BAB II

KAJIAN TEORI

A. Analisis Survival

Analisis survival atau analisis ketahanan hidup adalah metode yang berhubungan dengan jangka waktu, dari awal pengamatan sampai suatu kejadian terjadi (Collett, 2003). Jangka waktu dari awal dilakukan pengamatan pada suatu individu (*time origin*) sampai suatu kejadian terjadi (*end point atau failure event*) disebut dengan waktu survival. Kejadian yang diamati (*failure event*) tersebut dapat berupa kematian, sakit, sakit yang terulang kembali setelah pengobatan atau munculnya penyakit baru, kecelakaan, respon dari suatu percobaan, atau peristiwa lain yang dipilih sesuai dengan kepentingan peneliti. Analisis survival dapat diterapkan dalam bidang biologi, kedokteran, kesehatan umum seperti ketahanan hidup pasien kanker paru-paru, sosiologi, teknik seperti menganalisis masa hidup lampu pijar, ekonomi, demografi, dan epidemilogi.

1. Waktu Survival (*Survival Time*)

Dalam penentuan waktu survival ada tiga faktor yang dibutuhkan, antara lain:

a. Waktu awal pencatatan (*time origin* atau *start-point*) harus didefinisikan dengan tepat pada setiap subjek pengamatan, sebagai contoh awal mula pengamatan berupa tanggal perawatan pasien.

b. Waktu akhir pencatatan (*failure time* atau *end-point*) didefinisikan jelas untuk mengetahui subjek pengamatan mengalami kejadian atau tidak.
c. Skala waktu pengukuran yang jelas, misalnya skala diukur dalam tahun, bulan, atau hari.

2. Penyensoran

Penyensoran adalah salah satu langkah yang harus dilakukan untuk mengatasi ketidaklengkapan suatu data pengamatan. Data tersensor merupakan data pengamatan yang tidak dapat diamati secara lengkap karena subjek penelitian hilang atau sampai akhir penelitian subjek tersebut belum mengalami kejadian tertentu (Lee & Wang, 2003).

Menurut Kleinbaum & Klein (2012), ada 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.

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 subjek pengamatan mengalami suatu kejadian yang diinginkan dan sering kali membutuhkan biaya yang banyak.

Menurut Kleinbaum & Klein (2012) dalam analisis survival terdapat tiga tipe penyensoran, yaitu:

a. Sensor kanan *(right censoring)*

Sensor kanan terjadi karena subjek pengamatan belum mengalami kejadian hingga akhir pengamatan, sedangkan waktu awal dari subjek pengamatan dapat diamati secara penuh. Sebagai contoh, suatu pasien penderita kanker payudara diamati selama lima tahun dari awal pengamatan, kemudian pasien yang diamati pindah dari rumah sakit tersebut karena alasan tertentu pada tahun ketiga dan tidak dapat diamati lagi. Pasien tersebut masih memiliki waktu survival dalam penelitian setidaknya dua tahun, sehingga waktu pengamatan pasien tersebut dikatakan tersensor kanan.

b. Sensor kiri *(left censoring)*

Sensor kiri adalah sensor yang terjadi dikarenakan waktu kejadian dari subjek pengamatan tidak teramati pada pengamatan, sehingga waktu kejadian diasumsikan sama dengan waktu pengamatan terdekat dengan waktu kejadian. Sebagai contoh, seorang peneliti mengamati seseorang sampai orang tersebut positif terjangkit virus HIV. Peneliti dapat mencatat seseorang telah positif terjangkit virus HIV pada tes pertama, tetapi peneliti tidak mengetahui waktu yang tepat seseorang tersebut mulai terjangkit virus HIV, karena itu peneliti tidak tahu persis kapan seseorang telah positif terjangkit virus HIV, dengan demikian pasien tersebut tersensor kiri yaitu ketika mengalami kejadian pertama dengan hasil positif terjangkit virus HIV.

c. Sensor interval *(interval censoring)*

Sensor interval adalah sensor yang waktu survivalnya berada dalam suatu selang tertentu. Sebagai contoh, pada saat pemeriksaan terdapat seorang pasien berumur 50 tahun ditemukan menderita diabetes melitus. Apabila pada catatan medis mengindikasikan bahwa pada saat usia 45 tahun pasien belum menderita diabetes melitus, maka usia pasien didiagnosis menderita diabetes melitus antara 45-50 tahun.

