UNIVERSITAS INDONESIA

EXTENDED COX MODEL UNTUK *TIME-INDEPENDENT COVARIATE* YANG TIDAK MEMENUHI ASUMSI *PROPORTIONAL HAZARD* PADA MODEL *COX PROPORTIONAL HAZARD*

SKRIPSI

ISNA NUR AINI
0706261732

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

UNIVERSITAS INDONESIA

EXTENDED COX MODEL UNTUK *TIME-INDEPENDENT COVARIATE* YANG TIDAK MEMENUHI ASUMSI *PROPORTIONAL HAZARD* PADA MODEL *COX PROPORTIONAL HAZARD*

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana sains

ISNA NUR AINI
0706261732

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	: Isna Nur Aini
NPM	: 0706261732
Tanda Tangan	:
Tanggal	: 24 Juni 2011

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh	:
Nama	: Isna Nur Aini
NPM	: 0706261732
Program Studi	: Matematika
Judul Skripsi	: *Extended Cox Model* untuk *time-independent covariate* yang tidak memenuhi asumsi *proportional hazard* pada model *cox proportional hazard*

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

DEWAN PENGUJI

Pembimbing	: Sarini Abdullah, M.Stats.	( )
Penguji	: Dr. Dian Lestari	( )
Penguji	: Fevi Novkaniza, M.Si.	( )
Penguji	: Mila Novita, M.Si.	( )

Ditetapkan di	: Depok
Tanggal	: 30 Mei 2011

KATA PENGANTAR

Puji syukur dipanjatkan kehadirat Allah SWT, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan skripsi ini. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Sience 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 untuk menyelesaikan skripsi ini. Oleh karena itu, diucapkan terima kasih kepada:

(1) Sarini Abdullah, M.Stat, selaku dosen pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk memberikan bimbingan dan arahan dalam penyusunan skripsi ini
(2) Dra. Yahma selaku pembimbing akademik yang telah memberikan arahan dan masukan selama 4 tahun masa perkuliahan
(3) Ketua dan Sekretaris Departemen Matematika, Dr. Yudi Satria dan Rahmi Rusin, M.ScTech, atas segala bantuan serta dukungan yang telah diberikan
(4) Seluruh staf Tata Usaha, staf Perpustakaan, serta karyawan Departemen Matematika, atas segala bantuannya
(5) Papa dan Mama yang telah memberikan seluruh perhatian, nasihat, dan motivasinya dari semenjak kecil hingga penulis menjadi sarjana, juga atas kesabarannya membimbing penulis sehingga penulis bisa lulus tepat waktu
(6) Kakak, kakak ipar, adikku tersayang (Mas Amri, Mba Fitri, dan Dek Tia), dan seluruh saudara yang turut mendukung serta mendoakan sehingga penulis dapat menyelesaikan skripsi ini dengan baik dan lancar
(8) Teman-teman matematika angkatan 2007 yang telah memberikan motivasinya
(9) Sahabatku, Gamar, yang telah membantu membuat slide presentasi menjadi lebih bagus dan membantu kesulitan penulis dalam pengetikan tugas akhir ini
(10) Sahabat-sahabatku lainnya, Anjar yang telah menyediakan kamarnya untuk mengerjakan tugas akhir bersama, Shafa yang telah membantu mendapatkan

software R dan membantu memahami cara menggunakannya, Stefi, Wiwi, Sisca, Putri, dan Misda, yang telah mendoakan penulis

(11) Adik-adikku di matematika (Aci, Luthfa, Dewe, Icha, Vika) yang selalu memberi semangat kepada penulis. Segera menyusul ya..^^

(12) Saudari-saudariku di "Lingkaran Cahaya" yang selalu mendoakan dan memberikan semangat sehingga skripsi ini dapat terselesaikan dengan baik

Akhir kata, penulis berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga skripsi ini membawa manfaat bagi pengembangan ilmu.

Penulis

2011

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama	: Isna Nur Aini
NPM	: 0706261732
Program Studi	: Sarjana
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 :

Extended Cox Model untuk *time-independent covariate* yang tidak memenuhi asumsi *proportional hazard* pada model *cox proportional hazard*

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 : 24 Juni 2011

Yang menyatakan

(Isna Nur Aini)

ABSTRAK

Nama : Isna Nur Aini
Program Studi : Matematika
Judul : *Extended Cox Model* untuk *Time-Independent Covariate* yang Tidak Memenuhi Asumsi *Proportional Hazard* pada Model *Cox Proportional Hazard*

Extended cox model adalah perluasan dari model cox yaitu dengan melibatkan variabel yang bergantung pada waktu. Model ini digunakan untuk memperbaiki model cox proportional hazard apabila satu atau lebih *covariate* tidak memenuhi asumsi *proportional hazard*. *Covariate* yang tidak memenuhi asumsi *proportional hazard* dalam *extended cox model* diinteraksikan dengan fungsi waktu, sehingga diperoleh *covariate* yang bergantung pada waktu. Sehingga pada model terdapat *covariate* yang tidak bergantung pada waktu dan *covariate* yang bergantung pada waktu. Parameter-parameter dari *covariate* tersebut ditaksir dengan menggunakan metode maksimum partial likelihood. Untuk mengetahui apakah *extended cox model* adalah model yang sesuai untuk suatu data dalam kasus tertentu, digunakan uji ratio likelihood. Sebagai contoh penerapan digunakan data berupa waktu seorang pasien mengalami infeksi pada ginjal setelah dilakukan transplantasi ginjal, dengan *covariate* yang diperhatikan yaitu pemasangan *catheter* pada ginjal pasien. Diperoleh hasil bahwa model yang sesuai untuk data tersebut adalah *extended cox model*.

Kata Kunci : fungsi *hazard*, model *cox proportional hazard*, *extended cox model*, estimasi maksimum partial likelihood, uji ratio likelihood, sensor kanan tipe I.
xiv+77 halaman: 11 gambar; 34 tabel
Daftar Pustaka : 6 (1987-2005)

ABSTRACT

Name : Isna Nur Aini
Program Study : Mathematic
Title : Extended Cox Model for Time-Independent Covariates that Violate the Proportional Hazard Assumption in Cox Proportional Hazard Model

Extended cox model is an extension of cox model by constructing time-dependent covariates that can be added to the model. This model is used to adjust the cox proportional hazard model if one or more covariates do not satisfy the proportional hazard assumption. Covariates, which do not satisfy the proportional hazard assumption, in extended cox model, are interacted with time function, so that time-dependent variables are obtained. Therefore, the model contains both time-independent and time-dependent covariates. Parameters of these covariates are estimated by maximum partial likelihood method. To find out whether extended cox model is better than cox proportional hazard model, ratio likelihood test is used. As the example, data of the period of time a patient suffering kidney infection after having kidney transplantation, with the concerned covariate is placed catheter in patient's kidney. The result showed that extended cox model is appropriate for the data.

Key Words : hazard function, cox proportional hazard model, extended cox model, maximum partial likelihood estimator, ratio likelihood test, right cencoring type I.
xiv+77 pages : 11 pictures; 34 tables
Biblioraphy : 6 (1987-2005)

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Analisis *survival* adalah sekumpulan prosedur statistik untuk menganalisis data dengan variabel *respon* yang diperhatikan berupa waktu sampai terjadinya suatu peristiwa atau *event* (David G. Kleinbaum and Mitchel Klein, 2005). Variabel yang menyatakan waktu suatu objek telah *survive* setelah pengamatan atau studi dimulai sampai *event* yang diinginkan terjadi pada objek tersebut disebut *survival time* atau *failure time*. Waktu ini dapat dinyatakan dalam tahun, bulan, minggu, atau hari ; selain itu, waktu ini juga dapat dinyatakan dalam usia individu ketika suatu *event* terjadi. Sedangkan *event* yang diperhatikan biasanya kematian, munculnya suatu penyakit, kambuhnya suatu penyakit setelah dilakukan operasi, atau beberapa pengalaman negatif lainnya yang terjadi pada suatu objek. Namun, terdapat pula *event* yang positif seperti waktu sembuhnya seorang individu dari suatu penyakit setelah proses operasi.

Masalah yang seringkali dihadapi dalam menganalisis data *survival* yaitu bagaimana menyatakan fungsi *survival* jika terdapat informasi yang berkaitan (biasa disebut sebagai *covariate*, variabel *prediktor*, variabel penjelas, atau variabel bebas) yang berpengaruh terhadap *survival time*. *Covariate* ini dapat berupa variabel kuantitatif (tekanan darah, suhu, umur, dan berat badan), atau dapat juga berupa variabel kualitatif (jenis kelamin, status penyakit, atau beberapa perlakuan pada suatu percobaan (*treatment*)). Terdapat pula *covariate* yang bergantung pada waktu dan *covariate* yang tidak bergantung pada waktu. Dalam hal ini, akan dicari hubungan antara satu atau beberapa variabel bebas dengan *survival time*.

Salah satu metode yang digunakan untuk menyelesaikan permasalahan tersebut yaitu dengan analisis regresi. Analisis regresi adalah cabang dari statistika yang memperhatikan pola hubungan antara variabel *respon* terhadap satu atau beberapa *covariate*. Dalam analisis regresi, dikenal regresi parametrik, nonparametrik, dan semiparametrik.

Dalam regresi parametrik, variabel respon dari model yang digunakan berasal dari distribusi yang diketahui. Distribusi yang umumnya digunakan untuk *survival time* diantaranya distribusi weibull, eksponensial, log-logistik, dan log-normal. Untuk itu, dibutuhkan beberapa asumsi untuk error sehingga perlu dilakukan pengujian-pengujian terhadap asumsi-asumsi tersebut. Sebagai contoh, asumsi dari error yang harus dipenuhi untuk *survival time* yang berdistribusi log-normal yaitu error berdistribusi normal standar, identik, dan independent, $\varepsilon_i \sim NIID\ (0,1)$. Jika asumsi-asumsi dalam regresi parametrik ini dipenuhi maka model yang dihasilkan adalah model yang tepat sehingga penggunaan model tersebut sangat efisien dan berguna. Namun, ketika asumsi-asumsi dalam regresi parametrik ini tidak dipenuhi tetapi regresi parametrik tetap diterapkan pada data, maka akan menghasilkan taksiran model yang tidak sesuai sehingga menyebabkan interpretasi data yang menyesatkan, dan juga mengakibatkan hilangnya sejumlah informasi yang terkandung dalam data tersebut.

Untuk itu, terdapat regresi nonparametrik yang tidak membutuhkan asumsi apa-apa. Namun, taksiran yang didapat dari regresi nonparametrik memiliki kelemahan dalam konvergensi (Dorota M. Dabrowska, 1987).

Untuk mencegah hal ini, akan digunakan regresi semiparametrik. Salah satu model yang dapat digunakan dalam regresi semiparametrik yaitu model cox *proportional hazard*. Model ini menyatakan *hazard rate* suatu objek akan mengalami *event* pada waktu t dengan berbagai resiko **X** dimana **X** menyatakan kumpulan dari *covariate* yang akan dimodelkan untuk memprediksi *hazard rate* suatu objek.

Model cox *proportional hazard* ini melibatkan perkalian dua nilai, yaitu *baseline hazard function*, $h_0(t)$, yang melibatkan waktu t tetapi tidak melibatkan *covariate* **X** dan pernyataan eksponensial yang melibatkan *covariate* **X** tetapi tidak melibatkan waktu t. Bagian eksponensial ini menjamin bahwa model yang sesuai akan selalu memberikan taksiran *hazard* yang *non-negative*, sesuai dengan yang diharapkan karena variabel *respon*-nya berupa waktu sampai terjadinya *event* yang tidak mungkin bernilai negatif.

Model ini dikatakan semiparametrik karena *baseline hazard function*, $h_0(t)$, adalah fungsi yang tidak dapat ditentukan karena distribusi dari *survival time* tidak diketahui artinya bersifat non-parametrik, sedangkan fungsi eksponensial bersifat parametrik.

Meskipun bagian *baseline hazard function*, $h_0(t)$, tidak dapat ditentukan, masih memungkinkan untuk menaksir parameter, β, pada bagian eksponensial dari model. Taksiran β ini berguna untuk mengetahui seberapa besar pengaruh dari *covariate* yang diperhatikan. Pengaruh yang dapat dilihat berupa perbandingan dari dua objek dengan kondisi yang berbeda yang disebut dengan *hazard* rasio. *Hazard* rasio hanya menghasilkan bagian eksponensial yang melibatkan selisih dari **X** tetapi tidak melibatkan waktu, sehingga *hazard* rasio ini *proportional* sepanjang waktu. Artinya, perbandingan *hazard rate* dari dua objek dengan kondisi yang berbeda tetap sepanjang waktu. Hal ini disebut dengan asumsi *proportional hazard* di bawah model cox *proportional hazard*.

Namun, adakalanya asumsi ini tidak terpenuhi. Perubahan *hazard* rasio seiring berjalannya waktu mungkin saja terjadi, sehingga diperlukan adanya pengujian terhadap asumsi *proportional hazard* tersebut. Jika asumsi *proportional hazard* tidak terpenuhi, maka diperlukan model alternatif. Dalam tugas akhir ini model alternatif yang digunakan yaitu *extended cox model* (model cox yang melibatkan variabel yang bergantung pada waktu (*time-dependent covariate*)).

Extended cox model ini melibatkan unsur waktu. Jika *covariate* yang tidak bergantung pada waktu tidak memenuhi asumsi *proportional hazard*, maka diperlukan adanya interaksi antara fungsi waktu dan *covariate* tersebut. Dari model tersebut akan dicari taksirannya dengan menggunakan metode MPLE (*Maximum Partial Likelihood Estimator*).

Dalam suatu data, sering ditemukan adanya *ties*, yaitu terdapat lebih dari satu *event* yang terjadi dalam satu waktu. Jika hal ini terjadi, ada beberapa pendekatan yang dapat digunakan untuk membentuk *partial likelihood*-nya, yaitu pendekatan Efron, Breslow, dan Diskrit. Dalam tugas akhir ini pendekatan yang digunakan yaitu pendekatan Efron.

1.2 Perumusan Masalah dan Ruang Lingkup Masalah

Perumusan masalah yang diajukan pada tugas akhir ini adalah :

1. Bagaimana menentukan *extended cox model* untuk *covariate* yang tidak bergantung pada waktu yang tidak memenuhi asumsi *proportional hazard*?
2. Bagaimana mencari taksiran dari *extended cox model* tersebut?
3. Bagaimana penerapan *extended cox model* pada suatu kasus?

Ruang lingkup yang digunakan dalam tugas akhir ini meliputi :

1. Menggunakan data tersensor kanan yang *non-informative* tipe I.
2. Jika ada *ties* akan digunakan pendekatan Efron.
3. Tidak membahas metode pemilihan fungsi waktu.

1.3 Jenis Penelitian dan Metode yang Digunakan

Jenis penelitian yang digunakan dalam membuat tugas akhir ini yaitu studi literatur. Sedangkan metode yang digunakan untuk menaksir parameter pada *extended cox model* yaitu *maximum partial likelihood estimator*.

1.4 Tujuan Penulisan

Tujuan penulisan tugas akhir ini adalah :

1. Menentukan *extended cox model* untuk *covariate* yang tidak bergantung pada waktu yang tidak memenuhi asumsi *proportional hazard.*
2. Mencari taksiran dari *extended cox model* untuk *covariate* yang tidak bergantung pada waktu yang tidak memenuhi asumsi *proportional hazard.*
3. Memberikan contoh mengenai *extended cox model* tersebut dalam suatu data.

BAB 2
LANDASAN TEORI

Analisis *survival* merupakan sekumpulan prosedur statistika untuk keperluan analisis data dimana data yang digunakan berupa data waktu sampai terjadinya *event* tertentu. Oleh sebab itu, dalam analisis *survival* dibutuhkan beberapa hal berikut:

1. Waktu asal yang terdefinisi dengan baik (yaitu waktu ketika suatu objek masuk dalam studi/pengamatan)
2. Skala waktu pengukuran jelas, dan
3. Titik akhir atau *event* yang juga terdefinisi dengan baik.

2.1 *Survival Time*

Dalam analisis *survival*, variabel respon yang diperhatikan adalah waktu sampai terjadinya suatu *event*, sehingga analisis *survival* seringkali disebut sebagai analisis waktu kejadian (*time to event*). Waktu suatu objek telah bertahan selama periode pengamatan atau sampai terjadinya suatu *event* yang diinginkan disebut *survival time* atau *failure time*. Dengan perkataan lain, *survival time* adalah suatu variabel yang mengukur waktu dari sebuah titik awal tertentu (misalnya, waktu di mana suatu perlakuan/*treatment* dimulai) sampai dengan sebuah titik akhir yang ingin diperhatikan (misalnya, waktu kematian pada penderita kanker). Misalkan T adalah variabel random yang menunjukkan *survival time* dari sebuah populasi homogen, maka T bernilai non negatif, $T \geq 0$.

Kejadian (*event*) dapat dianggap sebagai suatu kegagalan (*failure*), karena kejadian yang diperhatikan biasanya adalah kematian, munculnya penyakit, atau beberapa kejadian negatif lain yang dapat terjadi pada suatu objek. Di samping itu, terdapat juga kasus kegagalan yang kejadiannya positif, seperti sembuhnya seseorang setelah dilakukan operasi. Dalam tugas akhir ini, *survival time* akan dianggap sama seperti *failure time*.

2.2 Kuantitas Dasar Analisis *Survival*

Dalam menggambarkan keadaan data *survival* digunakan kuantitas dasar yang seringkali dibahas pada analisis *survival* seperti *probability density function*, fungsi *survival*, dan fungsi *hazard*. Ketiga kuantitas dasar tersebut dibahas pada subbab berikut. Misalkan variabel random *T* menunjukkan *survival time* dari individu dalam populasi, dimana *T* merupakan variabel random non negatif dalam interval [0, ∞).

2.2.1 *Probability Density Function*

Menurut Lawless (1982), *probability density function* adalah probabilitas terjadinya suatu *event* dalam interval waktu dari *t* sampai $t+\Delta t$, dengan waktu *T* merupakan variabel random. Misalkan *T* adalah variabel *random* kontinu, *Probability density function* dinyatakan dengan :

$$f(t) = lim_{\Delta t \to 0} \left[\frac{P(t \leq T < (t + \Delta t))}{\Delta t} \right] \tag{2.1}$$

Survival time merupakan variabel random non negatif, sehingga waktu hidup diukur untuk nilai *t* yang non negatif dan diperoleh $f(t)=0$ untuk $t<0$ dan $\int_0^\infty f(t)dt = 1$.

2.2.2 Fungsi *Survival*

Kuantitas dasar yang juga digunakan untuk menggambarkan fenomena waktu kejadian adalah fungsi *survival*. Misalkan *T* adalah variabel random yang menggambarkan *survival time* dan t menyatakan beberapa nilai tertentu yang diperhatikan untuk variabel *T*. Maka, fungsi *survival* dapat didefinisikan sebagai probabilitas suatu objek bertahan melebihi suatu waktu tertentu *t*, atau dengan kata lain, probabilitas bahwa variabel random *T* lebih besar dari waktu yang ditentukan, yaitu *t* ($t > 0$). Secara matematis dapat dinyatakan sebagai :

$$S(t) = Pr\ (T > t) \tag{2.2}$$

Jika *T* adalah variabel random kontinu, maka fungsi *survival* merupakan komplemen dari fungsi distribusi kumulatif, dimana fungsi distribusi kumulatif adalah probabilitas bahwa variabel random *T* kurang dari atau sama dengan waktu *t* atau secara matematis dinyatakan dengan $F(t) = Pr\ (T \leq t)$. Fungsi *survival* yang merupakan komplemen dari fungsi distribusi kumulatif yaitu :

$$S(t) = Pr(T > t) = 1 - Pr(T \leq t) = 1 - F(t) \tag{2.3}$$

Fungsi *survival* juga dapat dinyatakan dalam bentuk *probability density function*, *f(t)*, yaitu :

$$S(t) = Pr(T > t) = \int_t^{\infty} f(t)dt \tag{2.4}$$

Jika T adalah variabel random diskret, maka fungsi *survival* diberikan sebagai berikut :

1. Misalkan nilai-nilai teramati untuk T adalah t_j, j=1, 2, ..., p dengan *probability mass function (p.m.f)*

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

   dimana $t_1 < t_2 < \cdots < t_p$
2. Fungsi *survival* untuk variabel random diskret T diberikan oleh,

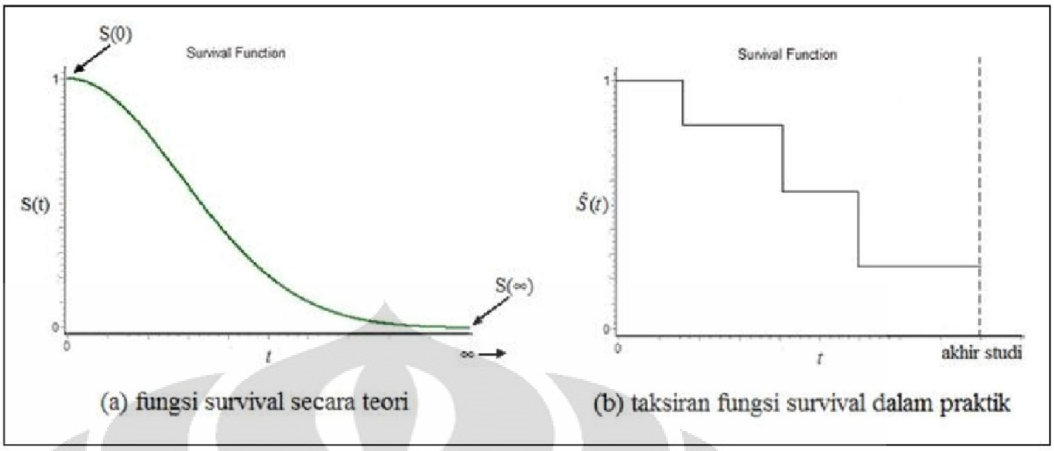
   $$S(t) = Pr(T > t) = \sum_{t_j > t} p(t_j) \tag{2.5}$$

(Klein and Moeschberger, 1997 : 26)

Secara teori, fungsi *survival* dapat diplot sebagai kurva *survival* yang menggambarkan probabilitas ketahanan suatu objek pada titik-titik waktu t, antara 0 sampai ∞. Semua fungsi *survival* memiliki karakteristik seperti berikut:

- Merupakan fungsi monoton tak naik
- Pada saat t=0, $S(t)$=$S(0)$=1; artinya, pada awal studi, karena belum ada individu yang telah mengalami *event* maka probabilitas *survival* pada saat itu adalah 1.
- Pada saat $t \to \infty$, $S(t) \to 0$; artinya, secara teori, jika periode studi bertambah tanpa batas, pada akhirnya tidak ada seorang pun yang akan bertahan hidup sehingga kurva *survival* akan menuju nol.

Tetapi dalam praktiknya, ketika digunakan data yang nyata (realistis), akan diperoleh grafik *survival* yang merupakan fungsi tangga. Lagipula, karena lamanya periode studi tidak mungkin sampai menuju tak berhingga, mungkin tidak setiap individu yang diamati mengalami *event* sehingga tidak semua fungsi *survival* (yang ditaksir) akan sama dengan nol pada akhir masa studi. Berikut ditunjukkan gambar kurva *survival* secara teoritis dan kurva *survival* dalam praktiknya:



Gambar 2.1. Kurva fungsi survival secara teori (a) dan secara praktik (b)

2.2.3 Fungsi *Hazard*

Suatu kuantitas dasar yang juga merupakan dasar dari analisis *survival* adalah fungsi *hazard*. Fungsi ini juga dikenal sebagai *hazard rate* yang dinotasikan dengan *h(t)*.

Fungsi *hazard* didefinisikan sebagai *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 dinyatakan sebagai :

$$h(t) = lim_{\Delta t \to 0} \frac{Pr(t \le T < t + \Delta t | T \ge t)}{\Delta t} \tag{2.6}$$

Jika *T* adalah suatu variabel random kontinu dan misalkan *f(t)* adalah *probability density function* pada waktu *t*, maka dari persamaan *h(t)* sebelumnya diperoleh :

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{Pr(t \le T < t + \Delta t | T \ge t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{Pr(t \le T < (t + \Delta t)) \cap (T \ge t)}{Pr(T \ge t) \,.\, \Delta t} \\ &= \lim_{\Delta t \to 0} \frac{Pr(t \le T < (t + \Delta t))}{Pr(T \ge t) \,.\, \Delta t} \\ &= \lim_{\Delta t \to 0} \frac{Pr(t \le T < (t + \Delta t))}{S(t). \Delta t} \end{aligned}$$

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

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

Pada subbab sebelumnya diketahui bahwa $S(t) = 1 - F(t)$, dan

$$f(t) = F'(t) = \frac{\partial(F(t))}{\partial t} = \frac{\partial(1-S(t))}{\partial t} = -\frac{\partial(S(t))}{\partial t} = -S'(t) \tag{2.8}$$

Maka,

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -S'(t).\frac{\partial lnS(t)}{\partial S(t)} = -\frac{\partial S(t)}{\partial t}.\frac{\partial lnS(t)}{\partial S(t)} = -\frac{\partial lnS(t)}{\partial t}$$

dari persamaan di atas diperoleh,

$$\int_0^t h(x)dx \quad = -\int_0^t \frac{dlnS(x)}{dx} dx$$

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

$$-\int_0^t h(x)dx \quad = lnS(x) \Big|_0^t$$

$$-\int_0^t h(x)dx \quad = lnS(t) - lnS(0)$$

Karena $S(0) = 1$, $ln\ S(0) = 0$, sehingga persamaan di atas menjadi

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

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

Dari uraian di atas, diperoleh hubungan antara $S(t)$, $h(t)$, dan $f(t)$, yaitu sebagai berikut :

a) $f(t) = -S'(t)$

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

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

Dengan demikian, jika fungsi *hazard*, *h(t)*, diketahui, maka *f(t)* dan *S(t)* bisa didapat, begitu pula jika *f(t)* ataupun *S(t)* yang diketahui.

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

$$\begin{aligned} h(t_j) &= Pr(T = t_j | T \geq t_j), j = 1, 2, \ldots, p \\ &= \frac{Pr(T = t_j \cap T \geq t_j)}{Pr\,(T \geq t_j)} \\ &= \frac{Pr(T = t_j)}{Pr\,(T \geq t_j)} \\ &= \frac{Pr(T = t_j)}{Pr\,(T > t_{j-1})} \end{aligned}$$

$$h(t_j) = \frac{p(t_j)}{S(t_{j-1})} \tag{2.10}$$

dimana $S(t_0) = 1$

Diketahui, $S(t_j) = Pr\,(T > t_j)$, dan

$$S(t_{j-1}) = Pr(T > t_{j-1}) = Pr(T = t_j) + Pr(T > t_j) = p(t_j) + S(t_j)$$

$$\leftrightarrow p(t_j) = S(t_{j-1}) - S(t_j) \tag{2.11}$$

Berdasarkan persamaan (2.10), diperoleh

$$h(t_j) = \frac{p(t_j)}{S(t_{j-1})} = \frac{S(t_{j-1}) - S(t_j)}{S(t_{j-1})} = 1 - \frac{S(t_j)}{S(t_{j-1})}\ ,\ j = 1, 2, \ldots, p \tag{2.12}$$

Fungsi *survival* dapat ditulis sebagai perkalian dari survival bersyarat,

$$S(t) = \prod_{t_j \leq t} \frac{S(t_j)}{S(t_{j-1})} \tag{2.13}$$

Jadi, berdasarkan persamaan (2.12) dan (2.13), hubungan fungsi survival dengan fungsi hazard yaitu :

$$S(t) = \prod_{t_j \leq t} [1 - h(t_j)] \tag{2.14}$$

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 atau ke bawah terhadap waktu *t*. Dengan perkataan lain, untuk suatu nilai tertentu *t*, fungsi *hazard h(t)* mempunyai karakteristik seperti berikut:

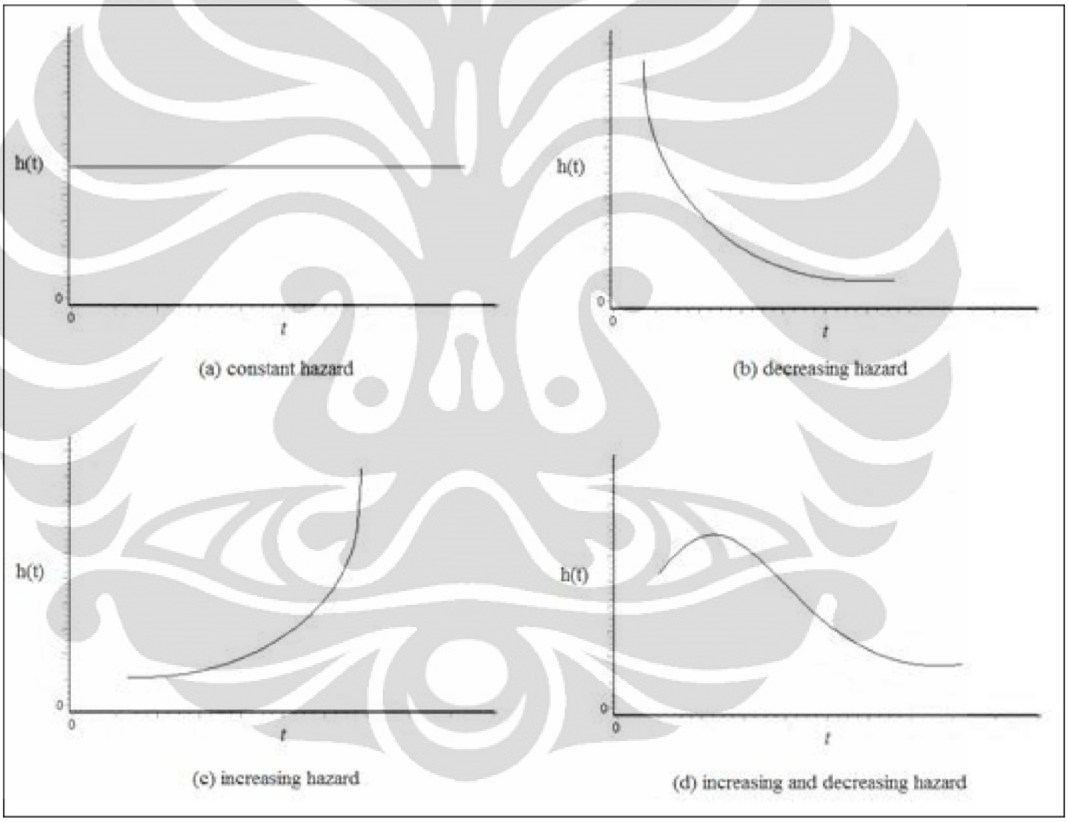
- Selalu bernilai non negatif, $h(t) \geq 0$

o Tidak memiliki batas atas.

Beberapa bentuk kurva fungsi *hazard* ditunjukkan pada gambar 2.2. Berikut penjelasan singkat mengenai bentuk kurva *hazard*.

o Dalam kehidupan nyata, untuk kasus di mana *hazard* bernilai konstan jarang ditemui.

o Untuk kurva *hazard* turun, misalnya saja tingkat kematian sesaat pada populasi bayi baru lahir. Cenderung tinggi pada awal setelah kelahiran dan seiring berjalannya waktu tingkat kematiannya akan semakin turun dan stabil.



Gambar 2.2. Bentuk kurva fungsi *hazard*. (a) konstan, (b) turun, (c) naik, (d) naik kemudian turun

o Untuk *hazard* naik, contohnya adalah tingkat kematian sesaat pada para penderita kanker. Pada awal terdeteksi, tingkat *hazard* masih rendah dan semakin lama akan semakin tinggi tingkat kematian pada penderita kanker tersebut.

- o *Hazard* naik dan turun, misalnya tingkat kematian pada individu setelah berhasilnya dilakukan operasi. Kemudian pada resiko awalnya dapat terjadi infeksi atau komplikasi lain sesaat setelah prosedur operasi berjalan, kemudian resikonya berkurang seiring dengan kesembuhan pasien.

Demikian telah dijelaskan mengenai tiga kuantitas dasar dalam analisis *survival*, tetapi dalam tugas akhir ini akan lebih terpusat hanya pada fungsi *hazard*.

2.3 Data Tersensor

Sehimpunan data yang digunakan dalam analisis *survival* dapat berupa data eksak ataupun data tersensor, dan mungkin juga data terpancung.

- Disebut data eksak apabila waktu tepatnya suatu *event* yang diinginkan terjadi diketahui.
- Sementara data dikatakan tersensor jika hanya diketahui beberapa informasi mengenai waktu sampai terjadinya *event* pada individu yang bersangkutan tetapi tidak diketahui waktu kejadiannya secara pasti. Data yang mengalami penyensoran hanya memuat sebagian informasi mengenai variabel random yang diperhatikan, namun berpengaruh terhadap perhitungan statistik.
- Pemancungan menentukan seorang individu masuk dalam studi/pengamatan atau tidak, terdiri dari pemancungan kiri dan pemancungan kanan. Pada pemancungan kiri; individu yang masuk dalam penelitian/studi adalah mereka yang belum mengalami *event*, sedangkan yang sudah mengalami *event* tidak diikutsertakan dalam studi. Sedangkan pada pemancungan kanan; individu yang masuk dalam penelitian adalah mereka yang sudah mengalami *event*.

Namun, dalam tugas akhir ini hanya akan digunakan data eksak dan data tersensor.

Data tersensor terdiri dari tersensor kiri, tersensor kanan, dan tersensor interval. Berikut akan dibahas masing-masing dari data tersensor tersebut.

2.3.1 Data Tersensor Kiri

Data tersensor kiri terjadi apabila *event* yang ingin diperhatikan pada individu ternyata sudah terjadi saat individu tersebut masuk dalam studi. Jadi, hanya diketahui bahwa waktu terjadinya *event* kurang dari suatu nilai tertentu.

Sebagai contoh, sebuah studi dilakukan untuk menentukan distribusi dari waktu saat usia berapa tahun untuk pertama kalinya anak laki-laki di suatu sekolah tinggi mulai menggunakan marijuana. Studi dilakukan selama 24 bulan. Pertanyaan yang diajukan adalah "Kapan pertama kali Anda menggunakan marijuana?" Jika terdapat anak laki-laki yang menjawab "Saya sudah menggunakannya tetapi tidak ingat kapan pertama kali saya pakai." Pernyataan ini menunjukkan bahwa *event* (yaitu pertama kali menggunakan marijuana) telah terjadi sebelum usia anak laki-laki tersebut saat *interview* tetapi usia tepatnya ia menggunakan marijuana untuk pertama kalinya tidak diketahui. Secara rinci, penjelasan di atas dapat ditulis sebagai berikut:

- Waktu asal (*time origin*): saat individu baru lahir,
- Skala waktu pengukuran: usia individu (dalam tahun),
- *Event* yang diamati: pertama kali memakai marijuana,
- Waktu akhir (*end of study*): setelah 24 bulan periode pengamatan,
- Waktu penyensoran kiri: usia individu saat diwawancara.

2.3.2 Data Tersensor Kanan

Data tersensor kanan merupakan jenis data tersensor yang paling umum dalam analisis *survival*, dan terjadi ketika hanya diketahui bahwa *survival time* melebihi suatu nilai tertentu. Secara lebih umum, data tersensor kanan dapat terjadi oleh karena adanya beberapa hal sebagai berikut,

- Seorang individu yang belum mengalami *event* setelah studi berakhir;
- Seorang individu yang keluar dari studi pada saat periode studi sedang berjalan;
- Seorang individu yang meninggal tetapi bukan karena alasan yang berhubungan dengan *event* yang ingin diperhatikan. Atau individu meninggal tetapi kematian bukan suatu *event* yang diperhatikan.

Sebagai contoh untuk data tersensor kanan, misalkan ingin diketahui berapa lama pasien-pasien suatu rumah sakit akan bertahan hidup setelah melakukan transplantasi ginjal. Jika direncanakan akan dilakukan sepuluh tahun pengamatan, beberapa kemungkinan dapat terjadi seperti:

- Seorang individu pindah dari rumah sakit tersebut sebelum masa studi berakhir. Dalam kasus ini tidak bisa diperoleh informasi yang berkaitan dengan waktu sampai terjadinya kematian (*event*). Namun, sebenarnya diketahui kapan waktu pasien tersebut pindah, dan waktu ini didefinisikan sebagai titik waktu di mana *survival time* yang sebenarnya tersensor kanan.
- Pada akhir periode studi, terdapat individu yang masih bertahan hidup. Dalam kasus ini, waktu ketika studi berakhir dapat dianggap sebagai titik data tersensor kanan untuk semua individu dalam studi yang masih hidup.

Alasan bahwa titik-titik data pada dua kemungkinan di atas dianggap sebagai data tersensor kanan adalah karena *survival time* eksak untuk individu-individu yang mengalami kemungkinan tersebut tidak diketahui, tetapi diketahui bahwa waktu kematian dari masing-masing individu akan terjadi pada suatu waktu setelah individu keluar dari pengamatan, atau setelah waktu studi berakhir.

Sehimpunan data tersensor kanan memuat sebuah variabel yang menunjukkan waktu seorang individu dalam studi dan sebuah indikator apakah waktu yang dimaksud diketahui secara pasti atau *survival time* yang tersensor kanan. Biasanya digunakan suatu variabel indikator, sebut λ, yang bernilai 1 jika *survival time* diketahui secara pasti dan 0 untuk waktu yang tersensor kanan.

Misalkan sehimpunan data sederhana yang memuat lima individu yang diikutsertakan dalam studi selama sepuluh tahun periode pengamatan. Diperoleh data sebagai berikut:

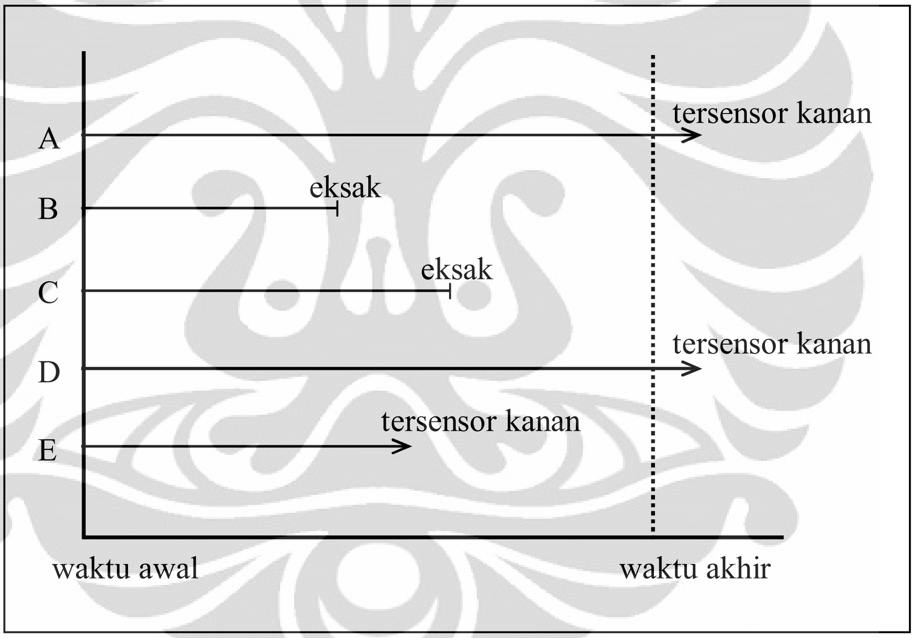
$$8^+, 10^+, 6, 9^+, 5.$$

Dalam himpunan data ini, terdapat tiga bilangan dengan tanda "+" yang biasanya digunakan sebagai petunjuk bahwa ketiganya merupakan titik-titik data tersensor kanan. Ketika dilakukan suatu analisis mengenai data tersebut dengan menggunakan perangkat lunak statistika, jika tidak ada tanda "+" berikan nilai indikator 1 dan nilai 0 apabila terdapat tanda "+". Himpunan data ini kemudian dapat ditulis sebagai (t_i , λ_i) :

$$(8,0), (10,0), (6,1), (9,0), (5,1).$$

Dalam bentuk ini, t_i adalah variabel yang menggambarkan waktu dari individu ke-i dan λ_i adalah indikator apakah *survival time* untuk individu i adalah eksak atau tersensor kanan.

Gambar 2.3 menunjukkan gambaran secara grafis dari informasi *survival* yang diketahui dari kelima individu dalam contoh sebelumnya. Informasi *survival* digambarkan dengan sebuah garis horizontal untuk setiap individu. Anak panah di akhir garis dari seorang individu menunjukkan *survival time* tersensor kanan.



Gambar 2.3. Himpunan data dengan *survival time* eksak dan tersensor kanan

Dari gambar di atas, *survival time* individu 3 dan 5 diketahui secara pasti, tetapi *survival time* untuk individu 1, 2, dan 4 tersensor kanan. Individu 1 dan 4 keluar atau hilang dari pengamatan pada suatu waktu sebelum studi berakhir. Individu 2 masih hidup pada akhir periode studi, sehingga individu ini memiliki *survival time* yang tersensor kanan.

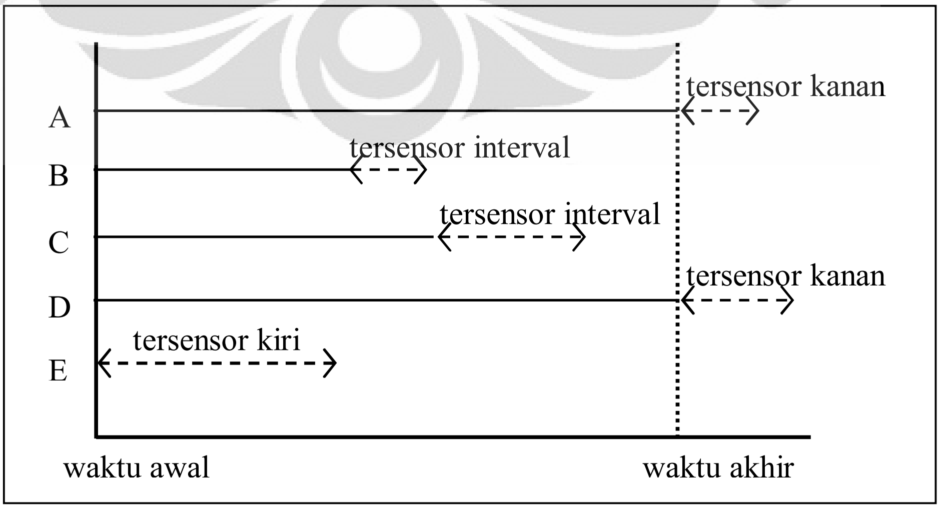
2.3.3 Data Tersensor Interval

Sensor interval terjadi ketika hanya diketahui bahwa suatu *event* yang diinginkan terjadi dalam suatu periode waktu. Data tersensor kiri dan tersensor kanan merupakan kasus khusus dari data tersensor interval.

Misalkan diambil contoh kasus yang dibahas pada data tersensor kanan (subbab 2.3.2). Tujuan dari studi ini adalah untuk mengamati berapa lama individu akan bertahan setelah melakukan transplantasi ginjal. Sebagai bagian dari studi, individu-individu yang terlibat dalam studi diminta untuk melakukan pemeriksaan berkala sebanyak satu kali dalam satu tahun. Maka berbagai kemungkinan bisa saja terjadi pada individu, seperti:

- Meninggal di antara dua waktu pemeriksaan berkala, yaitu setelah kunjungannya yang terakhir dan sebelum kunjungan berikutnya;
- Hilang dari pengamatan karena pindah rumah dan dikeluarkan dari studi;
- Meninggal karena kecelakaan mobil, yang tidak ada hubungannya dengan *event* yang diperhatikan. Dalam kasus ini, waktu kematian dianggap sebagai suatu titik waktu tersensor kanan.

Gambar 2.4 berikut ini memberikan contoh lain dengan lima individu dalam sampel. *Survival time* individu 1 dan 4 tersensor interval. Individu 2 dan 3 memiliki *survival time* tersensor kanan, karena mereka masih bertahan hidup pada akhir periode studi. Sedangkan individu 5 memiliki *survival time* tersensor kiri.



Gambar 2.4. Himpunan data dengan *survival time* tersensor interval

Dalam satu data dapat berlaku lebih dari satu skema *censoring*. Akan tetapi, pada tugas akhir ini, skema *censoring* yang digunakan adalah sensor kanan.

Dalam skema sensor kanan pun terdapat dua tipe sensor kanan, yaitu tipe I dan tipe II.

- Disebut data tersensor kanan tipe I apabila data *survival* dihasilkan setelah studi berjalan selama waktu yang telah ditentukan;
- Data tersensor kanan tipe II, jika studi diakhiri setelah sejumlah *event* tertentu telah terjadi. Sejumlah *event* tersebut telah ditentukan sebelumnya.

Pada tugas akhir ini, tipe sensor kanan yang digunakan adalah sensor kanan tipe I.

2.4 Model *Cox Proportional Hazard*

Salah satu tujuan analisis *survival* adalah untuk mengetahui hubungan antara waktu kejadian (*time to failure*) dan *covariate* yang terukur pada saat dilakukan penelitian. Analisis ini dapat dilakukan dengan metode regresi. Salah satu model regresi yang sering digunakan adalah Regresi *Cox Proportional Hazard* (PH). Bentuk model Cox PH adalah sebagai berikut:

$$h(t,\boldsymbol{X}) = h_0(t)exp^{\sum_{a=1}^{p}\beta_a X_a}, \boldsymbol{X} = (X_1, X_2, \ldots, X_p).$$

Model ini menyatakan *hazard rate* dari satu individu pada waktu *t* dengan diketahui variabel-variabel prediktor $\boldsymbol{X}$. Model ini terdiri dari dua kuantitas, yaitu:

Tabel 2.1. Dua kuantitas pembentuk *hazard rate*

$\boldsymbol{h_0(t)}$	$\boldsymbol{exp\left[\sum_{a=1}^{p}\beta_a X_a\right]}$
- Disebut baseline hazard - Mengandung *t*, tapi tidak mengandung *X*	- Eksponensial - Mengandung *X*, tapi tidak mengandung *t* - *X* tidak bergantung waktu (time-independent)

- **Fungsi *baseline hazard* ($\boldsymbol{h_0(t)}$)**

  Fungsi *baseline hazard* adalah hazard rate saat *X = 0.* $h_0(t)$ merupakan fungsi yang tidak diketahui karena distribusi dari *survival time* (*T*) tidak diketahui. Fungsi ini hanya bergantung waktu *t* dan tidak mengandung *X.*

- **Eksponensial ($\boldsymbol{exp\left[\sum_{a=1}^{p} \beta_a X_a\right]}$)**

  Kuantitas ini hanya bergantung pada *X* yang disebut *time-independent covariate*. Hal ini dikarenakan *X* tidak bergantung pada waktu. Jika *X* bergantung pada waktu, maka *X* disebut *time-dependent covariate*. Akan tetapi, apabila *X* bergantung pada waktu, maka diperlukan metode yang berbeda untuk memodelkan hazardnya, salah satunya adalah *Extended Cox Model*. Dalam tugas akhir ini, yang akan dibahas adalah model Cox-PH dengan *time-independent covariate*. *Time-independent covariate* adalah variabel yang nilainya tidak akan berubah sepanjang waktu. Contohnya, jenis kelamin, suku, dan warna kulit.

Model Cox-PH disebut model semiparametrik. Model ini berbeda dengan model parametrik, di mana $h_0(t)$ pada model parametrik mempunyai bentuk yang jelas. Misal, apabila model parametrik tersebut berdistribusi Weibull, maka $h_0(t) = \lambda p t^{p-1}$ dengan λ dan *p* adalah parameter pada distribusi Weibull. Dalam kenyataannya, data yang ada tidak diketahui distribusinya, sehingga bentuk $h_0(t)$-nya juga tidak dapat diketahui. Oleh karena itu, model semiparametrik lebih sering digunakan. Hal ini dikarenakan walaupun fungsi baseline hazardnya ($h_0(t)$) tidak diketahui bentuk fungsionalnya, akan tetapi model Cox-PH ini tetap dapat memberikan informasi yang berguna, berupa *Hazard Ratio* (HR) yang tidak bergantung pada $h_0(t)$. *Hazard Ratio* didefinisikan sebagai *hazard rate* dari satu individu dibagi dengan *hazard rate* dari individu lain. Hal ini dapat ditunjukkan sebagai berikut.

Misalnya, individu A memiliki *hazard rate* $h_A(t, X^*)$ di mana

$$X^* = (X_1^*, X_2^*, \dots, X_p^*)$$

dan individu B memiliki *hazard rate* $h_B(t, X)$ di mana

$$X = (X_1, X_2, \dots, X_p)$$

maka rasio hazardnya adalah

$$HR = \frac{h_A(t, X^*)}{h_B(t, X)} = \frac{h_0(t)\, exp\left[\sum_{a=1}^{p} \beta_a {X_a}^*\right]}{h_0(t)\, exp\left[\sum_{a=1}^{p} \beta_a X_a\right]}$$

$$= exp\left(\sum_{a=1}^{p} \beta_a {X_a}^* - \sum_{a=1}^{p} \beta_a X_a\right)$$

$$= exp[\sum_{a=1}^{p} \beta_a (X_a^* - X_a)]$$

Dari perumusan di atas, terlihat bahwa X tetap dapat menjelaskan "*survival experience*" dari suatu objek. Apabila nilai *hazard ratio* (HR) konstan sepanjang waktu, maka dapat dikatakan bahwa $X_1, X_2, \dots, X_p$ memenuhi asumsi PH.

Sifat dari model Cox PH adalah meskipun tidak diketahui bentuk $h_0(t)$-nya, akan tetapi tetap dapat ditaksir koefisien regresinya (β). Penaksiran diperlukan untuk mengetahui efek dari variabel-variabel penjelas.

2.5 Maximum Likelihood Estimator

Pandang suatu sampel acak $X_1, X_2, \dots, X_n$ dari suatu distribusi yang mempunyai pdf $f(x;\theta) : \theta \epsilon \boldsymbol{\Omega}$. Pdf bersama dari $X_1, X_2, \dots, X_n$ adalah

$$f(x_1;\theta)f(x_2;\theta)f(x_3;\theta) \dots\ f(x_n;\theta)$$

Pdf bersama ini dapat dipandang sebagai fungsi dari θ dan disebut sebagai fungsi likelihood L dari sampel acak, dinotasikan dengan :

$$L(\theta; x_1, x_2, \dots, x_n) = f(x_1;\theta)f(x_2;\theta)f(x_3;\theta) \dots\ f(x_n;\theta), \theta \epsilon \boldsymbol{\Omega}$$

Misalkan bahwa dapat ditemukan suatu fungsi nontrivial dari $x_1, x_2, \dots, x_n$, sebut $u(x_1, x_2, \dots, x_n)$ sedemikian sehingga ketika θ diganti dengan $u(x_1, x_2, \dots, x_n)$, maka fungsi likelihood L berharga maksimum. Yaitu $L[u(x_1, x_2, \dots, x_n); x_1, x_2, \dots, x_n]$ sedikitnya sebesar $L(\theta; x_1, x_2, \dots, x_n)$ untuk setiap $\theta \epsilon \boldsymbol{\Omega}$. Maka statistik $u(X_1, X_2, \dots, X_n)$ disebut maksimum likelihood estimator dari θ dan dinotasikan dengan simbol

$$\hat{\theta} = u(X_1, X_2, \dots, X_n)$$

Penaksiran maksimum likelihood dari θ didapat dengan menyelesaikan persamaan $\frac{dL(\theta)}{d\theta} = 0$ atau dengan menggunakan logaritma natural $\frac{dln\ L(\theta)}{d\theta} = 0$.

Misalkan ada *n* parameter yang tidak diketahui, maka penaksir maksimum likelihood dari θ_i diperoleh dengan menyelesaikan,

$$\frac{dln\,L(\theta_1, \theta_2, \dots, \theta_n)}{d\theta_i} = 0, dengan\ i = 1, 2, .., n$$

2.6 Uji *Ratio Likelihood*

Penggunaan dari besaran dari ratio dari dua pdf sebagai dasar dari pengujian terbaik atau *Uniformly Most Powerfull Test* (UMPT) dapat dimodifikasi dan memberikan metode untuk membentuk suatu pengujian dari suatu hipotesis majemuk terhadap hipotesis alternatif majemuk, atau membentuk suatu pengujian dari suatu hipotesis sederhana terhadap hipotesis alternatif majemuk ketika UMPT tidak ada. Metode ini membawa kepada suatu pengujian yang disebut Uji *Ratio Likelihood*. Uji *ratio likelihood* tidak perlu merupakan suatu UMPT, tetapi pengujian ini sering mempunyai sifat-sifat yang diinginkan.

Secara umum, misalkan $X_1, X_2, \dots, X_n$ menyatakan *n* variabel acak yang masing-masing mempunyai *probability density function,*

$$f_i(x_i; \theta_1, \theta_2, \dots, \theta_m), i = 1, 2, \dots, n.$$

Himpunan yang terdiri dari semua titik parameter $(\theta_1, \theta_2, \dots, \theta_m)$ dinotasikan dengan $\boldsymbol{\Omega}$ yang disebut ruang parameter. Misalkan ω adalah subset dari ruang parameter $\boldsymbol{\Omega}$. Diinginkan untuk menguji hipotesis (sederhana atau majemuk),

$H_0: (\theta_1, \theta_2, \dots, \theta_m) \epsilon\ \omega$

terhadap semua hipotesis alternatif.

Definisikan fungsi likelihood,

$$L(\omega) = \prod_{i=1}^{n} f_i(x_i; \theta_1, \theta_2, \dots, \theta_m), \qquad (\theta_1, \theta_2, \dots, \theta_m) \epsilon\ \omega$$

dimana $L(\omega)$ adalah fungsi likelihood di bawah H_0 benar, dan

$$L(\boldsymbol{\Omega}) = \prod_{i=1}^{n} f_i(x_i; \theta_1, \theta_2, \dots, \theta_m), \qquad (\theta_1, \theta_2, \dots, \theta_m) \epsilon\ \boldsymbol{\Omega}$$

dimana $L(\boldsymbol{\Omega})$ adalah fungsi likelihood dengan melibatkan seluruh parameter.

Misalkan $L(\widehat{\omega})$ dan $L(\widehat{\boldsymbol{\Omega}})$ adalah maksimum yang diasumsikan ada untuk kedua fungsi likelihood tersebut. Ratio dari $L(\widehat{\omega})$ terhadap $L(\widehat{\boldsymbol{\Omega}})$ disebut rasio likelihood dan dinotasikan dengan :

$$\lambda(t_1, t_2, \dots, t_n) = \lambda = \frac{L(\widehat{\omega})}{L(\widehat{\boldsymbol{\Omega}})}\ dimana\ 0 \leq \lambda \leq 1$$

Misalkan λ_0 adalah suatu bilangan pecahan positif. Prinsip uji *ratio likelihood* menyatakan bahwa hipotesis

$H_0: (\theta_1, \theta_2, \dots, \theta_m) \epsilon\ \omega$

ditolak jika dan hanya jika

$$\lambda(x_1, x_2, \dots, x_n) = \lambda \leq \lambda_0$$

Fungsi λ mendefinisikan suatu variabel acak $\lambda(X_1, X_2, \dots, X_n)$, dan signifikansi level dari pengujian diberikan oleh

$$\alpha = Pr[\lambda(X_1, X_2, \dots, X_n) \leq \lambda_0; H_0]$$

Uji ratio likelihood ini mempunyai pendekatan distribusi Chi-Square.

(Robert V. Hogg dan Allen T. Craig, 1978)

Rasio ini dapat dihitung pada interval kepercayaan yang diberikan, yaitu :

1. Hitung $\chi^2 = -2\, ln\, \lambda$. Semakin kecil λ, akan semakin besar χ^2 yang didapat
2. Untuk melihat kesignifikanan nilai χ^2 yaitu dengan membandingkan nilai χ^2 terhadap nilai persentil 100x(1-α) dari distribusi Chi-Square dengan *k* derajat bebas. χ^2 mempunyai pendekatan Chi-Square dengan *k* derajat bebas dan pendekatan ini baik meskipun sampelnya berukuran kecil
3. Aturan keputusannya yaitu H_0 ditolak jika χ^2 lebih besar daripada nilai persentil 100x(1- α) dari distribusi Chi-Square dengan *k* derajat bebas, dimana persentil bersesuaian dengan interval kepercayaan yang dipilih.

BAB 3
EXTENDED COX MODEL UNTUK *TIME-INDEPENDENT COVARIATE* YANG TIDAK MEMENUHI ASUMSI PH

Seperti yang dibahas pada bab sebelumnya, salah satu model semiparametrik yang digunakan untuk mengetahui pengaruh dari *covariate* yang diperhatikan yaitu model *Cox*. Model ini terdiri dari perkalian dua nilai yaitu nilai dari fungsi *baseline hazard* dan nilai eksponensial dari penjumlahan *covariate*-nya. Bagian fungsi *baseline hazard* melibatkan fungsi waktu tetapi tidak melibatkan *covariate*, sedangkan bagian eksponensial melibatkan *covariate* tetapi tidak melibatkan fungsi waktu.

Jika terdapat p *covariate*, misal, $X_1, X_2, ..., X_p$, yang diperhatikan, maka untuk mengetahui pengaruhnya terhadap suatu *event*, digunakan model *Cox Proportional Hazard* (biasa disingkat dengan Cox PH), yaitu :

$$h(t, \boldsymbol{X}) = h_0(t)\, exp\left[\sum_{a=1}^{p} \beta_a \boldsymbol{X}_a\right]$$

Berdasarkan model di atas, bagian fungsi *baseline hazard*, $h_0(t)$, tidak dapat ditentukan sehingga *hazard rate* dari masing-masing objek tidak dapat diketahui. Akan tetapi, dengan melihat perbandingan dua nilai *hazard rate* berdasarkan *covariate* yang berbeda, maka dapat dilihat pengaruh dari *covariate* tersebut.

Sebagai contoh, misalkan terdapat *p covariate* yang diperhatikan. *Covariate* untuk kondisi pertama, dengan nilai *covariate* $\boldsymbol{X^*}$, yaitu $X_1^*, X_2^*, \dots, X_p^*$, sedangkan untuk kondisi kedua, dengan nilai *covariate* $\boldsymbol{X}$, yaitu $X_1, X_2, ..., X_p$. Untuk melihat pengaruh dari *covariate* tersebut, akan dibandingkan nilai *hazard rate*-nya, sebagai berikut :

1. *Hazard rate* untuk kondisi pertama yaitu :

$$h_1(t, \boldsymbol{X}^*) = h_0(t)\, exp\left[\sum_{a=1}^{p} \beta_a \boldsymbol{X}_a^*\right]$$

2. *Hazard rate* untuk kondisi kedua yaitu :

$$h_2(t, \boldsymbol{X}) = h_0(t)\, exp\left[\sum_{a=1}^{p} \beta_a \boldsymbol{X}_a\right]$$

sehingga, perbandingannya yaitu :

$$HR = \frac{h_1(t, \boldsymbol{X}^*)}{h_2(t, \boldsymbol{X})} = \frac{h_0(t)\, exp\left[\sum_{a=1}^{p} \beta_a \boldsymbol{X}_a^*\right]}{h_0(t)\, exp\left[\sum_{a=1}^{p} \beta_a \boldsymbol{X}_a\right]} = exp\left[\sum_{a=1}^{p} \beta_a (\boldsymbol{X}_a^* - \boldsymbol{X}_a)\right]$$

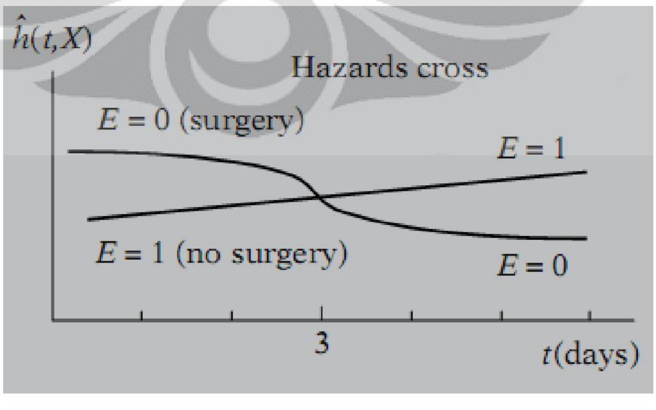
Perbandingan dari *hazard rate* kedua objek tersebut dinamakan ***Hazard Ratio*** (disingkat dengan HR).

Dari model *hazard ratio* di atas, terlihat bahwa nilainya tidak lagi bergantung pada waktu, artinya perbandingan kedua objeknya tetap sepanjang waktu. Hal ini disebut dengan asumsi ***Proportional Hazard***.

Hazard ratio tidak hanya membandingkan *hazard rate* dari dua objek, tetapi juga membandingkan dua *hazard rate* yang berbeda di waktu tertentu, sebut t_0, pada satu objek. Misalkan,

$$X = \begin{cases} x & t < t_0 \\ x + \Delta x & t \geq t_0 \end{cases}$$

Dalam beberapa kasus, *hazard ratio* dari suatu *covariate* tidak tetap sepanjang waktu. Contoh, terdapat suatu data dengan plot *hazard rate* terhadap waktu dari *covariate*, sebut *E*, sebagai berikut :



Gambar 3.1. Kurva *hazard rate* terhadap waktu dari *E*=1 dan *E*=0

Pada saat t<3, nilai *hazard rate* dari E=0 lebih besar daripada *hazard rate* dari E=1 atau $\frac{h(t,E=0)}{h(t,E=1)} > 1$, sedangkan pada saat t>3, nilai *hazard rate* dari E=0 lebih kecil daripada *hazard rate* dari E=1 atau $\frac{h(t,E=0)}{h(t,E=1)} < 1$. Ini menunjukkan bahwa nilai *hazard ratio*-nya tidak konstan.

Jika digunakan model *cox proportional hazard* untuk data tersebut, didapat taksiran *hazard ratio*-nya sebagai berikut :

$$\widehat{HR} = exp\left[\sum_{a=1}^{p} \hat{\beta}_a(\boldsymbol{X}_a^* - \boldsymbol{X}_a)\right]$$

Terlihat bahwa, taksiran *hazard ratio*-nya bernilai konstan sepanjang waktu, dan tidak sesuai dengan kondisi sebenarnya.

Jika *hazard ratio*-nya tidak tetap, maka asumsi *proportional hazard* tidak terpenuhi, sehingga model Cox PH tidak dapat menggambarkan kondisi yang sebenarnya.

Untuk melihat pengaruh dari *covariate* yang tidak memenuhi asumsi *proportional hazard*, *covariate* tersebut harus diubah dalam bentuk *covariate* yang telah melibatkan waktu, ***X(t)***, yang disebut sebagai *time dependent covariate*. Model Cox yang melibatkan *time dependent covariate* disebut ***Extended Cox Model.***

Time dependent covariate dibagi menjadi dua, yaitu :

1. ***X(t)*** dimana *covariate*-nya bergantung pada waktu

   Jika dibandingkan dua *hazard rate* dari suatu objek yang *covariate*-nya bergantung pada waktu, maka untuk waktu tertentu dapat diperoleh *hazard ratio* yang berbeda. Misalkan satu *covariate* diperhatikan,

   - *Hazard rate* pertama yaitu :

   $$h_1(t,\boldsymbol{X}) = h_0(t)\, exp[\beta_1 X_1(t)]$$

   - *Hazard rate* kedua yaitu :

   $$h_2(t,\boldsymbol{X}) = h_0(t)\, exp[\beta_2 X_2(t)]$$

   sehingga *hazard ratio*-nya adalah :

   $$HR = \frac{h_1(t,\boldsymbol{X}) = h_0(t)\, exp[\beta_1 X_1(t)]}{h_2(t,\boldsymbol{X}) = h_0(t)\, exp[\beta_1 X_2(t)]} = exp[\beta_1 \Delta X(t)]$$

2. $X_b(t)$ dimana *covariate*-nya tidak bergantung pada waktu namun karena tidak memenuhi asumsi PH, X_b harus diinteraksikan dengan fungsi waktu. Misalkan fungsi waktu untuk *covariate* ke-b yaitu $g_b(t)$, maka interaksi X_b dengan $g_b(t)$ dapat dibentuk sebagai perkalian X_b dengan $g_b(t)$. Jadi, $X_b(t)=X_b.g_b(t)$.

Pada tugas akhir ini, $X_b(t)$ yang dibahas yaitu $X_b(t) = X_b.g_b(t)$, dimana X_b adalah *covariate* yang tidak bergantung pada waktu.

3.1 *Extended Cox Model*

Berdasarkan penjelasan pada subbab sebelumnya, *time dependent covariate*, $X_b(t)$, yang dibahas yaitu $X_b(t) = X_b.g_b(t)$ dimana X_b adalah covariate yang tidak bergantung pada waktu, namun karena tidak memenuhi asumsi PH, X_b harus diinteraksikan dengan fungsi waktu.

Misalkan semua *p covariate* tidak memenuhi asumsi *proportional hazard,* sehingga sebanyak *p covariate* harus diinteraksikan dengan fungsi waktu. Misal $g_b(t)$ adalah fungsi waktu untuk *covariate* ke-b, maka bentuk *extended cox model*-nya yaitu :

$$h(t,\boldsymbol{X}(t)) = h_0(t)\,exp\left[\sum_{b=1}^{p}\beta_b X_b + \sum_{b=1}^{p}\delta_b X_b g_b(t)\right] \quad (3.1)$$

Secara umum, jika terdapat p_1 covariate yang memenuhi asumsi *proportional hazard* dan p_2 *covariate* yang tidak memenuhi asumsi *proportional hazard* dimana $p_1 + p_2 = p$, maka model di atas menjadi :

$$h(t,\boldsymbol{X}(t)) = h_0(t)\,exp\left[\sum_{a=1}^{p_1}\beta_a X_a + \sum_{b=P_1+1}^{p_2}\beta_b X_b + \sum_{b=P_1+1}^{p_2}\delta_b X_b g_b(t)\right] \quad (3.2)$$

Beberapa fungsi waktu, $g_b(t)$, yang dapat digunakan diantaranya :

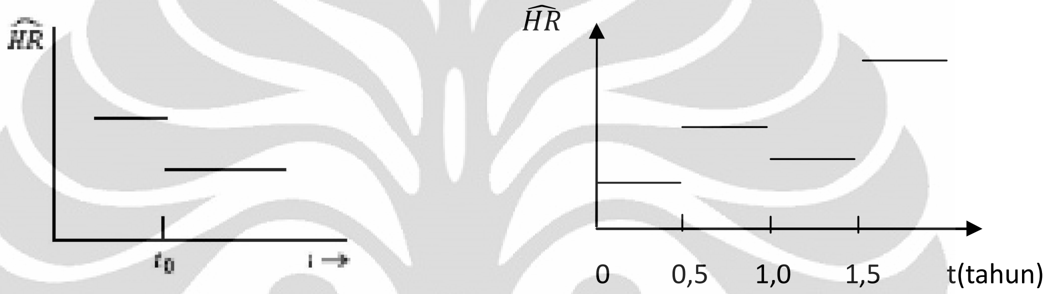
1. $g_b(t) = 0$
2. $g_b(t) = t$
3. $g_b(t) = \ln t$
4. $g_b(t)$ adalah fungsi *heaviside*

Selanjutnya, akan dibahas secara detail mengenai fungsi heaviside dan pembentukan *hazard ratio*-nya.

3.1.1 Fungsi *Heaviside*

Fungsi *heaviside* biasa disebut sebagai fungsi tangga, artinya bernilai konstan pada interval tertentu tetapi berbeda antar selang waktu. Fungsi ini digunakan jika *hazard ratio*-nya berubah hanya pada waktu tertentu saja. Jadi, ide dasar dari penggunaan fungsi *heaviside* adalah untuk mengakomodir perbedaan *hazard ratio* pada interval waktu yang berbeda, namun ditiap-tiap interval waktu, *hazard ratio*-nya bernilai sama.

Seperti yang ditunjukkan pada gambar berikut :

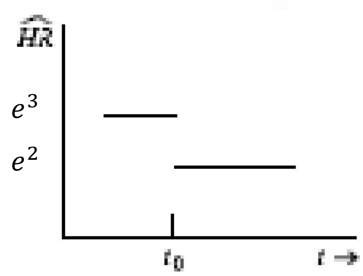


(a) (b)

Gambar 3.2. (a) Fungsi *heaviside* dengan dua interval waktu,
(b) Fungsi *heaviside* dengan empat interval waktu

Untuk lebih jelas, perhatikan ilustrasi berikut.

Dalam suatu data, dimisalkan satu *covariate*, sebut *X*, tidak memenuhi asumsi *proportional hazard*. Misalkan didapat kurva *hazard ratio* dari dua kondisi yang berbeda dari *covariate* tersebut sebagai berikut :



Gambar 3.3 Fungsi *heaviside* dengan dua interval waktu

Misalkan kondisi pertama dinotasikan dengan *X=1*, dan kondisi kedua dinotasikan dengan *X*=0. Maka, *hazard ratio* pada masing-masing interval dapat diperoleh dengan dua cara, yaitu :

1. Cara pertama

- $t < t_0 \rightarrow \widehat{HR} = \frac{h(t,X=1)}{h(t,X=0)} = \frac{h_0(t)\exp(\hat{\beta}.1)}{h_0(t)\exp(\hat{\beta}.0)} = \exp\hat{\beta} = \exp 3$. Jadi, $\hat{\beta} = 3$
- $t \geq t_0 \rightarrow \widehat{HR} = \frac{h(t,X=1)}{h(t,X=0)} = \frac{h_0(t)\exp(\hat{\beta}.1+\hat{\delta}.1)}{h_0(t)\exp(\hat{\beta}.0+\hat{\delta}.0)} = \exp\hat{\beta} + \hat{\delta} = \exp 2$. Jadi,

  $\hat{\beta} + \hat{\delta} = 2$. Diketahui $\hat{\beta} = 3$, maka $\hat{\delta} = -1$.

2. Cara kedua

- $t < t_0 \rightarrow \widehat{HR} = \frac{h(t,X=1)}{h(t,X=0)} = \frac{h_0(t)\exp(\hat{\delta}_1.1)}{h_0(t)\exp(\hat{\delta}_1.0)} = \exp\hat{\delta}_1 = \exp 3$. Jadi, $\hat{\delta}_1 = 3$
- $t \geq t_0 \rightarrow \widehat{HR} = \frac{h(t,X=1)}{h(t,X=0)} = \frac{h_0(t)\exp(\hat{\delta}_2.1)}{h_0(t)\exp(\hat{\delta}_2.0)} = \exp\hat{\delta}_2 = \exp 2$. Jadi, $\hat{\delta}_2 = 2$

Dari ilustrasi di atas, dapat terlihat bahwa baik cara pertama maupun cara kedua dapat diperoleh *hazard ratio* yang sama pada interval waktu yang sama. Sehingga, fungsi *heaviside* dapat dibentuk menjadi dua model, yaitu model asli dan model alternatif.

Untuk pembahasan fungsi *heaviside* selanjutnya, dimisalkan hanya satu *covariate* yang diperhatikan.

Fungsi *heaviside* dapat dikategorikan menjadi dua jenis, yaitu :

a. Fungsi *heaviside* dengan dua interval waktu (seperti yang ditunjukkan pada gambar 3.1(a))

   Fungsi *heaviside* ini dapat dibentuk dalam dua cara :

- Model asli, yaitu dengan memasukkan pengaruh utama dari *covariate* ke dalam model. Model yang terbentuk yaitu :

  $$h(t, \boldsymbol{X}(t)) = h_0(t)\, exp[\beta X + \delta X g(t)] \tag{3.3}$$

  dimana fungsi *heaviside*-nya adalah

  $$g(t) = \begin{cases} 1\; jika\; t \geq t_0 \\ 0\; jika\; t < t_0 \end{cases}$$

  Sehingga, untuk kedua interval tersebut, diperoleh model sebagai berikut :

Tabel 3.1. Bentuk model pada masing-masing interval waktu

Interval	Model
$t \geq t_0$	$h(t, \boldsymbol{X}) = h_0(t)\, exp[(\beta + \delta)X]$
$t < t_0$	$h(t, \boldsymbol{X}) = h_0(t)\, exp[\beta X]$

Selanjutnya, *hazard ratio* dapat dibentuk dengan membandingkan antara *hazard rate* saat $X = x$ dengan *hazard rate* saat $X = x+\Delta x$. Mengacu pada persamaan (3.4) dan (3.5), maka taksiran *hazard ratio* $(\widehat{HR}(x))$ pada masing-masing interval waktu adalah :

1. $t \geq t_0$,

$$\widehat{HR}(x) = \frac{h(t, \boldsymbol{X}=x)}{h(t, \boldsymbol{X}=x+\Delta x)} = \frac{h_0(t)\, exp[(\beta + \delta)\, x]}{h_0(t)\, exp[(\beta + \delta)(x+\Delta x)]} = exp\left((\hat{\beta} + \hat{\delta})\Delta x\right) \quad (3.6)$$

2. $t < t_0$,

$$\widehat{HR}(x) = \frac{h(t, \boldsymbol{X}=x)}{h(t, \boldsymbol{X}=x+\Delta x)} = \frac{h_0(t)\, exp[\beta(x)]}{h_0(t)\, exp[\beta(x+\Delta x)]} = exp(\hat{\beta}.\Delta x) \quad (3.7)$$

Dari persamaan (3.6) dan (3.7) dapat dilihat bahwa :

- Taksiran *hazard ratio* saat $t \geq t_0$ tidak bergantung waktu atau selalu tetap dengan nilai $exp\left((\hat{\beta} + \hat{\delta})\Delta x\right)$
- Taksiran *hazard ratio* saat $t < t_0$ tidak bergantung waktu atau selalu tetap dengan nilai $exp(\hat{\beta}.\Delta x)$
- Terlihat bahwa *hazard ratio* konstan pada masing-masing interval. Akan tetapi, pada interval yang berbeda dihasilkan *hazard ratio* yang berbeda pula

- Model alternatif, yaitu tanpa memasukkan pengaruh utama dari covariate ke dalam model. Model yang terbentuk yaitu :

$$h(t, \boldsymbol{X}(t)) = h_0(t)\, exp[\delta_1(X.g_1(t)) + \delta_2(X.g_2(t))] \quad (3.8)$$

dimana fungsi *heaviside* dari $g_1(t)$ dan $g_2(t)$ nya adalah

$$g_1(t) = \begin{cases} 1\; jika\; t \geq t_0 \\ 0\; jika\; t < t_0 \end{cases}$$

$$g_2(t) = \begin{cases} 1\; jika\; t < t_0 \\ 0\; jika\; t \geq t_0 \end{cases}$$

Sehingga, untuk kedua interval tersebut, diperoleh model sebagai berikut :

Tabel 3.2. Bentuk model pada masing-masing interval waktu

Interval	Model
$t \geq t_0$	$h(t,\boldsymbol{X}) = h_0(t)\, exp[\delta_1 X]$ (3.9)
$t < t_0$	$h(t,\boldsymbol{X}) = h_0(t)\, exp[\delta_2 X]$ (3.10)

Sama seperti pada model asli, hazard ratio dapat dibentuk dengan membandingkan antara hazard rate saat $X = x$ dengan hazard rate saat $X = x+\Delta x$. Mengacu pada persamaan (3.7) dan (3.8), maka $\widehat{HR}(x)$ pada masing-masing interval waktu adalah :

1. $t \geq t_0$, $\widehat{HR}(x) = \frac{h(t,\boldsymbol{X}=x)}{h(t,\boldsymbol{X}=x+\Delta x)} = \frac{h_0(t)\, exp[\delta_1.x]}{h_0(t)\, exp[\delta_1(x+\Delta x)]} = exp(\hat{\delta}_1.\Delta x)$ (3.11)
2. $t < t_0$, $\widehat{HR}(x) = \frac{h(t,\boldsymbol{X}=x)}{h(t,\boldsymbol{X}=x+\Delta x)} = \frac{h_0(t)\, exp[\delta_2.x]}{h_0(t)\, exp[\delta_2(x+\Delta x)]} = exp(\hat{\delta}_2.\Delta x)$ (3.12)

Dari persamaan (3.11) dan (3.12) dapat dilihat bahwa :

- Taksiran hazard ratio saat $t \geq t_0$ tidak bergantung waktu atau selalu tetap dengan nilai $exp(\hat{\delta}_1.\Delta x)$
- Taksiran hazard ratio saat $t < t_0$ tidak bergantung waktu atau selalu tetap dengan nilai $exp(\hat{\delta}_2.\Delta x)$
- Terlihat bahwa *hazard ratio* konstan pada masing-masing interval. Akan tetapi, pada interval yang berbeda dihasilkan *hazard ratio* yang berbeda pula

Model asli dan model alternatif ekivalen yaitu persamaan (3.3) sama dengan persamaan (3.8). Sehingga,

1. Untuk $t \geq t_0$

   Persamaan (3.4) sama dengan persamaan (3.9)

Tabel 3.3. Bentuk model asli dan model alternatif untuk interval waktu $t \geq t_0$

	Asli		Alternatif
$h_A(t,\boldsymbol{X})$	$h_0(t)\, exp[(\beta + \delta).x]$	=	$h_0(t)\, exp[\delta_1.x]$
$h_{\boldsymbol{B}}(t,\boldsymbol{X})$	$h_0(t)\, exp[(\beta + \delta)(x + \Delta x)]$	=	$h_0(t)\, exp[\delta_1(x + \Delta x)]$

Sehingga diperoleh $\hat{\beta} + \hat{\delta} = \hat{\delta}_1$. Hal ini mengakibatkan $\widehat{HR}(x)$ juga bernilai sama untuk interval waktu ini.

Tabel 3.4. Taksiran *hazard ratio* untuk model asli dan model alternatif untuk interval waktu $t \geq t_0$

	Asli		Alternatif
$\widehat{HR}(x)$	$exp\left((\hat{\beta} + \hat{\delta})\Delta x\right)$	=	$exp(\hat{\delta}_1.\Delta x)$

2. Untuk $t < t_0$

   Persamaan (3.5) sama dengan persamaan (3.10)

Tabel 3.5. Bentuk model asli dan model alternatif untuk interval waktu $t < t_0$

	Asli		Alternatif
$h_A(t, \boldsymbol{X})$	$h_0(t)\, exp[\beta.x]$	=	$h_0(t)\, exp[\delta_2.x]$
$h_B(t, \boldsymbol{X})$	$h_0(t)\, exp[\beta(x + \Delta x)]$	=	$h_0(t)\, exp[\delta_2(x + \Delta x)]$

Sehingga diperoleh $\hat{\beta} = \hat{\delta}_2$. Hal ini mengakibatkan $\widehat{HR}(x)$ juga bernilai sama untuk interval waktu ini.

Tabel 3.6. Taksiran *hazard ratio* untuk model asli dan model alternatif untuk interval waktu $t < t_0$

	Asli		Alternatif
$\widehat{HR}(x)$	$exp(\hat{\beta}.\Delta x)$	=	$exp(\hat{\delta}_2.\Delta x)$

Jadi, untuk interval waktu yang sama, bagaimanapun bentuk modelnya, baik model asli ataupun model alternatif lainnya, akan menghasilkan nilai taksiran *hazard ratio* yang sama.

b. Fungsi *heaviside* dengan lebih dari dua interval waktu

   Misalkan terdapat empat hazard ratio yang tetap dalam empat interval waktu (seperti yang ditunjukkan pada gambar 3.1(b)).

Terdapat empat interval waktu yaitu $0 \le t < 0.5$, $0.5 \le t < 1.0$, $1.0 \le t < 1.5$, dan $t \ge 1.5$. Fungsi heaviside nya juga dapat dibentuk dengan dua cara :

- Model asli, yaitu dengan memasukkan pengaruh utama dari covariate ke dalam model. Model yang terbentuk yaitu :

$$h(t, \boldsymbol{X}(t)) = h_0(t)\, exp[\beta X + \delta_1 X g_1(t) + \delta_2 X g_2(t) + \delta_3 X g_3(t)] \qquad (3.12)$$

dimana fungsi *heaviside*-nya adalah

$$g_1(t) = \begin{cases} 1\ jika\ 0.5 \le t < 1.0 \\ 0\ jika\ t\ yang\ lainnya \end{cases}$$

$$g_2(t) = \begin{cases} 1\ jika\ 1.0 \le t < 1.5 \\ 0\ jika\ t\ yang\ lainnya \end{cases}$$

$$g_3(t) = \begin{cases} 1\ jika\ t\ \ge 1.5 \\ 0\ jika\ t\ yang\ lainnya \end{cases}$$

Sehingga, untuk keempat interval tersebut, diperoleh model sebagai berikut :

Tabel 3.7. Bentuk model pada masing-masing interval waktu

Interval	Model
$0 \le t < 0.5$	$h(t, \boldsymbol{X}) = h_0(t)\, exp(\beta X)$ (3.14)
$0.5 \le t < 1.0$	$h(t, \boldsymbol{X}) = h_0(t)\, exp[(\beta + \delta_1)X]$ (3.15)
$1.0 \le t < 1.5$	$h(t, \boldsymbol{X}) = h_0(t)\, exp[(\beta + \delta_2)X]$ (3.16)
$t \ge 1.5$	$h(t, \boldsymbol{X}) = h_0(t)\, exp[(\beta + \delta_3)X]$ (3.17)

Selanjutnya, *hazard ratio* dapat dibentuk dengan membandingkan antara *hazard rate* saat $X = x$ dengan *hazard rate* saat $X = x + \Delta x$. Mengacu pada persamaan (3.14), (3.15), (3.16), dan (3.17), maka $\widehat{HR}(x)$ pada masing-masing interval waktu adalah :

1. $0 \le t < 0.5$,

$$\widehat{HR}(x) = \frac{h(t, \boldsymbol{X} = \boldsymbol{x})}{h(t, \boldsymbol{X} = \boldsymbol{x} + \Delta \boldsymbol{x})} = \frac{h_0(t)\, exp(\beta . x)}{h_0(t)\, exp(\beta(x + \Delta x))} = exp(\hat{\beta} . \Delta x) \qquad (3.18)$$

2. $0.5 \le t < 1.0$,

$$\widehat{HR}(x) = \frac{h(t, \boldsymbol{X} = \boldsymbol{x})}{h(t, \boldsymbol{X} = \boldsymbol{x} + \Delta \boldsymbol{x})} = \frac{h_0(t)\, exp[(\beta + \delta_1)x]}{h_0(t)\, exp[(\beta + \delta_1)(x + \Delta x)]} = exp\left((\hat{\beta} + \hat{\delta}_1)\Delta x\right) \qquad (3.19)$$

3. $1.0 \leq t < 1.5$,

$$\widehat{HR}(x) = \frac{h(t, \boldsymbol{X}=\boldsymbol{x})}{h(t, \boldsymbol{X}=\boldsymbol{x}+\Delta x)} = \frac{h_0(t)\, exp[(\beta + \delta_2)x]}{h_0(t)\, exp[(\beta + \delta_2)(x+\Delta x)]} = exp\left((\hat{\beta} + \hat{\delta}_2)\Delta x\right) \quad (3.20)$$

4. $t \geq 1.5$,

$$\widehat{HR}(x) = \frac{h(t, \boldsymbol{X}=\boldsymbol{x})}{h(t, \boldsymbol{X}=\boldsymbol{x}+\Delta x)} = \frac{h_0(t)\, exp[(\beta + \delta_3)x]}{h_0(t)\, exp[(\beta + \delta_3)(x+\Delta x)]} = exp\left((\hat{\beta} + \hat{\delta}_3)\Delta x\right) \quad (3.21)$$

Dari persamaan (3.18), (3.19), (3.20) dan (3.21) dapat dilihat bahwa :

- Taksiran *hazard ratio* saat *0≤ t<0.5* tidak bergantung waktu atau selalu tetap dengan nilai $exp(\hat{\beta}.\Delta x)$
- Taksiran *hazard ratio* saat *0.5≤ t<1.0* tidak bergantung waktu atau selalu tetap dengan nilai $exp\left((\hat{\beta} + \hat{\delta}_1)\Delta x\right)$
- Taksiran *hazard ratio* saat *0.5≤ t<1.0* tidak bergantung waktu atau selalu tetap dengan nilai $exp\left((\hat{\beta} + \hat{\delta}_2)\Delta x\right)$
- Taksiran *hazard ratio* saat *t ≥ 1.5* tidak bergantung waktu atau selalu tetap dengan nilai $exp\left((\hat{\beta} + \hat{\delta}_3)\Delta x\right)$
- Terlihat bahwa *hazard ratio* konstan pada masing-masing interval. Akan tetapi, pada interval yang berbeda dihasilkan *hazard ratio* yang berbeda pula

- Model alternatif, yaitu tanpa memasukkan pengaruh utama dari covariate ke dalam model. Model yang terbentuk yaitu :

$$h(t, \boldsymbol{X}(t)) = h_0(t)\, exp[\delta_1 X g_1(t) + \delta_2 X g_2(t) + \delta_3 X g_3(t) + \delta_4 X g_4(t)] \quad (3.22)$$

dimana fungsi *heaviside*-nya adalah

$$g_1(t) = \begin{cases} 1\ jika\ 0 \leq t < 0.5 \\ 0\ jika\ t\ yang\ lainnya \end{cases}$$

$$g_2(t) = \begin{cases} 1\ jika\ 0.5 \leq t < 1.0 \\ 0\ jika\ t\ yang\ lainnya \end{cases}$$

$$g_3(t) = \begin{cases} 1\ jika\ 1.0 \leq t < 1.5 \\ 0\ jika\ t\ yang\ lainnya \end{cases}$$

$$g_4(t) = \begin{cases} 1\ jika\ t \geq 1.5 \\ 0\ jika\ t\ yang\ lainnya \end{cases}$$

Sehingga, untuk keempat interval tersebut, diperoleh model sebagai berikut :

Tabel 3.8. Bentuk model pada masing-masing interval waktu

Interval	Model
$0 \leq t < 0.5$	$h(t, \boldsymbol{X}) = h_0(t)\, exp[\delta_1 X]$ (3.23)
$0.5 \leq t < 1.0$	$h(t, \boldsymbol{X}) = h_0(t)\, exp[\delta_2 X]$ (3.24)
$1.0 \leq t < 1.5$	$h(t, \boldsymbol{X}) = h_0(t)\, exp[\delta_3 X]$ (3.25)
$t \geq 1.5$	$h(t, \boldsymbol{X}) = h_0(t)\, exp[\delta_4 X]$ (3.26)

Sama seperti pada model asli, *hazard ratio* dapat dibentuk dengan membandingkan antara *hazard rate* saat $X = x$ dengan *hazard rate* saat $X=x+\Delta x$. $\widehat{HR}(x)$ pada keempat interval waktu tersebut adalah :

1. $0 \leq t < 0.5$,

$$\widehat{HR}(x) = \frac{h(t,\boldsymbol{X}=x)}{h(t,\boldsymbol{X}=x+\Delta x)} = \frac{h_0(t)\, exp[\delta_1 x]}{h_0(t)\, exp[\delta_1(x+\Delta x)]} = exp\left(\hat{\delta}_1 \Delta x\right) \qquad (3.27)$$

2. $0.5 \leq t < 1.0$,

$$\widehat{HR}(x) = \frac{h(t,\boldsymbol{X}=\boldsymbol{x})}{h(t,\boldsymbol{X}=\boldsymbol{x}+\Delta x)} = \frac{h_0(t)\, exp[\delta_2 x]}{h_0(t)\, exp[\delta_2(x+\Delta x)]} = exp\left(\hat{\delta}_2 \Delta x\right) \qquad (3.28)$$

3. $1.0 \leq t < 1.5$,

$$\widehat{HR}(x) = \frac{h(t,\boldsymbol{X}=\boldsymbol{x})}{h(t,\boldsymbol{X}=x+\Delta x)} = \frac{h_0(t)\, exp[\delta_3 x]}{h_0(t)\, exp[\delta_3(x+\Delta x)]} = exp\left(\hat{\delta}_3 \Delta x\right) \qquad (3.29)$$

4. $t \geq 1.5$,

$$\widehat{HR}(x) = \frac{h(t,\boldsymbol{X}=x)}{h(t,\boldsymbol{X}=\boldsymbol{x}+\Delta x)} = \frac{h_0(t)\, exp[\delta_4 x]}{h_0(t)\, exp[\delta_4(x+\Delta x)]} = exp\left(\hat{\delta}_4 \Delta x\right) \qquad (3.30)$$

Dari persamaan (3.18), (3.19), (3.20) dan (3.21) dapat dilihat bahwa :

- Taksiran *hazard ratio* saat $0 \leq t < 0.5$ tidak bergantung waktu atau selalu tetap dengan nilai $exp\left(\hat{\delta}_1 \Delta x\right)$
- Taksiran *hazard ratio* saat $0.5 \leq t < 1.0$ tidak bergantung waktu atau selalu tetap dengan nilai $exp\left(\hat{\delta}_2 \Delta x\right)$
- Taksiran *hazard ratio* saat $0.5 \leq t < 1.0$ tidak bergantung waktu atau selalu tetap dengan nilai $exp\left(\hat{\delta}_3 \Delta x\right)$
- Taksiran *hazard ratio* saat $t \geq 1.5$ tidak bergantung waktu atau selalu tetap dengan nilai $exp\left(\hat{\delta}_4 \Delta x\right)$

- Terlihat bahwa *hazard ratio* konstan pada masing-masing interval. Akan tetapi, pada interval yang berbeda dihasilkan *hazard ratio* yang berbeda pula

Model asli dan model alternatif ekivalen yaitu persamaan (3.12) sama dengan persamaan (3.22). Sehingga,

1. Untuk $0 \leq t < 0.5$

   Persamaan (3.14) sama dengan persamaan (3.23)

Tabel 3.9. Bentuk model asli dan model alternatif untuk interval waktu $0 \leq t < 0.5$

	Asli		Alternatif
$h_A(t, \boldsymbol{X})$	$h_0(t)\, exp(\beta . x)$	=	$h_0(t)\, exp[\delta_1 . x]$
$h_B(t, \boldsymbol{X})$	$h_0(t)\, exp(\beta(x + \Delta x))$	=	$h_0(t)\, exp[\delta_1(x + \Delta x)]$

Sehingga diperoleh $\hat{\beta} = \hat{\delta}_1$. Hal ini mengakibatkan $\widehat{HR}(x)$ juga bernilai sama untuk interval waktu ini.

Tabel 3.10. Taksiran *hazard ratio* untuk model asli dan model alternatif untuk interval waktu $0 \leq t < 0.5$

	Asli		Alternatif
$\widehat{HR}(x)$	$exp(\hat{\beta} . \Delta x)$	=	$exp(\hat{\delta}_1 . \Delta x)$

2. Untuk $0.5 \leq t < 1.0$

   Persamaan (3.15) sama dengan persamaan (3.24)

Tabel 3.11. Bentuk model asli dan model alternatif untuk interval waktu $0.5 \leq t < 1.0$

	Asli		Alternatif
$h_A(t, \boldsymbol{X})$	$h_0(t)\, exp[(\beta + \delta_1)x]$	=	$h_0(t)\, exp[\delta_2 x]$
$h_B(t, \boldsymbol{X})$	$h_0(t)\, exp[(\beta + \delta_1)(x + \Delta x)]$	=	$h_0(t)\, exp[\delta_2(x + \Delta x)]$

Sehingga diperoleh $\hat{\beta} + \hat{\delta}_1 = \hat{\delta}_2$. Hal ini mengakibatkan $\widehat{HR}(x)$ juga bernilai sama untuk interval waktu ini.

Tabel 3.12. Taksiran *hazard ratio* untuk model asli dan model alternatif untuk interval waktu 0.5≤ t<1.0

	Asli		Alternatif
$\widehat{HR}(x)$	$exp\left((\hat{\beta} + \hat{\delta}_1)\Delta x\right)$	=	$exp(\hat{\delta}_2.\Delta x)$

3. Untuk 1.0≤ t<1.5

   Persamaan (3.16) sama dengan persamaan (3.25)

Tabel 3.13. Bentuk model asli dan model alternatif untuk interval waktu 1.0≤ t<1.5

	Asli		Alternatif
$h_A(t,\boldsymbol{X})$	$h_0(t)\, exp[(\beta + \delta_2)x]$	=	$h_0(t)\, exp[\delta_3 x]$
$h_B(t,\boldsymbol{X})$	$h_0(t)\, exp[(\beta + \delta_2)(x + \Delta x)]$	=	$h_0(t)\, exp[\delta_3(x + \Delta x)]$

Sehingga diperoleh $\hat{\beta} + \hat{\delta}_2 = \hat{\delta}_3$. Hal ini mengakibatkan $\widehat{HR}(x)$ juga bernilai sama untuk interval waktu ini.

Tabel 3.14. Taksiran *hazard ratio* untuk model asli dan model alternatif untuk interval waktu 1.0≤ t<1.5

	Asli		Alternatif
$\widehat{HR}(x)$	$exp\left((\hat{\beta} + \hat{\delta}_2)\Delta x\right)$	=	$exp(\hat{\delta}_3.\Delta x)$

4. Untuk $t \geq 1.5$

   Persamaan (3.17) sama dengan persamaan (3.26)

Tabel 3.15. Bentuk model asli dan model alternatif untuk interval waktu $t \geq 1.5$

	Asli		Alternatif
$h_A(t,\boldsymbol{X})$	$h_0(t)\, exp[(\beta + \delta_3)x]$	=	$h_0(t)\, exp[\delta_4 x]$
$h_B(t,\boldsymbol{X})$	$h_0(t)\, exp[(\beta + \delta_3)(x + \Delta x)]$	=	$h_0(t)\, exp[\delta_4(x + \Delta x)]$

Sehingga diperoleh $\hat{\beta} + \hat{\delta}_3 = \hat{\delta}_4$. Hal ini mengakibatkan $\widehat{HR}(x)$ juga bernilai sama untuk interval waktu ini.

Tabel 3.16. Taksiran *hazard ratio* untuk model asli dan model alternatif untuk interval waktu $t \geq 1.5$

	Asli		Alternatif
$\widehat{HR}(x)$	$exp\left((\hat{\beta} + \hat{\delta}_3)\Delta x\right)$	=	$exp(\hat{\delta}_4.\Delta x)$

Jadi, untuk interval waktu yang sama, bagaimanapun bentuk modelnya, baik model asli ataupun model alternatif lainnya, akan menghasilkan nilai taksiran hazard ratio yang sama.

Dapat terlihat bahwa fungsi heaviside dapat dipakai untuk memberikan taksiran hazard ratio yang tetap konstan dalam setiap interval waktu, namun berbeda dalam interval waktu yang berlainan.

Dalam membentuk *extended cox model*, dapat digunakan berbagai bentuk fungsi waktu. Untuk menentukan fungsi waktu yang tetap untuk suatu data, dapat dilihat berdasarkan plot $\hat{S}$ terhadap waktu bagi covariate yang tidak memenuhi asumsi *proportional hazard*.

Diantara beberapa fungsi waktu tersebut yang bisa digunakan, fungsi waktu yang lebih sederhana untuk membentuk extended cox model adalah fungsi *heaviside*. Pada fungsi *heaviside* diperoleh *hazard ratio*-nya konstan untuk interval waktu tertentu. Untuk interval waktu yang lain, *hazard ratio*-nya juga konstan, namun nilainya berbeda dengan *hazard ratio* di interval waktu sebelumnya.

3.2 *Hazard Ratio* untuk *Extended Cox Model*

Karena *hazard ratio* yang didapat tidak lagi memenuhi asumsi PH (*Proportional Hazard*), artinya *hazard ratio* tidak bernilai sama sepanjang waktu, digunakan *extended cox model*.

Berdasarkan *hazard rate* dari *extended cox model*, perhatikan kembali model 3.2, akan dibandingkan *hazard rate* untuk dua kondisi yang berbeda. Misalkan ***Z(t)*** = ***(X*,X)*** dimana

$$\boldsymbol{X}^* = (X_1^*, X_2^*, \dots, X_p^*, X_1^*(t), X_2^*(t), \dots, X_p^*(t))$$

dan

$$\boldsymbol{X} = (X_1, X_2, \dots, X_p, X_1(t), X_2(t), \dots, X_p(t))$$

maka

1. Untuk pengamatan pertama, $\boldsymbol{X}^* \rightarrow h(t, \boldsymbol{X}^*(t))$

$$h\big(t, \boldsymbol{X}^*(t)\big) = h_0(t)\, exp\left[\sum_{a=1}^{p_1} \beta_a \boldsymbol{X}_{\boldsymbol{a}}^* + \sum_{b=p_1+1}^{p_2} \beta_b \boldsymbol{X}_b{}^* + \sum_{b=p_1+1}^{p_2} \delta_b \boldsymbol{X}_{\boldsymbol{b}}{}^*(t)\right]$$

2. Untuk pengamatan kedua, $\boldsymbol{X} \rightarrow h(t, \boldsymbol{X}(t))$

$$h\big(t, \boldsymbol{X}(t)\big) = h_0(t)\, exp\left[\sum_{a=1}^{p_1} \beta_a \boldsymbol{X}_{\boldsymbol{a}} + \sum_{b=p_1+1}^{p_2} \beta_b \boldsymbol{X}_{\boldsymbol{b}} + \sum_{b=p_1+1}^{p_2} \delta_b \boldsymbol{X}_{\boldsymbol{b}}(t)\right]$$

Sehingga, *hazard ratio* dari kedua pengamatan tersebut yaitu :

$$\widehat{HR}\big(t, \boldsymbol{Z}(t)\big) = \frac{h\big(t, \boldsymbol{X}^*(t)\big)}{h\big(t, \boldsymbol{X}(t)\big)}$$

$$= \frac{h_0(t)\, exp\left[\sum_{a=1}^{p_1} \beta_a \boldsymbol{X}_{\boldsymbol{a}}^* + \sum_{b=p_1+1}^{p_2} \beta_b \boldsymbol{X}_b{}^* + \sum_{b=p_1+1}^{p_2} \delta_b \boldsymbol{X}_b{}^*(t)\right]}{h_0(t)\, exp\left[\sum_{a=1}^{p_1} \beta_a \boldsymbol{X}_{\boldsymbol{a}} + \sum_{b=p_1+1}^{p_2} \beta_b \boldsymbol{X}_{\boldsymbol{b}} + \sum_{b=p_1+1}^{p_2} \delta_b \boldsymbol{X}_{\boldsymbol{b}}(t)\right]}$$

$$= exp\Big[\sum_{a=1}^{p_1} \hat{\beta}_a (\boldsymbol{X}_{\boldsymbol{a}}^* - \boldsymbol{X}_{\boldsymbol{a}}) + \sum_{b=p_1+1}^{p_2} \hat{\beta}_b (\boldsymbol{X}_{\boldsymbol{b}}{}^* - \boldsymbol{X}_{\boldsymbol{b}}) + \sum_{b=p_1+1}^{p_2} \hat{\delta}_b (\boldsymbol{X}_{\boldsymbol{b}}{}^*(t) - \boldsymbol{Xb}(t) \qquad (3.31)$$

Pada perumusan di atas, bentuk

$$\sum_{b=p_1+1}^{p_2} \hat{\delta}_b (\boldsymbol{X}_{\boldsymbol{b}}{}^*(t) - \boldsymbol{X}_{\boldsymbol{b}}(t))$$

menunjukkan bahwa taksiran hazard ratio bergantung pada waktu.

Sebagai contoh, misalkan terdapat satu *covariate* yang diperhatikan, sebut X. X tidak memenuhi asumsi PH sehingga perlu diinteraksikan dengan fungsi waktu. Misalkan fungsi waktu yang digunakan yaitu $g(t) = t$. Maka, modelnya adalah

$$h\big(t, \boldsymbol{X}(t)\big) = h_0(t)\, exp[\beta X + \delta(X.t)]$$

Hazard ratio untuk objek A dengan nilai $X = x$ terhadap objek B dengan nilai $X = x+\Delta x$ adalah

$$\widehat{HR}(t) = \frac{h_A\big(t, \boldsymbol{X}(t)\big)}{h_B\big(t, \boldsymbol{X}(t)\big)} = \frac{h_0(t)\, exp[(\beta + \delta t)x]}{h_0(t)\, exp[(\beta + \delta t)(x + \Delta x)\,]} = exp\big[\big(\hat{\beta} + \hat{\delta} t\big)\Delta x\big]$$

Berdasarkan persamaan di atas, dapat dilihat bahwa taksiran *hazard ratio* adalah fungsi dari waktu. Jika $\hat{\delta}$ bernilai positif, maka *hazard ratio* akan meningkat sejalan dengan meningkatnya waktu.

Secara umum, berdasarkan model (3.31), terlihat bahwa :

1. Jika $\hat{\delta}_i = 0 \rightarrow \widehat{HR}\big(t, \boldsymbol{Z}(t)\big)$ tidak mengandung fungsi waktu sehingga dapat digunakan model cox biasa yang memenuhi asumsi PH
2. Jika $\hat{\delta}_i \neq 0 \rightarrow \widehat{HR}\big(t, \boldsymbol{Z}(t)\big)$ mengandung fungsi waktu sehingga model yang digunakan yaitu *extended cox model.*

Pengecekan $\hat{\delta}_i$ ini adalah salah satu pengecekan asumsi PH dengan menggunakan *extended cox model*.

3.3 *Maximum Partial Likelihood Estimator* untuk *Extended Cox Model*

Pada subbab ini, akan digambarkan mengenai bagaimana parameter-parameter dari extended cox model didapat. Taksiran parameter dari covariate yang bersesuaian disebut taksiran maksimum likelihood dan dinyatakan dengan $\widehat{\boldsymbol{\beta}}\ dan\ \widehat{\boldsymbol{\delta}}_{\boldsymbol{i}}$.

Taksiran maksimum likelihood dari parameter *extended cox model* diturunkan dengan memaksimumkan fungsi likelihood, biasanya dinyatakan dengan *L*. Fungsi likelihood adalah pernyataan matematika yang menggambarkan probabilitas bersama dari data yang teramati pada seseorang dalam penelitian sebagai fungsi dari parameter-parameter dari model. Terkadang, *L* ditulis dengan notasi $L(\boldsymbol{\beta})$ dimana $\boldsymbol{\beta}$ menyatakan kumpulan dari parameter-parameter yang belum diketahui. Pada tugas akhir ini,

$$\boldsymbol{\beta} = \big(\beta_1, \beta_2, \dots, \beta_p, \delta_1, \delta_2, \dots, \delta_{p_2}\big).$$

Pada *extended cox model,* $L(\boldsymbol{\beta})$ yang digunakan yaitu fungsi partial likelihood, $L_p(\boldsymbol{\beta})$, karena :

1. Pada data survival, data dapat berupa data tersensor ataupun teramati. Sedangkan pada fungsi likelihood, probabilitas yang dipertimbangkan hanya kontribusi dari objek-objek yang *event*-nya teramati, kontribusi objek yang tersensor tidak dipertimbangkan secara langsung. Karena dalam pembentukan

fungsi likelihoodnya tidak menggunakan seluruh data yang ada, $L(\boldsymbol{\beta})$ yang digunakan yaitu fungsi partial likelihood, $L_p(\boldsymbol{\beta})$.

Contoh :

Misal sampel berukuran n, dimana :

- k objek mengalami kejadian
- n-k objek tersensor

Pada fungsi partial likelihood, hanya k objek yang mengalami kejadian yang berkontribusi langsung.

2. Pada fungsi partial likelihood, tidak diketahui distribusi dari variabel responnya. Oleh karena itu, probabilitas yang diperhitungkan pada L_p hanya bergantung pada statistik terurut dan data sampel.

Contoh,

Jika terdapat data survival seperti berikut, sebut data I.

Tabel 3.17. Data waktu sampai suatu objek mendapat *event*

j	1	2	3
t_j	3	8	9

dan terdapat data yang lain seperti berikut, sebut data II

Tabel 3.18. Data waktu sampai suatu objek mendapat *event*

j	1	2	3
t_j	2	5	8

dimana j = objek, t_j = *time to event* untuk objek ke-j.

Misalkan data tersebut memenuhi asumsi *equally likely* pada masing-masing waktu yang teramati, misal t_j.

- Untuk data I

  Pr(Objek mengalami event pada T=8) = Pr(Objek mengalami event pada T=8|T≥8) = 1/2
- Untuk data II

  Pr(Objek mengalami event pada T=8) = Pr(Objek mengalami event pada T=8|T≥8) = 1/1

Terlihat bahwa nilai probabilitas saat t=8 dari data I dan data II berbeda. Sementara, jika distribusi dari T diketahui, maka probabilitasnya akan sama untuk t=8, yaitu *f(8)*. Akan tetapi, (Robin dan Tsiatis) menunjukkan bahwa taksirannya *asymptotically* normal.

Fungsi partial likelihood dapat ditulis sebagai perkalian dari likelihood masing-masing objek yang *event*-nya teramati. Hal ini disebabkan karena objek dipilih secara acak, atau semua pengamatan berasal dari sampel acak.

Misalkan, $t_1 < t_2 < \cdots < t_n$ menyatakan *survival time* yang telah diurutkan. Misalkan hasil pengamatan seperti pada tabel berikut :

Tabel 3.19. Data *survival time* dan kontribusi *likelihood*-nya pada masing-masing waktu

j	T_j	λ_j	L_j
1	t_1	1	L_1
2	t_2	0	L_2
⋮	⋮	⋮	⋮
n	t_n	1	L_n

dimana,

- j = objek
- T_j = *survival time*
- $\lambda_j = \begin{cases} 1 \text{ untuk } event \text{ yang teramati} \\ 0 \text{ untuk } event \text{ yang tersensor} \end{cases}$

dan L_j menyatakan likelihood dari objek yang teramati pada waktu t_j.

Jika himpunan dari objek yang teramati dinotasikan dengan $D(t_j)$, maka fungsi *partial* likelihoodnya yaitu :

$$L_p(\beta) = \prod_{j \epsilon D(t_j)} L_j$$

Setelah fungsi partial likelihood dibentuk, langkah selanjutnya yaitu memaksimumkan fungsi partial likelihood ini. Secara umum diselesaikan dengan memaksimumkan natural log dari L, dimana perhitungannya lebih mudah.

Proses memaksimumkan dilakukan dengan mengambil turunan partial dari log L_p terhadap setiap parameter di dalam model, yaitu :

$$\frac{\partial lnL_p}{\partial \beta_a} = 0, \qquad a = 1, 2, \dots, p_1$$

$$\frac{\partial lnL_p}{\partial \beta_b} = 0, \qquad b = 1, 2, \dots, p_2$$

$$\frac{\partial lnL_p}{\partial \delta_b} = 0, \qquad b = 1, 2, \dots, p_2$$

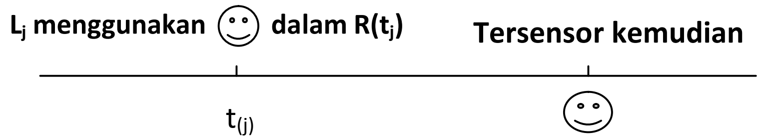
Setelah itu, sistem persamaan tersebut diselesaikan secara simultan dengan menggunakan iterasi.

3.3.1 Bentuk Fungsi *Partial Likelihood* untuk *Extended Cox Model*

Biasanya, pembentukan fungsi likelihood berdasarkan pada distribusi dari variabel responnya. Namun, pada *extended cox model,* distribusi dari variabel respon tidak diketahui. Inilah yang membedakannya dengan model parametrik.

Pembentukan fungsi partial likelihood dalam *extended cox model* berdasarkan pada urutan kejadian yang teramati. Setelah diurutkan, dibentuk likelihood dari masing-masing objek yang teramati. Likelihood dari masing-masing objek yang teramati menyatakan probabilitas objek tersebut mengalami *event* pada waktu t, bersyarat bahwa $T_j \geq t$. L_j bergantung pada himpunan objek-objek yang masih beresiko untuk mengalami *event* sampai waktu t_j yang disebut himpunan resiko, dinotasikan dengan $R(t_{(j)})$. Banyaknya anggota dari himpunan ini akan menjadi lebih kecil seiring dengan meningkatnya *failure time.*

Meskipun partial likelihood di atas tidak melibatkan objek yang tersensor, informasi dari objek yang tersensor tetap berkontribusi dalam partial likelihood tersebut. Kontribusinya yaitu dalam menentukan himpunan resiko pada waktu t_j. Terlihat pada gambar berikut :



Gambar 3.4. Suatu individu tersensor setelah waktu t_j

- Objek tersensor setelah waktu t_j → objek tersebut masih menjadi anggota dari himpunan resiko pada waktu t_j
- Objek tersensor sebelum waktu t_j → objek tersebut tidak lagi menjadi anggota dari himpunan resiko pada waktu t_j

Untuk menggambarkan fungsi partial likelihood dari *extended cox model*, perhatikan contoh berikut :

Misalkan Geri, Leri, dan Beri masing-masing diberikan satu tiket undian. Dari ketiga peserta tersebut, pemenang tiket dipilih pada waktu t_j (j = 1, 2, 3) dengan $t_1 < t_2 < t_3$. Dalam hal ini diasumsikan setiap individu pada akhirnya terpilih dan setelah satu individu terpilih maka dia tidak dapat dipilih lagi (artinya, individu tersebut telah keluar dari himpunan resiko). Berapa probabilitas terpilihnya tiket undian bila urutan terpilihnya pertama Beri, kemudian Geri, dan terakhir Leri (asumsikan probabilitas terambil setiap tiket adalah sama)?

Berdasarkan urutannya, dapat dilihat dalam tabel berikut :

Tabel 3.20. Probabilitas masing-masing individu di waktu t_j

Urutan	t_1	t_2	t_3
Tiket yang terpilih	Beri	Geri	Leri
Jumlah tiket yang dimiliki	1	1	1
Probabilitas	1/3	½	1/1

1. Probabilitas tiket Beri terpilih sebelum tiket Geri dan Leri terpilih yaitu 1/3 (karena anggota himpunan resikonya masih ada 3 individu). Setelah tiket Beri

terpilih, tiket Beri tidak dapat dipilih lagi sehingga keluar dari himpunan resiko.

2. Probabilitas tiket Geri terpilih sebelum tiket Leri terpilih yaitu ½ (karena anggota himpunan resikonya masih ada 2 individu). Setelah tiket Geri terpilih, tiket Geri tidak dapat dipilih lagi sehingga keluar dari himpunan resiko.
3. Probabilitas tiket Leri terpilih terakhir yaitu 1/1 (karena anggota himpunan resiko tinggal Leri sendiri). Setelah tiket Leri terpilih, tidak ada lagi anggota di dalam himpunan resiko.

   Jadi, probabilitas terpilihnya tiket undian dengan diberikan urutan kejadian seperti itu yaitu :

$$\frac{1}{3} \times \frac{1}{2} \times \frac{1}{1} = \frac{1}{6}$$

Sekarang, contoh di atas akan dikembangkan. Misalkan Beri mempunyai 4 tiket, Geri mempunyai 1 tiket, dan Leri mempunyai 2 tiket. Diasumsikan probabilitas terambilnya setiap tiket adalah sama. Berapa probabilitas terpilihnya tiket undian bila urutan terpilihnya pertama Beri, kemudian Geri, dan terakhir Leri?

Berdasarkan urutannya, dapat dilihat dalam tabel berikut :

Tabel 3.21. Probabilitas masing-masing individu di waktu t_j

Urutan	t_1	t_2	t_3
Tiket yang terpilih	Beri	Geri	Leri
Jumlah tiket yang dimiliki	4	1	2
Probabilitas	4/7	1/3	2/2

Jumlah tiket seluruhnya 7 tiket. Misalkan,

- Y_j menyatakan kejadian pada waktu t_j akan dipilih 1 tiket
- B menyatakan kejadian tiket Beri terpilih
- G menyatakan kejadian tiket Geri terpilih
- L menyatakan kejadian tiket Leri terpilih

maka :

1. Probabilitas tiket Beri terpilih pada waktu t_1 bersyarat pada waktu t_1 akan dipilih 1 tiket yaitu

$$P(B|Y_1) = \frac{P(B \cap Y_1)}{P(Y_1)} = \frac{P(B \cap Y_1)}{P(B \cup G \cup L)} = \frac{\frac{4}{7}}{\frac{4}{7} + \frac{1}{7} + \frac{2}{7}} = \frac{4}{7}$$

Setelah tiket Beri terpilih, jumlah tiket yang tersisa yaitu 7-4=3 tiket.

2. Probabilitas tiket Geri terpilih pada waktu t_2 bersyarat pada waktu t_2 akan dipilih 1 tiket yaitu

$$P(G|Y_2) = \frac{P(G \cap Y_2)}{P(Y_2)} = \frac{P(G \cap Y_2)}{P(G \cup L)} = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{2}{3}} = \frac{1}{3}$$

Setelah tiket Geri terpilih, jumlah tiket yang tersisa yaitu 3-1=2 tiket.

3. Probabilitas tiket Leri terpilih pada waktu t_3 bersyarat pada waktu t_3 akan dipilih 1 tiket yaitu

$$P(L|Y_3) = \frac{P(L \cap Y_3)}{P(Y_3)} = \frac{P(L \cap Y_3)}{P(L)} = \frac{\frac{2}{2}}{\frac{2}{2}} = \frac{2}{2} = 1$$

Setelah tiket Leri terpilih, jumlah tiket yang tersisa yaitu 2-2=0 tiket.

Jadi, probabilitas terpilihnya tiket undian dengan diberikan urutan kejadian seperti itu yaitu :

$$\frac{4}{7} \times \frac{1}{3} \times \frac{2}{2} = \frac{4}{21}$$

Dari contoh di atas, probabilitas dari urutan tertentu dipengaruhi oleh banyaknya tiket yang dimiliki setiap individu dan total banyaknya tiket pada tiap waktu pengundian.

Jika contoh-contoh tersebut dikaitkan dengan likelihood pada *extended cox model*, maka dapat dilihat pada tabel berikut :

Tabel 3.22. Probabilitas masing-masing individu yang dikaitkan dengan fungsi *hazard*

Pada contoh di atas				Notasi			Pada Extended Cox Model
Urutan waktu kejadian	t_1	t_2	t_3	t_1	t_2	t_3	*time to event*
Proba-bilitas	Beri	Geri	Leri	Beri	Geri	Leri	Hazard rate dari objek yang teramati
	4/7	1/3	2/2	$h_B(t_1)$	$h_G(t_2)$	$h_L(t_3)$	
Total Jumlah Tiket	7	3	2	$R(t_1)$	$R(t_2)$	$R(t_3)$	Himpunan Resiko
Probabilitas bersyarat	$P(B\|Y_1)$	$P(G\|Y_2)$	$P(L\|Y_3)$	$\frac{h_B(t_1)}{h(t_1)}$	$\frac{h_G(t_2)}{h(t_2)}$	$\frac{h_L(t_3)}{h(t_3)}$	Kontribusi likelihood

dengan,

- $h_B(t_1)$ = hazard rate Beri pada waktu t_1
- $h_G(t_2)$ = hazard rate Geri pada waktu t_2
- $h_L(t_3)$ = hazard rate Leri pada waktu t_3
- $h(t_1) = \sum_{i \epsilon R(t_1)} h_i(t_1)$
- $h(t_2) = \sum_{i \epsilon R(t_2)} h_i(t_2)$
- $h(t_3) = \sum_{i \epsilon R(t_3)} h_i(t_3)$

Untuk *extended cox model*, likelihood dari masing-masing objek sesuai urutan kejadian yang teramati dipengaruhi oleh covariate setiap objek. Untuk menggambarkan hubungan antara contoh di atas dengan pembentukan fungsi partial likelihood pada *extended cox model*, akan diberikan contoh sebagai berikut.

Pada contoh ini diberikan satu *covariate* yang diperhatikan, yaitu merokok, sebut X, dimana

$$X = \begin{cases} 1, \text{jika perokok} \\ 0, \text{jika bukan perokok} \end{cases}$$

- Beri mengalami *event* pertama kali yaitu pada waktu *t*=2 tahun dan dia adalah perokok.
- Selanjutnya, Geri mengalami *event* pada waktu *t*=3 tahun dan dia bukan perokok.
- Heri tersensor pada waktu *t*=5 tahun dan dia bukan perokok.
- Terakhir, Leri mengalami *event* pada waktu *t*=8 tahun dan dia adalah perokok.

Dalam bentuk tabel dapat dilihat sebagai berikut :

Tabel 3.23. *Survival time* dari Beri, Geri, Heri, dan Leri dan *covariate*-nya

ID	TIME	λ	MEROKOK
Beri	2	1	1
Geri	3	1	0
Heri	5	0	0
Leri	8	1	1

Misalkan diteliti waktu sampai seseorang sembuh setelah dilakukan operasi. *Covariate* yang diperhatikan yaitu merokok. Misalkan bahwa seseorang yang memiliki kebiasaan merokok akan semakin lama sembuhnya dibandingkan dengan seseorang yang tidak merokok. Namun, faktor usiapun memengaruhi perbandingan sembuhnya seseorang yang merokok dengan yang tidak merokok. Semakin tua usia seseorang, semakin besar perbandingan sembuhnya seseorang yang merokok dengan yang tidak merokok. Hal ini menyatakan bahwa covariate merokok tidak memenuhi asumsi PH, untuk memfasilitasi hal tersebut, perlu dilakukan modifikasi pada model cox PH, yaitu model cox dengan melibatkan interaksi antara covariate dan waktu, *g(t)*. Misalkan fungsi waktu yang dipilih yaitu *g(t)*=t, maka *extended cox model*-nya seperti berikut :

$$h(t) = h_0(t)\, exp(\beta_1 X + \delta_1 X_1)$$

dimana $X_1 = X.t$.

Dari model di atas, terlihat bahwa tidak hanya fungsi *baseline hazard* yang berubah terhadap waktu tetapi terdapat *covariate* yang nilainya juga berubah terhadap waktu, yaitu X_1. Berdasarkan contoh tersebut :

- Jika X=1, maka

$$h_a(t) = h_0(t)\, exp(\beta_1 + \delta_1 t)$$

- Jika X=0, maka

$$h_a(t) = h_0(t)\, exp(0)$$

Sehingga, *hazard ratio*-nya bergantung pada waktu, yaitu :

$$\widehat{HR}(t) = exp(\beta_1 + \delta_1 t)$$

Perhatikan ilustrasi berikut,

- Leri adalah seorang perokok, X=1, yang mengalami *event* pada waktu t=8 tahun. Karena $X.t = 1.t = t$, nilai *hazard* Leri pada setiap waktu kejadian, yaitu saat t = 2, 3, dan 8 tahun berubah-ubah. Dapat dilihat dari tabel berikut :

Tabel 3.24. *Hazard* Leri pada waktu yang teramati

TIME	*Hazard* Leri
2	$h_0(t)\, exp(\beta_1 + 2\delta_1)$
3	$h_0(t)\, exp(\beta_1 + 3\delta_1)$
8	$h_0(t)\, exp(\beta_1 + 8\delta_1)$

Seperti contoh sebelumnya, untuk menggambarkan *partial likelihood* untuk kasus ini, ditentukan probabilitas masing-masing individu yang teramati pada waktu t_j (j=1, 2, ...) bersyarat pada waktu t_j (j=1, 2, ...) tersebut terjadi suatu *event* dimana terdapat himpunan resiko, $R(t_{(j)})$. Himpunan individu-individu yang teramati dinotasikan dengan $D(t_j)$.

Tingkat *hazard* seseorang memainkan peran yang serupa dalam pembentukan fungsi partial likelihood seperti banyaknya tiket yang dimiliki setiap individu berperan untuk menghitung probabilitas terpilihnya tiket dalam contoh yang digambarkan sebelumnya.

Sesuai urutan waktu dimana individunya teramati, ditentukan :

1. Probabilitas Beri mengalami kejadian pada waktu t=2 tahun dimana anggota himpunan resikonya yaitu Beri, Geri, Heri, dan Leri, yaitu :

$$\frac{h_0(t)\,exp(\beta_1 + 2\delta_1)}{h_0(t)\,exp(\beta_1 + 2\delta_1) + h_0(t)\,exp\,0 + h_0(t)\,exp\,0 + h_0(t)\,exp(\beta_1 + 2\delta_1)}$$

Karena $h_0(t)$-nya sama di waktu t=2 maka dapat dicoret, sehingga

$$\frac{exp(\beta_1 + 2\delta_1)}{exp(\beta_1 + 2\delta_1) + exp(0) + exp(0) + exp(\beta_1 + 2\delta_1)}$$

2. Probabilitas Geri mengalami kejadian pada waktu t=3 tahun dimana anggota himpunan resikonya yaitu Geri, Heri, dan Leri, (Beri sudah keluar dari himpunan resiko karena diasumsikan individu yang sudah teramati atau tersensor tidak diamati lagi), yaitu :

$$\frac{h_0(t)\,exp(0)}{h_0(t)\,exp(0) + h_0(t)\,exp(0) + h_0(t)\,exp(\beta_1 + 3\delta_1)}$$

Karena $h_0(t)$-nya sama di waktu t=3 maka bisa dicoret, sehingga

$$\frac{exp(0)}{exp(0) + exp(0) + exp(\beta_1 + 3\delta_1)}$$

3. Heri tersensor pada waktu t=5 tahun. Karena likelihood yang dibentuk hanya mempertimbangkan individu yang mengalami *event*, Heri tidak diperhitungkan *likelihood*-nya.

4. Probabilitas Leri mengalami kejadian pada waktu t=8 tahun dimana dimana anggota himpunan resikonya yaitu Leri sendiri, (Geri dan Heri sudah keluar dari himpunan resiko karena diasumsikan individu yang sudah teramati atau tersensor tidak diamati lagi), yaitu :

$$\frac{h_0(t)\,exp(\beta_1 + 8\delta_1)}{h_0(t)\,exp(\beta_1 + 8\delta_1)}$$

Karena $h_0(t)$-nya sama di waktu t=8 maka bisa dicoret, sehingga

$$\frac{exp(\beta_1 + 8\delta_1)}{exp(\beta_1 + 8\delta_1)}$$

Seperti yang telah dibahas sebelumnya**,** *partial likelihood*-nya dapat dinyatakan dalam perkalian dari probabilitas Beri, Geri, dan Leri, yaitu :

$$L = \frac{exp(\beta_1 + 2\delta_1)}{exp(\beta_1 + 2\delta_1) + exp(0) + exp(0) + exp(\beta_1 + 2\delta_1)}$$
$$\times \frac{exp(0)}{exp(0) + exp(0) + exp(\beta_1 + 3\delta_1)} \times \frac{exp(\beta_1 + 8\delta_1)}{exp(\beta_1 + 8\delta_1)}$$

Berdasarkan contoh di atas, langkah-langkah untuk menentukan *partial likelihood* dari *extended cox model* yaitu :

1. Bentuk *extended cox model* yang dimaksud
2. Urutkan waktu kejadiannya
3. Konstruksi kontribusi masing-masing objek yang diamati pada himpunan *D(t_j)* terhadap fungsi partial likelihood. Sebut, individu Z pada waktu t_j dengan fungsi hazard $h_z(t_j, X(t_j))$, maka

$$L_{p_z} = \frac{h_z(t_j, X(t_j))}{\sum_{i \in R(t_{(j)})} h_i(t_j, X(t_j))}$$

dimana

L_{p_z} = Probabilitas bahwa individu A mengalami *event* di waktu t_j bersyarat bahwa terdapat suatu *event* di waktu t_j.

$h_z(t_j, X(t_j))$ = Pr(individu dengan *covariate X* mengalami *event* di waktu t_j)

$\sum_{i \in R(t_{(j)})} h_i(t_j, X(t_j))$ = Pr(terdapat suatu *event* di waktu t_j)

Mengacu pada persamaan (3.2),

$$L_{p_z} = \frac{h_0(t_j)\, exp\left[\sum_{a=1}^{p_1} \beta_{az}\boldsymbol{X}_{az} + \sum_{b=p_1+1}^{p_2} \beta_{bz}\boldsymbol{X}_{bz} + \sum_{b=p_1+1}^{p_2} \delta_{bz}\boldsymbol{X}_{bz}(t_j)\right]}{\sum_{i \in R(t_{(j)})} h_0(t_j)\, exp\left[\sum_{a=1}^{p_1} \beta_{ai}\boldsymbol{X}_{ai} + \sum_{b=p_1+1}^{p_2} \beta_{bi}\boldsymbol{X}_{bi} + \sum_{b=p_1+1}^{p_2} \delta_{bi}\boldsymbol{X}_{bi}(t_j)\right]}$$
$$= \frac{exp\left[\sum_{a=1}^{p_1} \beta_{az}\boldsymbol{X}_{az} + \sum_{b=p_1+1}^{p_2} \beta_{bz}\boldsymbol{X}_{bz} + \sum_{b=p_1+1}^{p_2} \delta_{bz}\boldsymbol{X}_{bz}(t_j)\right]}{\sum_{i \in R(t_{(j)})} exp\left[\sum_{a=1}^{p_1} \beta_{ai}\boldsymbol{X}_{ai} + \sum_{b=p_1+1}^{p_2} \beta_{bi}\boldsymbol{X}_{bi} + \sum_{b=p_1+1}^{p_2} \delta_{bi}\boldsymbol{X}_{bi}(t_j)\right]}$$

dengan,

$\boldsymbol{X}_{ai}$ = *Covariate* ke-a dari individu-i untuk *time independent covariate*

$\boldsymbol{X}_{bi}$ = *Covariate* ke-b dari individu-i untuk *time independent covariate*

$\boldsymbol{X}_{bi}(t_j)$ = *Covariate* ke-b dari individu-i pada waktu t_j untuk *time dependent covariate*

β_{ai} = Koefisien parameter dari *covariate* $\boldsymbol{X}_{ai}$

β_{bi} = Koefisien parameter dari *covariate* $\boldsymbol{X}_{bi}$

δ_{bi} = Koefisien parameter dari *covariate* $\boldsymbol{X}_{bi}(t_j)$

Proses yang sama dilakukan untuk objek lain yang menjadi anggota himpunan *D(tj)*.

4. *Partial likelihood* seluruhnya yaitu perkalian antara partial likelihood dari masing-masing objek yang teramati. Misalkan terdapat k objek yang menjadi anggota himpunan *D(tj)*

$$L_p(\boldsymbol{\beta}) = \prod_{j\epsilon D(t_j)} L_j = L_1 \times L_2 \times \ldots \times L_k$$

$$L_p(\boldsymbol{\beta}) = \prod_{j\epsilon D(t_j)} \frac{exp\left[\sum_{a=1}^{p_1}\beta_{aj}\boldsymbol{X}_{aj} + \sum_{b=p_1+1}^{p_2}\beta_{bj}\boldsymbol{X}_{bj} + \sum_{b=p_1+1}^{p_2}\delta_{bj}\boldsymbol{X}_{bj}(t_j)\right]}{\sum_{i\in R\left(t_{(j)}\right)} exp\left[\sum_{a=1}^{p_1}\beta_{ai}\boldsymbol{X}_{ai} + \sum_{b=p_1+1}^{p_2}\beta_{bi}\boldsymbol{X}_{bi} + \sum_{b=p_1+1}^{p_2}\delta_{bi}\boldsymbol{X}_{bi}(t_j)\right]}$$

Misalkan,

$$exp\left[\sum_{a=1}^{p_1}\beta_{aj}\boldsymbol{X}_{aj} + \sum_{b=p_1+1}^{p_2}\beta_{bj}\boldsymbol{X}_{bj} + \sum_{b=p_1+1}^{p_2}\delta_{bj}\boldsymbol{X}_{bj}(t_j)\right] = \psi_j$$

Sehingga, bentuk partial likelihood dapat disederhanakan menjadi :

$$L_p(\boldsymbol{\beta}) = \prod_{j\epsilon D(t_j)} \frac{\psi_j}{\sum_{i\epsilon R(t_j)}\psi_i}$$

5. Selanjutnya, mencari taksiran dari parameter-parameter dari model dengan metode *Maximum Partial Likelihood Estimator*.

- $\frac{\partial lnL_p}{\partial\beta_a} = 0,\ a = 1, 2, \ldots, p_1$

$$\frac{\partial}{\partial\beta_a}\left(ln\prod_{j\epsilon D(t_j)}\frac{\psi_j}{\sum_{i\epsilon R(t_j)}\psi_i}\right) = \frac{\partial}{\partial\beta_a}\left(\sum_{j\epsilon D(t_j)} ln\frac{\psi_j}{\sum_{i\epsilon R(t_j)}\psi_i}\right)$$

$$= \frac{\partial}{\partial\beta_a}\left(ln\frac{\psi_1}{\sum_{i\epsilon R(t_1)}\psi_i} + ln\frac{\psi_2}{\sum_{i\epsilon R(t_2)}\psi_i} + \cdots + ln\frac{\psi_d}{\sum_{i\epsilon R(t_d)}\psi_i}\right)$$

$$= \frac{\partial}{\partial \beta_a}\left(ln\,\psi_1 - ln \sum_{i\epsilon R(t_1)} \psi_i + ln\,\psi_2 - ln \sum_{i\epsilon R(t_2)} \psi_i + \cdots + ln\,\psi_d - ln \sum_{i\epsilon R(t_d)} \psi_i \right)$$

$$= \frac{\partial}{\partial \beta_a}\left(\sum_{j\epsilon D(t_j)} ln\,\psi_j - \sum_{j\epsilon D(t_j)} ln \sum_{i\epsilon R(t_j)} \psi_i \right)$$

$$= \frac{\partial}{\partial \beta_a} \sum_{j\epsilon D(t_j)} ln\,\psi_j - \frac{\partial}{\partial \beta_a} \sum_{j\epsilon D(t_j)} ln \sum_{i\epsilon R(t_j)} \psi_i$$

$$= 0$$

Misal, $\frac{\partial}{\partial \beta_a} \sum_{j\epsilon D(t_j)} \ln \psi_j$ ... (1) dan $\frac{\partial}{\partial \beta_a} \sum_{j\epsilon D(t_j)} ln \sum_{i\epsilon R(t_j)} \psi_i$ ... (2)

Untuk bagian (1), didapat,

$$\frac{\partial}{\partial \beta_a} \sum_{j\epsilon D(t_j)} ln\,\psi_j = \frac{\partial}{\partial \beta_a} \sum_{j\epsilon D(t_j)} ln\,exp\left[\sum_{a=1}^{p_1} \beta_{aj} X_{aj} + \sum_{b=p_1+1}^{p_2} \beta_{bj} X_{bj} + \sum_{b=p_1+1}^{p_2} \delta_{bj} X_{bj}(t_j) \right]$$

$$= \frac{\partial}{\partial \beta_a} \sum_{j\epsilon D(t_j)} \left[\sum_{a=1}^{p_1} \beta_{aj} X_{aj} + \sum_{b=p_1+1}^{p_2} \beta_{bj} X_{bj} + \sum_{b=p_1+1}^{p_2} \delta_{bj} X_{bj}(t_j) \right]$$

$$= \sum_{j\epsilon D(t_j)} \left[\sum_{a=1}^{p_1} X_{aj} + \sum_{b=p_1+1}^{p_2} \beta_{bj} X_{bj} + \sum_{b=p_1+1}^{p_2} \delta_{bj} X_{bj}(t_j) \right]$$

Untuk bagian (2), didapat,

$$\frac{\partial}{\partial \beta_a} \sum_{j\epsilon D(t_j)} ln \sum_{i\epsilon R(t_j)} \psi_i$$

$$= \frac{\partial}{\partial \beta_a} \sum_{j\epsilon D(t_j)} ln \sum_{i\epsilon R(t_j)} exp\left[\sum_{a=1}^{p_1} \beta_{ai} X_{ai} + \sum_{b=p_1+1}^{p_2} \beta_{bi} X_{bi} + \sum_{b=p_1+1}^{p_2} \delta_{bi} X_{bi}(t_j) \right]$$

$$= \sum_{j\epsilon D(t_j)} \left(\frac{1}{M} \sum_{i\epsilon R(t_j)} N \right)$$

dimana,

$$M = \sum_{i\epsilon R(t_j)} exp\left[\sum_{a=1}^{p_1} \beta_{ai} X_{ai} + \sum_{b=p_1+1}^{p_2} \beta_{bi} X_{bi} + \sum_{b=p_1+1}^{p_2} \delta_{bi} X_{bi}(t_j) \right]$$

$$N = \sum_{b=p_1+1}^{p_2} X_{ai}\, exp\left[\sum_{a=1}^{p_1} \beta_{ai} X_{ai} + \sum_{b=p_1+1}^{p_2} \beta_{bi} X_{bi} + \sum_{b=p_1+1}^{p_2} \delta_{bi} X_{bi}(t_j)\right]$$

Gabungan hasil turunan dari (1) dan (2) yaitu

$$= \frac{\partial}{\partial\beta_a} \sum_{j\epsilon D(t_j)} ln\,\psi_j - \frac{\partial}{\partial\beta_a} \sum_{j\epsilon D(t_j)} ln \sum_{i\epsilon R(t_j)} \psi_i$$

$$= \sum_{j\epsilon D(t_j)} \left[\sum_{a=1}^{p_1} X_{aj} + \sum_{b=p_1+1}^{p_2} \beta_{bj} X_{bj} + \sum_{b=p_1+1}^{p_2} \delta_{bj} X_{bj}(t_j)\right] - \sum_{j\epsilon D(t_j)} \left(\frac{1}{M} \sum_{i\epsilon R(t_j)} N\right) = 0 \quad (3.32)$$

- $\frac{\partial lnL_p}{\partial\beta_b} = 0,\ b = 1, 2, \dots, p_2$

$$\frac{\partial}{\partial\beta_b}\left(ln \prod_{j\epsilon D(t_j)} \frac{\psi_j}{\sum_{i\epsilon R(t_j)} \psi_i}\right) = \frac{\partial}{\partial\beta_a}\left(\sum_{j\epsilon D(t_j)} ln \frac{\psi_j}{\sum_{i\epsilon R(t_j)} \psi_i}\right)$$

$$= \frac{\partial}{\partial\beta_b}\left(ln \frac{\psi_1}{\sum_{i\epsilon R(t_1)} \psi_i} + ln \frac{\psi_2}{\sum_{i\epsilon R(t_2)} \psi_i} + \cdots + ln \frac{\psi_d}{\sum_{i\epsilon R(t_d)} \psi_i}\right)$$

$$= \frac{\partial}{\partial\beta_b}\left(ln\,\psi_1 - ln \sum_{i\epsilon R(t_1)} \psi_i + ln\,\psi_2 - ln \sum_{i\epsilon R(t_2)} \psi_i + \cdots + ln\,\psi_d - ln \sum_{i\epsilon R(t_d)} \psi_i\right)$$

$$= \frac{\partial}{\partial\beta_b}\left(\sum_{j\epsilon D(t_j)} ln\,\psi_j - \sum_{j\epsilon D(t_j)} ln \sum_{i\epsilon R(t_j)} \psi_i\right)$$

$$= \frac{\partial}{\partial\beta_b} \sum_{j\epsilon D(t_j)} ln\,\psi_j - \frac{\partial}{\partial\beta_b} \sum_{j\epsilon D(t_j)} ln \sum_{i\epsilon R(t_j)} \psi_i$$

$$= 0$$

Misal, $\frac{\partial}{\partial\beta_b} \sum_{j\epsilon D(t_j)} \ln\psi_j \dots (1)$ dan $\frac{\partial}{\partial\beta_b} \sum_{j\epsilon D(t_j)} ln \sum_{i\epsilon R(t_j)} \psi_i \dots (2)$

Untuk bagian (1), didapat,

$$\frac{\partial}{\partial\beta_b} \sum_{j\epsilon D(t_j)} ln\,\psi_j = \frac{\partial}{\partial\beta_b} \sum_{j\epsilon D(t_j)} ln\, exp\left[\sum_{a=1}^{p_1} \beta_{aj} X_{aj} + \sum_{b=p_1+1}^{p_2} \beta_{bj} X_{bj} + \sum_{b=p_1+1}^{p_2} \delta_{bj} X_{bj}(t_j)\right]$$

$$= \frac{\partial}{\partial \beta_b} \sum_{j \epsilon D(t_j)} \left[\sum_{a=1}^{p_1} \beta_{aj} X_{aj} + \sum_{b=p_1+1}^{p_2} \beta_{bj} X_{bj} + \sum_{b=p_1+1}^{p_2} \delta_{bj} X_{bj}(t_j) \right]$$

$$= \sum_{j \epsilon D(t_j)} \left[\sum_{a=1}^{p_1} \beta_{aj} X_{aj} + \sum_{b=p_1+1}^{p_2} X_{bj} + \sum_{b=p_1+1}^{p_2} \delta_{bj} X_{bj}(t_j) \right]$$

Untuk bagian (2), didapat,

$$\frac{\partial}{\partial \beta_b} \sum_{j \epsilon D(t_j)} ln \sum_{i \epsilon R(t_j)} \psi_i$$

$$= \frac{\partial}{\partial \beta_b} \sum_{j \epsilon D(t_j)} ln \sum_{i \epsilon R(t_j)} exp \left[\sum_{a=1}^{p_1} \beta_{ai} X_{ai} + \sum_{b=p_1+1}^{p_2} \beta_{bi} X_{bi} + \sum_{b=p_1+1}^{p_2} \delta_{bi} X_{bi}(t_j) \right]$$

$$= \sum_{j \epsilon D(t_j)} \left(\frac{1}{O} \sum_{i \epsilon R(t_j)} Y \right)$$

dimana,

$$O = \sum_{i \epsilon R(t_j)} exp \left[\sum_{a=1}^{p_1} \beta_{ai} X_{ai} + \sum_{b=p_1+1}^{p_2} \beta_{bi} X_{bi} + \sum_{b=p_1+1}^{p_2} \delta_{bi} X_{bi}(t_j) \right]$$

$$Y = \sum_{b=p_1+1}^{p_2} X_{bi}\, exp \left[\sum_{a=1}^{p_1} \beta_{ai} X_{ai} + \sum_{b=p_1+1}^{p_2} \beta_{bi} X_{bi} + \sum_{b=p_1+1}^{p_2} \delta_{bi} X_{bi}(t_j) \right]$$

Gabungan hasil turunan dari (1) dan (2) yaitu

$$= \frac{\partial}{\partial \beta_b} \sum_{j \epsilon D(t_j)} ln\, \psi_j - \frac{\partial}{\partial \beta_b} \sum_{j \epsilon D(t_j)} ln \sum_{i \epsilon R(t_j)} \psi_i$$

$$= \sum_{j \epsilon D(t_j)} \left[\sum_{a=1}^{p_1} \beta_{aj} X_{aj} + \sum_{b=p_1+1}^{p_2} X_{bj} + \sum_{b=p_1+1}^{p_2} \delta_{bj} X_{bj}(t_j) \right] - \sum_{j \epsilon D(t_j)} \left(\frac{1}{O} \sum_{i \epsilon R(t_j)} Y \right) = 0 \qquad (3.33)$$

- $\frac{\partial lnL_p}{\partial \delta_b} = 0,\ b = 1, 2, \ldots, p_2$

$$\frac{\partial}{\partial \delta_b} \left(\ln \prod_{j \epsilon D(t_j)} \frac{\psi_j}{\sum_{i \epsilon R(t_j)} \psi_i} \right) = \frac{\partial}{\partial \delta_b} \left(\sum_{j \epsilon D(t_j)} \ln \frac{\psi_j}{\sum_{i \epsilon R(t_j)} \psi_i} \right)$$

$$= \frac{\partial}{\partial \delta_b}\left(ln \frac{\psi_1}{\sum_{i\epsilon R(t_1)} \psi_i} + ln \frac{\psi_2}{\sum_{i\epsilon R(t_2)} \psi_i} + \cdots + ln \frac{\psi_d}{\sum_{i\epsilon R(t_d)} \psi_i} \right)$$

$$= \frac{\partial}{\partial \delta_b}\left(ln\, \psi_1 - ln \sum_{i\epsilon R(t_1)} \psi_i + ln\, \psi_2 - ln \sum_{i\epsilon R(t_2)} \psi_i + \cdots + ln\, \psi_d - ln \sum_{i\epsilon R(t_d)} \psi_i \right)$$

$$= \frac{\partial}{\partial \delta_b}\left(\sum_{j\epsilon D(t_j)} \ln \psi_j - \sum_{j\epsilon D(t_j)} ln \sum_{i\epsilon R(t_j)} \psi_i \right)$$

$$= \frac{\partial}{\partial \delta_b} \sum_{j\epsilon D(t_j)} \ln \psi_j - \frac{\partial}{\partial \delta_b} \sum_{j\epsilon D(t_j)} ln \sum_{i\epsilon R(t_j)} \psi_i$$

$$= 0$$

Misal, $\frac{\partial}{\partial \delta_b} \sum_{j\epsilon D(t_j)} \ln \psi_j \ldots (1)$ dan $\frac{\partial}{\partial \delta_b} \sum_{j\epsilon D(t_j)} ln \sum_{i\epsilon R(t_j)} \psi_i \ldots (2)$

Untuk bagian (1), didapat,

$$\frac{\partial}{\partial \delta_b} \sum_{j\epsilon D(t_j)} ln\, \psi_j = \frac{\partial}{\partial \delta_b} \sum_{j\epsilon D(t_j)} ln\, exp \left[\sum_{a=1}^{p_1} \beta_{aj} X_{aj} + \sum_{b=p_1+1}^{p_2} \beta_{bj} X_{bj} + \sum_{b=p_1+1}^{p_2} \delta_{bj} X_{bj}(t_j) \right]$$

$$= \frac{\partial}{\partial \delta_b} \sum_{j\epsilon D(t_j)} \left[\sum_{a=1}^{p_1} \beta_{aj} X_{aj} + \sum_{b=p_1+1}^{p_2} \beta_{bj} X_{bj} + \sum_{b=p_1+1}^{p_2} \delta_{bj} X_{bj}(t_j) \right]$$

$$= \sum_{j\epsilon D(t_j)} \left[\sum_{a=1}^{p_1} \beta_{aj} X_{aj} + \sum_{b=p_1+1}^{p_2} \beta_{bj} X_{bj} + \sum_{b=p_1+1}^{p_2} X_{bj}(t_j) \right]$$

Untuk bagian (2), didapat,

$$\frac{\partial}{\partial \delta_b} \sum_{j\epsilon D(t_j)} ln \sum_{i\epsilon R(t_j)} \psi_i$$

$$= \frac{\partial}{\partial \delta_b} \sum_{j\epsilon D(t_j)} ln \sum_{i\epsilon R(t_j)} exp \left[\sum_{a=1}^{p_1} \beta_{ai} X_{ai} + \sum_{b=p_1+1}^{p_2} \beta_{bi} X_{bi} + \sum_{b=p_1+1}^{p_2} \delta_{bi} X_{bi}(t_j) \right]$$

$$= \sum_{j\epsilon D(t_j)} \left(\frac{1}{F} \sum_{i\epsilon R(t_j)} G \right)$$

dimana,

$$F = \sum_{i \epsilon R(t_j)} exp\left[\sum_{a=1}^{p_1} \beta_{ai}X_{ai} + \sum_{b=p_1+1}^{p_2} \beta_{bi}X_{bi} + \sum_{b=p_1+1}^{p_2} \delta_{bi}X_{bi}(t_j)\right]$$

$$G = \sum_{b=p_1+1}^{p_2} X_{bi}(t_j)\, exp\left[\sum_{a=1}^{p_1} \beta_{ai}X_{ai} + \sum_{b=p_1+1}^{p_2} \beta_{bi}X_{bi} + \sum_{b=p_1+1}^{p_2} \delta_{bi}X_{bi}(t_j)\right]$$

Gabungan hasil turunan dari (1) dan (2) yaitu

$$= \frac{\partial}{\partial \delta_b} \sum_{j \epsilon D(t_j)} ln\, \psi_j - \frac{\partial}{\partial \delta_b} \sum_{j \epsilon D(t_j)} ln \sum_{i \epsilon R(t_j)} \psi_i$$

$$= \sum_{j \epsilon D(t_j)} \left[\sum_{a=1}^{p_1} \beta_{aj}X_{aj} + \sum_{b=p_1+1}^{p_2} \beta_{bj}X_{bj} + \sum_{b=p_1+1}^{p_2} X_{bj}(t_j)\right] - \sum_{j \epsilon D(t_j)} \left(\frac{1}{F} \sum_{i \epsilon R(t_j)} G\right) = 0 \qquad (3.34)$$

Persamaan (3.32), (3.33), dan (3.34) diselesaikan dengan iterasi numerik. Robbin (1989) dan Tsiatis (1990) telah menunjukkan bahwa taksiran yang didapat bersifat *asymptotically* normal, yaitu :

$$\widehat{\boldsymbol{\beta}} \sim N(\beta, (E[I(\beta)])^{-1})$$

dimana

$$I(\beta) = \begin{bmatrix} -\frac{\partial^2 I_p(\beta)}{\partial \beta_1^2} & \cdots & -\frac{\partial^2 I_p(\beta)}{\partial \beta_1 \partial \beta_p} \\ \vdots & \ddots & \vdots \\ -\frac{\partial^2 I_p(\beta)}{\partial \beta_p \partial \beta_1} & \cdots & -\frac{\partial^2 I_p(\beta)}{\partial \beta_p^2} \end{bmatrix}$$

dengan

$$I_p(\beta) = \sum_{j \epsilon D(t_j)} \left(x_j^T \beta - log \sum_{i \epsilon R(t_j)} e^{x_i^T \beta} \right)$$

3.3.2 Bentuk Fungsi *Partial Likelihood* jika Terdapat *Ties*

Pada subbab 3.4, *partial likelihood* yang dibentuk digunakan untuk data yang tidak ada *ties*, namun, adakalanya suatu data mengandung *ties*. Pada subbab ini, akan dibahas partial likelihood untuk data yang melibatkan *ties*.

Jika suatu data terdapat *ties*, maka akan menimbulkan permasalahan dalam membentuk partial likelihoodnya yaitu saat menentukan anggota dari himpunan resikonya. Untuk lebih jelas, perhatikan ilustrasi berikut :

Misalkan $t_1 < t_2 < \cdots < t_5$ menyatakan waktu yang teramati yang telah diurutkan. Berikut data mengenai objek yang mengalami event ataupun tersensor pada waktu ke-t_j :

Tabel 3.25. Data *survival time* dengan terdapat *ties* dan sensor

j	1	2	3	4	5
t_j	1	2+	4	4	5+

Pada waktu t=4, terdapat dua objek yang mengalami *event*. Tidak diketahui objek yang mana yang mengalami *event* terlebih dahulu.

Agar lebih sederhana penulisannya, fungsi hazard objek ke-j dinotasikan dengan ψ_j.

1. Misalkan dianggap objek ke-3 yang mengalami *event* terlebih dahulu, maka *partial likelihood*-nya yaitu :

$$L_{p1}(\boldsymbol{\beta}) = \frac{\psi_1}{(\psi_1+\psi_2+\psi_3+\psi_4+\psi_5)} \frac{\psi_3}{(\psi_3+\psi_4+\psi_5)} \frac{\psi_4}{(\psi_4+\psi_5)}$$

2. Bila dianggap objek ke-4 yang mengalami *event* terlebih dahulu, maka *partial likelihood*-nya yaitu :

$$L_{p2}(\boldsymbol{\beta}) = \frac{\psi_1}{(\psi_1+\psi_2+\psi_3+\psi_4+\psi_5)} \frac{\psi_4}{(\psi_3+\psi_4+\psi_5)} \frac{\psi_3}{(\psi_3+\psi_5)}$$

Nilai $\psi_3 \neq \psi_4$ yang menyebabkan perbedaan pada himpunan resikonya sehingga $L_{p1}(\boldsymbol{\beta}) \neq L_{p2}(\boldsymbol{\beta})$.

Partial likelihood yang diberikan ketika menganggap objek ke-3 yang terjadi lebih dahulu akan berbeda nilainya bila menganggap objek ke-4 yang terjadi lebih dahulu padahal untuk data yang sama seharusnya memberikan nilai yang sama. Inilah kenapa urutan sangat penting dalam pembentukan partial likelihood.

Untuk mengatasi permasalahan tersebut, dapat digunakan beberapa pendekatan, salah satunya yang akan dibahas pada subbab ini yaitu pendekatan Efron (1977).

Pada pendekatan Efron, himpunan resikonya diselesaikan dengan pengurangan terhadap rata-rata dari nilai ψ_3 dan ψ_4 karena tidak diketahui objek mana yang mengalami *event* terlebih dahulu. Bentuk partial likelihoodnya menjadi,

$$L_p(\boldsymbol{\beta}) \approx \frac{\psi_1}{(\psi_1+\psi_2+\psi_3+\psi_4+\psi_5)} \frac{\psi_3}{(\psi_3+\psi_4+\psi_5)} \frac{\psi_4}{\left(\psi_3+\psi_4+\psi_5-1/2(\psi_3+\psi_4)\right)}$$

Secara umum, partial likelihood untuk pendekatan Efron yaitu :

$$L_p(\boldsymbol{\beta}) \approx \prod_{j \epsilon D(t_j)} \frac{\prod_{i\epsilon D(t_{(j)})} \psi_i}{\prod_{k=1}^{d_j} \left[\sum_{i\epsilon R(t_{(j)})} \psi_i - {}^{(k-1)}/_{d_k} \left(\sum_{i\epsilon D(t_{(j)})} \psi_i\right)\right]}$$

dimana,

$D(t_j)$ = himpunan dari objek yang teramati

$R(t_j)$ = himpunan resiko

k = objek yang teramati

d_j = banyaknya pengamatan yang ties yang teramati pada waktu t_j.

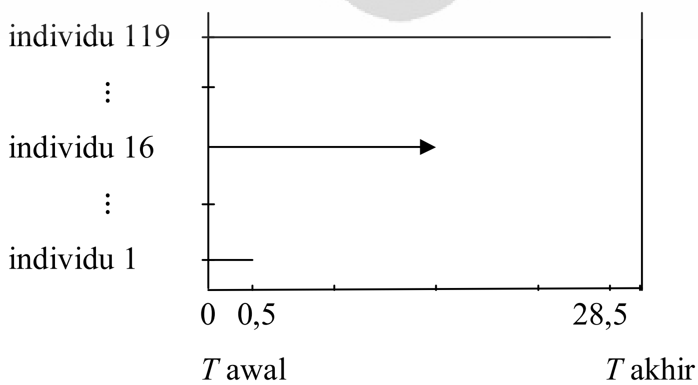
BAB 4
CONTOH PENERAPAN

Dalam bab ini, akan dibahas penerapan *extended cox model* pada data. Dalam pembentukan *extended cox model* akan digunakan dua fungsi waktu, yaitu log t dan fungsi *heaviside.* Kemudian akan dijelaskan interpretasi dari taksiran yang telah diperoleh. Selain itu, akan dibentuk juga model cox *proportional hazard*. Dengan uji ratio likelihood, *extended cox model* akan dibandingkan dengan model cox *proportional hazard* untuk melihat *extended cox model* atau model cox *proportional hazard* yang sesuai dengan data. Pengolahan data untuk mencari taksiran parameter dan nilai log partial likelihood dari *extended cox model* ataupun model cox *proportional hazard* diperoleh dengan perangkat lunak R versi 2.9.1 (free download).

4.1 Deskripsi Data

Data yang digunakan dalam tugas akhir ini adalah data yang diteliti oleh Nahman et al pada tahun 1992. Data ini mengenai waktu sampai seseorang yang terkena gagal ginjal pertama kali mengalami infeksi setelah pasien tersebut diberikan perlakuan (dalam bulan). Perlakuan yang diberikan yaitu pemasangan *catheter* pada ginjal pasien, sebut X_1. Perlakuan ini dibagi menjadi dua kategori, yaitu :

1. Pemasangan *catheter* yang ditempatkan dengan cara "surgically" (X_1=0)
2. Pemasangan *catheter* yang ditempatkan dengan cara "percutaneous" (X_1=1)



Gambar 4.1 Koordinat survival time dengan 119 individu

- T awal = waktu pemasangan *catheter* pada pasien
- *Event* yang terjadi = munculnya infeksi pertama kali
- T akhir = waktu saat pasien mengalami *event* atau saat penelitian berakhir
- Variabel respon (T) = waktu yang diperlukan sampai pasien tersebut mengalami *event*
- *Covariate* (X_1) = pemasangan *catheter* pada ginjal pasien.

  Dibagi menjadi dua kategori, yaitu :

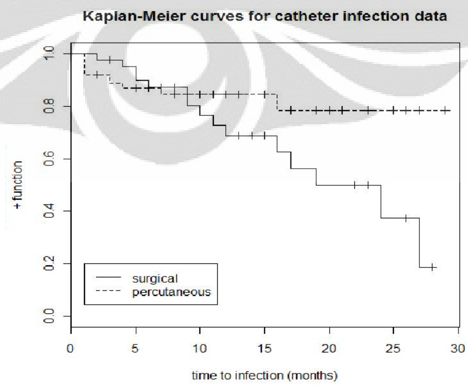
  1. X_1=0, untuk pemasangan *catheter* yang ditempatkan dengan cara "surgically"
  2. X_1=1, untuk pemasangan *catheter* yang ditempatkan dengan cara "percutaneous"

Data ini terdiri dari 119 pasien, dimana 43 pasien untuk X_1=0 dan 76 pasien untuk X_1=1. Dari ke-119 pasien, 26 pasien teramati, yaitu pasien yang mengalami kejadian sebelum masa penelitian berakhir, dan 93 pasien tersensor, yaitu tidak mengalami kejadian sampai masa penelitian berakhir atau pada saat penelitian, pasien menghilang, meninggal, atau lainnya. Hal ini dinyatakan dengan,

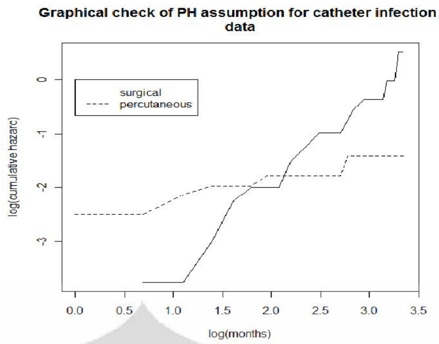
$$\text{status} = \begin{cases} 1 \text{ untuk pasien yang teramati} \\ 0 \text{ untuk pasien yang tersensor} \end{cases}$$

Data tersebut dapat dilihat pada lampiran 1.

Plot kurva survival dari data tersebut yaitu,



Gambar 4.2 Kurva Kaplan-Meier



Gambar 4.3 Grafik pengecekan asumsi PH

Dari plot di atas, terlihat bahwa kurva antara X_1=1 dan X_1=0 memotong, sehingga dicurigai bahwa asumsi proportional hazard dilanggar, artinya perbandingan antara hazard rate untuk X_1=1 dengan X_1=0 tidak konstan sepanjang waktu.

Untuk itu, akan dibentuk *extended cox model* untuk data tersebut. *Extended cox model* dengan satu *covariate* yang tidak memenuhi asumsi *proportional hazard* yaitu

$$h(t, \boldsymbol{X}) = h_0(t)\, exp(\beta_1 X_1 + \delta_1 X_1 g(t))$$

4.1.1 *Extended cox model* jika fungsi waktu yang digunakan yaitu *g(t)*=log t

Extended cox model-nya yaitu :

$$h(t, \boldsymbol{X}) = h_0(t)\, exp(\beta_1 X_1 + \delta_1 X_1\, log\, t)$$

Hazard rate untuk pemasangan *catheter* dengan cara percutaneous (X_1=1), sebut $h_p(t, \boldsymbol{X})$, yaitu :

$$h_p(t, 1) = h_0(t)\, exp(\beta_1 + \delta_1\, log\, t)$$

Sedangkan *hazard rate* untuk pemasangan *catheter* dengan cara surgical (X_1=0), sebut $h_s(t, \boldsymbol{X})$, yaitu :

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

Hazard ratio antara pemasangan *catheter* dengan cara percutaneous dan dengan cara surgical, yaitu :

$$\widehat{HR}(t) = \frac{h_p(t)}{h_s(t)} = \frac{h_0(t)\, exp(\beta_1 + \delta_1\, log\, t)}{h_0(t)} = exp(\hat{\beta}_1 + \hat{\delta}_1\, log\, t)$$

Dari perangkat lunak R versi 2.9.1 diperoleh :

Call:

coxph(formula = Surv(start.int1, stop.int1, event1) ~ cathtype1 + logt.cathtype1, method = "efron")

Tabel 4.1. Ringkasan untuk *extended cox model* dengan g(t) = log t

	coef	exp(coef)	se(coef)	Z	P
cathtype1	2.09	8.088	1.189	1.76	0.079
logt.cathtype1	-1.73	0.178	0.672	-2.57	0.010

Likelihood ratio test=14.5 on 2 df, p=0.000702 n= 1132

Log Likelihood = -96.97038

Dari tabel di atas, didapat nilai $\hat{\beta}_1 = 2.09$, sedangkan nilai $\hat{\delta}_1 = -1.73$ sehingga, *extended cox model*-nya yaitu :

$$h(t, \boldsymbol{X}) = h_0(t)\, exp(2.09X_1 - 1.73X_1\, log\, t)$$

dan *Hazard ratio*-nya yaitu :

$$\widehat{HR}(t) = \frac{h_p(t)}{h_s(t)} = exp(2.09 - 1.73\, log\, t)$$

Jadi, diperoleh bahwa $h_p(t) = exp(2.09 - 1.73\, log\, t)\, h_s(t)$. Dapat dinyatakan bahwa, pada waktu *t*, *hazard rate* untuk pemasangan *catheter* secara percutaneous sebesar $exp(2.09 - 1.73\, log\, t)$ kali *hazard rate* untuk pemasangan *catheter* secara surgical.

Artinya, semakin besar nilai *t*, semakin kecil nilai *hazard ratio*-nya, sehingga terdapat waktu tertentu dimana nilai perbandingannya berbanding terbalik. Misalkan,

1. Saat *t*=1, nilai *hazard ratio*-nya yaitu :

$$\widehat{HR}(t) = \frac{h_p(t)}{h_s(t)} = exp(2.09 - 1.73\, log\, 1) = exp(2.09) = 8.088$$

$$h_p(t) = 8{,}088\, h_s(t)$$

Artinya, nilai *hazard rate* dari pasien dengan pemasangan *catheter* secara percutaneous 8.088 kali nilai *hazard rate* dari pasien dengan pemasangan *catheter* secara surgical.

2. Saat *t*=100, nilai *hazard ratio*-nya yaitu :

$$\widehat{HR}(t) = \frac{h_p(t)}{h_s(t)} = exp(2.09 - 1.73\, log\, 100) = exp(-1.37) = 0.254$$

$$h_p(t) = 0{,}254 h_s(t)$$

Artinya, nilai *hazard rate* dari pasien dengan pemasangan *catheter* secara percutaneous 0.254 kali nilai *hazard rate* dari pasien dengan pemasangan *catheter* secara surgical.

Hal ini menunjukkan bahwa nilai *hazard ratio*-nya berubah secara signifikan pada titik waktu yang berlainan.

Pada subbab selanjutnya, akan dibahas *extended cox model* jika fungsi waktu yang digunakan adalah fungsi *heaviside*.

4.1.2 *Extended cox model* jika fungsi waktu yang digunakan yaitu fungsi *heaviside*

Dengan fungsi *heaviside,* nilai dari *hazard ratio*-nya berubah hanya pada waktu tertentu saja. Pada interval waktu tertentu, *hazard ratio* bernilai konstan, tetapi antar selang waktu, *hazard ratio*-nya berbeda.

Dalam kasus ini, akan dibagi menjadi dua selang waktu. Misalkan waktu yang membagi kedua selang waktu yaitu τ. Maka, model yang terbentuk yaitu :

$$h(t, \boldsymbol{X}) = h_0(t)\, exp(\beta_1 X_1 + \delta_1 X_1 g(t))$$

dimana,

$$g(t) = \begin{cases} 0\ untuk\ t \leq \tau \\ 1\ untuk\ t > \tau \end{cases}$$

Pertama-tama, akan dilihat nilai log *partial likelihood* dari masing-masing waktu yang teramati untuk mendapatkan waktu yang memisahkan kedua interval tersebut, τ. Secara algoritma, untuk mendapatkan nilai log *partial likelihood*, diawali dengan langkah-langkah sebagai berikut :

1. Daftarkan waktu-waktu saat pasien mengalami *event*
2. Urutkan waktu-waktu tersebut dari yang kecil hingga yang besar
3. Hitung nilai log partial likelihood dari masing-masing pasien dari yang terlebih dahulu mengalami *event*

Dari perangkat lunak R versi 2.9.1 diperoleh :

- cbind(event.times,logliks)

Tabel 4.2. Nilai loglikelihood dari masing-masing waktu yang teramati

	Event.times	Logliks
[1,]	0,5	-97,56271
[2,]	1,5	-99,93807
[3,]	2,5	-97,31866
[4,]	3,5	-97,20138
[5,]	4,5	-99,42334
[6,]	5,5	-100,24255
[7,]	6,5	-98,59629
[8,]	8,5	-100,20066
[9,]	9,5	-100,86155
[10,]	10,5	-101,44951
[11,]	11,5	-101,95374
[12,]	15,5	-100,61801
[13,]	16,5	-101,26864
[14,]	18,5	-101,85368
[15,]	23,5	-102,41635
[16,]	26,5	-103,02781

Dari tabel di atas terlihat bahwa, pada waktu t=3,5, log partial likelihoodnya bernilai maksimum, sehingga saat t=3,5, dapat dipilih sebagai waktu yang memisahkan kedua interval yaitu τ = 3,5.

Model yang terbentuk pada masing-masing interval sebagai berikut :

Tabel 4.3. Bentuk *extended cox model* pada masing-masing interval waktu

Interval waktu	*Hazard rate*
$t \leq 3.5$	$h(t,X) = h_0(t)\, exp(\beta_1 X_1)$
$t > 3.5$	$h(t,X) = h_0(t)\, exp((\beta_1 + \delta_1)X_1)$

Hazard rate untuk pemasangan *catheter* secara percutaneous, X_1=1, dan *hazard rate* untuk pemasangan *catheter* secara surgical, X_1=0, yaitu :

Tabel 4.4. Bentuk *extended cox model* dengan *covariate*-nya telah diubah

Interval waktu	*Hazard rate* saat X_1=1	*Hazard rate* saat X_1=0
$t \leq 3.5$	$h_p(t,1) = h_0(t)\, exp(\beta_1)$	$h_s(t,0) = h_0(t)\, exp(0)$
$t > 3.5$	$h_p(t,1) = h_0(t)\, exp(\beta_1 + \delta_1)$	$h_s(t,0) = h_0(t)\, exp(0)$

Hazard ratio antara pemasangan *catheter* secara percutaneous dan secara surgical, yaitu :

Tabel 4.5. Taksiran *hazard ratio* pada masing-masing interval

Interval waktu	Taksiran *Hazard ratio*
$t \leq 3.5$	$\widehat{HR}(t) = exp(\hat{\beta}_1)$
$t > 3.5$	$\widehat{HR}(t) = exp(\hat{\beta}_1 + \hat{\delta}_1)$

Dari perangkat lunak R versi 2.9.1, diperoleh :

Call:

coxph(formula = Surv(start.int1, stop.int1, event1) ~ cathtype1 + t.cathtype2, method = "efron")

n= 1132

Tabel 4.6. Ringkasan untuk *extended cox model* dengan g(t) yaitu fungsi *heaviside*

	Coef	exp (coef)	Se (coef)	Z	Pr(>|z|)	Exp (-coef)	Lower .95	Upper .95
Cathtype1	1.10023	3.0049	0.7829	1.405	0.15995	0.3328	0.6477	13.940
t.cathtyp2	-3.19436	0.0409	1.0909	-2.928	0.00341	24.395	0.0048	0.3478

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Rsquare= 0.012 (max possible= 0.168)
Likelihood ratio test= 14.06 on 2 df, p=0.0008845

Dari tabel di atas, terlihat bahwa, nilai $\hat{\beta}_1 = 1.10023$, sedangkan nilai $\hat{\delta}_1 = -3.19436$. Sehingga, *hazard ratio*-nya yaitu :

Tabel 4.7. Nilai taksiran *hazard ratio* pada masing-masing interval

Interval waktu	Taksiran *Hazard ratio*
$t \leq 3.5$	$\widehat{HR}(t) = exp(1.10023) = 3.0049$
$t > 3.5$	$\widehat{HR}(t) = exp(1.10023 - 3.19436) = exp(-2.09413) = 0.1232$

Interpretasi dari *hazard ratio* tersebut adalah :

1. Pada saat $t \leq 3.5$, $\widehat{HR}(t) = \frac{h_p(t)}{h_s(t)} = 3.0049$, *hazard rate* untuk pemasangan *catheter* secara percutaneous adalah 3.0049 kali dari *hazard rate* untuk pemasangan *catheter* secara surgical, artinya pasien dengan pemasangan *catheter* secara percutaneous lebih cepat mengalami infeksi dibandingkan dengan pasien dengan pemasangan *catheter* secara surgical.

   Jadi, pemasangan *catheter* secara surgical lebih baik dibandingkan dengan pemasangan *catheter* secara percutaneous.

2. Pada saat $t > 3.5$, $\widehat{HR}(t) = \frac{h_p(t)}{h_s(t)} = 0.1232$, *hazard rate* untuk pemasangan *catheter* secara percutaneous adalah 0.1232 kali dari *hazard rate* untuk pemasangan *catheter* secara surgical, artinya pasien dengan pemasangan *catheter* secara surgical lebih cepat mengalami infeksi dibandingkan dengan pasien dengan pemasangan *catheter* secara percutaneous.

   Jadi, pemasangan *catheter* secara percutaneous lebih baik dibandingkan dengan pemasangan *catheter* secara surgical.

4.1.3 Model cox *proportional hazard* untuk data

Jika dimisalkan bahwa *covariate* X_1 memenuhi asumsi proportional hazard, maka model yang terbentuk yaitu :

$$h(t, X) = h_0(t)\, exp(\beta_1 X_1)$$

Hazard rate untuk pemasangan *catheter* secara percutaneous, X_1=1, yaitu :

$$h_p(t,1) = h_0(t)\, exp(\beta_1)$$

Sedangkan *hazard rate* untuk pemasangan *catheter* secara surgical, X_1=0, yaitu :

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

Hazard ratio antara pemasangan *catheter* secara percutaneous dan secara surgical, yaitu :

$$\widehat{HR} = \frac{h_p(t)}{h_s(t)} = \frac{h_0(t)\, exp(\beta_1)}{h_0(t)} = exp(\hat{\beta}_1)$$

Dari perangkat lunak R versi 2.9.1 diperoleh :

Call:

coxph(formula = Surv(smonths, event) ~ cathtype, method = "efron")

Tabel 4.8. Ringkasan untuk model cox *proportional hazard*

	Coef	exp(coef)	se(coef)	Z	P
Cathtype	-0.613	0.542	0.398	-1.54	0.12

Likelihood ratio test=2.41 on 1 df, p=0.121 n= 119
Log Likelihood = -103.0278

Dari tabel di atas, terlihat bahwa, nilai $\hat{\beta}_1 = -0.613$, sehingga, model cox *proportional hazard*-nya yaitu :

$$h(t) = h_0(t)\, exp(-0.613X_1)$$

dan *Hazard ratio*-nya yaitu :

$$\widehat{HR} = \frac{h_p(t)}{h_s(t)} = exp(-0.613) = 0.542$$

Jadi, diperoleh bahwa $h_p(t) = 0.542\, h_s(t)$. Dapat dinyatakan bahwa *hazard rate* untuk pemasangan *catheter* secara percutaneous sebesar 0.542 kali *hazard rate* untuk pemasangan *catheter* secara surgical.

Artinya, untuk waktu berapapun, pasien dengan pemasangan *catheter* secara surgical selalu lebih cepat mengalami infeksi dibandingkan dengan pasien dengan pemasangan *catheter* secara percutaneous, atau pemasangan *catheter* secara percutaneous selalu lebih baik dibandingkan dengan pemasangan *catheter* secara surgical.

Hal ini tentu saja tidak sesuai jika dilihat berdasarkan grafik pada subbab sebelumnya dimana terjadi perpotongan. Untuk lebih menjamin bahwa *extended cox model* adalah model yang sesuai untuk data, dilakukan pengujian yang akan dijelaskan pada subbab selanjutnya.

4.1.4 Perbandingan antara *extended cox model* dengan model cox *proportional hazard*

Pada subbab ini, akan dilihat model mana yang sesuai untuk data tersebut, *extended cox model* ataukah model cox *proportional hazard*. Untuk menguji model mana yang sesuai, digunakan uji ratio likelihood. Dalam menggunakan uji ratio likelihood, perlu dihitung perbedaan antara *Log Likelihood* dari model cox *proportional hazard* (model yang tidak mengandung interaksi X_1 dengan $g(t)$), sebut lnL_R, dengan *Log Likelihood* dari *extended cox model* (model yang mengandung interaksi X_1 dengan $g(t)$), sebut lnL_F. Secara umum, uji *ratio likelihood* dapat ditulis dalam bentuk (-2 $ln\ L_R$) - (-2 $ln\ L_F$).

Bentuk model cox *proportional hazard* yaitu

$$h(t, X) = h_0(t)\, exp(\beta_1 X_1)$$

sedangkan, bentuk *extended cox model* yaitu

$$h(t, \boldsymbol{X}) = h_0(t)\, exp(\beta_1 X_1 + \delta_1 X_1 \log t)$$

sehingga, hipotesis yang diuji yaitu :

$H_0 : \delta_1 = 0$

$H_1 : \delta_1 \neq 0$

Dengan tingkat kepercayaan 95%, aturan keputusan dari hipotesis di atas yaitu H_0 ditolak jika nilai $p < \alpha = 0.05$, atau nilai *ratio likelihood* lebih besar daripada nilai persentil dari distribusi chi-square dengan 1 derajat bebas.

Dari model cox *proportional hazard*, diperoleh *Log Likelihood* ($ln\ L_R$) sebesar -103.0278. Sedangkan dari *extended cox model*, diperoleh *Log Likelihood* ($ln\ L_F$) sebesar -96.97038. Jadi, uji *ratio likelihood*-nya yaitu (-2.(-103.0278)) - (-2.(-96.97038)) = 206.0556 – 193.94076 = 12.11484, dengan nilai $p = 0.0005$ (diperoleh dari perangkat lunak R versi 2.9.1). Dari tabel distribusi *chi-square*, diperoleh nilai persentil dari distribusi *chi-square* dengan 1 derajat bebas yaitu 3.84.

Karena nilai *ratio likelihood* = 12.11484 > 3.84, yaitu nilai persentil dari distribusi chi-square dengan 1 derajat bebas, atau nilai $p = 0.0005 < 0.05 = \alpha$, maka H_0 ditolak. Artinya, dengan tingkat kepercayaan 95%, dipercaya bahwa $\delta_1 \neq 0$. Ini menunjukkan bahwa asumsi *proportional hazard* dilanggar, sehingga model yang sesuai untuk data tersebut yaitu *extended cox model*.

BAB 5
PENUTUP

5.1 Kesimpulan

Berdasarkan pembahasan pada bab-bab sebelumnya, dapat disimpulkan bahwa:

1. Jika terdapat *time-independent covariate* yang tidak memenuhi asumsi *proportional hazard* pada model *cox proportional hazard*, maka dapat diatasi dengan menggunakan *extended cox model*. Hal ini dapat dilakukan dengan menginteraksikan *time-independent covariate* yang tidak memenuhi asumsi *proportional hazard* dengan suatu fungsi waktu. *Time-independent covariate* yang diinteraksikan dengan fungsi waktu tersebut disebut *time-dependent covariate*, sehingga akan didapat *hazard ratio* yang bergantung pada waktu
2. Penaksiran parameter *extended cox model* dilakukan dengan metode *maximum partial likelihood*, dimana hanya objek yang mengalami event yang berkontribusi secara langsung terhadap likelihood, sementara kontribusi dari objek yang tersensor dapat dilihat pada himpunan resiko di tiap waktu yang teramati
3. Dalam contoh penerapan, digambarkan persoalan tentang bagaimana pengaruh dari variabel *treatment,* yaitu pemasangan *catheter* secara "*surgical*" dan secara "*percutaneous*" terhadap waktu sampai pasien mengalami infeksi pertama kali setelah dilakukan transplantasi ginjal. Dari hasil analisis diperoleh bahwa *extended cox model* dapat menjelaskan dengan lebih baik daripada model *cox proportional hazard*

5.2 Saran

Saran untuk pengembangan skripsi ini adalah

1. Dapat dijelaskan mengenai bentuk *extended cox model* dengan *covariate* yang diperhatikan yaitu *covariate* yang nilainya memang bergantung pada waktu
2. Untuk kasus data yang terdapat *ties*, dapat diterapkan pendekatan Breslow dan pendekatan Diskrit
3. Dapat dijelaskan bagaimana penentuan fungsi waktu yang tepat untuk *extended cox model* yang digunakan

DAFTAR PUSTAKA

1. *Dorota M, Dabrowska. 1987. Non-parametric Regression with Censored Survival Time Data.http://www.jstor.org/*
2. James, Robins. and Anastasios, A. Tsiatis. 1992. *Semiparametric Estimation of an Accelerated Failure Time Model with Time-Dependent Covariate*. Great Britain : Biometrika
3. Rober V. Hogg and Allen T. Craig. 1995. *Introduction to Mathematical Statistics*. New Jersey: Prentice-Hall
4. Klein, J.P. and Moeschberger, M.L. 1997. *Survival Analysis – Techniques for Censored and Truncated Data.* New York: Springer-Verlag
5. Daowen, Zhang. 2005. *Modelling Survival Data with Paranetric Regression Model. http://www.ncbi.nlm.nih.gov/pubmed/1480879*
6. Kleinbaum, D.G. and Klein, M. 2005. *Survival Analysis – A Self Learning Text, Second Edition.* New York: Springer

Lampiran 1

Data yang digunakan untuk contoh penerapan pada bab 4

Pid	smonths	Event	Cathtype	Pid	smonths	event	Cathtype
1	0,5	1	0	61	0,5	0	1
2	0,5	1	0	62	0,5	0	1
3	0,5	1	0	63	0,5	0	1
4	0,5	1	0	64	0,5	0	1
5	0,5	1	0	65	1,5	0	1
6	0,5	1	0	66	1,5	0	1
7	0,5	1	0	67	1,5	0	1
8	0,5	1	0	68	1,5	0	1
9	0,5	1	0	69	2,5	0	1
10	0,5	1	0	70	2,5	0	1
11	0,5	1	0	71	2,5	0	1
12	0,5	1	0	72	2,5	0	1
13	0,5	1	0	73	2,5	0	1
14	0,5	1	0	74	3,5	0	1
15	0,5	1	0	75	3,5	0	1
16	0,5	0	0	76	3,5	0	1
17	0,5	0	0	77	3,5	0	1
18	0,5	0	0	78	3,5	0	1
19	0,5	0	0	79	4,5	0	1
20	0,5	0	0	80	4,5	0	1
21	0,5	0	0	81	4,5	0	1
22	0,5	0	0	82	5,5	0	1
23	0,5	0	0	83	5,5	0	1
24	0,5	0	0	84	5,5	0	1
25	0,5	0	0	85	5,5	0	1
26	0,5	0	0	86	5,5	0	1
27	0,5	0	0	87	6,5	0	1
28	0,5	0	0	88	7,5	0	1
29	0,5	0	0	89	7,5	0	1
30	0,5	0	0	90	7,5	0	1
31	0,5	0	0	91	8,5	0	1
32	0,5	0	0	92	8,5	0	1
33	0,5	0	0	93	8,5	0	1
34	0,5	0	0	94	9,5	0	1
35	0,5	0	0	95	9,5	0	1

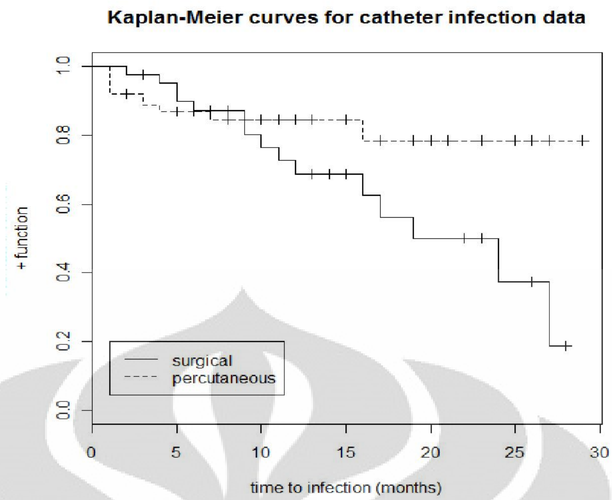
36	0,5	0	0	96	10,5	0	1
37	0,5	0	0	97	10,5	0	1
38	0,5	0	0	98	10,5	0	1
39	0,5	0	0	99	11,5	0	1
40	0,5	0	0	100	11,5	0	1
41	0,5	0	0	101	12,5	0	1
42	0,5	0	0	102	12,5	0	1
43	0,5	0	0	103	12,5	0	1
44	0,5	1	1	104	12,5	0	1
45	0,5	1	1	105	14,5	0	1
46	0,5	1	1	106	14,5	0	1
47	0,5	1	1	107	16,5	0	1
48	0,5	1	1	108	16,5	0	1
49	0,5	1	1	109	18,5	0	1
50	0,5	1	1	110	19,5	0	1
51	0,5	1	1	111	19,5	0	1
52	0,5	1	1	112	19,5	0	1
53	0,5	1	1	113	20,5	0	1
54	0,5	1	1	114	22,5	0	1
55	0,5	0	1	115	24,5	0	1
56	0,5	0	1	116	25,5	0	1
57	0,5	0	1	117	26,5	0	1
58	0,5	0	1	118	26,5	0	1
59	0,5	0	1	119	28,5	0	1
60	0,5	0	1				

R-versi 2.9.1

Sebelumnya pilih menu “package” → “load package” → pilih “survival” → OK

```
> data<-read.table("data bb4.txt",header=T)
> data1<-data.frame(data)
> attach(data1)
# Variabel yang diperlukan untuk analisis time-dependent covariate
>  smonths<-smonths+0.5
>  stop.int <- as.numeric(unlist(dimnames(table(smonths))))
> start.int <- c(0,stop.int[-length(stop.int)])
> npatient <- length(smonths)
```

```
> nrep <- rep(0,npatient)
> for (i in 1:npatient)
+  {
+  nrep[i] <- sum(stop.int<=smonths[i])
+  }
> start.int1 <- start.int[1:nrep[1]]
> stop.int1 <- stop.int[1:nrep[1]]
>  for (i in 2:npatient)
+  {
+  start.int1 <- c(start.int1,start.int[1:nrep[i]])
+ stop.int1 <- c(stop.int1,stop.int[1:nrep[i]])
+  }
>  pid1 <- rep(pid,nrep)
>  cathtype1 <- rep(cathtype,nrep)
>  event1 <- rep(0,length(stop.int1))
>  event1[cumsum(nrep)] <- event
# Time dependent covariate dengan covariate catheter type
>  t.cathtype1 <- stop.int1*cathtype1
>  logt.cathtype1 <- log(stop.int1)*cathtype1
# Analisis Kaplan-Meier untuk covariate catheter type
>  kmfit1 <- survfit(Surv(smonths,event)~cathtype)
> plot(kmfit1,lty=1:2,xlab="time to infection (months)",ylab="estimated survival
+ + function")
> leg.names <- c("surgical","percutaneous")
> legend(1,.2,leg.names,lty=1:2)
> title(main="Kaplan-Meier curves for catheter infection data",cex=.8)
```



Membentuk model untuk yang *no time dependent* covariate (model cox *proportional hazard*)

> ph.cath <- coxph(Surv(smonths,event)~cathtype,method="efron")

> ph.cath

Call:

coxph(formula = Surv(smonths, event) ~ cathtype, method = "efron")

coef exp(coef) se(coef) z p

cathtype -0.613 0.542 0.398 -1.54 0.12

Likelihood ratio test=2.41 on 1 df, p=0.121 n= 119

Nilai Log Likelihood

> e<-((ph.cath$loglik)[2])

> e

[1] -103.0278

Membentuk model untuk yang *time dependent* covariate (*extended cox model* dengan fungsi waktu, g(t)=log t)

> ph.logt<-

coxph(Surv(start.int1,stop.int1,event1)~cathtype1+logt.cathtype1,method="efron")

> ph.logt

Call:

coxph(formula = Surv(start.int1, stop.int1, event1) ~ cathtype1 +

logt.cathtype1, method = "efron")

coef exp(coef) se(coef) z p

```
cathtype1       2.09    8.088    1.189  1.76 0.079
logt.cathtype1 -1.73    0.178    0.672 -2.57 0.010
Likelihood ratio test=14.5  on 2 df, p=0.000702  n= 1132

# Nilai Log Likelihood
> d<-((ph.logt$loglik)[2])
> d
[1] -96.97038
# Nilai p untuk uji ratio likelihood
> pval.logt<-1-pchisq(2*(ph.logt$loglik-ph.cath$loglik)[2],1)
> pval.logt
[1] 0.0005002196
# