BAB II
TINJAUAN PUSTAKA

A. Analisis Survival

Analisis survival merupakan suatu analisis data dimana variabel yang diperhatikan adalah jangka waktu dari awal pengamatan sampai suatu *event* terjadi dengan melihat variabel-variabel yang mempengaruhi *event* tersebut. Di dalam analisis survival dibutuhkan beberapa faktor, yaitu:

1) Waktu awal pencatatan (*start point*) yang didefinisikan dengan baik.
2) Waktu akhir pencatatan (*end point*) yang terdefinisi dengan baik untuk mengetahui status tersensor maupun tidak tersensor suatu data.
3) Skala waktu pengukuran yang jelas. Skala diukur dalam hari, minggu atau tahun.

Analisis survival mempunyai beberapa tujuan dasar yaitu (1) memperkirakan fungsi survival dan fungsi *hazard*, (2) membandingkan antara fungsi survival dan fungsi *hazard* dan (3) melihat hubungan antara variabel-variabel terhadap waktu survival.

1. Waktu Survival

Pada analisis survival biasanya variabel waktu disebut juga sebagai waktu survival karena mengindikasikan bahwa seorang individu telah bertahan (*survived*) selama periode pengamatan. Waktu survival dapat didefinisikan pula sebagai suatu variabel yang mengukur waktu dari suatu titik awal (*start point*) tertentu sampai dengan titik akhir (*end point*) yang ditetapkan. Selain itu, suatu *event* dapat pula disebut dengan sebuah kegagalan (*failure*), untuk *event-event* yang diperhatikan semisal adalah kematian, munculnya suatu penyakit, atau peristiwa-peristiwa buruk lainnya yang menimpa suatu objek. Akan tetapi, suatu kegagalan (*failure*) tidak selamanya merupakan suatu peristiwa yang buruk, terdapat pula suatu peristiwa yang kegagalannya merupakan suatu peristiwa positif, misalnya sembuhnya seseorang dari suatu penyakit, seseorang mendapatkan suatu perkerjaan.

2. Penyensoran

Menurut Lee dan Wang (2003:2), data tersensor merupakan data yang tidak dapat diamati secara utuh dikarenakan subyek pengamatan hilang sehingga tidak dapat diambil datanya, atau sampai akhir penelitian subyek tersebut belum mengalami suatu *event* tertentu. Menurut Klein dan Kleinbaum (2005:6), terdapat tiga alasan terjadinya suatu penyensoran, yaitu:

a. Subyek pengamatan yang diamati tidak mengalami suatu *event* sampai penelitian berakhir (*loss to follow-up*).
b. Subyek pengamatan hilang selama penelitian.
c. Subyek pengamatan ditarik dari penelitian karena meninggal dimana meninggal merupakan suatu peristiwa yang tidak diperhatikan oleh peneliti atau alasan yang lain, misalnya reaksi obat yang buruk atau resiko yang lain.

Penyensoran merupakan suatu hal yang membedakan analisis survival dengan analisis statistik lainnya. Penyensoran dilakukan untuk mengatasi beberapa permasalahan dalam suatu analisis, misalnya peneliti membutuhkan waktu yang lama untuk mendapatkan data yang lengkap sampai subyek pengamatan mengalami suatu *event* yang diinginkan dan seringkali menelan biaya yang banyak.

Menurut David Collett (2004:2), terdapat tiga macam penyensoran di dalam analisis survival, yaitu:

a. Sensor kanan (*right censoring*)

   Data survival biasanya merupakan data yang tersensor kanan. Sensor kanan dapat terjadi karena beberapa alasan, yaitu: (1) subyek pengamatan belum mengalami suatu *event* sampai masa penelitian berakhir, (2) subyek pengamatan keluar pada saat masa penelitian berlangsung, (3) subyek pengamatan meninggal pada saat penelitian, akan tetapi penyebab meninggal tidak berhubungan dengan *event* yang diperhatikan. Sebagai contoh, akan dilakukan pengamatan kepada pasien-pasien penderita leukemia di suatu rumah sakit selama tujuh tahun, akan tetapi terdapat seorang individu pindah dari rumah sakit tersebut pada tahun kelima. Pasien ini masih mempunyai waktu survival selama dua tahun. Sehingga waktu pengamatan pasien tersebut dikatakan tersensor kanan.

b. Sensor kiri (*left censoring*)

Data tersensor kiri terjadi ketika subyek pengamatan tidak teramati pada awal waktu pengamatan, akan tetapi sebelum penelitan berakhir semua *event* sudah dapat diamati secara penuh. Atau dapat pula dikatakan bahwa *event* yang ingin diperhatikan pada subyek pengamatan tersebut sudah terjadi saat subyek pengamatan tersebut masuk ke dalam penelitian. Sebagai contoh, seorang dokter ingin mengetahui diagnosis usia seseorang saat terjangkit diabetes retinopati. Pada saat pemeriksaan terdapat seorang pasien berumur 55 tahun yang diketemukan terjangkit virus retinopati, akan tetapi tidak ada catatan tentang waktu tepatnya penyakit tersebut muncul, sehingga umur pasien pada saat pemeriksaan (sekitar 55 tahun) merupakan data yang tersensor kiri.

c. Sensor interval (*interval censoring*)

Sensor interval terjadi ketika suatu *event* yang diamati pada subyek pengamatan terjadi pada selang waktu tertentu. Sebagai contoh, pada saat pemeriksaan terdapat seorang pasien berumur 55 tahun yang diketemukan terjangkit virus retinopati. Apabila pada catatan medis mengindikasikan bahwa pada saat usia 50 tahun pasien belum terjangkit retinopati, maka usia pasien didiagnosis menderita diabetes retinopati antara 50 dan 55 tahun.

B. Dasar Teori Analisis Survival

Untuk T variabel random non negatif yang menunjukan waktu survival dari populasi yang homogen dan t merupakan beberapa nilai tertentu yang diperhatikan untuk variabel T. Menurut Lee dan Wang (2003: 8) terdapat tiga cara untuk menentukan distribusi dari T, yaitu fungsi kepadatan peluang (pdf), fungsi survival dan fungsi *hazard*.