B. Dasar Teori Analisis Survival

Misalkan variabel waktu survival T merupakan suatu variabel acak non negatif, maka nilai-nilai yang mungkin untuk T adalah $T \geq 0$. Menurut Lee dan Wang (2003), terdapat tiga fungsi untuk menentukan distribusi dari T, yaitu (1) fungsi kepadatan peluang, (2) fungsi survival, dan (3) fungsi *hazard.*

1. Fungsi Kepadatan Peluang

Fungsi kepadatan peluang atau *probability density function* (*pdf*) merupakan peluang suatu individu dalam waktu sesaat mengalami kejadian, 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 acak non negatif dalam interval $[0,\infty)$, dan $F(t)$ merupakan fungsi distribusi kumulatif kontinu dari T . Fungsi ini didefinisikan sebagai peluang suatu individu mengalami kejadian sebelum waktu t yang dapat dituliskan sebagai berikut:

$$F(t) = P(T < t) = \int_0^t f(x)dx \tag{2.2}$$

Dari Persamaan (2.2), dengan melakukan penurunan terhadap dt maka diperoleh:

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

2. Fungsi Survival

Fungsi survival dinotasikan dengan $S(t)$, yaitu peluang suatu individu mengalami kejadian setelah atau pada waktu t, yaitu sebagai berikut:

$$S(t) = P(T \geq t) \tag{2.4}$$

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

$$\begin{aligned} S(t) &= P(T \geq t) \\ &= \int_t^\infty f(x)\, dx \end{aligned}$$

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

$$\begin{aligned} S(t) &= P(T \geq t) \\ &= 1 - P(T < t) \\ &= 1 - F(t) \end{aligned}$$

diperoleh hubungan antara fungsi kepadatan peluang, fungsi distribusi kumulatif dan fungsi survival, yaitu:

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

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

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

$$f(t) = -S'(t) \tag{2.5}$$

maka diperoleh:

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

Menurut Klein dan Kleinbaum (2012), 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 $t = 0$, $S(0) = 1$.

   Pada awal dimulai pengamatan, karena belum ada individu yang mengalami kejadian maka peluang individu mengalami kejadian setelah atau pada saat $t = 0$ adalah 1.
c. Pada saat $t = \infty$, $S(t) \approx 0$.

   Secara teori, apabila periode penelitian meningkat tanpa batas, maka di akhir 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 kecepatan suatu individu mengalami kejadian pada interval waktu t sampai $(t + \Delta t)$ apabila diketahui individu tersebut masih bertahan hidup sampai dengan waktu t, dapat dituliskan dengan persamaan sebagai berikut:

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

Berdasarkan teorema peluang, jelas bahwa peluang kejadian A dengan syarat kejadian B, yaitu:

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

sehingga diperoleh persamaan untuk *hazard rate* sebagai berikut:

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

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

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

$$= \frac{1}{S(t)} \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}$$

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

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

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

$$= \frac{f(t)}{S(t)} \tag{2.9}$$

Apabila Persamaan (2.5) dihubungkan ke Persamaan (2.9), maka diperoleh:

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

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

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

$$= -\frac{d \ln S(t)}{dt} \tag{2.10}$$

Jika Persamaan (2.10) diintegralkan terhadap x dengan interval $[0, t]$, maka diperoleh:

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

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

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

$$= \ln S(t) - \ln S(0) \tag{2.11}$$

Karena $S(0) = 1$, maka $\ln S(0) = \ln 1 = 0$. Oleh karena itu, Persamaan (2.11) dapat ditulis menjadi:

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

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

$$S(t) = \exp\left[-\int_0^t h(x)dx\right] \quad (2.12)$$

Berdasarkan fungsi *hazard* yang diperoleh dari Persamaan (2.12), menurut Lee & Wang (2003) fungsi kumulatif *hazard* $H(t)$ yaitu

$$H(t) = \int_0^t h(x)dx \quad (2.13)$$

dan diperoleh hubungan dengan fungsi survival, yaitu:

$$S(t) = \exp[-H(t)] \quad (2.14)$$

C. Model Regresi Cox *Proportional Hazard*

Salah satu tujuan dari analisis survival adalah untuk mengetahui hubungan antara waktu survival dengan variabel-variabel yang diduga mempengaruhi waktu survival. Analisis ini dapat dilakukan dengan analisis regresi. Analisis survival mengenal tiga model, yaitu parametrik, semiparametrik, dan non parametrik. Model parametrik mengasumsikan bahwa distribusi yang mendasari waktu survival mengikuti suatu distribusi tertentu, misalnya distribusi *Weibull*, *gamma*, eksponensial. Model non parametrik mengasumsikan waktu survival tidak mengikuti suatu distribusi tertentu yang sudah ada, fungsi survival disusun dengan metode *Kaplan-Meier* dan *Nelson-Aalen*. Model semiparametrik dimana dalam model ini tidak memerlukan informasi tentang distribusi yang mendasari waktu survival dan fungsi *baseline hazard* tidak harus ditentukan untuk mengestimasi parameternya. Salah satu model semiparametrik adalah model regresi Cox *proportional hazard*.

Model regresi 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). Secara umum, bentuk dari model regresi Cox *proportional hazard* adalah:

$$h(t,X) = h_0(t)\exp\left(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p\right) \tag{2.15}$$

dengan:

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

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

$X_1, X_2, \dots, X_p$ = variabel-variabel bebas (kovariat)

Jika semua nilai $X = 0$, maka Persamaan (2.15) akan tereduksi menjadi fungsi *baseline hazard* $h_0(t)$. Fungsi *baseline hazard* ini merupakan fungsi awal atau dasar dari fungsi *hazard*, atau dapat dituliskan sebagai berikut:

$$h(t,0) = h_0(t)$$

Meskipun bentuk $h_0(t)$ pada model Cox tidak diketahui, tetapi 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 (Kleinbaum & Klein, 2012), yaitu sebagai berikut:

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

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

$$= \exp\left[\left(\beta_1 X_1^* + \beta_2 X_2^* + \cdots + \beta_p X_p^*\right) - \left(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p\right)\right]$$

$$= \exp\left[\sum_{j=1}^{p} \beta_j \left(X_j^* - X_j\right)\right] \tag{2.16}$$

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

1. Pendugaan Parameter Model Regresi Cox *Proportional Hazard*

Estimasi parameter β_j dengan $j = 1,2, \ldots, p$ pada model *Cox proportional hazard* dapat dilakukan salah satunya menggunakan metode *Maximum Partial Likelihood Estimation* (MPLE). Fungsi *likelihood* yang biasa dinotasikan dengan L terkadang dinotasikan juga dengan $L(\beta)$ dimana parameter-parameter β yang tidak diketahui nilainya yang menggambarkan peluang bersama dari data observasi (Kleinbaum & Klein, 2012).

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

$$L = L_1\, L_2\, L_3\, \ldots\, L_k = \prod_{j=1}^{k} L_j \tag{2.17}$$

Notasi L_j menunjukkan *likelihood* untuk kejadian saat ke- j dengan himpunan risiko $R(t_j)$.

Misalkan terdapat data untuk individu dan masing-masing mempunyai vektor kovariat $X_i = [X_{i1}, X_{i2}, \dots, X_{ip}]$. Dari n individu tersebut, misalkan terdapat r individu yang mengalami kejadian, maka terdapat $n - r$ individu yang tersensor. Apabila diurutkan, urutannnya menjadi $t_1 < t_2 < \cdots < t_i < \cdots < t_r$, dengan t_i merupakan urutan kejadian ke-i. Diasumsikan hanya terdapat satu individu yang meninggal 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:

$$P(\text{individu dengan variabel } x_i \text{ meninggal saat } t_i | \text{satu kematian saat } t_i) \quad (2.18)$$

Misalkan kejadian A adalah individu dengan variabel x_i meninggal saat t_i dan kejadian B adalah satu kematian saat t_i, maka peluang Persamaan (2.18) menjadi sebagai berikut:

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

$$= \frac{P(\text{individu dengan variabel } x_i \text{ meninggal saat } t_i)}{P(\text{satu kematian saat } t_i)}$$

(2.19)

Diasumsikan bahwa waktu kematian independen satu sama lain, maka penyebut Persamaan (2.19) merupakan penjumlahan dari peluang kematian semua individu yang berisiko mati pada waktu t_i. Apabila individu-individu tersebut dinotasikan dengan l dan $R(t_i)$ merupakan himpunan dari individu-individu yang berisiko pada waktu t_i, maka Persamaan (2.19) menjadi sebagai berikut:

$$P(A|B) = \frac{P(\text{individu dengan variabel } x_i \text{ meninggal saat } t_i)}{\sum_{l \in R(t_i)} P(\text{individu } l \text{ mati saat } t_i)}$$

(2.20)

Peluang individu yang mati saat t_i dapat diganti dengan peluang individu yang meninggal pada interval $(t_i, t_i + \Delta t)$, dan kemudian baik penyebut maupun pembilang pada Persamaan (2.20) dibagi dengan Δt, sehingga didapatkan persamaan sebagai berikut:

$$P(A|B) = \frac{P[\text{individu dengan variabel } x_i \text{ meninggal pada } (t_i, t_i + \Delta t)]/\Delta t}{\sum_{l \in R(t_i)} P[\text{individu } l \text{ meninggal pada } (t_i, t_i + \Delta t)]/\Delta t}$$

(2.21)

Nilai limit dari Persamaan (2.21) dimana $\Delta t \to 0$ merupakan rasio peluang dari Persamaan (2.20), sehingga,

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

Apabila individu ke-j mati pada saat t_i, fungsi *hazard* pada pembilang persamaan tersebut dapat ditulis $h_j(t_i)$, maka diperoleh:

$$h_j(t_i) = h_0(t) \exp\left(\sum_{j=1}^{p} \beta_j x_{j(i)}\right)$$

sedangkan penyebut persamaan tersebut merupakan penjumlahan dari fungsi *hazard* untuk individu yang berisiko pada waktu t_i yang dapat ditulis $h_l(t_i)$, maka diperoleh:

$$\sum_{l \in R(t_i)} h_l(t_i) = \sum_{l \in R(t_i)} h_0(t) \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)$$

sehingga peluang bersyarat untuk Persamaan (2.18) adalah sebagai berikut:

$$P(A|B) = \frac{h_j(t_i)}{\sum_{l\in R(t_i)} h_l(t_i)} \tag{2.22}$$

Karena pada Cox *likelihood* fungsi *baseline hazard* dapat dikeluarkan dari model, maka Persamaan (2.22) menjadi sebagai berikut:

$$P(A|B) = \frac{h_j(t_i)}{\sum_{l\in R(t_i)} h_l(t_i)}$$

$$= \frac{h_0(t)\exp\left(\sum_{j=1}^{p} \beta_j x_{j(i)}\right)}{\sum_{l\in R(t_i)} h_0(t)\exp\left(\sum_{j=1}^{p} \beta_j x_{jl})\right)}$$

$$= \frac{h_0(t)\exp\left(\sum_{j=1}^{p} \beta_j x_{j(i)}\right)}{h_0(t)\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl})\right)}$$

$$= \frac{\exp(\sum_{j=1}^{p} \beta_j x_{j(i)})}{\sum_{l\in R(t_i)} \exp\sum_{j=1}^{p} \beta_j x_{jl})} \tag{2.23}$$

dengan hasil peluang bersyarat pada Persamaan (2.23), maka bentuk fungsi *partial likelihood* adalah sebagai berikut:

$$L(\beta) = \prod_{i=1}^{r} \frac{\exp(\sum_{j=1}^{p} \beta_j x_{j(i)})}{\sum_{l\in R(t_i)} \exp(\sum_{j=1}^{p} \beta_j x_{jl})} \tag{2.24}$$

dari Persamaan (2.24), diperoleh fungsi *likelihood* sebagai berikut:

$$\ln L\,(\beta) = \ln \prod_{i=1}^{r} \frac{\exp(\sum_{j=1}^{p} \beta_j x_{j(i)})}{\sum_{l\in R(t_i)} \exp(\sum_{j=1}^{p} \beta_j x_{jl})} \tag{2.25}$$

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

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

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

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

$$\frac{\partial \ln L(\beta)}{\partial \beta_j} = \sum_{i=1}^{k}\left[\sum_{j=1}^{p} x_{j(i)} - \frac{\sum_{l\in R(t_i)} x_{jl} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)}{\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)}\right] \qquad (2.26)$$

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

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

diperoleh

$$\sum_{i=1}^{k}\left[\sum_{j=1}^{p} x_{j(i)} - \frac{\sum_{l\in R(t_i)} x_{jl} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)}{\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)}\right] = 0 \qquad (2.27)$$

Persamaan (2.27) dapat diselesaikan secara numerik, yaitu menggunakan metode *Newton-Raphson*. Selanjutnya akan dicari turunan kedua dari $\ln L(\beta)$, yaitu

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

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

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

$$\left. - \frac{\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)\left(\sum_{j=1}^{p} x_{jl}\right)^2}{\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)}\right]$$

$$= -\sum_{i=1}^{k}\left[\frac{\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)\left(\sum_{j=1}^{p} x_{jl}\right)^2}{\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)} - \frac{\left(\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)\sum_{j=1}^{p} x_{jl}\right)^2}{\left(\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)\right)^2}\right] \tag{2.28}$$

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

$$-\frac{\partial^2 \ln L(\beta)}{\partial^2 \beta_j} = -\left[-\sum_{i=1}^{k}\left[\frac{\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)\left(\sum_{j=1}^{p} x_{jl}\right)^2}{\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)} - \frac{\left(\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)\sum_{j=1}^{p} x_{jl}\right)^2}{\left(\sum_{l\in R(t_i)} \exp\left(\sum_{j=1}^{p} \beta_j x_{jl}\right)\right)^2}\right]\right] \tag{2.29}$$

Selanjutnya, untuk menggambarkan fungsi *partial likelihood* diberikan contoh sebagai berikut.

Akan diselidiki hubungan antara waktu kejadian individu mengalami kecelakaan dengan variabel-variabel yang diduga mempengaruhi individu untuk mengalami kecelakaan. Subjek yang diteliti sebanyak 4 individu dengan variabel yang diduga mempengaruhi adalah penggunaan sabuk pengaman (X_1) dan konsumsi alkohol (X_2). Data diperoleh dari keempat subjek seperti pada tabel berikut.

Tabel 2.1 Data Subjek Penelitian Kecelakaan Lalu Lintas

Subjek (i)	t	Status	rest_use (X_1)	dr_drink (X_2)
1	4	1	1	1
2	9	1	0	1
3	10	0	0	0
4	17	1	1	0

Pada Tabel 2.1 di atas, kolom subjek (i) adalah kolom individu dan kolom t adalah waktu survival (dalam bulan). Pada kolom status, angka 1 menunjukkan subjek mengalami kejadian dan angka 0 untuk subjek yang tersensor. Pada kolom rest_use (X_1), angka 1 menunjukkan subjek menggunakan sabuk pengaman dan angka 0 untuk subjek yang tidak menggunakan sabuk pengaman. Pada kolom dr_drink (X_2), angka 1 menunjukkan subjek mengonsumsi alkohol dan angka 0 untuk subjek yang tidak mengonsumsi alkohol.

Pada data dapat diketahui bahwa individu pertama mengalami kejadian pada waktu $t = 4$, individu kedua mengalami kejadian pada waktu $t = 9$, individu ketiga pada waktu $t = 10$ tersensor, dan individu keempat mengalami kejadian pada waktu $t = 17$. Individu pertama dan keempat menggunakan sabuk pengaman saat berkendara, sedangkan individu kedua dan ketiga tidak menggunakan sabuk pengaman saat berkendara. Selain itu, individu pertama dan kedua mengonsumsi alkohol saat berkendara, sedangkan individu ketiga dan keempat tidak mengonsumsi alkohol saat berkendara. Dari informasi tersebut, akan dicari fungsi *partial likelihood* menggunakan fungsi *hazard*. Persamaan model Cox dengan dua variabel tersebut adalah sebagai berikut:

$$h(t, X) = h_0(t)\exp(\beta_1 \text{sb\_use} + \beta_2 \text{alc\_drink}) \tag{2.30}$$

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

Tabel 2.2 Fungsi *Hazard* Subjek Penelitian Kecelakaan Lalu Lintas

i	Fungsi *Hazard*
1	$h_0(t)\exp(\beta_1 + \beta_2)$
2	$h_0(t)\exp(\beta_2)$
3	$h_0(t)\exp(0)$
4	$h_0(t)\exp(\beta_1)$

Pada data tersebut telah diketahui bahwa individu ketiga tersensor, sehingga *partial likelihood* data tersebut merupakan perkalian *likelihood* dari tiga individu yang lain, yaitu individu pertama, kedua, dan keempat. Individu pertama mengalami kejadian pada bulan ke-5, sehingga keempat subjek tersebut merupakan himpunan risikonya. Nilai *likelihood* kejadian pertama (L_1) adalah sebagai berikut:

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

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

Individu kedua mengalami kejadian pada waktu $t = 8$, sehingga himpunan risikonya adalah individu kedua, ketiga dan keempat. Nilai *likelihood* kejadian kedua (L_2) adalah sebagai berikut:

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

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

Individu keempat mengalami kejadian terakhir pada bulan ke-16, maka tidak ada individu lain yang masuk sebagai himpunan risiko kecuali dirinya sendiri. Nilai *likelihood* kejadian keempat (L_3) adalah sebagai berikut:

$$L_3 = \frac{h_0(t)\exp(\beta_1)}{h_0(t)\exp(\beta_1)} = 1$$

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

$$L = L_1\, L_2\, L_3$$

$$L = \frac{h_0(t)\exp(\beta_1+\beta_2)}{h_0(t)\exp(\beta_1+\beta_2) + h_0(t)\exp(\beta_2) + h_0(t)\exp(0) + h_0(t)\exp(\beta_1)}$$

$$\frac{h_0(t)\exp(\beta_2)}{h_0(t)\exp(\beta_2) + h_0(t)\exp(0) + h_0(t)\exp(\beta_1)} \qquad (2.31)$$

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

$$L = \frac{\exp(\beta_1 + \beta_2)}{\exp(\beta_1 + \beta_2) + \exp(\beta_2) + \exp(0) + \exp(\beta_1)}$$

$$\frac{\exp(\beta_2)}{\exp(\beta_2) + \exp(0) + \exp(\beta_1)} \quad (2.32)$$

2. Prosedur *Newton-Raphson*

Prosedur *Newton-Raphson* digunakan untuk memaksimalkan fungsi *partial likelihood* dalam penaksiran parameter model Cox. Misalkan $L(\beta)$ merupakan fungsi *partial likelihood* p dimensi vektor $\beta = (\beta_1, \beta_2, \dots, \beta_p)^2$. Misalkan $U(\beta)$ merupakan vektor berukuran p dari turunan parsial pertama $\ln L_p(\beta)$ sehingga,

$$U(\beta) = [U_1(\beta), U_2(\beta), \dots, U_p(\beta)]^t \quad (2.33)$$

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

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

$$I(\beta) = (I_{ij}(\beta)) \text{ dengan } i, j = 1,2, \dots, p \quad (2.34)$$

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

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

Algoritma pada metode *Newton-Raphson* adalah sebagai berikut:

$$\hat{\beta}_{c+1} = \hat{\beta}_c - I(\hat{\beta}_c)^{-1} U(\hat{\beta}_c) \quad (2.35)$$

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

Langkah iterasi dengan metode Newton-Raphson sebagai berikut:

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

Varians dari $\hat{\beta}_j$, yaitu:

$$Var\left(\hat{\beta}\right) = I\left(\hat{\beta}\right)^{-1} \tag{2.36}$$

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

$$SE\left(\hat{\beta}\right) = \sqrt{Var(\hat{\beta})} = \sqrt{I\left(\hat{\beta}\right)^{-1}} \tag{2.37}$$

Simpangan baku dapat digunakan untuk mencari selang kepercayaan $\hat{\beta}_j$, yaitu $(1-\alpha)100\%$, sedangkan kepercayaan untuk $\hat{\beta}_j$ adalah sebagai berikut:

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

3. Pengujian Parameter

Ada tiga tujuan statistik, yaitu (1) untuk menguji signifikansi parameter, (2) memperoleh estimasi titik, dan (3) memperoleh interval konfidensi. Menurut David W. Hosmer & Stanley Lemeshow (2008), terdapat tiga cara untuk menguji signifikansi parameter model Cox, yaitu dengan uji *partial likelihood ratio*, uji *Wald*, dan uji *Score*. Pengujian signifikansi parameter bertujuan untuk memeriksa apakah variabel bebas memiliki pengaruh yang nyata dalam model.

a. Uji *Partial Likelihood Ratio*

Uji *partial likelihood ratio* dinotasikan dengan G. Statistik uji ini digunakan untuk menguji hipotesis bahwa satu atau beberapa parameter regresi β_j untuk model pada Persamaan (2.15) adalah nol. Langkah-langkah dalam uji *partial likelihood ratio* adalah sebagai berikut:

1. Hipotesis:

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

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

2. Taraf signifikansi: α

3. Statistik uji:

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

   dengan,

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

   $\ln L(\hat{\beta}_j)$ = 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-value* $\leq \alpha$, dengan p adalah banyaknya variabel bebas.

5. Kesimpulan:

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

b. Uji *Wald*

Uji *Wald* digunakan untuk menguji pengaruh parameter secara terpisah yang dinotasikan dengan z. Statistik uji ini mengikuti distribusi *chi-square* dengan derajat bebas p. Langkah-langkah uji *Wald* untuk model pada Persamaan (2.15) sebagai berikut:

1. Hipotesis:

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

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

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

5. Kesimpulan:

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

c. Uji *Score*

Selain uji *partial likelihood ratio* dan uji *Wald*, terdapat pula uji *Score* untuk menguji signifikansi parameter. Statistik uji *Score* adalah rasio dari turunan *log partial likelihood* pada Persamaan (2.26), dengan akar kuadrat dari Persamaan (2.29). Statistik uji ini juga mengikuti sebaran distribusi *chi-square* dengan derajat bebas p . Langkah-langkah dalam uji *Score* untuk model pada Persamaan (2.15) sebagai berikut:

1. Hipotesis:

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

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

2. Taraf signifikansi: α

3. Statistik uji:

$$z^2 = \left(\frac{\frac{\partial L \beta_j}{\partial \beta_j}}{\sqrt{I(\beta_j)}} \middle| \beta_j = 0 \right)^2 \tag{2.40}$$

4. Kriteria keputusan:

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

5. Kesimpulan:

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

D. Pengujian Asumsi *Proportional Hazard*

Asumsi terpenting yang harus dipenuhi dalam regresi Cox adalah asumsi *proportional 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 (2012), jika rasio fungsi *hazard* tidak konstan dan asumsi *proportional hazard* tidak dipenuhi, maka model Cox tidak berlaku. Menurut David Collett (2003), 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.

1. Pengujian Asumsi *Proportional 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:

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

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 diintegralkan dari nol hingga t, maka diperoleh sebagai berikut:

$$\int_0^t h(t, X) dt = \exp\left(\sum_{j=1}^{p} \beta_i X_i\right) \int_0^t h_0(t) dt \tag{2.42}$$

menggunakan persamaan (2.11) dapat diperoleh:

$$H(t, X) = \exp\left(\sum_{j=1}^{p} \beta_i X_i\right) H_0(t) \tag{2.43}$$

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

$$\log H(t) = \log \exp\left(\sum_{j=1}^{p} \beta_i X_i\right) + \log H_0(t) \tag{2.44}$$

Persamaan di atas ekuivalen dengan persamaan berikut

$$\log[-\log(S(t))] = \sum_{j=1}^{p} \beta_i X_i + \log[-\log(S_0(t))] \tag{2.45}$$

Dari 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 diplotkan 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 nilainya perlu dikelompokkan menjadi variabel kategori. Jika pada plot *log-minus-log survival* menunjukkan kurva yang pararel, maka asumsi *proportional hazard* tidak terpenuhi. Kelemahan menggunakan pendekatan plot *log-minus-log survival* adalah bersifat subjektif, pararel atau tidaknya kurva sangat bergantung pada cara peneliti menilai (Collett, 2003).

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

Terdapat beberapa jenis residual yang dapat digunakan untuk melakukan pengujian asumsi *proportional hazard*, antara lain residual *Cox Snell*, residual *deviance*, dan residual *Schoenfeld* (Lee & Wang, 2003). Pada penelitian ini akan digunakan residual *Schoenfeld* untuk menguji asumsi *proportional hazard*. Residual *Schoenfeld* didefinisikan hanya pada waktu survival yang tidak tersensor. Residual *Schoenfeld* untuk individu ke-*i* pada kovariat ke-*j* adalah sebagai berikut:

$$R_{ji} = \delta_i \left(x_{ji} - \frac{\sum_{l \in R(t_{(i)})} x_{ji} \exp(\widehat{\beta}' x_{ji})}{\sum_{l \in R(t_{(i)})} \exp(\widehat{\beta}' x_{ji})} \right), j = 1,2, \dots, p \qquad (2.46)$$

dengan $\hat{\beta}$ merupakan estimator *partial likelihood* maksimum dari β. Karena $\hat{\beta}$ merupakan solusi dari persamaan turunan pertama fungsi *log likelihood* pada Persamaan (2.27), 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, sehingga residual *Schoenfeld* tidak berkorelasi dengan yang lainnya.

Adapun langkah-langkah pengujiannya sebagai berikut:

1. Mencari taksiran model Cox *proportional hazard* dan mencari residual *Schoenfeld* untuk masing-masing kovariat.
2. Membuat variabel rank waktu survival, yaitu waktu terjadi kejadian yang diurutkan. Individu yang mengalami kejadian pertama kali diberi nilai 1, individu yang mengalami kejadian selanjutnya diberi nilai 2, dan seterusnya.
3. Menguji korelasi antara variabel pada langkah pertama dan kedua. H_0 adalah korelasi antara residual *Schoenfeld* dan rank waktu survival adalah nol. Penolakan H_0 berarti asumsi *proportional hazard* tidak dipebuhi.

Menurut Therneau & Grambsch (2000), pada residual *Schoenfeld* jika *p-value* $< 0{,}05$, maka kovariat yang diuji tidak memenuhi asumsi *proportional hazard*. Langkah-langkah uji residual *Schoenfeld* adalah sebagai berikut:

1. Hipotesis:

   $H_0: \rho = 0$ (asumsi *proportional hazard* terpenuhi)

   $H_1: \rho \neq 0$ (asumsi *proportional hazard* tidak terpenuhi)
2. Taraf signifikansi: α
3. Statistik uji:

$$\chi^2_{hit} = \frac{\left\{\sum(g_j - \bar{g})R_{ji}\right\}^2}{d\, I_j \sum\left(g_j - \bar{g}\right)^2}$$

dengan,

R_{ji} = residual *Schoenfeld* untuk masing-masing kovariat

g_j = variabel rank waktu survival

$\bar{g}$ = rata-rata rank waktu survival

I_j = matriks elemen untuk kovariat

d = waktu kejadian

4. Kriteria keputusan:

   H_0 ditolak jika $\chi^2_{hit} \geq \chi^2_{(0,05;1)}$ atau *p-value* $< \alpha$

5. Kesimpulan:

   Jika H_0 ditolak maka $\rho \neq 0$, artinya terdapat korelasi antara residual *Schoenfeld* dengan variabel rank waktu survival sehingga asumsi *proportional hazard* tidak terpenuhi.