard* dengan Residual *Schoenfeld*

Terdapat beberapa jenis residual antara lain residual *Cox Snell*, residual *Martingale*, residual *Deviance*, residual *Schoenfeld*, dan residual *Score* (Lee & Wang, 2003: 331). Skripsi ini juga menggunakan residual *schoenfeld* untuk mengecek asumsi proporsional *hazard*. Residual *schoenfeld* didefinisikan hanya

pada waktu survival yang tidak tersensor. Residual *schoenfeld* untuk individu ke- i pada variabel bebas ke- j adalah sebagai berikut:

$$R_{ji} = \delta_i \left(x_{ji} - \frac{\sum_{l \in R(t_{(i)})} x_{jl} \exp(\hat{\beta}' x_{jl})}{\sum_{l \in R(t_{(i)})} \exp(\hat{\beta}' x_{jl})} \right) \quad (2.46)$$

dengan $j=1, 2, \dots, p$; $i=1, 2, \dots, p$, dimana $\hat{\beta}$ adalah *MPL*E dari β .

Menurut Grambsch dan Therneau (1994) *scaled residual schoenfeld* dapat dihitung dari estimator kovarian matriks $R_i = (R_{1i}, R_{2i}, \dots, R_{pi})$ dinotasikan $\hat{V}(R_i)$, yaitu sebagai berikut:

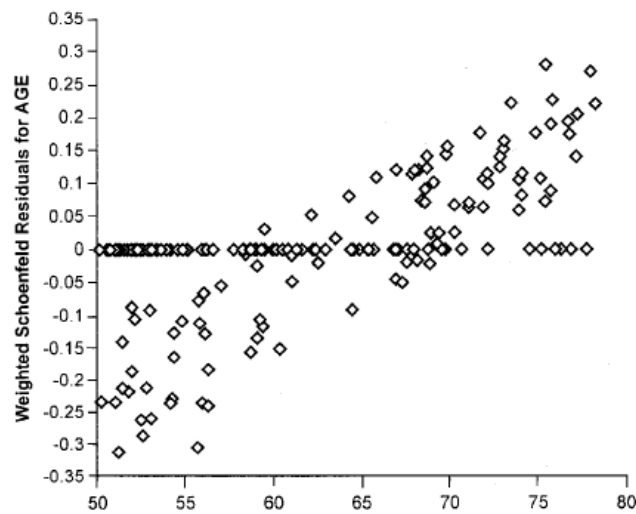
$$R_i^* = [\hat{V}(R_i)]^{-1} R_i \quad (2.47)$$

dengan $[\hat{V}(R_i)]^{-1} \simeq r \hat{V}(\hat{\beta})$, r adalah semua kejadian atau jumlah waktu survival yang tidak tersensor dan $\hat{V}(\hat{\beta})$ adalah estimasi kovarian matriks β , dapat ditulis sebagai berikut:

$$R_i^* = r \hat{V}(\hat{\beta}) R_i \quad (2.48)$$

Pada residual *schoenfeld*, apabila plot horizontal maka mengindikasikan bahwa koefisien dari X_j konstan dan asumsi proporsional *hazard* terpenuhi. Sebagai contoh pada kasus penyakit *Cardiovaskular (CVD)*. Diperoleh data pasien dari sebuah wawancara pemeriksaan pada 200 peserta dalam studi kasus penyakit *Cardiovaskular (CVD)*. Peserta berusia antara 50-79 tahun dan belum mengidap kardiovaskular pada saat pemeriksaan awal. *Body Mass Index (BMI)* pada penyakit *CVD* merupakan nilai dari salah satu sistem penilaian mengenai kondisi pasien *Cardiovaskular* seperti yang dikemukakan oleh *World*

Heart Federation. Ketika nilai *BMI* meningkat, maka risiko berpenyakit *CVD* juga meningkat. Pada gambar 2.4 terlihat bahwa residual *schoenfeld* pada variabel *BMI* memiliki kemiringan mendekati nol atau horizontal, sehingga menunjukkan bahwa asumsi proporsional *hazard* untuk variabel *BMI* terpenuhi. Berikut adalah gambar plot residual *schoenfeld* (Lee & Wang, 2003: 335).



Gambar 2. 4 Plot Residual *Schoenfeld* model *Cox Proportional Hazard* untuk Data *CVD*

E. Interpretasi Model *Cox Proportional Hazard*

Ketika model *Cox* telah terbentuk, maka langkah selanjutnya adalah melakukan interpretasi koefisien regresi. Diperlukan *hazard ratio* agar koefisien regresi dapat diinterpretasikan. Berdasarkan Persamaan (2.16), untuk variabel bebas X_0 dan X_1 dari dua individu diperoleh,

$$\frac{h(t, X_j)}{h(t, X_0)} = \frac{h_0 t \exp(\beta X_1)}{h_0 t \exp(\beta X_0)} = \frac{\exp(\beta X_1)}{\exp(\beta X_0)} = e^{(X_1 - X_0)\beta}, \forall t > 0 \quad (2.49)$$

Persamaan (2.49) menunjukkan besarnya rasio relatif dari individu dengan faktor risiko X_1 dibandingkan dengan faktor risiko X_0 dari individu lain (Lee & Wang, 2003: 299), dapat dituliskan sebagai berikut:

$$\begin{aligned}\frac{h(t, X_1)}{h(t, X_0)} &= e^{(X_1 - X_0)\beta} \\ \log\left(\frac{h(t, X_1)}{h(t, X_0)}\right) &= \log(e^{(X_1 - X_0)\beta}) \\ \log\left(\frac{h(t, X_1)}{h(t, X_0)}\right) &= (X_1 - X_0)\beta\end{aligned}\quad (2.50)$$

Nilai kovariat X_j lainnya tetap dapat diinterpretasikan seperti berikut:

$$\log\left(\frac{h(t, X_{j+1})}{h(t, X_j)}\right) = (X_{j+1} - X_j)\beta \quad (2.51)$$

Pada Persamaan(2.51), dapat disimpulkan bahwa setiap naiknya nilai β_j akan memperbesar nilai *log hazard ratio*, sehingga diperoleh,

$$\frac{h(t, X_{j+1})}{h(t, X_j)} = e^{(X_{j+1} - X_j)\beta}, \forall t > 0 \quad (2.52)$$

Dengan demikian nilai $\exp\left((X_{j+1} - X_j)\beta\right)$ merupakan *hazard ratio* yang dapat dihubungkan dengan kenaikan nilai x_j .

Jika $h(t, X_j) \approx P(t < T < t + \Delta t | T \geq t, X)$, Persamaan (2.52) dapat dituliskan sebagai berikut:

$$\frac{P(t < T < t + \Delta t | T \geq t, X_{j+1})}{P(t < T < t + \Delta t | T \geq t, X_j)} = e^{(X_{j+1} - X_j)\beta}, \forall t > 0 \quad (2.53)$$

Nilai $\exp((X_{j+1} - X_j)\beta)$ juga dapat diinterpretasikan sebagai rasio dua peluang bersyarat dari *hazard* individu yang diketahui masih hidup pada saat t .

Persamaan (2.52) ekuivalen dengan berikut ini:

$$\frac{h(t, X_{j+1})}{h(t, X_j)} - \frac{h(t, X_j)}{h(t, X_j)} = e^{(X_{j+1} - X_j)\beta} - 1, \forall t > 0$$

$$\frac{h(t, X_{j+1}) - h(t, X_j)}{h(t, X_j)} = e^{(X_{j+1} - X_j)\beta} - 1, \forall t > 0 \quad (2.54)$$

Oleh karena itu $e^{(X_{j+1} - X_j)\beta} - 1$ dapat diinterpretasikan sebagai presentasi perubahan naik atau turunnya nilai *hazard* dari setiap naiknya nilai x_j , dengan menganggap kovariat yang lain tetap.

Terdapat 3 macam ketentuan tentang bertambahnya atau berkurangnya nilai *hazard* yaitu sebagai berikut:

1. Jika $(X_{j+1} - X_j)\beta > 0$, maka kenaikan nilai x_j akan memperbesar nilai *hazard* atau semakin besar risiko seorang individu untuk meninggal atau mengalami kejadian.
2. Jika $(X_{j+1} - X_j)\beta < 0$, maka kenaikan nilai x_j akan memperkecil nilai *hazard* atau semakin kecil risiko seorang individu untuk meninggal atau mengalami kejadian.
3. Jika $(X_{j+1} - X_j)\beta = 0$, maka besar risiko seorang individu untuk hidup sama dengan risiko seorang individu untuk meninggal atau mengalami kejadian.

F. Kejadian Berulang

Pada umumnya, analisis survival hanya memperhatikan kejadian tunggal yang dialami subjek atau setiap subjek hanya mengalami satu kali kejadian, namun tidak menutup kemungkinan bahwa kejadian yang diinginkan peneliti mungkin terjadi lebih dari satu kali dalam suatu subjek. (Hosmer & Lemeshow, 2008: 307).

Kejadian berulang dalam analisis survival terdiri dari dua macam menurut Kleinbaum dan Klein (2005: 335), yaitu kejadian berulang identik dan kejadian berulang tidak identik. Kejadian berulang dikatakan identik jika urutan kejadian berulang tidak menimbulkan efek perbedaan tertentu. Sebagai contoh kasus serangan jantung, baik kasus serangan jantung yang pertama, kedua, dan selanjutnya dianggap sama dan tidak mengalami tingkat keparahan yang berbeda. Kejadian berulang dikatakan tidak identik jika ada urutan kejadian berulang atau perbedaan kategori yang lain yang menyebabkan efek perbedaan tertentu pada kejadian berulang. Misal pada kasus penyakit kanker, pada kekambuhan kanker pertama, kedua, dan selanjutnya menunjukkan tingkat keparahan yang berbeda.

Analisis yang digunakan pada kejadian berulang identik adalah model *Counting Process* yang dikembangkan oleh Anderson Gill (Andersen et al., 1993), sedangkan pada kejadian berulang tidak identik, analisis survival menggunakan pendekatan model Cox Stratifikasi salah satunya adalah model Cox Stratifikasi *PWP-Total Time* yang dikemukakan oleh Prentice, Wilian dan Peterson.