1. Fungsi Kepadatan Peluang (PDF)

Fungsi kepadatan peluang merupakan peluang suatu individu mengalami *event*, gagal atau mati dalam interval waktu t sampai $(t + \Delta t)$ yang dinotasikan dengan $f(t)$. Fungsi ini dirumuskan sebagai berikut:

$$f(t) = \lim_{\Delta t \to 0} \frac{P\big(t \leq T < (t + \Delta t)\big)}{\Delta t} = \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t}. \qquad (2.1)$$

T merupakan variabel random non negatif dalam interval $[0, \infty)$, $F(t)$ merupakan fungsi distribusi kumulatif (cdf) dari T. Fungsi ini didefinisikan sebagai peluang suatu individu mengalami *event* sampai dengan waktu t yang dapat dituliskan sebagai berikut:

$$\begin{aligned} F(t) &= P(T \leq t) \\ &= \int_0^t f(x)\, dx. \end{aligned} \qquad (2.2)$$

Dari persamaan (2.2) di atas, dengan melakukan penurunan terhadap dt pada kedua sisi diperoleh:

$$F'(t) = D_t\left(\int_0^t f(x)\,dx\right) = f(t). \qquad (2.3)$$

2. Fungsi Survival

Menurut Klein dan Moeschberger (2003: 22) fungsi survival merupakan suatu kuantitas dasar yang digunakan untuk menggambarkan fenomena waktu kejadian. Fungsi survival dapat dinotasikan dengan $S(t)$, yaitu peluang suatu individu bertahan hidup lebih dari waktu t, yaitu sebagai berikut:

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

$$= \int_t^{\infty} f(x)\,dx. \qquad (2.4)$$

Dengan menggunakan definisi fungsi distribusi kumulatif $F(t) = P(T \leq t)$, fungsi survival dapat dituliskan sebagai berikut:

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

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

$$= 1 - F(t)$$

$$F(t) = 1 - S(t)$$

$$\frac{d\big(F(t)\big)}{dt} = \frac{d\big(1 - S(t)\big)}{dt}$$

$$f(t) = -\frac{d\big(S(t)\big)}{dt} = -S'(t). \qquad (2.5)$$

Menurut Klein dan Kleinbaum (2005: 9), secara teori fungsi survival dapat diplot sebagai kurva survival yang menggambarkan peluang ketahanan suatu individu pada waktu t dalam interval 0 sampai ∞. Fungsi survival mempunyai beberapa karakteristik, yaitu sebagai berikut:

a. Fungsi survival merupakan fungsi monoton tak naik.
b. Pada saat $t = 0, S(t) = S(0) = 1$.

   Pada awal dimulainya penelitian, karena belum ada individu yang mengalami *event* maka probabilitas survival pada saat $t = 0$ adalah satu.
c. Pada saat $t = \infty, S(t) \approx 0$.

   Secara teori, apabila periode penelitian meningkat tanpa batas, maka diakhir waktu tidak ada seorang individu yang akan bertahan hidup, sehingga kurva survival akan bergerak menuju nol.

3. Fungsi *Hazard*

Fungsi *hazard* dikenal juga sebagai *hazard rate* yang dinotasikan dengan $h(t)$. Fungsi ini didefinisikan sebagai kelajuan suatu individu untuk mengalami *event* pada interval waktu t sampai $(t + \Delta t)$ apabila diketahui individu tersebut belum mengalami *event* sampai dengan waktu t. Secara matematis dapat dituliskan sebagai berikut:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < (t + \Delta t) | T \geq t)}{\Delta t}. \qquad (2.6)$$

Untuk T merupakan suatu variabel acak dan (t) merupakan fungsi padat peluang dari T, dengan menggunakan teorema peluang bersyarat maka diperoleh persamaan untuk *hazard rate* sebagai berikut:

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

$$= \lim_{\Delta t \to 0} \frac{P\Big(\big(t \leq T < (t + \Delta t)\big) \cap (T \geq t)\Big)}{P(T \geq t). \Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{P\big(t \leq T < (t + \Delta t)\big)}{S(t). \Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{S(t). \Delta t}$$

$$= \frac{1}{S(t)} \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t}$$

$$= \frac{F'(t)}{S(t)}$$

$$= \frac{f(t)}{S(t)}. \tag{2.7}$$

Pada persamaan (2.5) di atas, telah diketahui bahwa

$f(t) = -\frac{d(S(t))}{dt}$, sehingga $h(t)$ dapat dinyatakan sebagai berikut:

$$h(t) = \frac{f(t)}{S(t)}$$

$$= -\frac{d\big(S(t)\big)}{dt} . \frac{1}{S(t)}$$

$$= -\frac{dS(t)}{dt} . \frac{d \ln S(t)}{dS(t)}$$

$$= -\frac{d \ln S(t)}{dt}. \tag{2.8}$$

Dari persamaan (2.8) di atas diperoleh:

$$\int_0^t h(x)dx = -\int_0^t \frac{d \ln S(x)}{dx} dx$$

$$-\int_0^t h(x)dx = \int_0^t \frac{d}{dx} \ln S(x)\, dx$$

$$= \ln S(x) \Big|_0^t$$

$$= \ln S(t) - \ln S(0).$$

Karena $S(0) = 1$, maka $\ln S(0) = \ln 1 = 0$. Oleh karena itu, persamaan di atas dapat ditulis menjadi:

$$-\int_0^t h(x)dx = \ln S(t)$$

$$\exp\left[-\int_0^t h(x)dx\right] = \exp[\ln S(t)]$$

$$S(t) = \exp\left[-\int_0^t h(x)dx\right]. \tag{2.9}$$

Pada persamaan (2.9) di atas dapat digunakan untuk mendefinisikan fungsi *hazard* kumulatif $H(t)$, yaitu sebagai berikut:

$$H(t) = \int_0^t h(x)dx. \tag{2.10}$$

Persamaan (2.9) di atas dapat ditulis sebagai berikut:

$$S(t) = \exp[-H(t)]. \qquad (2.11)$$

Menurut Klein dan Kleinbaum (2005: 12) fungsi *hazard* juga dapat diplot sebagai kurva fungsi *hazard* terhadap t seperti fungsi survival. Akan tetapi, terdapat perbedaan antara kedua fungsi tersebut. Pada fungsi *hazard*, kurva $h(t)$ tidak harus dimulai dari satu dan bergerak menuju nol, tetapi kurva $h(t)$ dapat dimulai dari nilai berapapun dengan syarat $h(t) \geq 0$ dan dapat bergerak ke atas maupun ke bawah terhadap waktu t. Fungsi *hazard* juga mempunyai karakteristik, antara lain sebagai berikut:

a. Fungsi *hazard* selalu bernilai positif, $h(t) \geq 0$.
b. Fungsi *hazard* tidak mempunyai batas atas.

C. Model Cox *Proportional Hazard*

Salah satu tujuan dari analisis survival adalah untuk menyelidiki hubungan antara waktu survival dengan variabel-variabel yang diduga mempengaruhi waktu survival. Analisis ini dapat menggunakan analasis regresi. Analisis regresi adalah analisis statistika yang memanfaatkan hubungan antara dua atau lebih peubah kuantitatif sehingga salah satu peubah dapat diramalkan dari peubah lainnya. Salah satu analisis regresi yang sering digunakan untuk menganalisis data survival adalah regresi Cox. Regresi Cox termasuk dalam metode semiparametrik, dimana di dalam metode ini tidak memerlukan informasi tentang distribusi yang mendasari waktu survival dan fungsi *baseline hazard* tidak harus ditentukan untuk mengestimasi parameternya. Selain metode semiparametrik, terdapat metode lainnya yang dapat digunakan menganalisis data survival, yaitu metode parametrik, metode nonparametrik dan metode semiparametrik. Metode parametrik mengasumsikan bahwa distribusi yang mendasari waktu survival mengikuti suatu distribusi tertentu, misalnya distribusi *Weibull*, *gamma*, eksponensial. Metode nonparametrik digunakan apabila data yang digunakan tidak mengikuti suatu distribusi tertentu yang sudah ada, yaitu metode *Kaplan-Meier* dan *Nelson-Aalen*.

Secara umum, bentuk dari model Cox adalah sebagai berikut:

$$\begin{aligned} h(t,X) &= h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k) \\ &= h_0(t)\mathrm{e}^{(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)} \end{aligned} \tag{2.12}$$

dimana:

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

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

$X_1, X_2, \dots, X_k$: variabel-variabel penjelas (kovariat)

Pada persamaan (2.12) di atas apabila semua nilai $X = 0$, maka rumus tersebut akan tereduksi menjadi fungsi *baseline hazard*. Fungsi *baseline hazard* ini merupakan fungsi awal atau dasar dari fungsi *hazard*, atau dapat dituliskan sebagai berikut:

$$h(t,0) = h_0(t). \tag{2.13}$$

Meskipun $h_0(t)$ pada model Cox tidak diketahui, namun model ini dapat digunakan dengan memanfaatkan *hazard ratio* (HR) yang tidak bergantung pada $h_0(t)$. *Hazard ratio* didefinisikan sebagai perbandingan antara fungsi *hazard* individu satu dengan fungsi *hazard* untuk individu yang lain (Klein dan Kleinbaum, 2005 :100), yaitu sebagai berikut:

Misalkan individu pertama mempunyai *hazard rate* $h_1(t, X^*)$ dimana $X^* = (X_1^*, X_2^*, \ldots, X_k^*)$ dan individu kedua mempunyai *hazard rate* $h_2(t, X)$ dimana $X = (X_1, X_2, \ldots, X_k)$, maka diperoleh bentuk *hazard ratio* sebagai berikut:

$$HR = \frac{h_1(t, X^*)}{h_2(t, X)} = \frac{h_0(t)\exp(\beta_1 X_1^* + \beta_2 X_2^* + \cdots + \beta_k X_k^*)}{h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}$$

$$= \exp[(\beta_1 X_1^* + \beta_2 X_2^* + \cdots + \beta_k X_k^*) - (\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)]$$

$$= \exp\left[\sum_{j=1}^{k} \beta_j \left(X_j^* - X_j\right)\right]$$

$$= e^{\sum_{j=1}^{k} \beta_j \left(X_j^* - X_j\right)}. \tag{2.14}$$

Nilai *hazard ratio* di atas konstan, dengan kata lain rasio *hazard rate* suatu individu dengan individu yang lain adalah *proportional*. Asumsi inilah yang menyebabkan model Cox dikenal juga sebagai model Cox *proportional hazard*.

Fungsi survival dapat dinyatakan ke dalam fungsi *hazard* sama seperti fungsi *hazard* yang dapat dinyatakan melalui fungsi survival. Berdasarkan persamaan (2.11) yang telah diperoleh di atas, fungsi survival dapat dinyatakan ke dalam bentuk sebagai berikut:

$$S(t, X) = \exp[-H(t, X)]. \tag{2.15}$$

D. Fungsi *Likelihood* Model Cox

Menurut Klein dan Kleinbaum (2005:98) fungsi *likelihood* adalah fungsi dari parameter-parameter (β) yang tidak diketahui nilainya yang menggambarkan probabilitas bersama dari data observasi. Fungsi ini dinotasikan dengan $L(\beta)$. Pada model Cox, bentuk dari Cox *likelihood* memperhatikan urutan *event* yang diamati, sehingga fungsi Cox *likelihood* disebut dengan "*partial*" *likelihood*. Istilah "*partial*" *likelihood* digunakan karena *likelihood* yang diperhatikan hanya probabilitas untuk subyek yang mengalami *event* positif saja.

Partial likelihood adalah perkalian dari *likelihood-likelihood* sebanyak n *event* yang teramati, untuk $i = 1,2, \dots, n$. Dengan *likelihood* suatu *event* merupakan probabilitas suatu objek yang mengalami *event* pada waktu t_i dan bergantung pada himpunan dari objek-objek yang beresiko untuk mengalami *event* sampai waktu t_i. Himpunan tersebut disebut dengan himpunan resiko, yang dinotasikan dengan $R(t_i)$. Meskipun dalam pembentukan *partial likelihood* tidak melibatkan data yang tersensor, akan tetapi informasi dari data yang tersensor berperan pada saat pembentuk himpunan resikonya. Untuk memperjelas penyusunan fungsi *likelihood* model Cox, diberikan contoh sebagai berikut:

Terdapat data survival untuk mengetahui pengaruh penggunaan sabuk pengaman dan penggunaan alkohol terhadap kecelakaan lalu lintas dimana $t_1 < t_2 < t_3 < t_4$ seperti terlihat pada tabel 2.1.

Tabel 2.1 Cotoh Data Survival

i	t	Status	Rest_use	Dr_drink
1	2	1	1	1
2	3	0	0	0
3	4	1	1	0
4	5	1	0	1

Keterangan:

i : individu.

t : waktu survival (dalam tahun).

Status : 1 untuk *event* dan 0 untuk tersensor.

Rest_use : 1 untuk menggunakan sabuk pengaman dan 0 untuk tidak menggunakan sabuk pengaman.

Dr_drink : 1 untuk mengkonsumsi alkohol dan 0 tidak mengkonsumsi alkohol.

