



UNIVERSITAS INDONESIA

**PENGECEKAN ASUMSI *PROPORTIONAL HAZARD*
PADA MODEL COX PH**

SKRIPSI

**RONI TUA YOHANES
0606067793**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
PROGRAM STUDI SARJANA MATEMATIKA
DEPOK
JUNI 2011**



UNIVERSITAS INDONESIA

**PENGECEKAN ASUMSI *PROPORTIONAL HAZARD*
PADA MODEL COX PH**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
sarjana sains**

**RONI TUA YOHANES
0606067793**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
PROGRAM STUDI SARJANA MATEMATIKA
DEPOK
JUNI 2011**

HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.

Nama : Roni Tua Yohanes

NPM : 0606067793

Tanda Tangan : 

Tanggal : 10 Juni 2011

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :
Nama : Roni Tua Yohanes
NPM : 0606067793
Program Studi : Matematika
Judul Skripsi : Pengecekan Asumsi *Proportional Hazard*
pada Model Cox PH

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia

DEWAN PENGUJI

Pembimbing	: Sarini Abdullah, S.Si., M.Stats.	()
Penguji	: Fevi Novkaniza, S.Si, M.Si.	()
Penguji	: Mila Novita, S.Si, M.Si.	()
Penguji	: Dra. Saskya Mary Soemartojo, M.Si.	()

Ditetapkan di : Depok
Tanggal : 10 Juni 2011

KATA PENGANTAR

Segala puji syukur kepada Tuhan Yesus atas kasih, berkat, dan penyertaan-Nya di dalam sepanjang hidup. Puji syukur juga kepada-Nya atas kekuatan yang diberikan sehingga skripsi ini dapat diselesaikan. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Sains Jurusan Matematika pada Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia.

Penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi penulis untuk menyelesaikan skripsi ini. Oleh karena itu, penulis mengucapkan terima kasih kepada berbagai pihak sebagai berikut:

1. Dosen pembimbing penulis, Sarini Abdullah, S.Si, M.Stats, yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan penulis dalam penyusunan skripsi ini. Terima kasih juga untuk nasehat, doa, dukungan, serta kesabaran yang telah diberikan di dalam suka dan duka selama penyusunan skripsi ini.
2. Dr. Kiki Ariyanti Sugeng, selaku pembimbing akademik penulis yang telah memberikan arahan, masukan, dan dukungan selama lima tahun masa perkuliahan penulis.
3. Ketua dan Sekretaris Departemen Matematika, Dr. Yudi Satria dan Rahmi Rusin, M.ScTech, atas segala bantuan serta dukungan yang telah diberikan.
4. Dosen-dosen di Matematika, terima kasih atas ilmu yang diberikan kepada penulis selama masa kuliah.
5. Seluruh staf Tata Usaha, staf Perpustakaan, serta karyawan Departemen Matematika, terima kasih atas segala bantuannya.
6. Papa dan M a m a , yang telah memberikan bantuan material, dukungan dan doa, terima kasih atas segala perhatian, kasih sayang, kesabaran, dan berbagai nasehat yang telah diberikan kepada penulis.

7. Adik-adik serta keluarga penulis lainnya terima kasih atas doa dan dukungannya.
8. Teman-teman terdekat (Novi, Lani, Tika, Ranti, Stefani). Terima kasih atas bantuan selama perkuliahan dan waktu-waktu menyenangkan bersama.
9. Teman-teman angkatan 2006, terima kasih atas bantuan dan kebersamaan selama perkuliahan. Semoga sukses untuk kita semua.
10. Teman-teman lain dan pihak-pihak yang tidak dapat disebutkan satu-persatu, terima kasih atas dukungan dan doanya.

Semoga Tuhan Yang Maha Esa membalas segala kebaikan semua pihak yang telah membantu. Semoga skripsi ini membawa manfaat bagi pengembangan ilmu pengetahuan.

Penulis
2011

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Indonesia, saya yang bertanda tangan di bawah ini:

Nama : Roni Tua Yohanes
NPM : 0606067793
Program Studi : S1
Departemen : Matematika
Fakultas : MIPA (Matematika dan Ilmu Pengetahuan Alam)
Jenis karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Indonesia Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul :

Pengecekan Asumsi *Proportional Hazard* pada Model Cox PH

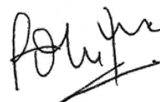
beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Indonesia berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Depok

Pada tanggal : 10 Juni 2011

Yang menyatakan



(Roni Tua Yohanes)

ABSTRAK

Nama : Roni Tua Yohanes
Program Studi : Matematika
Judul : Pengecekan Asumsi *Proportional Hazard* pada Model Cox PH

Pada penelitian survival, kadang kala *survival time* dipengaruhi oleh faktor-faktor lain. Untuk kondisi tersebut dapat digunakan metode analisis regresi dengan model Cox PH. Dari model Cox ini, diperoleh bahwa *hazard ratio* untuk dua individu dengan nilai kovariat yang berbeda tidak akan dipengaruhi oleh waktu. Atau dengan perkataan lain, hazard untuk satu individu proporsional dengan hazard individu lainnya dengan keproporsionalan yang konstan, tidak dipengaruhi oleh waktu. Oleh sebab itu, ketika diterapkan pada data harus diperiksa apakah asumsi tersebut terpenuhi. Pengecekan asumsi *proportional hazard* akan dilakukan dengan dua pendekatan. Pendekatan yang pertama dengan grafik, yaitu grafik log-log, dan yang kedua dengan pengujian goodness-of-fit. Metode grafik memberikan hasil yang subjektif sedangkan pengujian goodness-of-fit memberikan hasil yang objektif berdasarkan pengujian statistik.

Kata Kunci : analisis survival, *survival time*, model Cox, *proportional hazard*, grafik log-log, goodness-of-fit.

xiii+70 halaman : 7 gambar; 7 tabel

Daftar Pustaka : 10 (1971-2010)

ABSTRACT

Name : Roni Tua Yohanes
Program Study : Mathematics
Title : Checking the Proportional Hazard Assumption in Cox PH Model

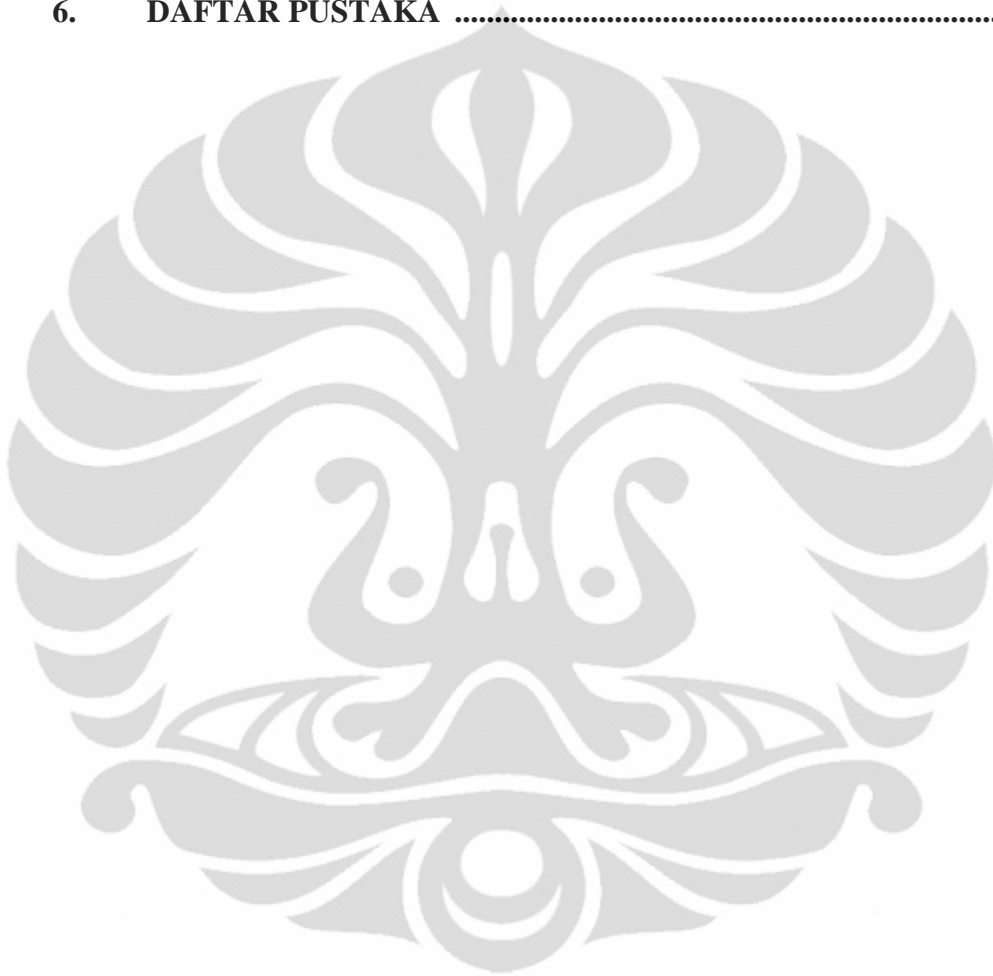
In survival studies, sometimes the time until the occurrence of an event is influenced by other factors. Cox PH model, which is one of the semi parametric regression method can be applied for the data analysis. From this Cox model, it is found that the hazard ratio for two individuals with different covariates not be affected by time. In other words, hazard for an individual is proportional with hazard from another individual with constant proportionality, not affected by the time. Therefore, when applied to the data, it should be checked whether the assumptions are met. The assumption of proportional hazard will be checked using two approaches. The first approach using graph, the log-log graph, and the second by testing the goodness-of-fit. Graphical method gives subjective results while the goodness-of-fit testing gives objective results based on statistical testing.

Key Words : survival analysis, survival time, Cox model, proportional hazard, log-log graph, goodness-of-fit.
xiii+70 pages : 7 pictures; 7 tables
Bibliography : 10 (1971-2010)

DAFTAR ISI

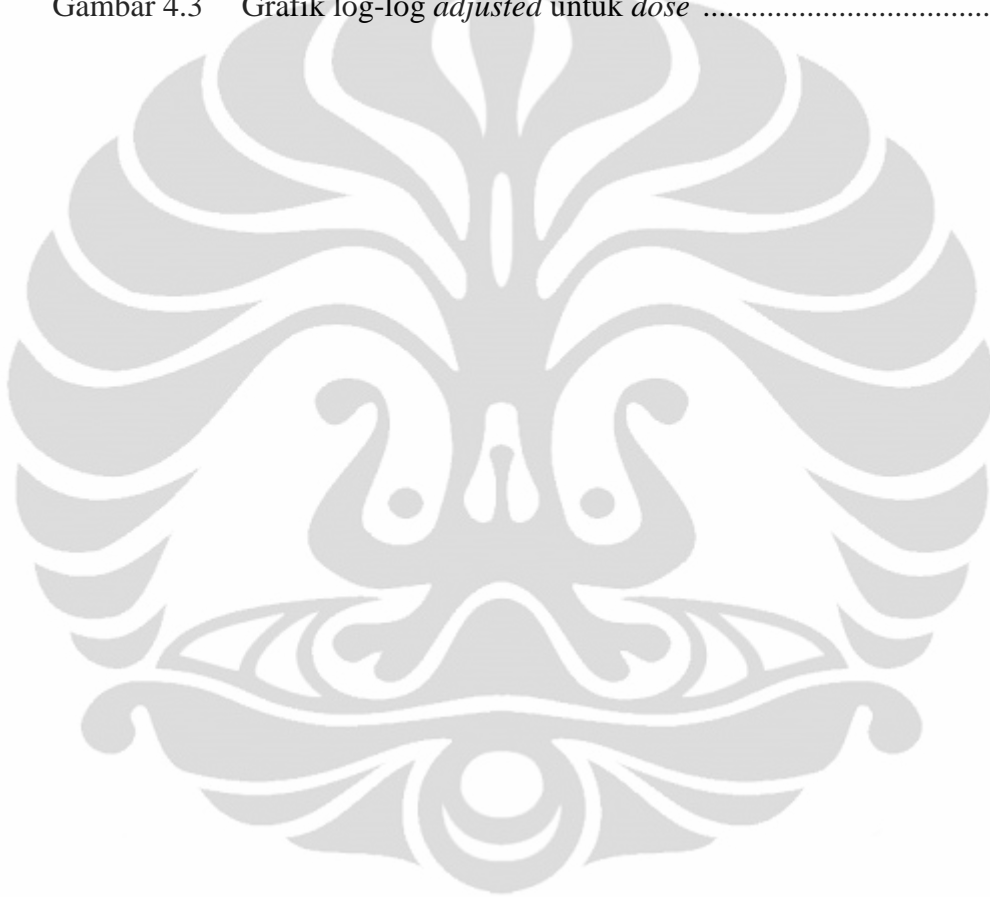
HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS	ii
LEMBAR PENGESAHAN	iii
KATA PENGANTAR	iv
LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH	vi
ABSTRAK	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR LAMPIRAN	xiii
1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Tujuan Penulisan	2
1.4 Pembatasan Masalah	3
1.5 Sistematika Penulisan	3
2. LANDASAN TEORI	5
2.1 <i>Survival Time</i>	5
2.2 <i>Cumulative Distribution Function</i>	5
2.3 Fungsi Survival	6
2.4 Fungsi Hazard	7
2.5 Penyensoran Data	10
2.5.1 Data Tersensor Kanan	10
2.5.2 Data Tersensor Kiri	11
2.5.3 Data Tersensor Interval	12
2.6 Taksiran Kaplan-Meier	14
2.7 Model Cox	18
2.7.1 Definisi dan Karakteristik Model	18
2.7.2 Partial Likelihood	20
2.8 Koefisien Korelasi	23
3. Pengecekan Asumsi Proportional Hazard	24
3.1 Pendekatan Grafik: Grafik Log-log Survival	24
3.1.1 Pengecekan untuk Variabel Kontinu	26
3.1.2 Pengecekan untuk Beberapa Variabel	28
3.2 Pendekatan dengan Pengujian Goodness of Fit (GOF)	30
3.2.1 Schoenfeld Residual	30
3.2.2 Koefisien Korelasi Rank Spearman	32
3.2.3 Pengujian Goodness of Fit	37
4. Contoh Penerapan	39
4.1 Data	39

4.2	Analisis Data	39
4.2.1	Pengecekan dengan Grafik Log-log	41
4.2.2	Pengecekan dengan Goodness of Fit	46
5.	PENUTUP	49
5.1	Kesimpulan	49
5.2	Saran	49
6.	DAFTAR PUSTAKA	50



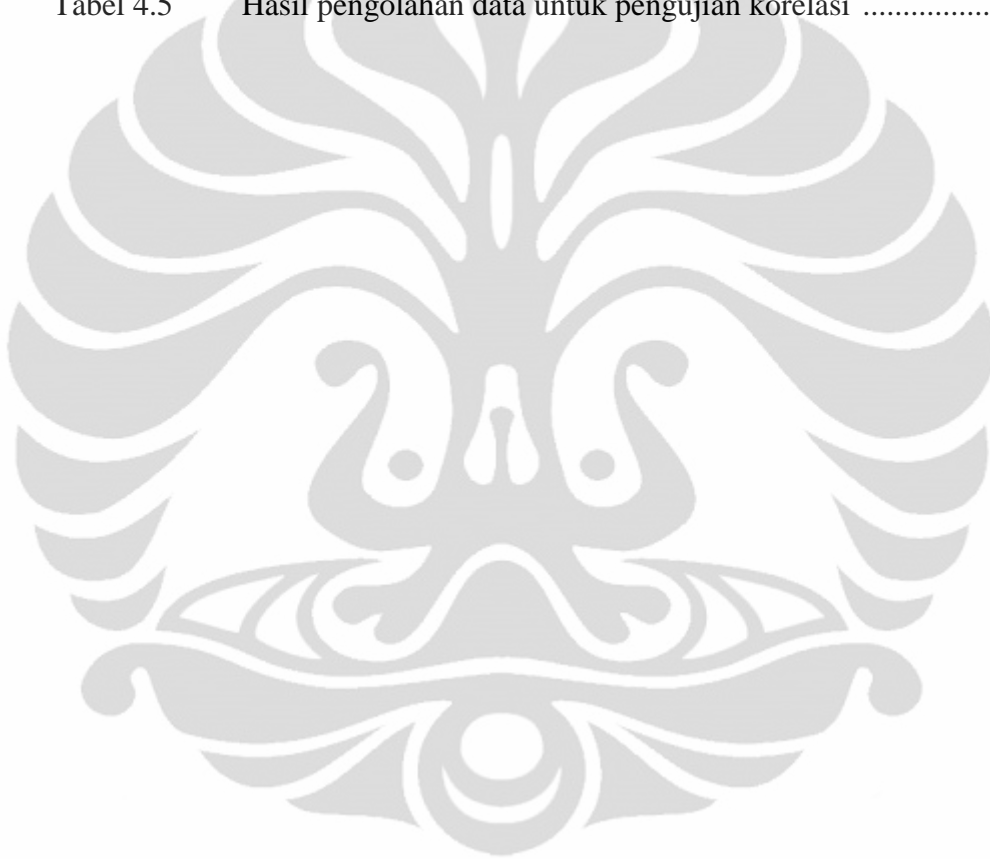
DAFTAR GAMBAR

Gambar 2.1	Contoh data eksak dan tersensor kanan	11
Gambar 2.2	Contoh data eksak dan tersensor kiri	12
Gambar 2.3	Contoh data tersensor interval	13
Gambar 3.1	Grafik log-log survival terhadap waktu	26
Gambar 4.1	Grafik log-log <i>adjusted</i> untuk <i>clinic</i>	43
Gambar 4.2	Grafik log-log <i>adjusted</i> untuk <i>prison</i>	44
Gambar 4.3	Grafik log-log <i>adjusted</i> untuk <i>dose</i>	45



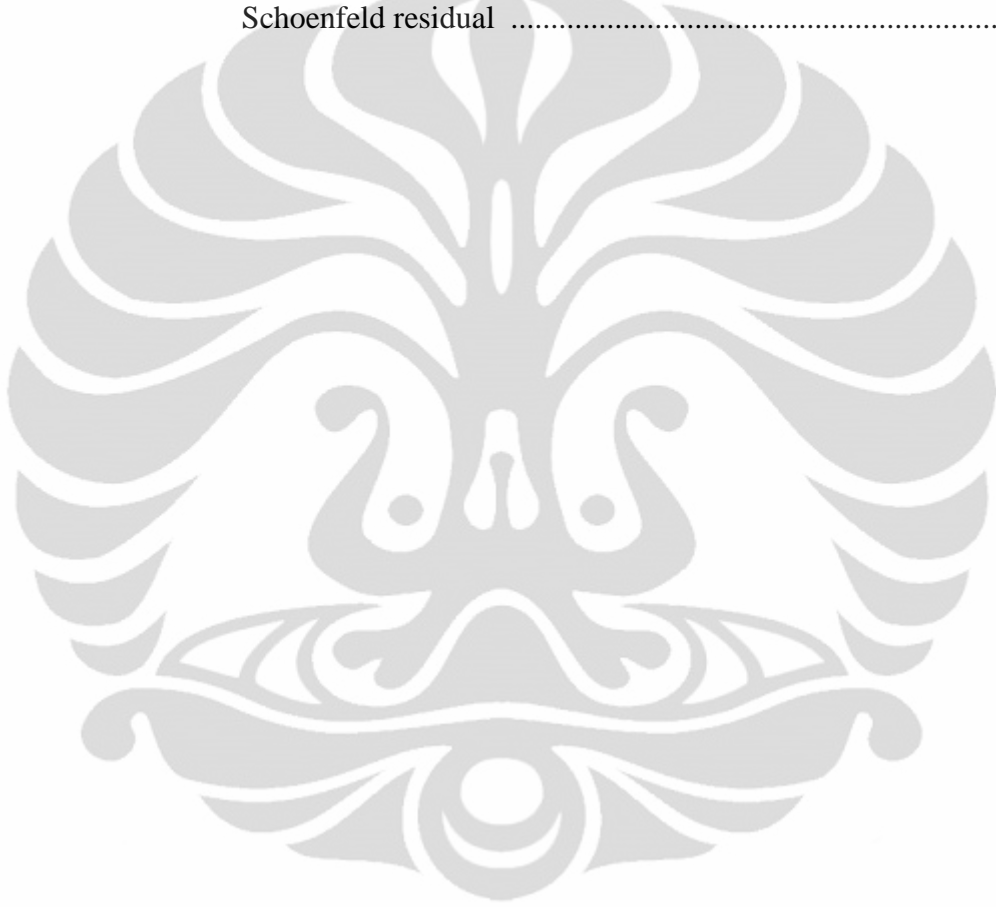
DAFTAR TABEL

Tabel 3.1	Komponen t_i , R_i , (X_1, \dots, X_p) , dan (r_1, \dots, r_p) untuk sampel n individu dengan k individu yang <i>survival time</i> -nya teramati ..38
Tabel 3.2	Komponen t_i dan r_i yang akan dilakukan pengecekan asumsi ...38
Tabel 4.1	Hasil pengolahan data untuk pengujian <i>likelihood ratio</i>40
Tabel 4.2	Hasil pengolahan data untuk pengujian Wald41
Tabel 4.3	Rata-rata untuk masing-masing variabel42
Tabel 4.4	Hasil pengolahan data untuk Schoenfeld residual46
Tabel 4.5	Hasil pengolahan data untuk pengujian korelasi48



DAFTAR LAMPIRAN

Lampiran 1	Pembuktian persamaan 3.29	52
Lampiran 2	Pembuktian $\text{var}(R(Y_a)) = (n^2 - 1)/12$	56
Lampiran 3	Pembuktian $\text{cov}(R(Y_a), R(Y_b)) = -(n+1)/12$	57
Lampiran 4	Tabel data Caplehorn dari 238 pecandu heroin	59
Lampiran 5	Tabel data Caplehorn dari 238 pecandu heroin dengan Schoenfeld residual	65



BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pada studi survival, biasanya peneliti tertarik untuk melihat waktu hingga terjadinya suatu event tertentu. Akan tetapi, pengamatan hanya berdasarkan waktu kurang memberikan informasi yang memadai. Terkadang, waktu survival tersebut dipengaruhi oleh faktor-faktor lain, misalnya umur, jenis kelamin, konsumsi alkohol, tekanan darah, level gula darah, dan lain-lain.

Beberapa metode analisis tersedia untuk mendapatkan informasi dari data survival, misalnya metode non parametrik dengan grafik fungsi survival menggunakan estimasi Kaplan-Meier. Akan tetapi, kelemahannya adalah hanya melihat waktu survival saja tidak mengakomodir keberadaan informasi dari pengukuran lainnya. Metode parametrik, misalnya dengan regresi, memberikan hasil yang lebih baik. Hal ini dikarenakan pada regresi parametrik dapat diketahui pola dan kekuatan hubungan antara waktu survival dengan variabel-variabel lain (dikenal kovariat). Akan tetapi, kelemahannya adalah asumsi regresi parametrik yang sulit terpenuhi pada data sebenarnya.

Dengan kondisi-kondisi tersebut, akan digunakan metode lain untuk memodelkan antara waktu survival dengan kovariat-kovariatnya. Memodelkan pengaruh kovariat-kovariat pada survival dapat dilakukan dengan memodelkan *conditional hazard* sebagai fungsi dari kovariat-kovariat. Pada tugas akhir ini akan digunakan model *Cox proportional hazard* (model Cox PH). Pada model Cox PH, *conditional hazard* dari individu dengan kovariat-kovariat merupakan produk dari *baseline hazard* dan fungsi eksponen kovariat-kovariatnya.

Fungsi *baseline hazard* merupakan fungsi dari waktu (t) tetapi tidak melibatkan kovariat (X) sedangkan bagian ekspresi eksponensial melibatkan kovariat tetapi tidak melibatkan waktu. Kovariat (X) di sini disebut *time-independent* kovariat. Variabel-variabel X yang *time-independent* merupakan variabel-variabel yang seiring dengan berjalannya waktu tidak akan mengalami perubahan nilai. Dimungkinkan juga untuk menganggap variabel-variabel X melibatkan waktu. Variabel yang demikian disebut variabel *time-dependent*. Jika

menggunakan variabel *time-dependent*, model Cox dapat digunakan, tetapi model tersebut tidak memenuhi asumsi *proportional hazard* dan disebut model extended Cox.

Dari model Cox, kita mengenal *hazard ratio* (HR) yaitu perbandingan hazard dari dua individu dengan nilai kovariat-kovariat yang berbeda. Untuk dua individu dengan masing-masing nilai kovariatnya, maka *hazard ratio* dari kedua individu untuk mengalami event pada suatu waktu t , hanya bergantung pada nilai-nilai dari kovariatnya. Dapat dikatakan, hazard untuk satu individu proporsional dengan hazard individu lainnya dengan keproporsionalan yang konstan, tidak dipengaruhi oleh waktu. *Hazard ratio* konstan terhadap waktu inilah yang disebut sebagai *proportional hazard* (PH).

Mengingat kemudahan dan keunggulan dari model Cox PH seperti dijelaskan di atas, maka model ini seringkali digunakan dalam analisis survival. Ada beberapa hal yang harus dipenuhi dalam penggunaan model Cox PH, diantaranya adalah bahwa data memang memenuhi asumsi *proportional hazard*. Oleh karena itu, perlu diperiksa lebih lanjut apakah memang asumsi tersebut dipenuhi.

Jadi, pada tugas akhir ini akan dibahas dua pendekatan yang dapat dipakai untuk memeriksa asumsi *proportional hazard*. Pendekatan yang pertama dengan menggunakan grafik dan yang kedua dengan pengujian goodness-of-fit. Dengan metode grafik, digunakan grafik log-log.

1.2 Perumusan Masalah

Perumusan masalah yang diajukan pada tugas akhir ini adalah:

Bagaimana mengecek asumsi *proportional hazard* dalam model Cox PH?

1.3 Tujuan Penulisan

Tujuan penulisan tugas akhir ini adalah mengecek asumsi *proportional hazard* dalam model Cox PH dengan

- 1) metode grafik
- 2) metode goodness-of-fit

1.4 Pembatasan Masalah

Permasalahan pada tugas akhir ini dibatasi untuk:

- 1) data tersensor kanan tipe I yang non informatif,
- 2) variabel penjelas (kovariat) yang *time-independent*,
- 3) tidak ada yang *ties* pada data *time to event*.

1.5 Sistematika Penulisan

Bab 1: Bab 1 berisi Pendahuluan.

Terdiri dari latar belakang, perumusan masalah, tujuan penulisan, pembatasan masalah, dan sistematika penulisan.

Bab 2: Bab 2 berisi Landasan Teori.

Pada bab ini dijelaskan mengenai *survival time*, *cumulative distribution function*, fungsi survival, fungsi hazard, penyensoran data, taksiran Kaplan-Meier fungsi survival, model Cox, dan koefisien korelasi. Pada penyensoran data dibahas data tersensor kanan dan tersensor kiri. Pada model Cox dibahas definisi dan karakteristik model, *hazard ratio*, dan *partial likelihood*.

Bab 3: Pada bab 3 dijelaskan mengenai Pengecekan Asumsi Proporsional Hazard pada Model Cox PH.

Penjelasannya meliputi pendekatan grafik dan pengujian goodness-of-fit. Pada pendekatan grafik digunakan grafik log-log.

Bab 4: Bab 4 berisi Contoh Penerapan pada Data.

Terdiri dari penjelasan mengenai data yang digunakan pada bab ini dan pengecekan asumsi pada model Cox berdasarkan data.

Pengecekan dilakukan dengan pendekatan grafik log-log dan goodness-of-fit.

Bab 5: Bab 5 berisi Penutup.

Terdiri dari kesimpulan yang diperoleh dari tugas akhir ini dan saran untuk pengembangan tugas akhir ini.



BAB 2

LANDASAN TEORI

2.1 *Survival Time*

Data survival berhubungan dengan waktu sampai terjadinya kejadian (event) yang diperhatikan. *Survival time* merupakan variabel yang mengukur waktu dari suatu titik awal sampai ke suatu titik akhir yang menjadi perhatian. Dalam menyatakan *survival time*, digunakan skala pengukuran waktu seperti tahun, bulan, hari, jam, menit, detik, atau lainnya.

Dalam pengamatan, perlu diperhatikan pendefinisian titik awal (waktu awal pengamatan) dan titik akhir (waktu akhir pengamatan). Pendefinisian waktu akhir termasuk mudah contohnya kambuhnya rasa sakit, sembuh dari penyakit, atau kematian. Mendefinisikan awal kejadian yang menjadi patokan waktu awal pengamatan lebih susah untuk dilakukan. Misalnya dalam bidang medis untuk pengamatan waktu kematian. Waktu awal dapat didefinisikan sebagai waktu mulai terdiagnosis penyakitnya tetapi ada juga yang mendefinisikan waktu awalnya adalah waktu mulai terinfeksi penyakitnya.

Survival time selalu non negatif dan dapat berupa data eksak, data tersensor, atau data terpancung.

2.2 *Cumulative Distribution Function*

Misalkan variabel random T menunjukkan *survival time* dari individu dalam populasi, T merupakan variabel random non negatif dalam interval $[0, \infty)$. *Cumulative distribution function* atau c.d.f. dari variabel T , dinyatakan $F(t)$, adalah probabilitas bahwa variabel akan lebih kecil atau sama dengan nilai t apapun yang dipilih. Maka,

$$F(t) = \Pr(T \leq t). \quad (2.1)$$

Dikenal juga *probability density function* atau p.d.f. yaitu probabilitas variabel pada saat bernilai t dinyatakan $f(t)$. p.d.f juga merupakan turunan dari c.d.f.

2.3 Fungsi Survival

Fungsi survival dapat didefinisikan sebagai probabilitas individu dapat *survive* lebih dari waktu t . Secara matematis dinyatakan sebagai

$$S(t) = \Pr(T > t). \quad (2.2)$$

Jika T merupakan variabel random kontinu, maka fungsi survival merupakan komplemen dari c.d.f. yaitu

$$S(T) = \Pr(T > t) = 1 - \Pr(T \leq t) = 1 - F(t). \quad (2.3)$$

Selain itu, fungsi survival dinyatakan dalam p.d.f. sebagai integral dari p.d.f., $f(t)$, yaitu

$$S(t) = \Pr(T > t) = \int_t^{\infty} f(u) du \quad (2.4)$$

karena p.d.f. merupakan turunan dari c.d.f. maka,

$$f(t) = \frac{dF(t)}{dt} = \frac{d(1 - S(t))}{dt} = -\frac{dS(t)}{dt}. \quad (2.5)$$

Untuk T merupakan variabel diskrit, misalkan T memiliki nilai pada $t_j, j = 1, 2, \dots, n$ dengan *probability mass function*

$$p(t_j) = \Pr(T = t_j), j = 1, 2, \dots, n \quad (2.6)$$

dengan $t_1 < t_2 < \dots < t_n$ Fungsi survival untuk variabel random T adalah

$$S(t) = \Pr(T > t) = \sum_{t_j > t} p(t_j). \quad (2.7)$$

Fungsi survival, $S(t)$, dapat diplot pada grafik yang menggambarkan probabilitas individu akan *survive* pada beberapa titik waktu t antara 0 sampai ∞ . Fungsi survival memiliki sifat-sifat sebagai berikut:

- fungsi monoton tak naik,
- saat $t = 0, S(t) = 1$; $t = 0$ menunjukkan waktu awal pengamatan. Pada awal pengamatan belum ada individu yang mengalami kejadian sehingga probabilitas survival pada saat itu adalah 1.

- saat $t \rightarrow \infty$, $S(t) \rightarrow 0$; $t \rightarrow \infty$ menunjukkan waktu pengamatan yang berlangsung tanpa batas yang pada akhirnya tidak ada seorangpun yang *survive* sehingga fungsi survival menuju 0.

2.4. Fungsi Hazard

Fungsi hazard didefinisikan sebagai *instaneous rate* suatu individu untuk mengalami event dalam interval waktu dari t sampai $t+\Delta t$ jika diketahui individu tersebut masih dapat bertahan hidup sampai dengan waktu t .

Secara matematis dapat dituliskan sebagai berikut:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \quad (2.8)$$

Jika T adalah suatu variabel random kontinu, hazard dapat dinyatakan dalam bentuk lain dengan menghubungkan p.d.f dan fungsi survival yaitu:

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

Berdasarkan pembahasan sebelumnya (2.5) diketahui bahwa

$$f(t) = -\frac{dS(t)}{dt} \text{ sehingga } h(t) \text{ dapat dinyatakan sebagai berikut:}$$

$$\begin{aligned}
 h(t) &= \frac{f(t)}{S(t)} \\
 &= -\frac{dS(t)}{dt} \cdot \frac{1}{S(t)} \\
 &= -\frac{dS(t)}{dt} \cdot \frac{d \ln S(t)}{dS(t)} \\
 h(t) &= -\frac{d \ln S(t)}{dt}. \tag{2.10}
 \end{aligned}$$

Dengan mengintegalkan kedua sisi persamaan di atas akan memberikan bentuk fungsi survival yang dinyatakan dalam fungsi hazard:

$$\begin{aligned}
 h(t) &= -\frac{d \ln S(t)}{dt} \\
 \Leftrightarrow \int_0^t h(u) du &= \int_0^t -\frac{d \ln S(u)}{du} du \\
 \Leftrightarrow \int_0^t h(u) du &= -\ln S(u) \Big|_0^t \\
 \Leftrightarrow -\int_0^t h(u) du &= \ln S(t) - \ln S(0) \\
 \Leftrightarrow -\int_0^t h(u) du &= \ln S(t) \\
 S(t) &= \exp \left[-\int_0^t h(u) du \right]. \tag{2.11}
 \end{aligned}$$

Dikenal juga *cumulative hazard function* $H(t)$, yang didefinisikan sebagai

$$H(t) = \int_0^t h(u) du. \tag{2.12}$$

Berdasarkan hasil sebelumnya (2.11), diperoleh persamaan

$$S(t) = \exp[-H(t)] \tag{2.13}$$

atau

$$H(t) = -\ln[S(t)] \tag{2.14}$$

Jika T adalah variabel random diskret, maka fungsi hazard diberikan oleh:

$$\begin{aligned}
 h(t_j) &= \Pr(T = t_j | T \geq t_j), \quad j = 1, 2, \dots, n \\
 &= \frac{\Pr(T = t_j \cap T \geq t_j)}{\Pr(T \geq t_j)} \\
 &= \frac{\Pr(T = t_j)}{\Pr(T \geq t_j)} \\
 &= \frac{\Pr(T = t_j)}{\Pr(T > t_{j-1})} \\
 h(t_j) &= \frac{p(t_j)}{S(t_{j-1})}, \tag{2.15}
 \end{aligned}$$

dengan $S(t_0) = 1$.

Diketahui bahwa $S(t_j) = \Pr(T > t_j)$ sehingga

$$\begin{aligned}
 S(t_{j-1}) &= \Pr(T > t_{j-1}) \\
 &= \Pr(T \geq t_j) \\
 &= \Pr(T = t_j) + \Pr(T > t_j) \\
 &= p(t_j) + S(t_j)
 \end{aligned}$$

maka,

$$p(t_j) = S(t_{j-1}) - S(t_j). \tag{2.16}$$

Berdasarkan hasil tersebut, $h(t_j)$ dapat dinyatakan sebagai

$$h(t_j) = \frac{p(t_j)}{S(t_{j-1})} = \frac{S(t_{j-1}) - S(t_j)}{S(t_{j-1})} = 1 - \frac{S(t_j)}{S(t_{j-1})}, \quad j = 1, 2, \dots, n \tag{2.17}$$

Seperti fungsi survival, fungsi hazard juga dapat diplot sebagai kurva fungsi hazard terhadap nilai t . Namun, berbeda dengan fungsi survival, kurva $h(t)$ tidak harus dimulai dari 1 dan bergerak ke bawah menuju 0, tetapi kurva $h(t)$ bisa dimulai dari nilai berapapun ($h(t) \geq 0$) dan bergerak ke atas dan ke bawah terhadap waktu t . Dengan perkataan lain, untuk suatu nilai tertentu t , fungsi hazard $h(t)$ mempunyai karakteristik seperti berikut:

- selalu bernilai non negatif, $h(t) \geq 0$,
- tidak memiliki batas atas.

2.5 Penyensoran Data

Data survival dapat berupa data eksak, data tersensor, atau data terpancung. Data eksak terjadi ketika waktu terjadinya kejadian (event) tepat diketahui. Data tersensor terjadi ketika waktu terjadinya kejadian tidak diketahui dan hanya diketahui beberapa informasi mengenai waktu sampai terjadinya kejadian. Data terpancung terjadi karena adanya penyaringan beberapa subyek sehingga pengamat tidak memperhatikan mereka. Hanya individu-individu dengan pengalaman tertentu, mengalami kejadian sebelum kejadian yang diperhatikan, yang akan diamati.

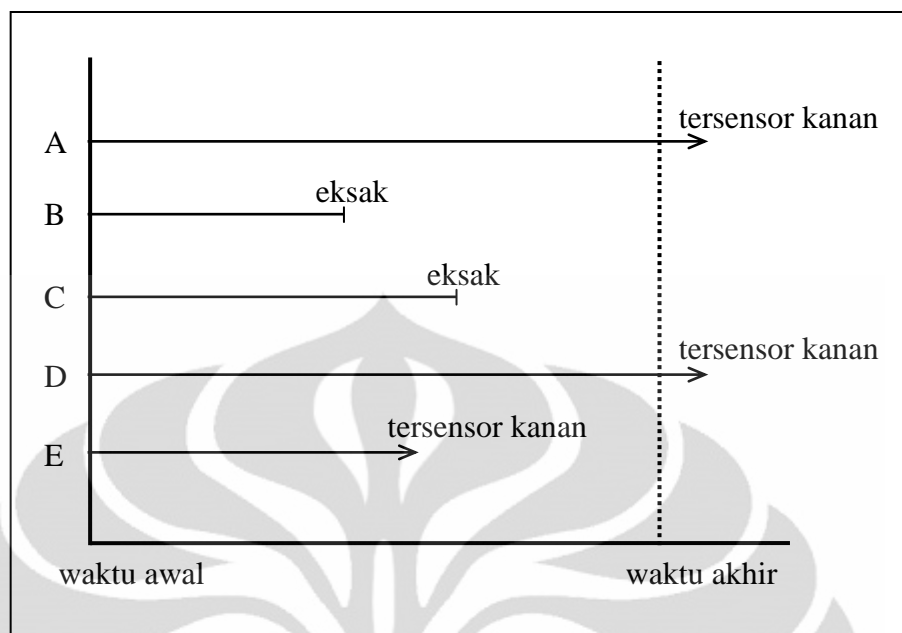
Pada subbab ini akan difokuskan untuk membahas data tersensor yang terdiri dari tersensor kanan, tersensor kiri, dan tersensor interval.

2.5.1 Data tersensor kanan

Data tersensor kanan terjadi ketika individu memiliki *survival time* melebihi suatu nilai tertentu. Secara umum terdapat beberapa alasan sehingga terjadinya sensor kanan

- individu belum mengalami kejadian setelah pengamatan berakhir,
- individu tidak melanjutkan pengamatan yang sedang berlangsung,
- individu dikeluarkan dari pengamatan karena mengalami kejadian tetapi bukan kejadian yang menjadi perhatian.

Sebagai contoh, (Klein & Moeschberger, 1997) eksperimen hewan pada *National Center for Toxicological Research* (NCTR) dengan sekelompok tikus diberi zat karsinogen. Tujuannya adalah untuk mengetahui efek zat karsinogen terhadap waktu hidup. Tikus-tikus diamati dari awal pengamatan hingga mati atau hingga akhir waktu pengamatan selesai. Contoh tersebut diilustrasikan pada gambar berikut:



Gambar 2.1 Contoh data eksak dan tersensor kanan

Dari gambar di atas, *survival time* B dan C diketahui secara pasti tetapi *survival time* A, D, dan E tersensor kanan. E keluar atau hilang dari pengamatan sebelum waktu akhir pengamatan. A dan D masih hidup hingga waktu akhir pengamatan sehingga memiliki *survival time* yang tersensor kanan.

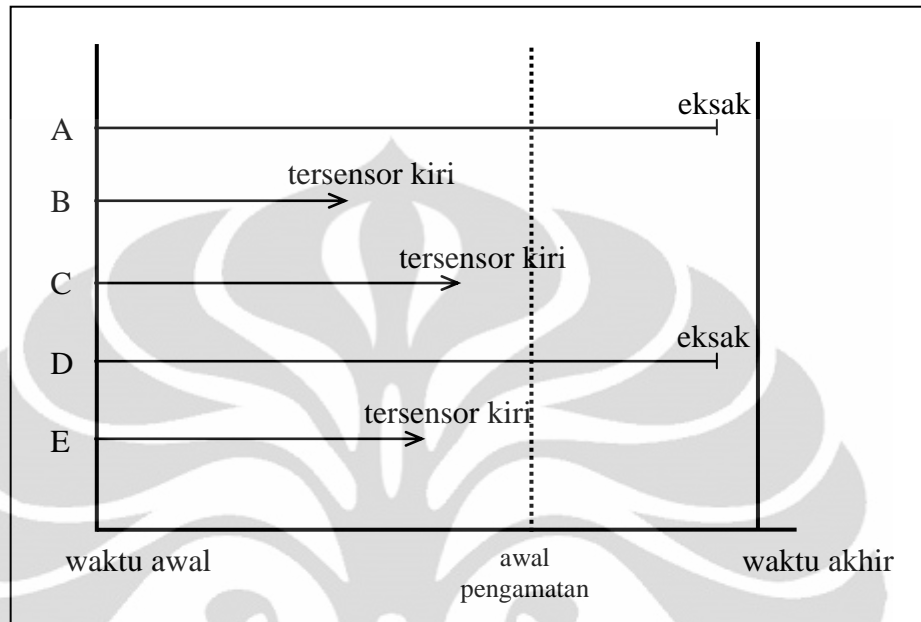
2.5.2 Data Tersensor Kiri

Data tersensor kiri terjadi ketika *survival time* dari individu kurang dari suatu nilai tertentu atau kejadian yang diperhatikan sudah dialami individu sebelum pengamatan dilakukan.

Sebagai contoh, pada suatu pusat pembelajaran anak usia dini, pengamatan difokuskan pada pengujian anak-anak untuk menentukan kapan seorang anak belajar untuk melakukan suatu tugas tertentu (misalkan berbicara). Ketika pengamatan berlangsung mungkin akan diketahui bahwa ada anak-anak yang sudah dapat berbicara.

Dari contoh tersebut, waktu awal merupakan saat anak-anak lahir dan waktu akhir adalah setelah (misalkan) dua tahun periode pengamatan. Maka,

waktu yang tersensor kiri merupakan waktu dari lahir hingga dapat berbicara. Contoh tersebut diilustrasikan pada gambar berikut:



Gambar 2.2 Contoh data eksak dan tersensor kiri

Dari gambar di atas, *survival time* A dan D diketahui secara pasti tetapi *survival time* B, C, dan E tersensor kiri. B, C, dan E sudah mengalami event, dalam contoh ini dapat berbicara, sebelum pengamatan dilakukan. A dan D mengalami event, dapat berbicara, dalam jangka waktu pengamatan sehingga memiliki *survival time* yang eksak.

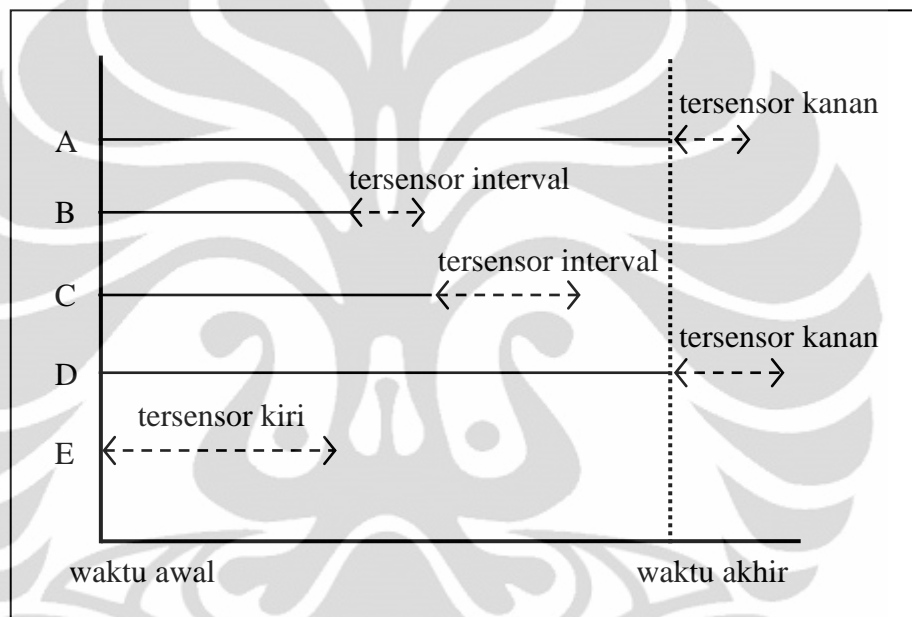
2.5.3 Data Tersensor Interval

Data tersensor interval terjadi ketika event yang menjadi perhatian terjadi pada suatu interval tertentu. Penyensoran interval ini biasanya terjadi pada pengamatan longitudinal yang memiliki pengamatan *follow-up* secara periodik.

Sebagai contoh, (Klein & Moeschberger, 1997) pada *Framingham Heart Study*, usia dari pasien ketika pertama kali mendapatkan *coronary heart disease* (CHD) biasanya diketahui secara pasti. Namun, usia ketika terjadinya pertama

kali subkategori *angina pectoris* hanya diketahui pada dua pemeriksaan klinis, kurang lebih berjarak dua tahun.

Dari contoh tersebut, kejadian yang ingin diamati adalah waktu hingga terjadi pertama kali subkategori *angina pectoris*. Untuk itu, pasien melakukan pemeriksaan klinis secara periodik setiap dua tahun. Ketika seorang pasien mengalami kejadian, maka waktu pastinya tidak diketahui tetapi berada di antara interval dua pengamatan. Contoh tersebut diilustrasikan pada gambar berikut:



Gambar 2.3 Contoh data tersensor interval

Penyensoran interval merupakan generalisasi dari penyensoran kanan atau kiri. Penyensoran interval yang dinyatakan sebagai (L_i, R_i) , ketika batas kiri merupakan 0 dan batas kanan suatu nilai tertentu, didapatkan sensor kiri, dan ketika batas kiri merupakan suatu nilai tertentu dan batas kanan tak terhingga, didapatkan sensor kanan.

2.6. Taksiran Kaplan-Meier

Metode penaksiran Kaplan-Meier merupakan metode penaksiran nonparametrik fungsi survival yang banyak digunakan. Penaksiran ini mencakup keseluruhan data yang ada, baik yang tidak tersensor maupun yang tersensor.

Misalkan pada data berukuran n terdapat k survival time yang berbeda, $t_1 < t_2 < \dots < t_k$. Pada setiap t_i , terdapat sebanyak Y_i individu yang beresiko mengalami event. Beresiko mengalami event berarti mereka belum mengalami event atau tidak tersensor sebelum t_i . Jika terdapat individu yang tersensor pada tepat t_i , mereka juga dianggap beresiko mengalami event pada t_i . Misalkan d_i banyaknya individu yang mengalami event pada waktu t_i . Taksiran Kaplan-Meier didefinisikan sebagai:

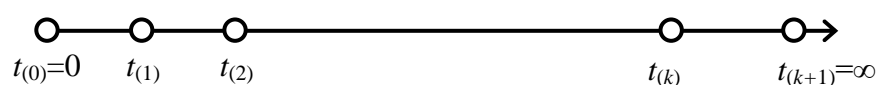
$$\hat{S}(t) = \begin{cases} 1 & t < t_1 \\ \prod_{t_i \leq t} \left[1 - \frac{d_i}{Y_i} \right] & t_1 \leq t \end{cases} \quad (2.18)$$

Penaksiran Kaplan-Meier disebut juga sebagai *product-limit estimator* dan berikut akan ditunjukkan.

Misalkan:

- $t_{(1)} < t_{(2)} < \dots < t_{(k)}$ menyatakan *survival time* tidak tersensor pada data, dengan $t_{(0)} = 0$ dan $t_{(k+1)} = \infty$.
- $t_{i,1}, t_{i,2}, \dots, t_{i,m_i}$ menyatakan *survival time* untuk individu yang tersensor dalam interval $[t_{(i)}, t_{(i+1)})$ dan m_i adalah banyaknya individu yang tersensor dalam $[t_{(i)}, t_{(i+1)})$.
- d_i menyatakan banyaknya individu yang mengalami event pada waktu $t_{(i)}$.

Misalkan diberikan gambar ilustrasi sebagai berikut:



- Misalkan interval untuk gambar di atas dibagi menjadi dua kelompok, yaitu $[t_{(0)}, t_{(1)})$ kelompok pertama dan $[t_{(i)}, t_{(i+1)})$ dengan $i = 1, 2, \dots, k-1$ kelompok kedua.
- Pada interval $[t_{(0)}, t_{(1)})$ belum ada individu yang mengalami event dan hanya ada individu-individu yang mengalami penyensoran.



- Karena event terjadi pertama kali saat $t_{(1)}$, kemudian $t_{(2)}$, dan seterusnya sampai $t_{(k)}$, maka mulai pada interval $[t_{(1)}, t_{(2)})$ dan seterusnya sampai interval $[t_{(k-1)}, t_{(k)})$ terdapat sebanyak d_i individu yang mengalami event saat $t_{(i)}$, $i = 1, 2, \dots, k$, dan juga terdapat individu-individu yang tersensor pada interval $[t_{(i)}, t_{(i+1)})$ dengan $i = 1, 2, \dots, k-1$.



Dari ilustrasi di atas, dapat dibentuk fungsi likelihood sebagai berikut:

$$L = \left\{ \prod_{j=1}^{m_0} \Pr(T > t_{0,j}) \right\} \prod_{i=1}^k \left\{ \Pr(T = t_i)^{d_i} \prod_{j=1}^{m_i} \Pr(T > t_{i,j}) \right\} \quad (2.19)$$

Jika probabilitas untuk suatu observasi tersensor $t_{i,j}$ diberikan sebagai observasi tak tersensor berikutnya t_{i+1} , maka probabilitasnya menjadi $\Pr(T = t_{(i+1)})$, dan dengan menganggap $\Pr(T > t_{i,j})$ sama untuk semua j , maka persamaan di atas dapat ditulis:

$$\begin{aligned}
L &= 1 \cdot \prod_{i=1}^k \left\{ \left(S(t_{i-1}) - S(t_i) \right)^{d_i} \prod_{j=1}^{m_i} S(t_i) \right\} \\
&= \prod_{i=1}^k \left\{ \left(S(t_{i-1}) - S(t_i) \right)^{d_i} S(t_i)^{m_i} \right\} \\
&= \prod_{i=1}^k \left\{ \left(S(t_{i-1}) - S(t_i) \right)^{d_i} \frac{S(t_{i-1})^{d_i}}{S(t_{i-1})^{d_i}} S(t_i)^{m_i} \right\} \\
&= \prod_{i=1}^k \left\{ \left(\frac{S(t_{i-1}) - S(t_i)}{S(t_{i-1})} \right)^{d_i} S(t_{i-1})^{d_i} S(t_i)^{m_i} \right\} \\
&= \prod_{i=1}^k \left\{ \left(1 - \frac{S(t_i)}{S(t_{i-1})} \right)^{d_i} S(t_{i-1})^{d_i} S(t_i)^{m_i} \right\} \\
&= \prod_{i=1}^k \left\{ \left(1 - \frac{S(t_i)}{S(t_{i-1})} \right)^{d_i} S(t_{i-1})^{d_i} \frac{S(t_{i-1})^{m_i}}{S(t_{i-1})^{m_i}} S(t_i)^{m_i} \right\} \\
&= \prod_{i=1}^k \left\{ \left(1 - \frac{S(t_i)}{S(t_{i-1})} \right)^{d_i} S(t_{i-1})^{d_i} S(t_{i-1})^{m_i} \frac{S(t_i)^{m_i}}{S(t_{i-1})^{m_i}} \right\} \\
&= \prod_{i=1}^k \left\{ \left(1 - \frac{S(t_i)}{S(t_{i-1})} \right)^{d_i} S(t_{i-1})^{d_i+m_i} \frac{S(t_i)^{m_i}}{S(t_{i-1})^{m_i}} \right\} \\
L &= \prod_{i=1}^k \left\{ \lambda_i^{d_i} S(t_{i-1})^{d_i+m_i} (1-\lambda_i)^{m_i} \right\} \tag{2.20}
\end{aligned}$$

dengan $\lambda_i = 1 - \frac{S(t_i)}{S(t_{i-1})}$.

$$\lambda_i = 1 - \frac{S(t_i)}{S(t_{i-1})} \quad \rightarrow \quad 1 - \lambda_i = \frac{S(t_i)}{S(t_{i-1})}$$

$$S(t_i) = \frac{S(t_i)}{S(t_{i-1})} \frac{S(t_{i-1})}{S(t_{i-2})} \dots \frac{S(t_1)}{S(t_0)} = (1 - \lambda_i)(1 - \lambda_{i-1}) \dots (1 - \lambda_1)$$

$$\text{sehingga } S(t_{(i-1)}) = \begin{cases} 1 & , i = 1 \\ \prod_{j=1}^{i-1} (1 - \lambda_j) & , i = 2, \dots, k \end{cases} \quad (2.21)$$

maka persamaan (2.20) dapat ditulis

$$\begin{aligned} L &= \lambda_1^{d_1} S(t_{(0)})^{d_1+m_1} (1-\lambda_1)^{m_1} \prod_{i=2}^k \left\{ \lambda_i^{d_i} (1-\lambda_i)^{m_i} S(t_{(i-1)})^{d_i+m_i} \right\} \\ &= \lambda_1^{d_1} (1-\lambda_1)^{m_1} \prod_{i=2}^k \left\{ \lambda_i^{d_i} (1-\lambda_i)^{m_i} \left(\prod_{j=1}^{i-1} (1-\lambda_j) \right)^{d_i+m_i} \right\} \\ &= \left\{ \prod_{i=1}^k \lambda_i^{d_i} (1-\lambda_i)^{m_i} \right\} \left\{ \prod_{i=2}^k \left[\left(\prod_{j=1}^{i-1} (1-\lambda_j) \right)^{d_i+m_i} \right] \right\} \\ &= \left\{ \prod_{i=1}^k \lambda_i^{d_i} (1-\lambda_i)^{m_i} \right\} (1-\lambda_1)^{d_2+m_2} [(1-\lambda_1)(1-\lambda_2)]^{d_3+m_3} \\ &\quad \cdots [(1-\lambda_1)(1-\lambda_2)\cdots(1-\lambda_{k-1})]^{d_k+m_k} \\ &= \left\{ \prod_{i=1}^k \lambda_i^{d_i} (1-\lambda_i)^{m_i} \right\} (1-\lambda_1)^{d_2+m_2+d_3+m_3+\cdots+d_k+m_k} (1-\lambda_2)^{d_3+m_3+\cdots+d_k+m_k} \\ &\quad \cdots (1-\lambda_{k-1})^{d_k+m_k} \\ &= \left\{ \prod_{i=1}^k \lambda_i^{d_i} (1-\lambda_i)^{m_i} \right\} \prod_{i=1}^k (1-\lambda_i)^{d_{i+1}+m_{i+1}+\cdots+d_k+m_k} \\ &= \prod_{i=1}^k \lambda_i^{d_i} (1-\lambda_i)^{m_i} (1-\lambda_i)^{d_{i+1}+m_{i+1}+\cdots+d_k+m_k} \\ &= \prod_{i=1}^k \lambda_i^{d_i} (1-\lambda_i)^{m_i+d_{i+1}+m_{i+1}+\cdots+d_k+m_k} \\ L &= \prod_{i=1}^k \lambda_i^{d_i} (1-\lambda_i)^{Y_i-d_i} \quad (2.22) \end{aligned}$$

dengan $Y_i = d_i + m_i + d_{i+1} + m_{i+1} + \dots + d_k + m_k$.

Kemudian akan dimaksimumkan

$$\log L = \sum_{i=1}^k [d_i \log \lambda_i + (Y_i - d_i) \log (1 - \lambda_i)] \quad (2.23)$$

untuk masing-masing $\lambda_1, \dots, \lambda_k$. Turunkan $\log L$ terhadap λ_i diperoleh

$$\frac{d \log L}{d \lambda_i} = \frac{d_i}{\lambda_i} - \frac{(Y_i - d_i)}{(1 - \lambda_i)}. \quad (2.24)$$

Dengan menyamakan turunan tersebut dengan nol dan menyelesaikan persamaannya akan diperoleh solusi untuk λ_i adalah $\frac{d_i}{Y_i}$ untuk $i = 1, 2, \dots, k$.

Maka, $\hat{\lambda}_i = \frac{d_i}{Y_i}$ merupakan taksiran yang akan memaksimumkan fungsi

likelihood L dan $\log L$. Sehingga didapat,

$$\hat{S}(t_{(i)}) = \prod_{j=1}^i \left(1 - \frac{d_j}{Y_j} \right), j = 1, 2, \dots, k$$

Untuk t secara umum, penaksir Kaplan-Meier didefinisikan sebagai

$$\hat{S}(t) = \begin{cases} 1 & t < t_{(1)} \\ \prod_{t_{(j)} \leq t} \left[1 - \frac{d_j}{Y_j} \right] & t_{(1)} \leq t \end{cases} \quad (2.25)$$

2.7 Model Cox

Dalam analisis survival, ada kalanya peneliti ingin melihat hubungan *survival time* dengan faktor-faktor lain (kovariat), untuk itu akan digunakan penaksiran dengan pemodelan regresi. Model regresi untuk masalah survival yang sering digunakan adalah model Cox. Model Cox adalah model semiparametrik yang artinya tidak diketahui distribusi dari data sehingga tidak diketahui bentuk fungsional dari fungsi *baseline hazard*. Akan tetapi parameter-parameter, β , dari model dapat diketahui distribusinya. Pada tugas akhir ini akan digunakan model *Cox proportional hazard*.

2.7.1 Definisi dan karakteristik model

Model Cox PH dirumuskan sebagai berikut:

$$h(t, \mathbf{X}) = h_0(t) \exp \left(\sum_{l=1}^p \beta_l X_l \right) \quad (2.26)$$

dengan

- $h_0(t)$ merupakan fungsi *baseline hazard*,
- X_1, \dots, X_p merupakan variabel-variabel penjelas (kovariat),
- β_1, \dots, β_p merupakan koefisien model Cox.

Model Cox, jika semua X -nya bernilai nol maka bentuknya akan tereduksi menjadi fungsi *baseline hazard*. Hal ini menunjukkan bahwa fungsi *baseline hazard* merupakan fungsi awal sebelum dikenakan dengan faktor-faktor X . Fungsi *baseline hazard*, $h_0(t)$, tidak perlu diketahui distribusinya.

Dari model Cox PH, diketahui bahwa fungsi *baseline hazard* merupakan fungsi dari t tetapi tidak mengandung X . Sedangkan, ekspresi eksponensial dari model Cox PH mengandung X tetapi tidak mengandung t . X yang tidak bergantung pada t disebut variabel yang *time-independent*. Dimungkinkan juga untuk menganggap bahwa X bergantung terhadap t , X yang demikian disebut variabel yang *time-dependent*. Jika terdapat variabel *time-dependent*, model Cox tetap dapat digunakan akan tetapi tidak memenuhi asumsi PH dan modelnya disebut model extended Cox.

Pada model Cox PH dikenal *hazard ratio* (HR) yaitu perbandingan hazard dari dua individu dengan nilai kovariat-kovariat yang berbeda. Untuk individu pertama dengan kovariat X^* dan individu kedua dengan kovariat X , *hazard ratio* dapat dituliskan sebagai berikut:

$$\begin{aligned}
 HR &= \frac{h(t, \mathbf{X}^*)}{h(t, \mathbf{X})} \\
 &= \frac{h_0(t) \exp\left(\sum_{l=1}^p \beta_l X_l^*\right)}{h_0(t) \exp\left(\sum_{l=1}^p \beta_l X_l\right)} \\
 &= \frac{\cancel{h_0(t)} \exp\left(\sum_{l=1}^p \beta_l X_l^*\right)}{\cancel{h_0(t)} \exp\left(\sum_{l=1}^p \beta_l X_l\right)} \\
 HR &= \exp\left(\sum_{l=1}^p \beta_l (X_l^* - X_l)\right). \tag{2.27}
 \end{aligned}$$

Dari hasil tersebut diperoleh bahwa perbandingan hazard untuk setiap individu dengan individu lain dengan nilai kovariat berbeda akan konstan tidak dipengaruhi waktu. Dengan perkataan lain, hazard untuk satu individu

proporsional dengan hazard individu lainnya dengan keproporsionalan yang konstan, tidak dipengaruhi oleh waktu. Kondisi inilah yang disebut dengan *proportional hazard* (PH).

Pada bagian sebelumnya fungsi hazard dapat dinyatakan melalui fungsi survival. Dengan menggunakan persamaan (2.13), fungsi survival dapat dinyatakan sebagai

$$S(t, \mathbf{X}) = \exp[-H(t, \mathbf{X})]. \quad (2.28)$$

Fungsi *cumulative hazard* pada (2.12) dapat dinyatakan

$$\begin{aligned} H(t, \mathbf{X}) &= \int_0^t h(u, \mathbf{X}) du \\ &= \int_0^t h_0(u) \exp\left(\sum_{l=1}^p \beta_l X_l\right) du \\ &= \exp\left(\sum_{l=1}^p \beta_l X_l\right) \int_0^t h_0(u) du \\ &= \exp\left(\sum_{l=1}^p \beta_l X_l\right) H_0(t). \end{aligned} \quad (2.29)$$

Substitusi persamaan (2.29) ke (2.28), diperoleh

$$\begin{aligned} S(t, \mathbf{X}) &= e^{\left[-\exp\left(\sum_{l=1}^p \beta_l X_l\right) H_0(t)\right]} \\ &= \left[e^{-H_0(t)}\right]^{\exp\left(\sum_{l=1}^p \beta_l X_l\right)} \\ &= S_0(t) \exp\left(\sum_{l=1}^p \beta_l X_l\right) \end{aligned} \quad (2.30)$$

yang merupakan fungsi survival dari model Cox dengan $S_0(t) = e^{-H_0(t)}$ adalah fungsi *baseline survival*.

2.7.2 Partial likelihood

Seperti halnya pada model regresi lain, akan ditaksir parameter-parameter β . Untuk menaksir parameter dengan metode maksimum likelihood akan dituliskan probabilitas (p.d.f.) dari data sebagai fungsi dari parameter pada model.

Misalkan $f(t_i, \mathbf{X})$ merupakan p.d.f. untuk event pada saat $t_i, i = 1, 2, \dots, n$. Untuk individu i dengan event yang teramati pada saat t_i , berkontribusi $f(t_i, \mathbf{X})$ pada likelihood. Untuk individu i yang tersensor pada saat t_i , hanya diketahui bahwa individu tersebut *survive* sampai saat t_i dan observasi tersebut berkontribusi $S(t_i, \mathbf{X})$ pada likelihood. Maka, fungsi likelihood untuk data dinyatakan sebagai

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n [f(t_i, \mathbf{X})]^{c_i} [S(t_i, \mathbf{X})]^{1-c_i} \quad (2.31)$$

dengan c_i merupakan indikator penyensoran, 1 jika *survival time* untuk individu ke- i teramati dan 0 jika tersensor.

Berdasarkan persamaan (2.9) diketahui $h(t) = f(t)/S(t)$ sehingga $f(t) = h(t)S(t)$. Dengan informasi ini, substitusikan ke persamaan (2.31)

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^n [h(t_i, \mathbf{X}) S(t_i, \mathbf{X})]^{c_i} [S(t_i, \mathbf{X})]^{1-c_i} \\ &= \prod_{i=1}^n [h(t_i, \mathbf{X})]^{c_i} S(t_i, \mathbf{X}) \end{aligned}$$

Dengan mensubstitusikan persamaan model Cox (2.26) dan persamaan fungsi survival berdasarkan model Cox, (2.30), menghasilkan fungsi likelihood

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left[h_0(t_i) \exp\left(\sum_{l=1}^p \beta_l X_{il}\right) \right]^{c_i} S_0(t_i)^{\exp\left(\sum_{l=1}^p \beta_l X_{il}\right)}. \quad (2.32)$$

Akan dimaksimumkan fungsi likelihood (2.32) yang akan memberikan taksiran parameter $\boldsymbol{\beta}$, juga taksiran fungsi *baseline hazard*, dan fungsi *baseline survival* tetapi masalah ini tidak mudah didapatkan. Untuk itu, pada model Cox digunakan penaksiran partial likelihood untuk menaksir parameter-parameter pada modelnya.

Misalkan terdapat n individu, $i = 1, \dots, n$ dan masing-masing memiliki p -vektor kovariat $\mathbf{X}_i = [X_{i1} \dots X_{ip}]'$. Dari n individu tersebut, misalkan k individu mengalami event sehingga terdapat $n-k$ individu yang tersensor. Misalkan D

adalah himpunan yang beranggotakan individu-individu yang mengalami event. Model Cox PH untuk individu $i \in D$ dinyatakan sebagai:

$$h_i(t) = h_0(t) \exp \boldsymbol{\beta}' \mathbf{X}_i. \quad (2.33)$$

dengan $\boldsymbol{\beta}' = [\beta_1 \dots \beta_p]$ adalah koefisien model Cox.

Misalkan $t_1 < t_2 < \dots < t_k$ merupakan *survival time* untuk event yang teramati dan R_i himpunan individu yang beresiko untuk mengalami event pada saat t_i . Berdasarkan Cox, likelihood di waktu t_i dinyatakan

$$L_i = \frac{\exp \boldsymbol{\beta}' \mathbf{X}_i}{\sum_{q \in R_i} \exp \boldsymbol{\beta}' \mathbf{X}_q}. \quad (2.34)$$

Rasio tersebut menyatakan bahwa hazard untuk individu pada saat t_i relatif terhadap *cumulative hazard* untuk semua individu yang beresiko pada saat event terjadi untuk individu ke- i . Sehingga partial likelihood untuk semua individu adalah

$$\begin{aligned} L(\boldsymbol{\beta}) &= L_1 L_2 \dots L_k \\ &= \prod_{i=1}^k \frac{\exp \boldsymbol{\beta}' \mathbf{X}_i}{\sum_{q \in R_i} \exp \boldsymbol{\beta}' \mathbf{X}_q}. \end{aligned} \quad (2.35)$$

Log dari partial likelihood (2.35) adalah

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^k \boldsymbol{\beta}' \mathbf{X}_i - \sum_{i=1}^k \log \left(\sum_{q \in R_i} \exp \boldsymbol{\beta}' \mathbf{X}_q \right). \quad (2.36)$$

Untuk mendapatkan taksiran maksimum partial likelihood, akan diturunkan persamaan (2.36) terhadap $\boldsymbol{\beta}$ yaitu

$$\frac{\partial \log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^k \mathbf{X}_i - \sum_{i=1}^k \frac{\sum_{q \in R_i} X_q \exp \boldsymbol{\beta}' \mathbf{X}_q}{\sum_{q \in R_i} \exp \boldsymbol{\beta}' \mathbf{X}_q}. \quad (2.37)$$

Dengan menyamakan masing-masing turunan pada (2.37) dengan nol dan menyelesaikan persamaannya akan diperoleh taksiran untuk parameter yang tidak diketahui.

2.8 Koefisien Korelasi

Koefisien korelasi merupakan ukuran kekuatan hubungan (linier) antara dua variabel. Koefisien korelasi hanya memperhatikan hubungan dua variabel dan tidak menyimpulkan adanya sebab-akibat. Untuk dua variabel random X dan Y , koefisien korelasi antara dua variabel tersebut memiliki karakteristik seperti berikut:

- nilainya berada di antara -1 sampai 1,
- nilainya mendekati -1, berarti terdapat korelasi negatif yang kuat,
- nilainya mendekati 1, berarti terdapat korelasi positif yang kuat,
- nilainya mendekati 0, berarti tidak terdapat korelasi.

Koefisien korelasi untuk populasi dinyatakan sebagai berikut

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}} = \frac{E[(X - E(X))(Y - E(Y))]}{\sqrt{E(X - E(X))^2 E(Y - E(Y))^2}}. \quad (2.38)$$

Jika korelasi tersebut ditaksir menggunakan data sampel berukuran n , dapat digunakan *Pearson's product moment correlation coefficient* yang didefinisikan sebagai

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}. \quad (2.39)$$

r merupakan variabel random dan memiliki suatu distribusi.

BAB 3

PENGECEKAN ASUMSI PROPORTIONAL HAZARD

Pada tugas akhir ini akan dibahas dua pendekatan untuk pengecekan asumsi *proportional hazard* yaitu dengan grafik dan pengujian goodness-of-fit. Pada pendekatan dengan grafik akan digunakan grafik log-log.

3.1 Pendekatan Grafik: Grafik Log-log Survival

Secara singkat, langkah-langkah pengecekan asumsi *proportional hazard* dengan grafik log-log survival adalah:

1. mencari taksiran fungsi survival, $\hat{S}(t, \mathbf{X})$, berdasarkan model Cox,
2. mencari nilai log-log survival, $\ln[-\ln \hat{S}(t, \mathbf{X})]$,
3. membentuk grafik dengan sumbu- x merupakan *survival time* dan sumbu- y nilai $\ln[-\ln \hat{S}(t, \mathbf{X})]$,
4. asumsi *proportional hazard* dipenuhi jika untuk dua individu dengan kovariat yang berbeda, kurva pada grafiknya akan paralel.

Pada bab sebelumnya telah didapatkan fungsi survival dari model Cox pada persamaan (2.30) yaitu $S(t, \mathbf{X}) = S_0(t) \exp\left(\sum_{i=1}^p \beta_i X_i\right)$. Formula log-log mengharuskan ambil log dari fungsi survival dua kali. Karena fungsi survival merupakan fungsi probabilitas, $0 \leq S(t, \mathbf{X}) \leq 1$, sehingga log dari $S(t, \mathbf{X})$ selalu bernilai negatif, $-\infty < \ln S(t, \mathbf{X}) \leq 0$. Untuk itu perlu dinegasikan log yang pertama agar bernilai positif sehingga dapat diambil log untuk yang kedua, $0 \leq -\ln S(t, \mathbf{X}) < \infty$, $-\infty < \ln[-\ln S(t, \mathbf{X})] < \infty$.

Transformasi log-log fungsi survival adalah sebagai berikut:

$$S(t, \mathbf{X}) = S_0(t) \exp\left(\sum_{i=1}^p \beta_i X_i\right)$$
$$-\ln S(t, \mathbf{X}) = -\ln \left[S_0(t) \exp\left(\sum_{i=1}^p \beta_i X_i\right) \right]$$

$$\begin{aligned}
&= \exp\left(\sum_{l=1}^p \beta_l X_l\right) \times [-\ln S_0(t)] \\
\ln[-\ln S(t, \mathbf{X})] &= \ln\left[\exp\left(\sum_{l=1}^p \beta_l X_l\right) \times [-\ln S_0(t)]\right] \\
&= \ln\left[\exp\left(\sum_{l=1}^p \beta_l X_l\right)\right] + \ln[-\ln S_0(t)] \\
&= \exp\left(\sum_{l=1}^p \beta_l X_l\right) + \ln[-\ln S_0(t)] \tag{3.1}
\end{aligned}$$

Akan ditunjukkan bahwa grafik taksiran log-log dapat digunakan untuk mengevaluasi asumsi *proportional hazard*. Misalkan untuk dua individu memiliki karakteristik pada vektor kovariat \mathbf{X} masing-masing \mathbf{X}_1 dan \mathbf{X}_2 .

$$\begin{aligned}
\mathbf{X}_1 &= (X_{11}, X_{12}, \dots, X_{1p}) \\
\mathbf{X}_2 &= (X_{21}, X_{22}, \dots, X_{2p})
\end{aligned}$$

Masing-masing kurva log-log dari dua individu tersebut adalah

$$\begin{aligned}
\ln[-\ln S(t, \mathbf{X}_1)] &= \sum_{l=1}^p \beta_l X_{1l} + \ln[-\ln S_0(t)] \\
\ln[-\ln S(t, \mathbf{X}_2)] &= \sum_{l=1}^p \beta_l X_{2l} + \ln[-\ln S_0(t)]
\end{aligned} \tag{3.2}$$

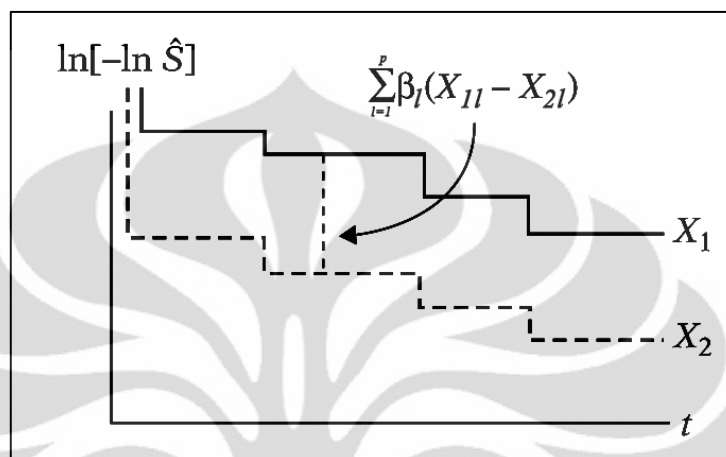
Kurva log-log survival individu pertama dikurangi kurva log-log survival individu kedua

$$\begin{aligned}
&\ln[-\ln S(t, \mathbf{X}_1)] - (\ln[-\ln S(t, \mathbf{X}_2)]) \\
&= \sum_{l=1}^p \beta_l X_{1l} + \ln[-\ln S_0(t)] - \left(\sum_{l=1}^p \beta_l X_{2l} + \ln[-\ln S_0(t)]\right) \\
&= \sum_{l=1}^p \beta_l X_{1l} - \sum_{l=1}^p \beta_l X_{2l} \\
&= \sum_{l=1}^p \beta_l (X_{1l} - X_{2l}) \tag{3.3}
\end{aligned}$$

Hasil yang diperoleh menunjukkan bahwa hasil pengurangannya tidak lagi bergantung pada waktu t .

Sekarang ditampilkan hasil dari taksiran kurva log-log survival dari kedua individu pada satu grafik dengan sumbu- x merupakan *survival time* dan sumbu- y

nilai $\ln[-\ln \hat{S}(t, \mathbf{X})]$ dan didapatkan kedua kurvanya paralel. Jarak antara dua kurva untuk dua individu berbeda merupakan bentuk ekspresi linier yang tidak mengandung waktu.



Gambar 3.1 Grafik log-log survival terhadap waktu

Keparalelan dari grafik log-log survival untuk model Cox PH memberikan pendekatan secara grafik untuk mendapatkan asumsi *proportional hazard*. Yaitu, jika model Cox memenuhi asumsi *proportional hazard* untuk kovariat-kovariat yang diberikan, diharapkan plot untuk kurva log-log survival individu-individu yang berbeda akan hampir paralel.

Namun, terdapat kelemahan jika menggunakan metode grafik log-log survival yaitu tidak diketahui secara jelas sejauh mana jarak paralel dari kurva akan disebut paralel. Peneliti akan memberikan pertimbangan masing-masing yang juga disesuaikan dengan banyaknya data yang digunakan. Solusi yang direkomendasikan adalah mengambil keputusan dengan mengasumsikan bahwa asumsi *proportional hazard* dipenuhi kecuali memang terlihat jelas bahwa grafik log-log survival tidak paralel (misalnya berpotongan).

3.1.1 Pengecekan untuk Variabel Kontinu

Hal yang perlu diperhatikan dalam pengecekan asumsi PH dengan grafik log-log adalah pengecekan asumsi untuk kovariat (variabel prediktor) yang

merupakan variabel kontinu. Agar dapat digambarkan pada grafik log-log survival dengan baik, variabel kontinu tersebut diubah menjadi beberapa kategori.

Pemilihan kategori atau pengkategorianya tergantung dari peneliti dengan pertimbangan banyaknya data yang ada.

Jika jumlah kategori yang dipilih cukup banyak, data akan mengecil untuk setiap kategori, menyebabkan kesulitan untuk membandingkan untuk tiap-tiap kurva. Untuk itu pengkategorian dari variabel kontinu tersebut disarankan banyaknya kategori sesedikit mungkin (dua atau tiga). Pemilihan kategorinya juga yang akan memberikan makna dengan jelas terhadap variabel tersebut.

Terdapat dua cara mencari fungsi log-log dari variabel kontinu. Pertama dengan dibentuk model menggunakan $c-1$ variabel *dummy* untuk variabel X dengan c kategori. Sebagai contoh, terdapat tiga kovariat, X_1 , X_2 , dan X_3 , dengan X_3 variabel kontinu. Kovariat X_3 akan dibentuk 3 kategori (misal, rendah, sedang, dan tinggi) sehingga terdapat 2 *dummy* variabel sebut, X'_3 dan X''_3 . Model Cox yang terbentuk adalah

$$h(t, \mathbf{X}) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X'_3 + \beta_4 X''_3) \quad (3.4)$$

dengan

$$X'_3 = \begin{cases} 1 & \text{jika tinggi} \\ 0 & \text{lainnya} \end{cases} \quad \text{dan} \quad X''_3 = \begin{cases} 1 & \text{jika sedang} \\ 0 & \text{lainnya} \end{cases}$$

sehingga tinggi = (1,0); sedang = (0,1); dan rendah = (0,0).

Taksiran fungsi survival-nya

$$\begin{aligned} \hat{S}(t, \mathbf{X}_{\text{tinggi}}) &= \hat{S}_0(t)^{\exp(\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X'_3)} \\ \hat{S}(t, \mathbf{X}_{\text{sedang}}) &= \hat{S}_0(t)^{\exp(\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_4 X''_3)} \\ \hat{S}(t, \mathbf{X}_{\text{rendah}}) &= \hat{S}_0(t)^{\exp(\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2)} \end{aligned} \quad (3.5)$$

Bentuk log-log survival-nya

$$\begin{aligned} \ln[-\ln \hat{S}(t, \mathbf{X}_{\text{tinggi}})] &= (\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X'_3) + \ln[-\ln S_0(t)] \\ \ln[-\ln \hat{S}(t, \mathbf{X}_{\text{sedang}})] &= (\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_4 X''_3) + \ln[-\ln S_0(t)] \\ \ln[-\ln \hat{S}(t, \mathbf{X}_{\text{rendah}})] &= (\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2) + \ln[-\ln S_0(t)] \end{aligned} \quad (3.6)$$

Cara kedua adalah dengan dibentuk model menggunakan kovariat kontinu yang diperhatikan (X_3). Taksiran fungsi survival-nya dicari untuk masing-masing kategori dengan menggunakan nilai (misalkan) rata-rata dari X_3 pada tiap kategori. Dengan menggunakan contoh di atas, model Cox yang terbentuk adalah

$$h(t, \mathbf{X}) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3). \quad (3.7)$$

Taksiran fungsi survival-nya

$$\begin{aligned} \hat{S}(t, \mathbf{X}_{\text{tinggi}}) &= \hat{S}_0(t) \exp(\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 \bar{X}_{3(\text{tinggi})}) \\ \hat{S}(t, \mathbf{X}_{\text{sedang}}) &= \hat{S}_0(t) \exp(\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 \bar{X}_{3(\text{sedang})}) \\ \hat{S}(t, \mathbf{X}_{\text{rendah}}) &= \hat{S}_0(t) \exp(\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 \bar{X}_{3(\text{rendah})}) \end{aligned} \quad (3.8)$$

Bentuk log-log survival-nya

$$\begin{aligned} \ln[-\ln \hat{S}(t, \mathbf{X}_{\text{tinggi}})] &= (\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 \bar{X}_{3(\text{tinggi})}) + \ln[-\ln S_0(t)] \\ \ln[-\ln \hat{S}(t, \mathbf{X}_{\text{sedang}})] &= (\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 \bar{X}_{3(\text{sedang})}) + \ln[-\ln S_0(t)] \\ \ln[-\ln \hat{S}(t, \mathbf{X}_{\text{rendah}})] &= (\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 \bar{X}_{3(\text{rendah})}) + \ln[-\ln S_0(t)] \end{aligned} \quad (3.9)$$

3.1.2 Pengecekan untuk Beberapa Variabel

Hal lain yang perlu diperhatikan adalah bagaimana mengecek asumsi *proportional hazard* untuk beberapa variabel yang ada. Solusi pertama adalah dengan mengkategorikan semua variabel secara terpisah, membentuk kombinasi dari kategor-kategori tersebut, dan membandingkan kurva log-log untuk semua kombinasi pada grafik.

Akibat dengan cara tersebut, jumlah data pada masing-masing kategori akan mengecil jika terdapat banyak kombinasi dari kategori-kategori. Selain itu, meskipun terdapat jumlah data yang cukup untuk kombinasi kategori, akan sulit untuk menentukan variabel apa yang memberikan ketidakparalelan pada grafik. Misalkan pada pengamatan terdapat dua kovariat, yaitu X_1 dengan 2 kategori dan X_2 dengan 3 kategori. Maka terdapat enam kombinasi kategori dan enam kurva log-log survival yang diplot pada satu grafik.

Kovariat 2 \ Kovariat 1	Kategori 1	Kategori 2
Kategori 1	Kov.1 kat.1 dan kov.2 kat.1	Kov.1 kat.2 dan kov.2 kat.1
Kategori 2	Kov.1 kat.1 dan kov.2 kat.2	Kov.1 kat.2 dan kov.2 kat.2
Kategori 3	Kov.1 kat.1 dan kov.2 kat.2	Kov.1 kat.2 dan kov.2 kat.2

Solusi kedua untuk menghadapi masalah ini adalah dengan mengecek asumsi *proportional hazard* untuk satu kovariat disesuaikan dengan kovariat lain yang diasumsikan memenuhi asumsi *proportional hazard*. Misalkan pengamatan dengan dua kovariat X_1 dan X_2 dengan taksiran model Cox

$$\hat{h}(t, \mathbf{X}) = \hat{h}_0(t) \exp(-2X_1 + 3X_2). \quad (3.10)$$

Maka, log-log fungsi survival-nya adalah

$$\ln[-\ln \hat{S}(t, \mathbf{X})] = (2X_1 + 3X_2) + \ln[-\ln S_0(t)]. \quad (3.11)$$

Jika diperiksa asumsi *proportional hazard* untuk kovariat pertama, diasumsikan kovariat kedua telah memenuhi asumsi. Sehingga bentuk log-log fungsi survival-nya akan berubah dengan nilai X_2 yang disesuaikan menjadi rata-ratanya yaitu

$$\ln[-\ln S(t, \mathbf{X})] = (2X_1 + 3\bar{X}_2) + \ln[-\ln S_0(t)], \quad (3.12)$$

begitu pula sebaliknya.

Pada model Cox, distribusi dari data tidak diketahui sehingga tidak diketahui nilai dan fungsi *baseline hazard* maupun *baseline survival*. Untuk itu agar dapat menghitung kurva log-log survival perlu ditaksir nilai dari fungsi *baseline survival*. Penghitungan pada perangkat lunak akan mendapatkan hasil langsung tetapi akan bermasalah jika dilakukan penghitungan secara manual. Solusinya adalah dapat digunakan taksiran Kaplan-Meier, pada rumus (2.18), untuk mendapatkan taksiran fungsi *baseline survival* apabila dilakukan penghitungan lalu membentuk grafik log-log survival secara manual.

3.2 Pendekatan dengan Pengujian Goodness of Fit (GOF)

Pada bagian sebelumnya, digunakan pendekatan dengan metode grafik untuk mengecek asumsi *proportional hazard*. Pada subbab ini akan digunakan pengecekan dengan pengujian goodness of fit. Metode ini akan memberikan hasil yang lebih obyektif karena menggunakan pengujian secara statistik.

Pengujian asumsi *proportional hazard* secara statistik akan digunakan Schoenfeld residual. Untuk setiap kovariat pada model, Schoenfeld residual terdefinisi untuk setiap individu yang mengalami event. Pengujian asumsi *proportional hazard* ini didasarkan pada perkiraan bahwa asumsi *proportional hazard* terpenuhi untuk suatu kovariat jika Schoenfeld residual untuk kovariat tersebut tidak berkorelasi dengan waktu.

Adapun langkah-langkah pengujiannya sebagai berikut:

1. Mencari taksiran model Cox PH dan mencari Schoenfeld residual untuk masing-masing kovariat.
2. Membuat variabel rank *survival time* yaitu waktu terjadi event (*survival time*) yang diurutkan. Individu yang mengalami event pertama kali diberi nilai 1, mengalami event selanjutnya diberi nilai 2, dan seterusnya.
3. Menguji korelasi antara variabel pada langkah pertama dan kedua.

Hipotesis nullnya korelasi antara Schoenfeld residual dan rank *survival time* adalah nol. Penolakan hipotesis null berarti asumsi PH tidak dipenuhi.

3.2.1 Schoenfeld Residual

Misalkan terdapat n individu, $i = 1, \dots, n$ dan masing-masing memiliki p -vektor kovariat $\mathbf{X}_i = [X_{i1} \dots X_{ip}]'$. Dari n individu tersebut, misalkan sebanyak k mengalami event dan sebanyak $(n - k)$ yang tersensor, sehingga *time-to-event*-nya adalah $t_1 < t_2 < \dots < t_k$. Sebut himpunan yang memuat individu yang mengalami event adalah D . Model Cox PH untuk individu $i \in D$ dinyatakan sebagai:

$$h_i(t) = h_0(t) \exp \boldsymbol{\beta}' \mathbf{X}_i \quad (3.13)$$

dengan $\boldsymbol{\beta}' = [\beta_1 \dots \beta_p]$ adalah koefisien model Cox.

Untuk masing-masing kovariat, (Schoenfeld, 1982) Schoenfeld residual untuk individu ke- i kovariat ke- l didefinisikan sebagai

$$r_{il} = X_{il} - E(X_{il} | R_i). \quad (3.14)$$

Residual ke- i kovariat ke- l merupakan beda antara nilai observasi X_{il} dan *conditional expectation*-nya jika diketahui R_i .

Berikut adalah penjelasan untuk Schoenfeld residual. Misalkan R_i merupakan himpunan individu yang beresiko untuk mengalami event pada saat t_i . Berdasarkan Cox, likelihood untuk semua individu yang mengalami event dinyatakan

$$L(\boldsymbol{\beta}) = \prod_{i=1}^k \frac{\exp \boldsymbol{\beta}' \mathbf{X}_i}{\sum_{q \in R_i} \exp \boldsymbol{\beta}' \mathbf{X}_q}. \quad (3.15)$$

Log dari partial likelihood (3.15) adalah

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^k \boldsymbol{\beta}' \mathbf{X}_i - \sum_{i=1}^k \log \left(\sum_{q \in R_i} \exp \boldsymbol{\beta}' \mathbf{X}_q \right). \quad (3.16)$$

Turunan persamaan (3.16) terhadap $\boldsymbol{\beta}$ yaitu

$$\frac{\partial \log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^k \mathbf{X}_i - \sum_{i=1}^k \frac{\sum_{q \in R_i} X_q \exp \boldsymbol{\beta}' \mathbf{X}_q}{\sum_{q \in R_i} \exp \boldsymbol{\beta}' \mathbf{X}_q}. \quad (3.17)$$

Pada model ini X_i merupakan variabel random dengan

$$E(X_{il} | R_i) = \frac{\sum_{q \in R_i} X_{ql} \exp \boldsymbol{\beta}' \mathbf{X}_q}{\sum_{q \in R_i} \exp \boldsymbol{\beta}' \mathbf{X}_q}. \quad (3.18)$$

Dengan menyamakan masing-masing turunan pada (3.17) dengan nol dan menyelesaikan persamaannya akan diperoleh taksiran untuk parameter $\boldsymbol{\beta}$ yang tidak diketahui. Dengan kata lain, taksiran untuk parameter $\boldsymbol{\beta}$ merupakan solusi dari

$$\sum_{i=1}^k \{X_{il} - E(X_{il} | R_i)\} = 0 \quad (3.19)$$

dengan $X_{il} - E(X_{il} | R_i)$ merupakan Schoenfeld residual.

3.2.2 Koefisien Korelasi Rank Spearman

Pada bab 2, telah dijelaskan mengenai koefisien korelasi dan taksiran koefisien korelasi untuk sampel dengan menggunakan *Pearson's product moment correlation coefficient*. *Pearson's product moment correlation coefficient*, r , yang didefinisikan pada (2.39) merupakan variabel random dan memiliki suatu distribusi. Fungsi distribusi dari r bergantung pada fungsi distribusi bivariat dari (X, Y) . Jika fungsi distribusi dari (X, Y) tidak diketahui akan menyebabkan r tidak memiliki nilai sebagai statistik uji pada pengujian statistik.

Untuk kondisi tersebut, terdapat ukuran korelasi dengan fungsi distribusi yang tidak bergantung pada fungsi distribusi bivariat dari (X, Y) , jika X dan Y independent, sehingga dapat digunakan sebagai statistik uji pada pengujian nonparametrik. Ukuran korelasi yang akan dibahas merupakan fungsi dari rank observasi. Pada tugas akhir ini akan digunakan *Spearman's rank correlation coefficient* atau *Spearman's rho*.

Data terdiri dari sampel random bivariat berukuran n , $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$. Misalkan $R(X_i)$ merupakan rank dari X_i yang dibandingkan dengan X yang lainnya untuk $i = 1, 2, \dots, n$. Yaitu $R(X_i) = 1$ jika X_i merupakan yang paling kecil dari X_1, X_2, \dots, X_n , $R(X_i) = 2$ jika X_i merupakan yang kedua paling kecil, dan selanjutnya dengan rank n diberikan untuk X_i yang paling besar. Begitu juga halnya untuk $R(Y_i)$ bernilai sama dengan $1, 2, \dots$, atau n bergantung pada besar dari Y_i yang dibandingkan dengan Y_1, Y_2, \dots, Y_n untuk setiap i .

Untuk data kembar, masing-masing diberikan nilai rata-rata dari rank yang akan diberikan jika tidak kembar.

Rumus dari *Spearman's rank correlation coefficient*, r_s , didapatkan dari *Pearson's product moment correlation coefficient* dengan nilai X_i dan Y_i merupakan rank-nya. Sehingga didapatkan nilai-nilai sebagai berikut:

$$\begin{aligned}
 - \sum_{i=1}^n R(X_i) &= 1 + 2 + \dots + n = \sum_{i=1}^n i = \frac{n(n+1)}{2} \\
 \text{dan } \sum_{i=1}^n R(Y_i) &= \frac{n(n+1)}{2}, \tag{3.20}
 \end{aligned}$$

$$- \overline{R(X)} = \frac{1}{n} \sum_{i=1}^n R(X_i) = \frac{n+1}{2} \text{ dan } \overline{R(Y)} = \frac{n+1}{2}, \quad (3.21)$$

$$- \sum_{i=1}^n R(X_i)^2 = 1^2 + 2^2 + \dots + n^2 = \sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$$

dan $\sum_{i=1}^n R(Y_i)^2 = \frac{n(n+1)(2n+1)}{6}, \quad (3.22)$

$$- \sum_{i=1}^n \left(R(X_i) - \overline{R(X)} \right)^2 = \sum_{i=1}^n \left(i - \frac{n+1}{2} \right)^2$$

$$= \sum_{i=1}^n \left[i^2 - i(n+1) + \left(\frac{n+1}{2} \right)^2 \right]$$

$$= \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{2} + \frac{n(n+1)^2}{4}$$

$$= \frac{2n(n+1)(2n+1) - 6n(n+1)^2 + 3n(n+1)^2}{12}$$

$$= \frac{2n(n+1)(2n+1) - 3n(n+1)^2}{12}$$

$$= \frac{n(n+1)[2(2n+1) - 3(n+1)]}{12}$$

$$= \frac{n(n+1)[4n+2-3n-3]}{12}$$

$$= \frac{n(n+1)(n-1)}{12}$$

$$= \frac{n(n^2-1)}{12}$$

dan $\sum_{i=1}^n \left(R(Y_i) - \overline{R(Y)} \right)^2 = \frac{n(n^2-1)}{12}, \quad (3.23)$

$$- \sum_{i=1}^n R(X_i) - R(Y_i) = \sum_{i=1}^n R(X_i) - \overline{R(X)} + \overline{R(X)} - R(Y_i)$$

$$= \sum_{i=1}^n R(X_i) - \overline{R(X)} + \overline{R(Y)} - R(Y_i)$$

$$= \sum_{i=1}^n \left(R(X_i) - \overline{R(X)} \right) - \left(R(Y_i) - \overline{R(Y)} \right), \quad (3.24)$$

$$- \sum_{i=1}^n \left[R(X_i) - R(Y_i) \right]^2 = \sum_{i=1}^n \left[\left(R(X_i) - \overline{R(X)} \right) - \left(R(Y_i) - \overline{R(Y)} \right) \right]^2$$

$$\begin{aligned}
&= \sum_{i=1}^n \left(R(X_i) - \overline{R(X)} \right)^2 + \left(R(Y_i) - \overline{R(Y)} \right)^2 \\
&\quad - 2 \left(R(X_i) - \overline{R(X)} \right) \left(R(Y_i) - \overline{R(Y)} \right) \\
&= \sum_{i=1}^n \left(R(X_i) - \overline{R(X)} \right)^2 + \sum_{i=1}^n \left(R(Y_i) - \overline{R(Y)} \right)^2 \\
&\quad - 2 \sum_{i=1}^n \left(R(X_i) - \overline{R(X)} \right) \left(R(Y_i) - \overline{R(Y)} \right) \quad (3.25)
\end{aligned}$$

$$\begin{aligned}
&\sum_{i=1}^n \left(R(X_i) - \overline{R(X)} \right) \left(R(Y_i) - \overline{R(Y)} \right) \\
&= \frac{1}{2} \left\{ \sum_{i=1}^n \left(R(X_i) - \overline{R(X)} \right)^2 + \sum_{i=1}^n \left(R(Y_i) - \overline{R(Y)} \right)^2 - \sum_{i=1}^n \left[R(X_i) - R(Y_i) \right]^2 \right\} \quad (3.26)
\end{aligned}$$

Dengan mensubstitusi hasil-hasil tersebut ke dalam persamaan (2.39) maka,

$$\begin{aligned}
r_s &= \frac{\sum_{i=1}^n \left[R(X_i) - \overline{R(X)} \right] \left[R(Y_i) - \overline{R(Y)} \right]}{\sqrt{\sum_{i=1}^n \left[R(X_i) - \overline{R(X)} \right]^2 \sum_{i=1}^n \left[R(Y_i) - \overline{R(Y)} \right]^2}} \\
&= \frac{\sum_{i=1}^n \left(R(X_i) - \overline{R(X)} \right)^2 + \sum_{i=1}^n \left(R(Y_i) - \overline{R(Y)} \right)^2 - \sum_{i=1}^n \left[R(X_i) - R(Y_i) \right]^2}{2 \sqrt{\sum_{i=1}^n \left[R(X_i) - \overline{R(X)} \right]^2 \sum_{i=1}^n \left[R(Y_i) - \overline{R(Y)} \right]^2}} \\
&= \frac{\frac{n(n^2-1)}{12} + \frac{n(n^2-1)}{12} - \sum_{i=1}^n \left[R(X_i) - R(Y_i) \right]^2}{2 \sqrt{\frac{n(n^2-1)}{12} \frac{n(n^2-1)}{12}}} \\
&= \frac{2 \frac{n(n^2-1)}{12} - \sum_{i=1}^n \left[R(X_i) - R(Y_i) \right]^2}{2 \frac{n(n^2-1)}{12}} \\
&= 1 - \frac{\sum_{i=1}^n \left[R(X_i) - R(Y_i) \right]^2}{n(n^2-1)} \\
r_s &= 1 - \frac{6 \sum_{i=1}^n \left[R(X_i) - R(Y_i) \right]^2}{n(n^2-1)} \quad (3.27)
\end{aligned}$$

Bentuk lain dari r_s dapat dinyatakan sebagai berikut

$$\begin{aligned}
 r_s &= \frac{\sum_{i=1}^n [R(X_i) - \overline{R(X)}][R(Y_i) - \overline{R(Y)}]}{\sqrt{\sum_{i=1}^n [R(X_i) - \overline{R(X)}]^2 \sum_{i=1}^n [R(Y_i) - \overline{R(Y)}]^2}} \\
 &= \frac{\sum_{i=1}^n [R(X_i)R(Y_i) - \overline{R(X)}R(Y_i) - R(X_i)\overline{R(Y)} + \overline{R(X)}\overline{R(Y)}]}{\sqrt{\sum_{i=1}^n [R(X_i) - \overline{R(X)}]^2 \sum_{i=1}^n [R(Y_i) - \overline{R(Y)}]^2}} \\
 &= \frac{\sum_{i=1}^n R(X_i)R(Y_i) - \sum_{i=1}^n \overline{R(X)}R(Y_i) - \sum_{i=1}^n R(X_i)\overline{R(Y)} + \sum_{i=1}^n \overline{R(X)}\overline{R(Y)}}{\sqrt{\sum_{i=1}^n [R(X_i) - \overline{R(X)}]^2 \sum_{i=1}^n [R(Y_i) - \overline{R(Y)}]^2}} \\
 &= \frac{\sum_{i=1}^n R(X_i)R(Y_i) - n\overline{R(X)}\sum_{i=1}^n \frac{R(Y_i)}{n} - n\overline{R(Y)}\sum_{i=1}^n \frac{R(X_i)}{n} + n\overline{R(X)}\overline{R(Y)}}{\sqrt{\sum_{i=1}^n [R(X_i) - \overline{R(X)}]^2 \sum_{i=1}^n [R(Y_i) - \overline{R(Y)}]^2}} \\
 &= \frac{\sum_{i=1}^n R(X_i)R(Y_i) - n\overline{R(X)}\overline{R(Y)} - n\overline{R(Y)}\overline{R(X)} + n\overline{R(X)}\overline{R(Y)}}{\sqrt{\sum_{i=1}^n [R(X_i) - \overline{R(X)}]^2 \sum_{i=1}^n [R(Y_i) - \overline{R(Y)}]^2}} \\
 &= \frac{\sum_{i=1}^n R(X_i)R(Y_i) - n\overline{R(X)}\overline{R(Y)}}{\sqrt{\sum_{i=1}^n [R(X_i) - \overline{R(X)}]^2 \sum_{i=1}^n [R(Y_i) - \overline{R(Y)}]^2}} \\
 &= \frac{\sum_{i=1}^n R(X_i)R(Y_i) - n\left(\frac{n+1}{2}\right)\left(\frac{n+1}{2}\right)}{\sqrt{\frac{n(n^2-1)}{12} \frac{n(n^2-1)}{12}}} \\
 &= \frac{\sum_{i=1}^n R(X_i)R(Y_i) - n\frac{(n+1)^2}{4}}{\frac{n(n^2-1)}{12}} \\
 &= \frac{12\sum_{i=1}^n R(X_i)R(Y_i) - 3n(n+1)^2}{n(n^2-1)} \\
 &= \frac{12\sum_{i=1}^n R(X_i)R(Y_i)}{n(n^2-1)} - 3\left(\frac{n+1}{n-1}\right) \tag{3.28}
 \end{aligned}$$

Pengujian hipotesis untuk koefisien korelasi rank Spearman didefinisikan sebagai berikut:

- Uji dua arah
 H_0 : X_i dan Y_i saling bebas.
 H_1 : X_i dan Y_i tidak saling bebas.
 Dengan tingkat signifikansi α , H_0 ditolak jika $|r_s| > r_{s,\alpha/2}$
- Uji satu arah untuk korelasi positif
 H_0 : X_i dan Y_i saling bebas.
 H_1 : X_i dan Y_i berhubungan positif.
 Dengan tingkat signifikansi α , H_0 ditolak jika $r_s > r_{s,\alpha}$.
- Uji satu arah untuk korelasi negatif
 H_0 : X_i dan Y_i saling bebas.
 H_1 : X_i dan Y_i berhubungan negatif.
 Dengan tingkat signifikansi α , H_0 ditolak jika $r_s < -r_{s,\alpha}$.

Untuk ukuran sampel dengan besar, (Hollander and Wolfe, 1999) $n > 30$, digunakan aproksimasi kenormalan (*standardized*) dari r_s . Bentuk terstandarisasi dari r_s adalah

$$r_s^* = (n-1)^{1/2} r_s \quad (3.29)$$

(Persamaan 3.29 dibuktikan pada lampiran 1)

Jika H_0 benar, r_s^* memiliki distribusi asimptotik $N(0,1)$ ketika n menuju tak hingga. Sehingga prosedur pengujiannya adalah

- Uji dua arah
 H_0 : X_i dan Y_i saling bebas.
 H_1 : X_i dan Y_i tidak saling bebas.
 Dengan tingkat signifikansi α , H_0 ditolak jika $|r_s^*| > z_{\alpha/2}$
- Uji satu arah untuk korelasi positif
 H_0 : X_i dan Y_i saling bebas.
 H_1 : X_i dan Y_i berhubungan positif.
 Dengan tingkat signifikansi α , H_0 ditolak jika $r_s^* > z_\alpha$.

- Uji satu arah untuk korelasi negatif

H_0 : X_i dan Y_i saling bebas.

H_1 : X_i dan Y_i berhubungan negatif.

Dengan tingkat signifikansi α , H_0 ditolak jika $r_s^* < -z_\alpha$.

3.2.3 Pengujian goodness of fit

Pengujian statistik untuk pengecekan asumsi PH adalah dengan menguji korelasi antara Schoenfeld residual dengan rank *survival time* untuk masing-masing kovariat. Schoenfeld residual terdefinisi untuk *survival time* yang teramati (eksak). Sehingga untuk n individu dengan sebanyak k individu yang memiliki *survival time* yang eksak, $t_1 < t_2 < \dots < t_k$, dan kovariat pada saat t_i ,

$\mathbf{X}_i = [X_{i1} \dots X_{ip}]'$, terdapat Schoenfeld residual pada t_i yang merupakan vektor

$\mathbf{r}_i = [r_{i1} \dots r_{ip}]'$, $i = 1, \dots, k$.

Untuk pengecekan asumsi *proportional hazard* dengan goodness of fit, akan dilihat korelasi antara Schoenfeld residual dengan rank *survival time*.

Pengecekan asumsi pada kovariat pertama, X_1 , akan dicek korelasi antara rank *survival time*, (t_1, \dots, t_k) dengan rank (r_{11}, \dots, r_{k1}) , pengecekan asumsi pada

kovariat kedua, X_2 , akan dicek korelasi antara rank *survival time*, (t_1, \dots, t_k)

dengan rank (r_{12}, \dots, r_{k2}) , demikian selanjutnya hingga pengecekan asumsi pada

kovariat ke- p , X_p , akan dicek korelasi antara rank *survival time*, (t_1, \dots, t_k) dengan

rank (r_{1p}, \dots, r_{kp}) . Untuk itu dapat diilustrasikan sebagai berikut

Tabel 3.1 Komponen t_i , R_i , (X_1, \dots, X_p) , dan (r_1, \dots, r_p) untuk sampel n individu dengan k individu yang *survival time*-nya teramati

t_i	R_i	X_1	r_1	X_2	r_2	...	X_p	r_p
t_1	R_1	X_{11}	r_{11}	X_{12}	r_{12}	...	X_{1p}	r_{1p}
t_2	R_2	X_{21}	r_{21}	X_{22}	r_{22}		X_{2p}	r_{2p}
...	...							
t_k	R_k	X_{k1}	r_{k1}	X_{k2}	r_{k2}		X_{kp}	r_{kp}

Tabel 3.2 Komponen t_i dan r_i yang akan dilakukan pengecekan asumsi

t_i	r_1	t_i	r_2	...	t_i	r_p
t_1	r_{11}	t_1	r_{12}	...	t_1	r_{1p}
t_2	r_{21}	t_2	r_{22}		t_2	r_{2p}
...		
t_k	r_{k1}	t_k	r_{k2}		t_k	r_{kp}
cek asumsi untuk kovariat X_1		cek asumsi untuk kovariat X_2			cek asumsi untuk kovariat X_p	

Koefisien korelasi antara *survival time* dengan Schoenfeld residual dihitung dengan menggunakan *Spearman's rank correlation coefficient* dan pengujian hipotesisnya digunakan seperti yang sudah dinyatakan pada subbab sebelumnya.

BAB 4

CONTOH PENERAPAN

Dalam bab ini, akan dibahas pengecekan asumsi *proportional hazard* pada data. Pengecekan asumsi akan dilakukan dengan dua pendekatan yaitu secara grafik, yaitu dengan grafik log-log, dan pengujian goodness-of-fit. Pengolahan data untuk mencari taksiran model Cox PH, grafik log-log, dan pengujian korelasi digunakan perangkat lunak statistik.

4.1 Data

Data yang digunakan pada tugas akhir ini adalah data penelitian di Australia pada tahun 1991 oleh Caplehorn et al. Data penelitian ini dibandingkan 238 pecandu heroin yang diberikan dua jenis pengobatan methadone yang berbeda untuk diteliti waktu lamanya berada dalam klinik. *Survival time* dari pasien adalah waktu, dalam hari, hingga keluar dari klinik atau tersensor. Dari 238 individu terdapat 150 individu dengan *survival time* yang teramati dan 88 individu dengan *survival time* yang tersensor.

Selain itu, terdapat dua kovariat tambahan yaitu, catatan penjara dan dosis maksimum methadone, yang dipercaya mempengaruhi *survival time*. Keterangan untuk variabel-variabel yang terlibat diberikan sebagai berikut:

- *clinic* (1 atau 2),
- *survival status* (0 = *censored*, 1 = *departed from clinic*),
- *survival time (days)*
- *prison record* (0 = *none*, 1 = *any*),
- *maximum methadone dose (mg/day)*.

4.2 Analisis Data

Terlebih dahulu dibentuk model Cox yang melihat hubungan *survival time* dengan jenis pengobatan methadone, catatan penjara, dan dosis maksimum methadone.

Model umum Cox yang akan dibuat berdasarkan data adalah sebagai berikut:

$$h(t, \mathbf{X}) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3) \quad (4.1)$$

dengan

t adalah *survival time* (dalam hari),

X_1 adalah *clinic*,

X_2 adalah *prison*,

X_3 adalah *dose*.

Untuk pengujian *likelihood ratio* yang menguji signifikansi parameter, diperoleh nilai -2 Log likelihood untuk β_0 , $-2LL(\beta_0)$ adalah 1411.324.

Sedangkan untuk model Cox (4.1) diperoleh $-2LL(\beta) = 1346.805$. Sehingga nilai dari statistik uji *likelihood ratio*, X_{LR}^2 , adalah 64.519 yang berdistribusi chi-square dengan derajat bebas 3.

Tabel 4.1 Hasil pengolahan data untuk pengujian *likelihood ratio*

-2 Log Likelihood β_0	-2 Log Likelihood β	Chi-square	df	Sig
1411.324	1346.805	64.519	3	.000

Dengan mengambil nilai $\alpha = 0.05$, dari tabel 4.1 diperoleh nilai sig. yang lebih kecil dari α . Jadi, dapat disimpulkan bahwa parameter β secara bersama-sama berkontribusi secara signifikan pada model Cox.

Untuk pengujian Wald yang menguji signifikansi masing-masing parameter, diperoleh nilai yang diberikan pada tabel berikut

Tabel 4.2 Hasil pengolahan data untuk pengujian Wald

Parameter	Koefisien	Wald	df	Sig
β_1	-1.009	22.045	1	0.000
β_2	0.327	3.813	1	0.051
β_3	-0.035	30.785	1	0.000

Dengan mengambil nilai $\alpha = 0.05$, dari tabel 4.2 diperoleh nilai sig. yang lebih kecil dari α untuk parameter β_1 dan β_3 . Sedangkan parameter β_2 signifikan untuk nilai $\alpha = 0.10$. Jadi, variabel *clinic*, *prison*, dan *dose* memberikan kontribusi secara signifikan pada model Cox.

Sehingga taksiran model Cox kita adalah

$$\hat{h}(t, \mathbf{X}) = \hat{h}_0(t) \exp(-1.009x_1 + 0.327x_2 - 0.035x_3) \quad (4.2)$$

Setelah kita mendapatkan taksiran model Cox, kita akan mengecek apakah asumsi *proportional hazard* terpenuhi. Pengecekan asumsi *proportional hazard* akan dilakukan dengan menggunakan grafik log-log dan pengujian goodness-of-fit.

4.2.1 Pengecekan dengan Grafik Log-log

Setelah kita mendapatkan taksiran model Cox (4.2), kita akan mengecek asumsi *proportional hazard* untuk masing-masing variabel. Karena kita memiliki tiga variabel penjelas, untuk mengecek asumsi masing-masing variabel akan digunakan grafik log-log *adjusted*. Untuk mengecek asumsi variabel *clinic*, akan diasumsikan bahwa variabel *prison* dan *dose* memenuhi asumsi begitu juga untuk pengecekan asumsi pada variabel *prison* dan *dose* diasumsikan bahwa variabel lain, yang tidak sedang diperiksa, telah memenuhi asumsi PH.

Berdasarkan model Cox (4.2), taksiran fungsi survival-nya adalah

$$\hat{S}(t, \mathbf{X}) = \hat{S}_0(t)^{\exp(-1.009x_1 + 0.3327x_2 - 0.035x_3)} \quad (4.3)$$

Kemudian dicari nilai log-log survival, diperoleh

$$\ln(-\ln \hat{S}(t, \mathbf{X})) = (-1.009x_1 + 0.327x_2 - 0.035x_3) + \ln(-\ln \hat{S}_0(t)). \quad (4.4)$$

Untuk mengecek asumsi *proportional hazard* kita akan memplot dengan

$\ln(-\ln \hat{S}(t, \mathbf{X}))$ sebagai sumbu-y dengan *survival time* sebagai sumbu-x.

Dengan menggunakan grafik log-log *adjusted*, untuk pengecekan asumsi variabel *clinic* akan diasumsikan variabel *prison* dan *dose* memenuhi asumsi dan bentuk log-log survival-nya pun akan berubah dengan disesuaikan nilai rata-rata x_2 dan x_3 . Nilai rata-rata untuk masing-masing variabel adalah sebagai berikut

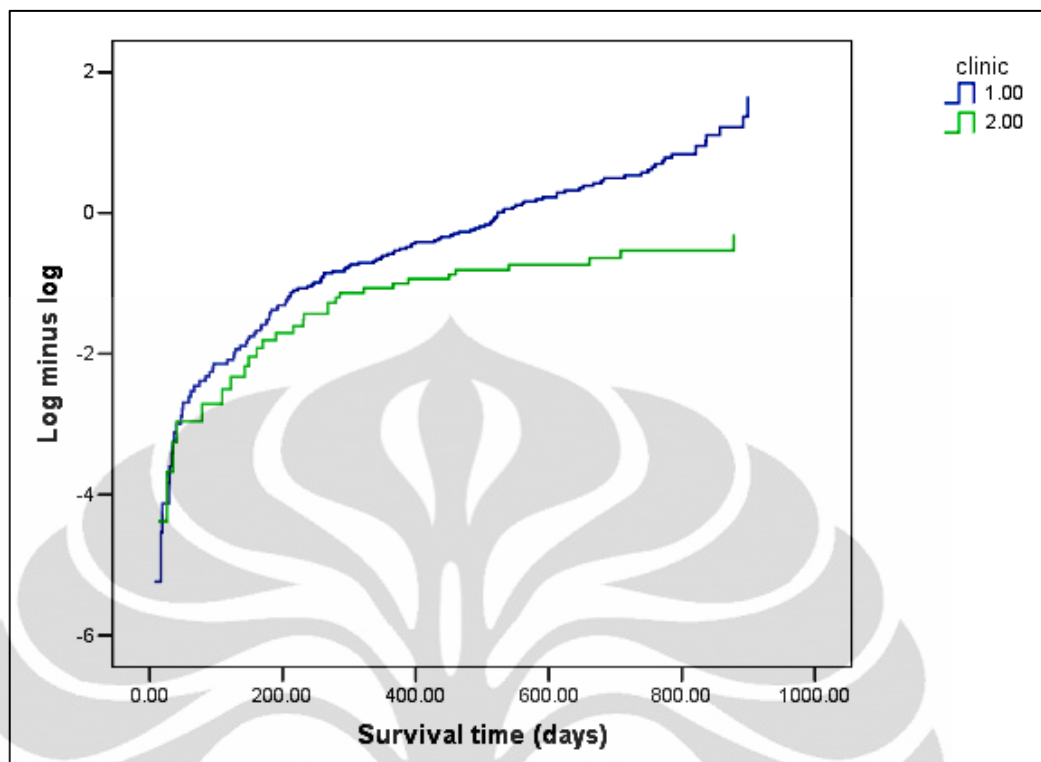
Tabel 4.3 Rata-rata untuk masing-masing variabel

Kovariat	Rata-rata
<i>Clinic</i>	1.314
<i>Prison</i>	0.462
<i>Dose</i>	60.487

Sehingga akan diperoleh bentuk log-log survival

$$\begin{aligned} \ln(-\ln \hat{S}(t, \mathbf{X})) &= (-1.009x_1 + 0.327\bar{x}_2 - 0.035\bar{x}_3) + \ln(-\ln \hat{S}_0(t)) \\ &= (-1.009x_1 + 0.327(0.462) - 0.035(60.487)) + \ln(-\ln \hat{S}_0(t)) \\ &= (-1.009x_1 + 0.151 - 2.117) + \ln(-\ln \hat{S}_0(t)) \end{aligned} \quad (4.5)$$

Hasil dari perangkat lunak ditampilkan sebagai berikut



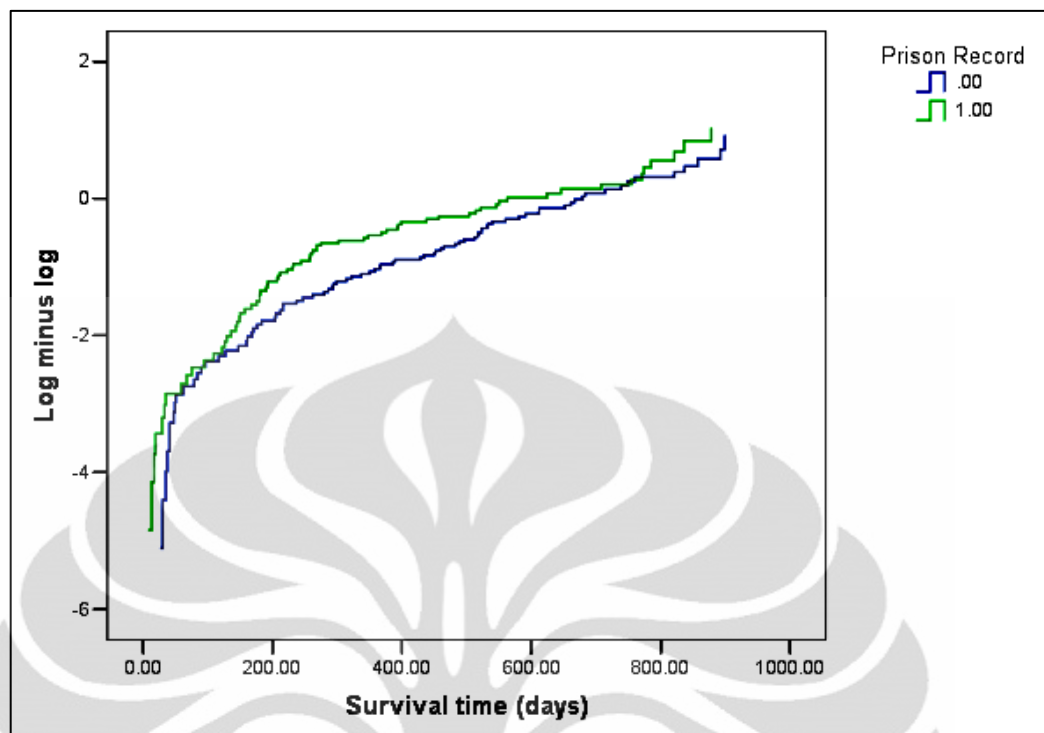
Gambar 4.1 Grafik log-log *adjusted* untuk *clinic*

Dari gambar tersebut terlihat bahwa terjadi perpotongan di beberapa titik antara waktu 0 dan 200. Karena perpotongan tersebut terjadi pada awal penelitian dan dibandingkan dengan keseluruhan data, hal tersebut boleh dianggap tidak mempengaruhi keparalelan kedua kurva tersebut. Maka, asumsi *proportional hazard* terpenuhi untuk variabel *clinic* setelah mempertimbangkan keberadaan *prison* dan *dose*.

Kemudian, pengecekan asumsi untuk variabel *prison* disesuaikan *clinic* dan *dose*. Bentuk log-log survival-nya adalah

$$\ln(-\ln \hat{S}(t, \mathbf{X})) = (-1.3258 + 0.327x_2 - 2.117) + \ln(-\ln \hat{S}_0(t)) \quad (4.6)$$

Hasil dari perangkat lunak ditampilkan sebagai berikut



Gambar 4.2 Grafik log-log *adjusted* untuk *prison*

Dari gambar tersebut terlihat bahwa terjadi perpotongan di beberapa titik antara waktu 0 dan 200 kemudian grafiknya agak paralel lalu terjadi perpotongan lagi di antara 600 dan 800. Meskipun demikian, dapat dilihat bahwa terlihat kedua kurva tersebut memiliki jarak yang hampir berdekatan dan hampir paralel maka, dapat disimpulkan asumsi *proportional hazard* terpenuhi untuk variabel *prison* setelah mempertimbangkan keberadaan *clinic* dan *dose*.

Pada pengecekan asumsi untuk variabel *dose* disesuaikan *clinic* dan *prison*, karena variabel *dose* kontinu akan digunakan cara kedua pada subbab 3.1.1. Bentuk model Cox tidak berubah seperti dinyatakan pada persamaan (4.2).

Kategorisasinya adalah

- kecil untuk dosis $x_3 \leq 40$,
- sedang untuk dosis $40 < x_3 < 70$,
- besar untuk dosis $x_3 \geq 70$.

Sehingga taksiran fungsi survival untuk masing-masing kategori adalah

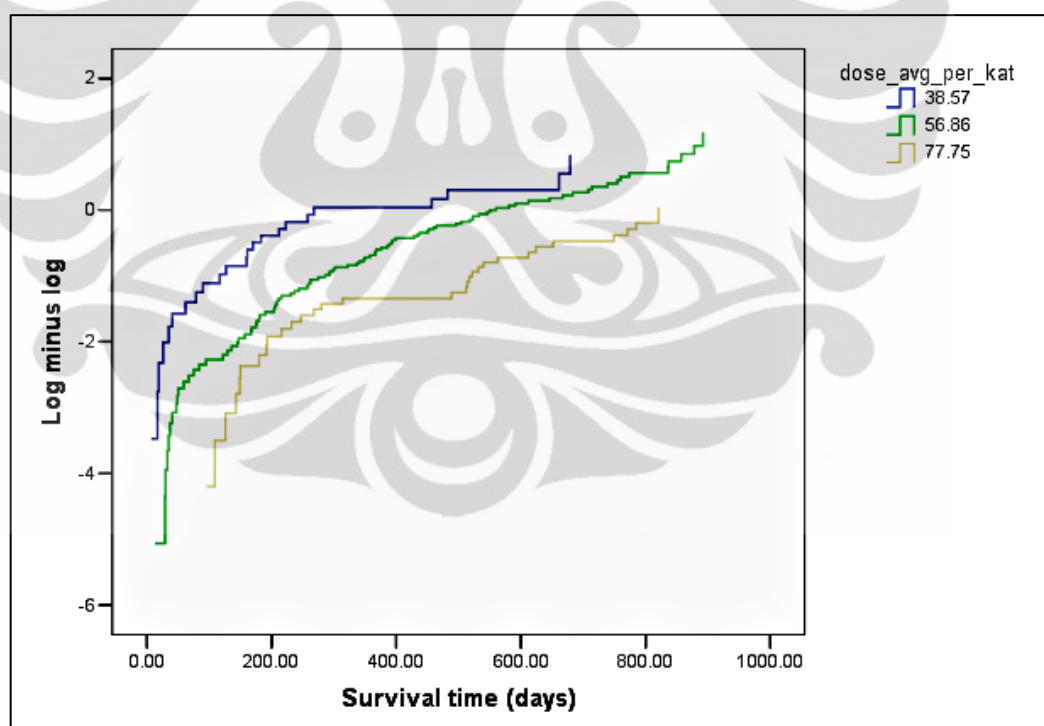
$$\begin{aligned}
\hat{S}(t, \mathbf{X}_{\text{kecil}}) &= \hat{S}_0(t) \exp(-1.009x_1 + 0.327x_2 - 0.035\bar{x}_3(\text{kecil})) \\
\hat{S}(t, \mathbf{X}_{\text{sedang}}) &= \hat{S}_0(t) \exp(-1.009x_1 + 0.327x_2 - 0.035\bar{x}_3(\text{sedang})) \\
\hat{S}(t, \mathbf{X}_{\text{besar}}) &= \hat{S}_0(t) \exp(-1.009x_1 + 0.327x_2 - 0.035\bar{x}_3(\text{besar}))
\end{aligned} \tag{4.7}$$

dan bentuk log-log survivalnya

$$\begin{aligned}
\ln[-\ln \hat{S}(t, \mathbf{X}_{\text{kecil}})] &= (-1.009x_1 + 0.327x_2 - 0.035\bar{x}_3(\text{kecil})) + \ln[-\ln S_0(t)] \\
\ln[-\ln \hat{S}(t, \mathbf{X}_{\text{sedang}})] &= (-1.009x_1 + 0.327x_2 - 0.035\bar{x}_3(\text{sedang})) + \ln[-\ln S_0(t)] \\
\ln[-\ln \hat{S}(t, \mathbf{X}_{\text{besar}})] &= (-1.009x_1 + 0.327x_2 - 0.035\bar{x}_3(\text{besar})) + \ln[-\ln S_0(t)]
\end{aligned} \tag{4.8}$$

Kemudian, dibentuk grafik log-log dan dibandingkan untuk masing-masing kategori.

Hasil dari perangkat lunak ditampilkan sebagai berikut



Gambar 4.3 Grafik log-log *adjusted* untuk *dose*

Dari gambar tersebut terlihat bahwa tidak terjadi perpotongan sehingga dapat disimpulkan asumsi *proportional hazard* terpenuhi untuk variabel *prison* setelah mempertimbangkan keberadaan *clinic* dan *dose*.

Pengecekan asumsi *proportional hazard* dengan menggunakan pendekatan grafik log-log *adjusted* untuk ketiga kovariat, diperoleh hasil bahwa variabel *clinic*, *prison* dan *dose* memenuhi asumsi *proportional hazard*. Namun, hasil ini subjektif tergantung bagaimana kita melihat dan menginterpretasikan kurva-kurva pada masing-masing grafik. Untuk mendapatkan hasil yang lebih objektif akan diperiksa asumsi *proportional hazard* dengan pengujian goodness-of-fit.

4.2.2 Pengecekan dengan Goodness-of-fit

Pada metode ini akan dilihat korelasi antara Schoenfeld residual dengan *survival time*. Asumsi *proportional hazard* dipenuhi jika tidak terdapat korelasi antara Schoenfeld residual dengan *survival time*. Untuk data Caplehorn terdapat 238 individu dan tiga kovariat. Dari data tersebut, yang *survival time*-nya eksak ada 150. Pengecekan asumsi pada metode ini, terdapat tiga pengujian yang akan diuji korelasinya antara *survival time*, t_i , dengan Schoenfeld residual masing-masing kovariat yaitu r_{i1}, r_{i2}, r_{i3} , $i = 1, \dots, 150$.

Penghitungan pada perangkat lunak diperoleh hasil demikian

Tabel 4.4 Hasil pengolahan data untuk Schoenfeld residual

t_i	rank	R_i	X_1	r_1	X_2	r_2	X_3	r_3
7	1	236	1	-0.13374	1	0.48394	40	-13.77672
13	2	235	2	0.86474	1	0.48946	60	6.06619
...								
899	150	9	1	-0.50698	0	-0.32898	60	-13.39413
...								
1076	-	1	2	-	1	-	80	-

Pada $t_1 = 7$, r_{11} atau Schoenfeld residual saat t_1 untuk kovariat X_1 adalah

$$\begin{aligned}
 r_{11} &= X_{11} - E(X_{11} | R_1) \\
 &= X_{11} - \frac{\sum_{q=1}^{236} X_{q1} \exp(-1.009x_{q1} + 0.327x_{q2} - 0.035x_{q3})}{\sum_{q=1}^{236} \exp(-1.009x_{q1} + 0.327x_{q2} - 0.035x_{q3})} \\
 &= X_{11} - \left\{ \frac{X_{11} \exp(-1.009(1) + 0.327(1) - 0.035(40)) + \dots}{\exp(-1.009(1) + 0.327(1) - 0.035(40)) + \dots} \right. \\
 &\quad \left. + \frac{X_{236,1} \exp(-1.009(2) + 0.327(1) - 0.035(80))}{\exp(-1.009(2) + 0.327(1) - 0.035(80))} \right\} \\
 &= 1 - \left\{ \frac{1 \exp(-1.009(1) + 0.327(1) - 0.035(40)) + \dots}{\exp(-1.009(1) + 0.327(1) - 0.035(40)) + \dots} \right. \\
 &\quad \left. + \frac{2 \exp(-1.009(2) + 0.327(1) - 0.035(80))}{\exp(-1.009(2) + 0.327(1) - 0.035(80))} \right\} \\
 &= 1 - 1.13374 \\
 &= -0.13374.
 \end{aligned}$$

Proses yang sama dapat dilakukan untuk penghitungan Schoenfeld residual untuk individu dan kovariat yang lain.

Setelah didapatkan Schoenfeld residual, kemudian akan dilakukan pengujian korelasi antara variabel rank *survival time* dengan Schoenfeld residual dengan hipotesis sebagai berikut:

H_0 : tidak terdapat korelasi antara rank *survival time* dan Schoenfeld residual.

H_1 : terdapat korelasi antara rank *survival time* dan Schoenfeld residual.

Tingkat signifikansi $\alpha = 0.05$.

Tabel 4.5 Hasil pengolahan data untuk pengujian korelasi

	Koefisien korelasi rank Spearman	Sig
<i>Survival time</i> dengan Schoenfeld residual <i>clinic</i>	-0.773	0.000
<i>Survival time</i> dengan Schoenfeld residual <i>prison</i>	0.227	0.005
<i>Survival time</i> dengan Schoenfeld residual <i>dose</i>	0.071	0.386

Dari tabel 4.5 diperoleh nilai sig. yang lebih kecil dari α untuk pengujian korelasi antara *survival time* dengan kovariat *clinic* dan *prison* yang berarti penolakan hipotesis null. Sedangkan untuk pengujian korelasi antara *survival time* dengan kovariat *dose* diperoleh nilai sig. yang lebih besar dari α sehingga hipotesis null tidak ditolak. Jadi, dapat disimpulkan bahwa kovariat *clinic* dan *prison* memiliki korelasi terhadap *survival time* secara signifikan dan tidak terbukti bahwa terdapat korelasi antara *survival time* dengan kovariat *dose*. Sehingga asumsi *proportional hazard* tidak dipenuhi untuk kovariat *clinic* dan *prison* dan hanya kovariat *dose* yang memenuhi asumsi *proportional hazard* pada model Cox.

BAB 5 PENUTUP

5.1 Kesimpulan

Kesimpulan yang diperoleh dari penulisan tugas akhir ini adalah:

1. Pengecekan asumsi *proportional hazard* dengan menggunakan grafik log-log dilihat dari keparalelan kurva pada grafik log-log survival. Dengan menggunakan metode ini, keputusan yang diambil akan subjektif tergantung bagaimana seseorang menginterpretasikan keparalelan kurva pada grafik log-log survival.
2. Pengecekan asumsi *proportional hazard* dilakukan dengan menguji korelasi antara rank *survival time* dengan Schoenfeld residual. Dengan metode ini, keputusan yang diambil akan objektif karena dilakukan dengan pengujian statistik.
3. Pengecekan asumsi *proportional hazard* dengan grafik log-log dan pengujian goodness-of-fit dapat memberikan hasil yang berbeda. Namun, pengecekan dengan grafik log-log dapat dilakukan sebagai dugaan sementara sebelum akhirnya diuji dengan menggunakan goodness-of-fit.

5.2 Saran

Saran untuk pengembangan tulisan ini adalah:

1. Dapat digunakan metode lain dalam pengecekan asumsi *proportional hazard*, misalnya dengan grafik *observed versus expected*, *extended Cox*, atau yang lainnya.
2. Dapat dibahas jika menggunakan data tersensor yang informatif dan dengan skema penyensoran yang lain.
3. Dapat dibahas jika terdapat data yang *ties*.
4. Dapat digunakan koefisien korelasi selain rank Spearman.

DAFTAR PUSTAKA

- Conover, W.J. 1971. *Practical Nonparametric Statistics*. New York: John Wiley & Sons, Inc.
- Fox, John. 2001. *Introduction to Survival Analysis*.
<http://socserv.mcmaster.ca/jfox/Courses/soc761/survival-analysis.pdf>.
8 Oktober 2010. pk. 13.08.
- Friel, C.M. 2001. *Cox Proportional Hazard Model*.
http://www.shsu.edu/~icc_cmf/cj_789/coxRegression.doc. 14 Februari
2011. pk. 14.29
- Guo, S. 2010. *Survival Analysis – Pocket Guides to Social Work Research Methods*. New York: Oxford University Press, Inc.
- Hosmer, D.W. and Lemeshow, S. 1999. *Applied Survival Analysis – Regression Modelling of Time to Event Data*. New York: John Wiley & Sons, Inc.
- Hollander, M. and Wolfe, D.A. 1999. *Nonparametric Statistical Methods*, Second Edition. New York: John Wiley & Sons, Inc.
- Kleinbaum, D.G. and Klein, M. 2005. *Survival Analysis – A Self Learning Text*, Second Edition. New York: Springer.
- Klein, J.P. and Moeschberger, M.L. 1997. *Survival Analysis – Techniques for Censored and Truncated Data*. New York: Springer-Verlag.
- Schoenfeld, D. 1982. Partial Residuals for The Proportional Hazards Regression Model. *Biometrika*, 69 (1), 239-241.

Zulfitri. 2009 *Koefisien Korelasi Rank Spearman: r_s* .

http://pksm.mercubuana.ac.id/new/elearning/files_modul/94020-14-258272392854.doc. 10 April 2011. pk 00.01.



LAMPIRAN

Lampiran 1

Pembuktian persamaan 3.29

$$r_S^* = (n-1)^{1/2} r_S$$

Bukti:

Di bawah kondisi H_0 , X_i dan Y_i saling bebas, rank $(R(X_1), R(X_2), \dots, R(X_n))$ dan rank $(R(Y_1), R(Y_2), \dots, R(Y_n))$ independen dan masing-masing terdistribusi uniform dalam himpunan $n!$ permutasi dari $(1, 2, \dots, n)$. Sehingga variabel random $\sum_{i=1}^n R(X_i)R(Y_i)$ dan $\sum_{j=1}^n jR(Y_j)$ memiliki distribusi yang sama di bawah kondisi H_0 . Sehingga berdasarkan persamaan (3.28), diperoleh

$$r_S = \frac{12 \sum_{j=1}^n jR(Y_j)}{n(n^2-1)} - 3 \left(\frac{n+1}{n-1} \right)$$

dan

$$\begin{aligned} E(r_S) &= E \left[\frac{12 \sum_{j=1}^n jR(Y_j)}{n(n^2-1)} - 3 \left(\frac{n+1}{n-1} \right) \right] \\ &= \frac{12 \sum_{j=1}^n jE(R(Y_j))}{n(n^2-1)} - 3 \left(\frac{n+1}{n-1} \right) \end{aligned}$$

Karena untuk setiap $j=1, \dots, n$, $R(Y_j)$ memiliki distribusi yang uniform pada himpunan $(1, 2, \dots, n)$, sehingga

$$E(R(Y_j)) = \sum_{v=1}^n v \left(\frac{1}{n} \right) = (1/n) \sum_{v=1}^n v = (1/n) \left(\frac{n(n+1)}{2} \right) = \frac{(n+1)}{2}$$

untuk $j=1, 2, \dots, n$.

(lanjutan)

Maka,

$$\begin{aligned}
 E(r_S) &= \frac{12 \sum_{j=1}^n j E(R(Y_j))}{n(n^2-1)} - 3 \left(\frac{n+1}{n-1} \right) \\
 &= \frac{12 \sum_{j=1}^n j \frac{n+1}{2}}{n(n^2-1)} - 3 \left(\frac{n+1}{n-1} \right) \\
 &= \frac{12 \frac{n+1}{2} \sum_{j=1}^n j}{n(n^2-1)} - 3 \left(\frac{n+1}{n-1} \right) \\
 &= \frac{12 \left(\frac{n+1}{2} \right) \frac{n(n+1)}{2}}{n(n^2-1)} - 3 \left(\frac{n+1}{n-1} \right) \\
 &= \frac{3n(n+1)(n+1)}{n(n+1)(n-1)} - 3 \left(\frac{n+1}{n-1} \right) \\
 &= \frac{3(n+1)}{(n-1)} - 3 \left(\frac{n+1}{n-1} \right) \\
 &= 0
 \end{aligned}$$

Di bawah kondisi H_0 , maka variansi r_S adalah

$$\begin{aligned}
 \text{var}(r_S) &= \text{var} \left[\frac{12 \sum_{j=1}^n j R(Y_j)}{n(n^2-1)} - 3 \left(\frac{n+1}{n-1} \right) \right] \\
 &= \left(\frac{12}{n(n^2-1)} \right)^2 \text{var} \left(\sum_{j=1}^n j R(Y_j) \right)
 \end{aligned}$$

dengan

$$\begin{aligned}
 \text{var} \left(\sum_{j=1}^n j R(Y_j) \right) &= \sum_{j=1}^n \text{var} (j R(Y_j)) + \sum_{j=1}^n \sum_{v=1, v \neq j}^n \text{cov} (j R(Y_j), v R(Y_v)) \\
 &= \sum_{j=1}^n j^2 \text{var} (R(Y_j)) + \sum_{j=1}^n \sum_{v=1, v \neq j}^n jv \text{cov} (R(Y_j), R(Y_v))
 \end{aligned}$$

karena distribusi bersama $(R(Y_j), R(Y_v))$ sama untuk setiap $j \neq v = 1, \dots, n$ dan

distribusi marginal $R(Y_j)$ sama untuk setiap $j = 1, \dots, n$, sehingga

(lanjutan)

$$\begin{aligned}\operatorname{var}\left(\sum_{j=1}^n jR(Y_j)\right) &= \operatorname{var}(R(Y_a))\sum_{j=1}^n j^2 + \operatorname{cov}(R(Y_a), R(Y_b))\sum_{j=1}^n \sum_{v=1, v \neq j}^n jv \\ &= \operatorname{var}(R(Y_a))\sum_{j=1}^n j^2 + \operatorname{cov}(R(Y_a), R(Y_b))\left\{\left(\sum_{j=1}^n j\right)\left(\sum_{v=1}^n v\right) - \sum_{j=1}^n j^2\right\}\end{aligned}$$

Dengan menggunakan

$$\sum_{j=1}^n j^2 = \frac{n(n+1)(2n+1)}{6}$$

dan

$$\begin{aligned}\sum_{j=1}^n \sum_{v=1, v \neq j}^n jv &= \left(\sum_{j=1}^n j\right)\left(\sum_{v=1}^n v\right) - \sum_{j=1}^n j^2 \\ &= \left(\frac{n(n+1)}{2}\right)^2 - \frac{n(n+1)(n+2)}{6} \\ &= \frac{n(n^2-1)(3n+2)}{12}\end{aligned}$$

diperoleh

$$\begin{aligned}\operatorname{var}\left(\sum_{j=1}^n jR(Y_j)\right) &= \operatorname{var}(R(Y_a))\sum_{j=1}^n j^2 + \operatorname{cov}(R(Y_a), R(Y_b))\sum_{j=1}^n \sum_{v=1, v \neq j}^n jv \\ &= \frac{n(n+1)(2n+1)}{6} \operatorname{var}(R(Y_a)) \\ &\quad + \frac{n(n^2-1)(3n+2)}{12} \operatorname{cov}(R(Y_a), R(Y_b))\end{aligned}$$

Dapat ditunjukkan bahwa $\operatorname{var}(R(Y_a)) = (n^2 - 1)/12$ dan

$$\operatorname{cov}(R(Y_a), R(Y_b)) = -(n+1)/12.$$

Maka kita mendapatkan

$$\begin{aligned}\operatorname{var}\left(\sum_{j=1}^n jR(Y_j)\right) &= \operatorname{var}(R(Y_a))\sum_{j=1}^n j^2 + \operatorname{cov}(R(Y_a), R(Y_b))\sum_{j=1}^n \sum_{v=1, v \neq j}^n jv \\ &= \frac{n(n+1)(2n+1)}{6} \left(\frac{n^2-1}{12}\right) + \frac{n(n^2-1)(3n+2)}{12} \left(-\frac{n+1}{12}\right)\end{aligned}$$

(lanjutan)

$$\begin{aligned}
 &= \frac{n(n+1)(2n+1)(n^2-1)}{72} - \frac{n(n^2-1)(3n+2)(n+1)}{144} \\
 &= \frac{2n(n+1)(2n+1)(n^2-1)}{144} - \frac{n(n^2-1)(3n+2)(n+1)}{144} \\
 &= \frac{n(n+1)(n^2-1)}{144} (2(2n+1) - (3n+2)) \\
 &= \frac{n(n+1)(n^2-1)}{144} (n) \\
 &= \frac{n^2(n+1)(n^2-1)}{144}
 \end{aligned}$$

Sehingga,

$$\begin{aligned}
 \text{var}(r_s) &= \left(\frac{12}{n(n^2-1)} \right)^2 \text{var} \left(\sum_{j=1}^n jR(Y_j) \right) \\
 &= \left(\frac{12}{n(n^2-1)} \right)^2 \frac{n^2(n+1)(n^2-1)}{144} \\
 &= \left(\frac{144}{n^2(n^2-1)^2} \right) \frac{n^2(n+1)(n^2-1)}{144} \\
 &= \frac{n+1}{n^2-1} \\
 &= \frac{n+1}{(n+1)(n-1)} \\
 &= \frac{1}{n-1}
 \end{aligned}$$

Di bawah kondisi H_0 , bentuk terstandarisasi r_s adalah

$$r_s^* = \frac{r_s - E(r_s)}{\{\text{var}(r_s)\}^{1/2}} = \frac{r_s - 0}{\{1/(n-1)\}^{1/2}} = (n-1)^{1/2} r_s.$$

Lampiran 2

Pembuktian $\text{var}(R(Y_a)) = (n^2 - 1)/12$.

Untuk setiap $j = 1, \dots, n$, $R(Y_j)$ memiliki distribusi yang uniform pada himpunan $(1, 2, \dots, n)$ dan distribusi bersama $(R(Y_j), R(Y_v))$ sama untuk setiap $j \neq v = 1, \dots, n$. Sehingga

$$\begin{aligned}
 \text{var}(R(Y_a)) &= E\left\{\left[R(Y_a) - E(R(Y_a))\right]^2\right\} \\
 &= E\left\{\left(R(Y_a) - \frac{n+1}{2}\right)^2\right\} \\
 &= E\left\{R(Y_a)^2 - 2R(Y_a)\frac{n+1}{2} + \left(\frac{n+1}{2}\right)^2\right\} \\
 &= E\left\{R(Y_a)^2\right\} - (n+1)E(R(Y_a)) + \frac{(n+1)^2}{4} \\
 &= E\left\{R(Y_a)^2\right\} - (n+1)\frac{n+1}{2} + \frac{(n+1)^2}{4} \\
 &= E\left\{R(Y_a)^2\right\} - \frac{(n+1)^2}{4} \\
 &= \sum_{j=1}^n j^2 \left(\frac{1}{n}\right) - \frac{(n+1)^2}{4} \\
 &= \frac{1}{n} \sum_{j=1}^n j^2 - \frac{(n+1)^2}{4} \\
 &= \frac{1}{n} \left(\frac{n(n+1)(2n+1)}{6}\right) - \frac{(n+1)^2}{4} \\
 &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\
 &= \frac{(n+1)[2(2n+1) - 3(n+1)]}{12} \\
 &= \frac{(n+1)(n-1)}{12} \\
 &= \frac{(n^2 - 1)}{12}
 \end{aligned}$$

Lampiran 3

Pembuktian $\text{cov}(R(Y_a), R(Y_b)) = -(n+1)/12$.

Untuk setiap $j = 1, \dots, n$, $R(Y_j)$ memiliki distribusi yang uniform pada himpunan $(1, 2, \dots, n)$ dan distribusi bersama $(R(Y_j), R(Y_v))$ sama untuk setiap $j \neq v = 1, \dots, n$. Sehingga

$$\begin{aligned} \text{cov}(R(Y_a), R(Y_b)) &= E\left\{\left[R(Y_a) - E(R(Y_a))\right]\left[R(Y_b) - E(R(Y_b))\right]\right\} \\ &= E\left\{\left(R(Y_a) - \frac{n+1}{2}\right)\left(R(Y_b) - \frac{n+1}{2}\right)\right\} \\ &= \sum_{j=1}^n \sum_{v=1, v \neq j}^n \left(j - \frac{n+1}{2}\right)\left(v - \frac{n+1}{2}\right)\left(\frac{1}{n(n-1)}\right) \\ &= \sum_{j=1}^n \sum_{v=1}^n \left(j - \frac{n+1}{2}\right)\left(v - \frac{n+1}{2}\right)\left(\frac{1}{n(n-1)}\right) \\ &\quad - \sum_{j=1}^n \left(j - \frac{n+1}{2}\right)^2 \left(\frac{1}{n(n-1)}\right) \\ &= \left(\frac{1}{n(n-1)}\right) \sum_{j=1}^n \left(j - \frac{n+1}{2}\right) \sum_{v=1}^n \left(v - \frac{n+1}{2}\right) \\ &\quad - \left(\frac{1}{n(n-1)}\right) \sum_{j=1}^n \left(j - \frac{n+1}{2}\right)^2 \end{aligned}$$

Untuk

$$\begin{aligned} \sum_{j=1}^n \left(j - \frac{n+1}{2}\right) &= \sum_{j=1}^n j - \sum_{j=1}^n \frac{n+1}{2} \\ &= \frac{n(n+1)}{2} - \frac{n(n+1)}{2} \\ &= 0 \end{aligned}$$

dan

$$\begin{aligned} \sum_{j=1}^n \left(j - \frac{n+1}{2}\right)^2 &= \sum_{j=1}^n j^2 - j(n+1) + \left(\frac{n+1}{2}\right)^2 \\ &= \sum_{j=1}^n j^2 - (n+1) \sum_{j=1}^n j + n \left(\frac{n+1}{2}\right)^2 \end{aligned}$$

(lanjutan)

$$\begin{aligned}
 &= \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)(n+1)}{2} + \frac{n(n+1)(n+1)}{4} \\
 &= n(n+1) \frac{4n+2-6n-6+3n+3}{12} \\
 &= \frac{n(n+1)(n-1)}{12}
 \end{aligned}$$

sehingga

$$\begin{aligned}
 \text{cov}(R(Y_a), R(Y_b)) &= \left(\frac{1}{n(n-1)} \right) \sum_{j=1}^n \left(j - \frac{n+1}{2} \right) \sum_{v=1}^n \left(v - \frac{n+1}{2} \right) \\
 &\quad - \left(\frac{1}{n(n-1)} \right) \sum_{j=1}^n \left(j - \frac{n+1}{2} \right)^2 \\
 &= - \frac{1}{n(n-1)} \frac{n(n+1)(n-1)}{12} \\
 &= - \frac{n+1}{12}
 \end{aligned}$$

Lampiran 4

Tabel data Caplehorn dari 238 pecandu heroin.

Keterangan: *survival time* (hari), *status* (1 = teramati, 0 = tersensor), *clinic* (1 atau 2), *prison* (0 = tidak ada, 1 = ada), *dose* (mg/hari).

<i>surv. time</i>	<i>status</i>	<i>clinic</i>	<i>prison</i>	<i>dose</i>
428	1	1	0	50
275	1	1	1	55
262	1	1	0	55
183	1	1	0	30
259	1	1	1	65
714	1	1	0	55
438	1	1	1	65
796	0	1	1	60
892	1	1	0	50
393	1	1	1	65
161	0	1	1	80
836	1	1	1	60
523	1	1	0	55
612	1	1	0	70
212	1	1	1	60
399	1	1	1	60
771	1	1	1	75
514	1	1	1	80
512	1	1	0	80
624	1	1	1	80
209	1	1	1	60
341	1	1	1	60
299	1	1	0	55
826	0	1	0	80
262	1	1	1	65
566	0	1	1	45
368	1	1	1	55
302	1	1	1	50
602	0	1	0	60
652	1	1	0	80
293	1	1	0	65
564	0	1	0	60
394	1	1	1	55
755	1	1	1	65
591	1	1	0	55
787	0	1	0	80

739	1	1	0	60
550	1	1	1	60
837	1	1	0	60
612	1	1	0	65
581	0	1	0	70
523	1	1	0	60
504	1	1	1	60
785	1	1	1	80
774	1	1	1	65
560	1	1	0	65
160	1	1	0	35
482	1	1	0	30
518	1	1	0	65
683	1	1	0	50
147	1	1	0	65
563	1	1	1	70
646	1	1	1	60
899	1	1	0	60
857	1	1	0	60
180	1	1	1	70
452	1	1	0	60
760	1	1	0	60
496	1	1	0	65
258	1	1	1	40
181	1	1	1	60
386	1	1	0	60
439	0	1	0	80
563	0	1	0	75
337	1	1	0	65
613	0	1	1	60
192	1	1	1	80
405	0	1	0	80
667	1	1	0	50
905	0	1	0	80
247	1	1	0	70
821	1	1	0	80
821	1	1	1	75
517	0	1	0	45
346	0	1	1	60
294	1	1	0	65
244	1	1	1	60
95	1	1	1	60
376	1	1	1	55
212	1	1	0	40
96	1	1	0	70
532	1	1	0	80
522	1	1	1	70
679	1	1	0	35

408	0	1	0	50
840	0	1	0	80
148	0	1	1	65
168	1	1	0	65
489	1	1	0	80
541	0	1	0	80
205	1	1	0	50
475	0	1	1	75
237	1	1	0	45
517	1	1	0	70
749	1	1	0	70
150	1	1	1	80
465	1	1	0	65
708	1	2	1	60
713	0	2	0	50
146	0	2	0	50
450	1	2	0	55
555	0	2	0	80
460	1	2	0	50
53	0	2	1	60
122	1	2	1	60
35	1	2	1	40
532	0	2	0	70
684	0	2	0	65
769	0	2	1	70
591	0	2	0	70
769	0	2	1	40
609	0	2	1	100
932	0	2	1	80
932	0	2	1	80
587	0	2	0	110
26	1	2	0	40
72	0	2	1	40
641	0	2	0	70
367	0	2	0	70
633	0	2	0	70
661	1	2	0	40
232	1	2	1	70
13	1	2	1	60
563	0	2	0	70
969	0	2	0	80
1052	0	2	0	80
944	0	2	1	80
881	0	2	0	80
190	1	2	1	50
79	1	2	0	40
884	0	2	1	50
170	1	2	0	40

286	1	2	0	45
358	0	2	0	60
326	0	2	1	60
769	0	2	1	40
161	1	2	0	40
564	0	2	1	80
268	1	2	1	70
611	0	2	1	40
322	1	2	0	55
1076	0	2	1	80
2	0	2	1	40
788	0	2	0	70
575	0	2	0	80
109	1	2	1	70
730	0	2	1	80
790	0	2	0	90
456	0	2	1	70
231	1	2	1	60
143	1	2	1	70
86	0	2	1	40
1021	0	2	0	80
684	0	2	1	80
878	1	2	1	60
216	1	2	0	100
808	0	2	0	60
268	1	2	1	40
222	0	2	0	40
683	0	2	0	100
496	0	2	0	40
389	1	2	0	55
126	1	1	1	75
17	1	1	1	40
350	1	1	0	60
531	0	2	1	65
317	0	1	1	50
461	0	1	1	75
37	1	1	0	60
167	1	1	1	55
358	1	1	0	45
49	1	1	0	60
457	1	1	1	40
127	1	1	0	20
7	1	1	1	40
29	1	1	1	60
62	1	1	0	40
150	0	1	1	60
223	1	1	1	40
129	0	1	1	40

204	0	1	1	65
129	1	1	1	50
581	1	1	0	65
176	1	1	0	55
30	1	1	0	60
41	1	1	0	60
543	0	1	0	40
210	0	1	1	50
193	1	1	1	70
434	1	1	0	55
367	1	1	0	45
348	1	1	1	60
28	0	1	0	50
337	0	1	0	40
175	0	1	1	60
149	1	2	1	80
546	1	1	1	50
84	1	1	0	45
283	0	1	1	80
533	1	1	0	55
207	1	1	1	50
216	1	1	0	50
28	0	1	0	50
67	1	1	1	50
62	0	1	1	60
111	0	1	0	55
257	1	1	1	60
136	1	1	1	55
342	0	1	0	60
41	1	2	0	40
531	0	2	1	45
98	0	1	0	40
145	1	1	1	55
50	1	1	0	50
53	0	1	0	50
103	0	1	1	50
2	0	1	1	60
157	1	1	1	60
75	1	1	1	55
19	1	1	1	40
35	1	1	0	60
394	0	2	1	80
117	1	1	0	40
175	1	1	1	60
180	1	1	1	60
314	1	1	0	70
480	0	1	0	50
325	0	1	1	60

280	1	2	0	90
204	1	1	0	50
366	1	2	0	55
531	0	2	1	50
59	1	1	1	45
33	1	1	1	60
540	1	2	0	80
551	0	2	0	65
90	1	1	0	40
47	1	1	0	45



Lampiran 5

Tabel data Caplehorn dari 238 pecandu heroin dengan Schoenfeld residual.

<i>ranked surv. time</i>	<i>Schoenfeld resid. clinic</i>	<i>Schoenfeld resid. prison</i>	<i>Schoenfeld resid. dose</i>
2	.	.	.
2	.	.	.
7	-0.13374	0.48394	-13.7767
13	0.86474	0.48946	6.06619
17	-0.13349	0.49047	-13.9214
19	-0.13503	0.49613	-14.0823
26	0.86339	-0.49806	-14.2469
28	.	.	.
28	.	.	.
29	-0.13555	0.49439	5.65754
30	-0.13635	-0.5027	5.69082
33	-0.13693	0.49516	5.71508
35	0.86225	0.4981	-14.251
35	-0.13775	-0.5019	5.74904
37	-0.13452	-0.50186	5.71067
41	-0.1351	-0.50404	5.73548
41	0.8649	-0.50404	-14.2645
47	-0.13289	-0.50788	-9.28562
49	-0.13389	-0.51169	5.64475
50	-0.13448	-0.51395	-4.33025
53	.	.	.
53	.	.	.
59	-0.13424	0.48057	-9.37284
62	-0.13567	-0.51428	-14.4732
62	.	.	.
67	-0.13781	0.484	-4.57346
72	.	.	.
75	-0.13495	0.4908	0.31429
79	0.864	-0.50537	-14.6833
84	-0.13297	-0.50714	-9.73469
86	.	.	.
90	-0.12979	-0.50885	-14.8866
95	-0.13107	0.48615	4.96712
96	-0.13195	-0.51056	15.00069
98	.	.	.
103	.	.	.
109	0.86495	0.4873	14.85205
111	.	.	.
117	-0.13432	-0.51491	-15.1223

122	0.8643	0.47978	4.72172
126	-0.13348	0.48102	19.73387
127	-0.13404	-0.51698	-35.1839
129	.	.	.
129	-0.13692	0.47189	-5.94121
136	-0.14043	0.484	-1.24443
143	0.85831	0.48833	13.74444
145	-0.14004	0.48927	-1.22915
146	.	.	.
147	-0.13886	-0.50775	8.74197
148	.	.	.
149	0.8596	0.49306	23.83882
150	-0.13921	0.49374	23.87177
150	.	.	.
157	-0.14083	0.49948	3.99343
160	-0.14194	-0.4966	-20.9752
161	.	.	.
161	0.85609	-0.5035	-16.2668
167	-0.14081	0.4963	-1.24269
168	-0.14217	-0.4989	8.7453
170	0.85713	-0.50136	-16.2117
175	.	.	.
175	-0.13913	0.49646	3.71757
176	-0.14144	-0.4953	-1.22067
180	-0.14246	0.50114	13.77054
180	-0.14246	0.50114	3.77054
181	-0.14452	0.50837	3.88457
183	-0.14576	-0.48727	-26.0822
190	0.85159	0.50387	-6.55642
192	-0.14453	0.50616	23.41376
193	-0.14516	0.50835	13.51519
204	.	.	.
204	-0.14606	-0.48849	-6.40088
205	-0.14851	-0.48915	-6.39525
207	-0.1499	0.50628	-6.45494
209	-0.15186	0.51292	3.46046
210	.	.	.
212	-0.15533	0.52462	3.40536
212	-0.15533	-0.47538	-16.5946
216	0.84098	-0.47703	43.20409
216	-0.15902	-0.47703	-6.79591
222	.	.	.
223	-0.1557	0.51538	-16.925
231	0.84115	0.52582	2.732
232	0.84422	0.52774	12.74198
237	-0.15361	-0.4709	-12.2253
244	-0.15552	0.52322	2.62219

247	-0.15712	-0.4714	12.64913
257	-0.15794	0.52613	2.7153
258	-0.15959	0.53162	-17.2564
259	-0.16305	0.54313	7.37014
262	-0.16452	-0.45197	-2.56343
262	-0.16452	0.54803	7.43657
268	0.83242	0.54879	12.48066
268	0.83242	0.54879	-17.5193
275	-0.15839	0.55485	-2.62762
280	0.83949	-0.43771	32.33716
283	.	.	.
286	0.83945	-0.43503	-12.5052
293	-0.15628	-0.43724	7.43125
294	-0.15736	-0.44027	7.48263
299	-0.15845	-0.44333	-2.46527
302	-0.16004	0.55223	-7.48997
314	-0.16274	-0.43846	12.38378
317	.	.	.
322	0.83346	-0.43147	-2.67167
325	.	.	.
326	.	.	.
337	-0.16161	-0.42355	7.35723
337	.	.	.
341	-0.16585	0.56533	2.08597
342	.	.	.
346	.	.	.
348	-0.17178	0.57607	2.16054
350	-0.17407	-0.41626	2.1893
358	.	.	.
358	-0.17576	-0.4203	-12.7895
366	0.82426	-0.42889	-2.99668
367	.	.	.
367	-0.17218	-0.43074	-13.0096
368	-0.173	0.56072	-3.20274
376	-0.17595	0.57028	-3.25735
386	-0.179	-0.41983	1.68614
389	0.81911	-0.42428	-3.29599
393	-0.1771	0.57376	6.68873
394	-0.17932	0.58094	-3.22754
394	.	.	.
399	-0.18033	0.59311	1.77336
405	.	.	.
408	.	.	.
428	-0.18714	-0.40641	-8.20992
434	-0.19023	-0.41313	-3.3457
438	-0.1929	0.58106	6.60729
439	.	.	.

450	0.80327	-0.41332	-3.17279
452	-0.19251	-0.41549	1.81054
456	.	.	.
457	-0.19136	0.58196	-18.1158
460	0.8018	-0.39723	-8.76355
461	.	.	.
465	-0.19491	-0.39355	6.34868
475	.	.	.
480	.	.	.
482	-0.20298	-0.39888	-28.5662
489	-0.21121	-0.41504	20.27655
496	-0.21267	-0.41792	5.41746
496	.	.	.
504	-0.20684	0.57256	0.27385
512	-0.21106	-0.41575	20.27944
514	-0.2126	0.58121	20.42759
517	.	.	.
517	-0.21478	-0.41285	10.63657
518	-0.22279	-0.42826	5.37302
522	-0.22573	0.56609	10.44389
523	-0.22924	-0.4251	-4.39363
523	-0.22924	-0.4251	0.60637
531	.	.	.
531	.	.	.
531	.	.	.
532	-0.2114	-0.42135	20.23959
532	.	.	.
533	-0.20971	-0.42682	-4.54178
540	0.78586	-0.43586	20.3621
541	.	.	.
543	.	.	.
546	-0.22157	0.54221	-10.1253
550	-0.23003	0.56291	-0.51182
551	.	.	.
555	.	.	.
560	-0.22894	-0.42569	4.57167
563	-0.23285	0.56703	9.64981
563	.	.	.
563	.	.	.
564	.	.	.
564	.	.	.
566	.	.	.
575	.	.	.
581	.	.	.
581	-0.24713	-0.4064	4.4739
587	.	.	.
591	-0.25511	-0.42197	-5.20391

591	.	.	.
602	.	.	.
609	.	.	.
611	.	.	.
612	-0.24238	-0.43164	9.27917
612	-0.24238	-0.43164	4.27917
613	.	.	.
624	-0.26248	0.57185	19.53702
633	.	.	.
641	.	.	.
646	-0.25666	0.57687	0.07155
652	-0.26754	-0.39868	20.07458
661	0.72837	-0.40478	-19.6184
667	-0.2545	-0.41429	-10.0797
679	-0.26657	-0.43394	-25.5576
683	-0.28995	-0.47199	-12.799
683	.	.	.
684	.	.	.
684	.	.	.
708	0.71175	0.49957	-3.18971
713	.	.	.
714	-0.25851	-0.50074	-8.53471
730	.	.	.
739	-0.26388	-0.52124	-3.79832
749	-0.27582	-0.54483	6.02974
755	-0.28488	0.43729	1.22761
760	-0.30123	-0.53761	-3.70191
769	.	.	.
769	.	.	.
769	.	.	.
771	-0.21917	0.49658	8.31981
774	-0.23033	0.52187	-1.25651
785	-0.24835	0.5627	13.6452
787	.	.	.
788	.	.	.
790	.	.	.
796	.	.	.
808	.	.	.
821	-0.25485	-0.38336	14.41053
821	-0.25485	0.61664	9.41053
826	.	.	.
836	-0.30197	0.62946	-3.34324
837	-0.3521	-0.26605	-3.89822
840	.	.	.
857	-0.42874	-0.32396	-3.30987
878	0.49806	0.62074	-3.87494
881	.	.	.

884	.	.	.
892	-0.34475	-0.22371	-15.9082
899	-0.50698	-0.32898	-13.3941
905	.	.	.
932	.	.	.
932	.	.	.
944	.	.	.
969	.	.	.
1021	.	.	.
1052	.	.	.
1076	.	.	.