Pada data tersebut mengidikasikan bahwa individu pertama mengalami *event* pada waktu $t = 2$, individu kedua pada waktu $t = 3$ menyatakan mundur dari penelitian dan tidak ingin diteliti lagi sehingga individu kedua tersensor pada waktu $t = 3$, individu ketiga mengalami *event* pada waktu $t = 4$ dan individu keempat mengalami *event* pada waktu $t = 5$. Individu pertama dan ketiga menggunakan sabuk pengaman saat berkendara, sedangkan individu kedua dan keempat tidak menggunakan sabuk pengaman saat berkendara. Selain itu, individu pertama dan keempat tidak mengkonsumsi alkohol saat berkendara, sedangkan individu kedua dan ketiga mengkonsumsi alkohol saat berkendara. Dari informasi tersebut, akan dicari fungsi *partial likelihood* dengan menggunakan fungsi *hazard*. Persamaan model Cox dengan dua variabel tersebut adalah sebagai berikut:

$$h(t,X) = h_0(t)e^{\beta_1 \text{rest\_use} + \beta_2 \text{dr\_drink}}.$$

Berdasarkan informasi di atas fungsi *hazard* untuk masing-masing individu adalah sebagai berikut:

Tabel 2.2 Fungsi *Hazard* Masing-masing Individu

i	*Hazard*
1	$h_0(t)e^{\beta_1+\beta_2}$
2	$h_0(t)e^{0}$
3	$h_0(t)e^{\beta_1}$
4	$h_0(t)e^{\beta_2}$

Pada data tersebut telah diketahui bahwa individu kedua tersensor, sehingga *partial likelihood* data tersebut merupakan perkalian *likelihood* dari tiga individu yang lain, yaitu individu pertama, ketiga dan keempat. Individu pertama mengalami *event* pada waktu $t = 2$, sehingga keempat subyek tersebut merupakan himpunan resikonya. Nilai *likelihood event* pertama (L_1) adalah sebagai berikut:

$$L_1 = \frac{h_0(t)e^{\beta_1+\beta_2}}{h_0(t)e^{\beta_1+\beta_2} + h_0(t)e^{0} + h_0(t)e^{\beta_1} + h_0(t)e^{\beta_2}}.$$

Pembilang nilai *likelihood* tersebut merupakan fungsi *hazard* dari individu pertama dan penyebutnya merupakan penjumlahan dari fungsi *hazard* empat subyek.

Individu ketiga mengalami *event* pada waktu $t = 4$, sehingga himpunan resikonya hanya individu ketiga dan keempat. Nilai *likelihood event* ketiga (L_2) adalah sebagai berikut:

$$L_2 = \frac{h_0(t)e^{\beta_1}}{h_0(t)e^{\beta_1} + h_0(t)e^{\beta_2}}.$$

Pembilang nilai *likelihood* tersebut merupakan fungsi *hazard* dari individu ketiga dan penyebutnya merupakan penjumlahan dari fungsi *hazard* individu ketiga dan keempat.

Individu keempat mengalami *event* terakhir pada waktu $t = 5$, maka tidak ada individu yang lain yang masuk sebagai himpunan resiko kecuali dirinya sendiri. Nilai *likelihood event* keempat (L_3) adalah sebagai berikut:

$$L_3 = \frac{h_0(t)e^{\beta_2}}{h_0(t)e^{\beta_2}} = 1.$$

Fungsi *partial likelihood* dapat pula dikatakan sebagai perkalian dari *likelihood-likelihood* yang *event*-nya teramati. *Partial likelihood* dari contoh di atas adalah sebagai berikut:

$$L = L_1.L_2.L_3$$

$$= \frac{h_0(t)e^{\beta_1+\beta_2}}{h_0(t)e^{\beta_1+\beta_2} + h_0(t)e^{0} + h_0(t)e^{\beta_1} + h_0(t)e^{\beta_2}}.\frac{h_0(t)e^{\beta_1}}{h_0(t)e^{\beta_1} + h_0(t)e^{\beta_2}}.1$$

Pada Cox *likelihood*, fungsi *baseline hazard* $\left(h_0(t)\right)$ dapat dikeluarkan dari model, sehingga *partial likelihood*-nya menjadi sebagai berikut:

$$L = \frac{e^{\beta_1+\beta_2}}{e^{\beta_1+\beta_2} + e^{0} + e^{\beta_1} + e^{\beta_2}} \cdot \frac{e^{\beta_1}}{e^{\beta_1} + e^{\beta_2}}.$$

Meskipun untuk mengilustrasikan fungsi *partial likelihood* di atas menggunakan data yang kecil, namun dapat dibentuk ke dalam bentuk umum. 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 *event* sehingga terdapat $n - r$ individu yang tersensor, sehingga apabila diurutkan urutannnya menjadi $t_1 < t_2 < \cdots < t_i < \cdots < t_r$, dengan t_i merupakan urutan *event* 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:

$$P(\text{individu dengan variabel } x_i \text{ mati saat } t_i | \text{satu kematian saat } t_i). \quad (2.16)$$

Misalkan *event* A adalah individu dengan variabel x_i mati saat t_i dan *event* B adalah satu kematian saat t_i, maka peluang persamaan (2.16) menjadi sebagai berikut:

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

$$= \frac{P(\text{individu dengan variabel } x_i \text{ mati saat } t_i)}{P(\text{satu kematian saat } t_i)}. \quad (2.17)$$

Diasumsikan bahwa waktu kematian independen satu sama lain, maka penyebut persamaan (2.17) di atas 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 (2.17) menjadi sebagai berikut:

$$\frac{\text{P(individu dengan variabel } x_i \text{ mati saat } t_i)}{\sum_{l \in R(t_i)} \text{P(individu } l \text{ mati saat } t_i)}. \qquad (2.18)$$

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 persaamaan (2.18) tersebut dibagi dengan Δt, sehingga didapatkan persamaan sebagai berikut:

$$\frac{\text{P[individu dengan variabel } x_i \text{ mati pada } (t_i, t_i + \Delta t)]/\Delta t}{\sum_{l \in R(t_i)} \text{P[individu } l \text{ mati pada } (t_i, t_i + \Delta t)]\ /\Delta t}. \qquad (2.19)$$

Nilai limit dari persamaan (2.19) dimana $\Delta t \rightarrow 0$ merupakan rasio peluang dari persamaan (2.18). Berdasarkan persamaan (2.6), limit tersebut merupakan fungsi *hazard* kematian pada waktu t_i, 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 (2.16) adalah sebagai berikut:

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

Karena pada Cox *likelihood* fungsi *baseline hazard* dapat dikeluarkan dari model, maka persamaan di atas menjadi sebagai berikut:

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

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

$$L(\beta) = \prod_{i=1}^{r} \frac{\exp\left(\sum_{j=1}^{k} \beta_j X_{j(i)}\right)}{\sum_{l \in R(t_i)} \exp\left(\sum_{j=1}^{k} \beta_j X_{jl}\right)} \qquad (2.16)$$

Misalkan, $\psi_i = \exp\left(\sum_{j=1}^{k} \beta_j X_{j(i)}\right)$, maka

$$L(\beta) = \prod_{i=1}^{r} \frac{\psi_i}{\sum_{l \in R(t_i)} \psi_l}. \qquad (2.17)$$

E. Estimasi Parameter Model Cox

Untuk mengestimasi parameter β_j pada model Cox *proportional hazard* digunakan metode *Maximum Partial Likelihood Estimator* (MPLE) yang dapat diselesaikan dengan memaksimumkan natural log dari fungsi *partial likelihood*. Dari persamaan (2.17) didapatkan fungsi *log partial likelihood,* yaitu sebagai berikut:

$$\begin{aligned}
\ln L(\beta) &= \ln \prod_{i=1}^{r} \frac{\psi_i}{\sum_{l \in R(t_i)} \psi_l} \\
&= \sum_{i=1}^{r} \left[\ln \psi_i - \ln \sum_{l \in R(t_i)} \psi_l \right] \\
&= \sum_{i=1}^{r} \left[\ln \left(\exp \left(\sum_{j=1}^{k} X_{ji} \beta_j \right) \right) - \ln \left(\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^{k} X_{jl} \beta_j \right) \right) \right] \\
&= \sum_{i=1}^{r} \left[\left(\sum_{j=1}^{k} X_{ji} \beta_j \right) - \ln \left(\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^{k} X_{jl} \beta_j \right) \right) \right] \qquad (2.18)
\end{aligned}$$

Setelah diperoleh fungsi *log partial likelihood*, maka langkah selanjutnya adalah mencari turunan pertama dari $\ln L(\beta)$ terhadap β_j, yaitu sebagai berikut:

$$\frac{\partial \ln L(\beta)}{\partial \beta_j} = \frac{\partial \left(\sum_{i=1}^{r} \left[\left(\sum_{j=1}^{k} X_{ji}\beta_j \right) - \ln \left(\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^{k} X_{jl}\beta_j \right) \right) \right] \right)}{\partial \beta_j}$$

$$= \sum_{i=1}^{r} \left[\sum_{j=1}^{k} X_{ji} - \frac{\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^{k} X_{jl}\beta_j \right) \sum_{j=1}^{k} X_{jl}}{\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^{k} X_{jl}\beta_j \right)} \right]. \qquad (2.19)$$

Pendugaan β_j dapat diperoleh dengan memaksimukan turunan pertama fungsi *log partial likelihood*, yaitu sebagai berikut:

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

$$\sum_{i=1}^{r} \left[\sum_{j=1}^{k} X_{j(i)} - \frac{\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^{k} X_{jl}\beta_j \right) \sum_{j=1}^{k} X_{jl}}{\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^{k} X_{jl}\beta_j \right)} \right] = 0. \qquad (2.20)$$

Persamaan (2.20) di atas dapat diselesaikan dengan menggunakan iterasi *Newton-Raphson*.Turunan kedua dari $\ln L(\beta)$ terhadap β_j, yaitu sebagai berikut:

$$\frac{\partial^2 \ln L(\beta)}{\partial^2 \beta_j} = \frac{\partial}{\partial \beta_j} \left(\frac{\partial \ln L(\beta)}{\partial \beta_j} \right)$$

$$= \frac{\partial}{\partial \beta_j} \left[\sum_{i=1}^{r} \left[\sum_{j=1}^{k} X_{j(i)} - \frac{\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^{k} X_{jl}\beta_j \right) \sum_{j=1}^{k} X_{jl}}{\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^{k} X_{jl}\beta_j \right)} \right] \right]$$

$$= \sum_{i=1}^{r} \left[\frac{\left(\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^{k} X_{jl}\beta_j \right) \sum_{j=1}^{k} X_{jl} \right)^2}{\left(\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^{k} X_{jl}\beta_j \right) \right)^2} - \frac{\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^{k} X_{jl}\beta_j \right) \left(\sum_{j=1}^{k} X_{jl} \right)^2}{\sum_{l \in R(t_i)} \exp \left(\sum_{j=1}^{k} X_{jl}\beta_j \right)} \right]$$

$$= -\sum_{i=1}^{r}\left[\frac{\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{k} X_{jl}\beta_j\right)\left(\sum_{j=1}^{k} X_{jl}\right)^2}{\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{k} X_{jl}\beta_j\right)} - \frac{\left(\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{k} X_{jl}\beta_j\right)\sum_{j=1}^{k} X_{jl}\right)^2}{\left(\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{k} X_{jl}\beta_j\right)\right)^2}\right]. \tag{2.21}$$

Negatif turunan kedua dari *log partial likelihood* yaitu sebagai berikut:

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

$$= -\left[-\sum_{i=1}^{r}\left[\frac{\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{k} X_{jl}\beta_j\right)\left(\sum_{j=1}^{k} X_{jl}\right)^2}{\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{k} X_{jl}\beta_j\right)} - \frac{\left(\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{k} X_{jl}\beta_j\right)\sum_{j=1}^{k} X_{jl}\right)^2}{\left(\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{k} X_{jl}\beta_j\right)\right)^2}\right]\right]. \tag{2.22}$$

F. Prosedur *Newton Raphson*

Prosedur *Newton-Raphson* digunakan untuk memaksimalkan fungsi *partial likelihood*. Misalkan $L(\beta)$ merupakan fungsi *partial likelihood* k dimensional vektor $\beta = (\beta_1, \beta_2, \dots, \beta_k)^t$ dan $\mathrm{U}(\beta)$ merupakan vektor berukuran k dari turunan parsial pertama $L(\beta)$, sehingga

$$\mathrm{U}(\beta) = [\mathrm{U}_1(\beta), \dots, \mathrm{U}_k(\beta)]^t \tag{2.23}$$

dengan memisalkan $\mathrm{U}_j(\beta) = \frac{\partial \ln L(\beta)}{\partial \beta_j}, j = 1,2, \dots, k$.

Misalkan $\mathrm{I}(\beta)$ merupakan matrik *Hessian* berukuran $k \times k$ dari turunan *partial likelihood* kedua $\ln L(\beta)$, yaitu

$$\mathrm{I}(\beta) = \left(\mathrm{I}_{tj}(\beta)\right), t, j = 1,2, \dots, k \tag{2.24}$$

dengan memisalkan $\mathrm{I}_{tj}(\beta) = \frac{\partial^2 \ln L(\beta)}{\partial \beta_t \beta_j}$, maka

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

Algoritma yang digunakan dalam metode *Newton Raphson* adalah sebagai berikut:

$$\hat{\beta}_{c+1}=\hat{\beta}_c-\mathrm{I}\left(\hat{\beta}_c\right)^{-1}\mathrm{U}\left(\hat{\beta}_c\right) \tag{2.25}$$

dengan memisalkan $c=0,1,2,\ldots$ dan $\mathrm{I}\left(\hat{\beta}_c\right)^{-1}$ adalah invers dari $\mathrm{I}\left(\hat{\beta}_c\right)$.

Langkah-langkah iterasi dengan menggunakan metode *Newton Raphson* adalah sebagai berikut:

1. Menentukan nilai awal, $\hat{\beta}_0=0$.
2. Memasukan ke dalam persamaan (2.25), yaitu: $\hat{\beta}_1=\hat{\beta}_0-\mathrm{I}\left(\hat{\beta}_0\right)^{-1}\mathrm{U}\left(\hat{\beta}_0\right)$.
3. Iterasi dilakukan sampai diperoleh nilai yang konvergen, $\hat{\beta}_{c+1}\cong\hat{\beta}_c$.

Menurut Hosmer dan Lemeshow (2008:72) varians dari $\hat{\beta}_j$ dapat didefinisikan sebagai:

$$Var\left(\hat{\beta}\right)=\mathrm{I}\left(\hat{\beta}\right)^{-1}, \tag{2.26}$$

sedangkan standar deviasi dari $\hat{\beta}_j$ merupakan akar kuadrat dari varians $\hat{\beta}_j$, yaitu sebagai berikut:

$$SE\left(\hat{\beta}\right)=\sqrt{Var\left(\hat{\beta}\right)}=\sqrt{\mathrm{I}\left(\hat{\beta}\right)^{-1}}. \tag{2.27}$$

Standar deviasi dapat digunakan untuk mencari selang kepercayaan $\hat{\beta}_j$ yaitu $(1-\alpha)100\%$. Selang kepercayaan untuk $\hat{\beta}_j$ adalah sebagai berikut:

$$\left(\hat{\beta}_j - Z_{\frac{\alpha}{2}} SE(\hat{\beta}), \hat{\beta}_j + Z_{\frac{\alpha}{2}} SE(\hat{\beta})\right). \quad (2.28)$$

G. Pengujian Parameter

Terdapat tiga tujuan statistik, yaitu sebagai berikut: (1) untuk menguji signifikansi parameter, (2) memperoleh estimasi titik dan (3) memperoleh interval konfidensi. Menurut Hosmer dan Lemeshow (2008:29) terdapat tiga cara untuk menguji signifikansi parameter model Cox, yaitu dengan uji *partial likelihood ratio*, uji *Wald* serta uji *Score*. Pengujian parameter bertujuan untuk memeriksa apakah variabel bebas mempunyai pengaruh yang nyata dalam model.

1. Uji *Partial Likelihood Ratio*

Uji *Partial Likelihood Ratio* dinotasikan dengan G. Statistik uji ini mengikuti distribusi *chi-square* dengan derajat bebas p. Langkah-langkah dalam uji *Partial Likelihood Ratio* adalah sebagai berikut:

a. Hipotesis

$H_0: \forall\beta_j = 0, j = 1,2,\ldots,p$ (variabel bebas tidak berpengaruh terhadap waktu survival)

$H_1: \exists\beta_j \neq 0, j = 1,2,\ldots,p$ (variabel bebas berpengaruh terhadap waktu survival)

b. Taraf Signifikansi : α

c. Statistik Uji : $G = -2\left[\ln L(0) - \ln L(\hat{\beta}_j)\right]$

dengan:

$\ln L(0)$ merupakan *log partial likelihood* dari model tanpa varibel bebas (medel nol).

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

d. Kriteria Keputusan : H_0 ditolak jika $G \geq \chi^2_{(\alpha,p)}$ atau *p-value* $\leq \alpha$

e. Kesimpulan :

Jika H_0 ditolak maka $\beta_j \neq 0$, yang artinya variabel bebas berpengaruh terhadap waktu survival (variabel dependen).

2. Uji *Wald*

Uji ini digunakan untuk menguji pengaruh parameter secara terpisah, yang dinotasikan dengan z. Statistik uji ini mengikuti distribusi *chi-square* dengan derajat bebas p. Langkah-langkah dalam uji *Wald* adalah sebagai berikut:

a. Hipotesis

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

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

b. Taraf Signifikansi : α

c. Statistik Uji :

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

d. Kriteria Keputusan :

H_0 ditolak jika $z^2 \geq \chi^2_{(\alpha,p)}$ atau *p-value*$< \alpha$

e. Kesimpulan :

Jika H_0 ditolak maka $\beta_j \neq 0$, yang artinya variabel bebas berpengaruh terhadap waktu survival (variabel dependen).

3. Uji *Score*

Selain uji *Partial Likelihood Ratio* dan uji *Wald*, tedapat pula uji *Score* untuk menguji signifikansi parameter. Statistik uji *Score* adalah rasio dari turunan *log partial likelihood* pada persamaan (2.19) dengan akar kuadrat dari persamaan (2.22). Statistik uji ini juga mengikuti sebaran distribusi *chi*-*square* dengan derajat bebas p. Langkah-langkah dalam uji S*core* adalah sebagai berikut:

a. Hipotesis

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

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

b. Taraf Signifikansi : α

c. Statistik Uji :

$$z^2 = \left(\frac{\frac{\partial L\beta_j}{\partial \beta_j}}{\sqrt{\mathrm{I}(\beta_j)}} \middle| \beta_j = 0 \right)^2$$

d. Kriteria Keputusan :

H_0 ditolak jika $z^2 \geq \chi^2_{(\alpha,p)}$ atau *p-value* $\leq \alpha$

e. Kesimpulan :

Jika H_0 ditolak maka $\beta_j \neq 0$, yang artinya variabel bebas berpengaruh terhadap waktu survival (variabel dependen).

H. Pemilihan Model Terbaik Cox

Untuk mendapatkan model terbaik Cox pertama-tama dilakukan pemilihan variabel yang masuk atau keluar dari model. Terdapat tiga metode untuk melakukan pemilihan variabel tersebut. Ketiga metode tersebut adalah prosedur seleksi *forward*, prosedur seleksi *backward* dan prosedur seleksi *stepwise* (David Collett, 2004). Prosedur seleksi *forward* atau disebut dengan seleksi maju merupakan suatu proses penambahan satu variabel yang terpilih dan ditambahkan ke dalam model pada setiap langkahnya. Prosedur seleksi *backward* atau disebut dengan seleksi mundur adalah suatu proses eliminasi dimana pada awalnya semua variabel dimasukan ke dalam model, kemudian melakukan eliminasi terhadap variabel tersebut satu per satu berdasarkan kriteria keputusannya. Sedangkan, prosedur seleksi *stepwise* merupakan kombinasi dari prosedur seleksi *forward* dan prosedur seleksi *backward*.

Pada penelitian ini, pemilihan model terbaik Cox dilakukan dengan menggunakan prosedur seleksi *backward*. Prosedur seleksi *backward* dihentikan apabila semua variabel yang masuk ke dalam model sudah signifikan. Untuk memeriksa setiap variabel yang akan dikeluarkan dilakukan sebuah pengujian. Langkah-langkah pengujian yang dilakukan dalam prosedur seleksi *backward* adalah sebagai berikut:

- Hipotesis

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

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

- Taraf Signifikansi : α
- Statistik Uji :

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

- Kriteria Keputusan: H_0 ditolak jika $z^2 \geq \chi^2_{(\alpha,p)}$ atau *p-value* $\leq \alpha$
- Kesimpulan :

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

I. Residual Model Cox *Proportional Hazard*

Menurut David Collett (2004), untuk menguji asumsi *proportional hazard* dalam suatu model Cox *proportional hazard* terdapat dua macam cara untuk mengujinya. Kedua cara tersebut adalah pendekatan grafik menggunakan plot *log-minus-log survival* dan menggunakan residual *Schoenfeld*. Apabila pada dua pengujian asumsi *proportional hazard* tersebut ditemukan variabel yang tidak memenuhi asumsi *proportional hazard*, maka asumsi *proportional hazard* dilanggar atau dapat disebut dengan Cox *nonproportional hazard*. Untuk mengatasi permasalahan tersebut, terdapat tiga macam pilihan, yaitu mengeluarkan variabel bebas yang tidak memenuhi asumsi dari model, menggunakan model Cox stratifikasi dan menggunakan perluasan model Cox.

1. **Pengujian Asumsi *Proportional Hazard* dengan Grafik *Log-Minus-Log Survival***

Pendekatan grafik yang digunakan adalan grafik *log-minus-log survival*. Menurut model regresi Cox, fungsi *hazard* kumulatif pada persamaan (2.10) dapat pula dinyatakan sebagai berikut:

$$H(t,X) = \int_0^t h(x,X)dx$$

$$= \int_0^t h_0(x)\exp\left(\sum_{j=1}^{k} \beta_j X_j\right)dx$$

$$= \exp\left(\sum_{j=1}^{k} \beta_j X_j\right)\int_0^t h_0(x)dx$$

$$= \exp\left(\sum_{j=1}^{k}\beta_j X_j\right) H_0(t). \qquad (2.29)$$

Persamaan (2.29) disubstitusikan ke dalam persamaan (2.15), sehingga diperoleh hasil sebagai berikut:

$$S(t,X) = e^{\left[-\exp\left(\sum_{j=1}^{k}\beta_j X_j\right) H_0(t)\right]}$$

$$= \left[\mathrm{e}^{-H_0(t)}\right]^{\exp\left(\sum_{j=1}^{k}\beta_j X_j\right)}$$

$$= S_0(t)^{\exp\left(\sum_{j=1}^{k}\beta_j X_j\right)} \qquad (2.30)$$

Hasil persamaan (2.30) tersebut merupakan fungsi survival dari model Cox. Selanjutnya, dilakukan operasi logaritma asli terhadap fungsi survival tersebut untuk kedua sisi, sehingga diperoleh sebagai berikut:

$$S(t,X) = S_0(t)^{\exp\left(\sum_{j=1}^{k}\beta_j X_j\right)}$$

$$-\ln S(t,X) = -\ln\left[S_0(t)^{\exp\left(\sum_{j=1}^{k}\beta_j X_j\right)}\right]$$

$$= \exp\left(\sum_{j=1}^{k}\beta_j X_j\right)[-\ln S_0(t)]$$

$$\ln[-\ln S(t,X)] = \ln\left[\exp\left(\sum_{j=1}^{k}\beta_j X_j\right)[-\ln S_0(t)]\right]$$

$$= \ln\left[\exp\left(\sum_{j=1}^{k}\beta_j X_j\right)\right] + \ln[-\ln S_0(t)]$$

$$= \left(\sum_{j=1}^{k} \beta_j X_j \right) + \ln[-\ln S_0(t)]. \quad (2.31)$$

Fungsi *log-minus-log survival* pada pesamaan (2.31) di atas menunjukan bahwa fungsi tersebut tidak bergantung terhadap waktu, sehingga fungsi *log-minus-log survival* pada model Cox *proportional hazard* pada persamaan (2.12) akan berlaku apabila diplotkan melawan waktu survival. Oleh karena itu, kurva yang terbentuk akan paralel. Jika terdapat variabel kontinu pada data survival yang digunakan, maka nilainya perlu dikelompokan menjadi variabel kategori. Menurut David Collett (2004), plot *log-minus-log survival* merupakan sebuah plot dari logaritma estimasi fungsi *hazard* kumulatif terhadap waktu survival yang menghasilkan kurva paralel apabila laju *proportional hazard* diseluruh kelompok berbeda. Akan tetapi, pendekatan grafik ini mempunyai kelemahan yaitu bersifat subyektif, paralel atau tidaknya grafik tergantung dari cara pandang peneliti.

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

Terdapat beberapa jenis residual yang dapat digunakan untuk melakukan pengujian asumsi *proportional hazard*. Menurut David Collett (2004), residual tersebut antara lain residual *Martingale*, residual *Deviance*, residual *Schoenfeld*, residual *Cox-Snell* dan residual *Score*. Pada penelitian ini akan digunakan residual *Schoenfeld* untuk menguji asumsi *proportional hazard*. Menurut Lee dan Wang

(2004:331-332) residual *Schoenfeld* didefinisikan sebagai residual yang setiap individu dan setiap variabel bebasnya berdasarkan turunan pertama dari fungsi *log likelihood* pada persamaan (2.20). Residual *Schoenfeld* untuk individu ke-i pada variabel bebas ke-j adalah sebagai berikut:

$$R_{ji} = \delta_i \left(x_{ji} - \frac{\sum_{l \in R(t_i)} x_{jl} \exp\left(\hat{\beta}' x_l\right)}{\sum_{l \in R(t_i)} \exp\left(\hat{\beta}' x_l\right)} \right), j = 1,2, \dots, p \qquad (2.32)$$

dengan $\hat{\beta}$ merupakan estimator *partial likelihood* maksimum dari β. Karena $\hat{\beta}$ merupakan solusi dari persamaan turunan pertama fungsi *log likelihood* pada persamaan (2.20), maka jumlah residual *Schoenfeld* adalah nol atau dengan kata lain residual *Schoenfeld* mempunyai rata-rata nol. Apabila jumlah sampel besar, nilai harapan dari R_{ji} adalah nol, sehingg residual *Schoenfeld* tidak berkorelasi dengan yang lainnya.

Scaled residual Schoenfeld dapat dihitung dengan menggunakan invers dari estimator matrik kovarian $R_i = \left(R_{1i}, \dots, R_{pi}\right)'$ yang dinotasikan dengan $\hat{V}(R_i)$, sehingga diperoleh

$$R_i^* = \left[\hat{V}(R_i)\right]^{-1} R_i. \qquad (2.33)$$

Untuk menyederhanakan perhitungan, Therneau dan Grambsch (1994) mengusulkan bahwa perkiraan dari $\left[\hat{V}(R_i)\right]^{-1}$ pada persamaan (2.33) adalah sebagai berikut:

$$\left[\hat{V}(R_i)\right]^{-1} \cong r\hat{V}\left(\hat{\beta}\right), \qquad (2.34)$$

Dimana r adalah jumlah *event* dan $\hat{V}(\hat{\beta})$ adalah estimator matrik kovarian dari $\hat{\beta}$ pada persamaan (2.26). Dengan menggunakan perkiraan tersebut, *scaled residual Schoenfeld* pada persamaan (2.33) dapat ditulis sebagai berikut:

$$R_i^* = r\hat{V}(\hat{\beta})R_i. \tag{2.35}$$

Grafik antara residual *Schoenfeld* dengan waktu survival dapat digunakan untuk memeriksa kelengkapan model *proportional hazard*. Cox *proportional hazard* dikatakan proporsional apabila *hazard ratio*-nya independen terhadap waktu. Apabila terdapat variabel bebas yang tergantung pada waktu, maka asumsi *proportional hazard* tidak terpenuhi.

J. Interpretasi Model Cox *Proportional Hazard*

Interpretasi dapat dilakukan apabila persamaan regresi Cox $h(t, X_j) = h_0 t \exp(\beta X_j)$ telah didapatkan dengan menggunakan *hazard ratio*. Persamaan (2.12) dapat dituliskan d alam bentuk lain, yaitu sebagai berikut:

$$\ln h(t, X) = \ln\left[\left(h_0(t)\right)e^{\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k}\right]$$

$$\ln \frac{h(t, X)}{h_0(t)} = \ln\left[e^{\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k}\right]$$

$$\ln \frac{h(t, X)}{h_0(t)} = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

$$\ln[HR(x)] = (\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k). \tag{2.36}$$

Variabel bebas $X_1, X_2, \dots, X_k$ dan koefisien β_j pada model persamaan (2.36) di atas mengimplikasikan peningkatan pada *log hazard ratio* untuk peningkatan satu satuan variabel bebas X_j, dengan asumsi nilai variabel bebas yang lain adalah konstan. Apabila variabel bebas dengan *hazard ratio* kurang dari $1(\beta < 0)$, maka peningkatan nilai variabel bebas berhubungan dengan penurunan resiko. Apabila *hazard ratio* lebih besar dari $1(\beta > 0)$, maka peningkatan nilai variabel bebas berhubungan dengan peningkatan resiko.

Menurut Lee dan Wang (2003:298), berdasarkan persamaan (2.36), untuk variabel bebas X_0 dan X_1 dari dua individu diperoleh hasil sebagai berikut:

$$\frac{h(t, X_j)}{h(t, X_0)} = \frac{h_0 t \exp(\beta X_1)}{h_0 t \exp(\beta X_0)}$$

$$= \frac{\exp(\beta X_1)}{\exp(\beta X_0)}$$

$$= e^{(X_1 - X_0)\beta}, \forall t > 0. \qquad (2.37)$$

Persamaan (2.37) di atas menunjukan besar rasio relatif dari dua individu dengan faktor resiko X_1 dibandingkan dengan faktor resiko X_0 dari individu yang lain. Persamaan (2.38) dapat dituliskan sebagai berikut:

$$\frac{h(t, X_j)}{h(t, X_0)} = e^{(X_1 - X_0)\beta}$$

$$\ln\left[\frac{h(t, X_j)}{h(t, X_0)}\right] = \ln\left[e^{(X_1 - X_0)\beta}\right]$$

$$\ln\left[\frac{h(t, X_j)}{h(t, X_0)}\right] = (X_1 - X_0)\beta. \qquad (2.38)$$

Untuk nilai kovariat X_j tetap dapat diinterpretasikan sebagai berikut:

$$\ln\left[\frac{h(t,X_{j+1})}{h(t,X_j)}\right] = (X_{j+1} - X_j)\beta_j. \tag{2.39}$$

Dari persamaan (2.39) di atas dapat diambil kesimpulan bahwa setiap naiknya nilai β_j maka akan memperbesar nilai logaritma *hazard ratio*, sehingga diperoleh:

$$\ln\left[\frac{h(t,X_{j+1})}{h(t,X_j)}\right] = (X_{j+1} - X_j)\beta_j$$

$$\left[\frac{h(t,X_{j+1})}{h(t,X_j)}\right] = e^{(X_{j+1}-X_j)\beta_j}, \forall t > 0. \tag{2.40}$$

Dengan demikian nilai $e^{(X_{j+1}-X_j)\beta_j}$ merupakan *hazard ratio* yang dapat dihubungkan dengan kenaikan nilai x_j.

Berdasarkan persamaan $h(t,X_j) \approx P(t < T < (t+\Delta t)|T \geq t, X)$, persamaan (2.40) dapat dituliskan sebagai berikut:

$$\frac{P(t < T < (t+\Delta t)|T \geq t, X_{j+1})}{P(t < T < (t+\Delta t)|T \geq t, X_j)} = e^{(X_{j+1}-X_j)\beta_j}, \forall t > 0. \tag{2.41}$$

Nilai $e^{(X_{j+1}-X_j)\beta_j}$ dapat diinterpretasikan pula sebagai rasio dua probabilitas bersyarat dari gagalnya individu yang diketahui tersebut masih hidup pada saat t. Persamaan (2.41) di atas ekuivalen dengan

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

$$\frac{h(t,X_{j+1}) - h(t,X_j)}{h(t,X_j)} = e^{(X_{j+1}-X_j)\beta_j} - 1, \forall t > 0. \tag{2.42}$$

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

Terdapat tiga macam ketentuan tentang bertambah atau berkurangnya nilai *hazard* yaitu sebagai berikut:

1. Apabila $(X_{j+1}-X_j)\beta_j > 0$, maka setiap kenaikan nilai X_j akan membuat nilai *hazard* akan bertambah besar atau dengan kata lain semakin besar resiko seorang individu untuk meninggal atau gagal.
2. Apabila $(X_{j+1}-X_j)\beta_j < 0$, maka setiap kenaikan nilai X_j akan memperkecil nilai *hazard* atau dengan kata lain semakin kecil resiko seorang individu untuk meninggal atau gagal.
3. Apabila $(X_{j+1}-X_j)\beta_j = 0$, maka besar resiko seorang individu untuk hidup sama dengan besarnya resiko seorang individu untuk meninggal atau gagal.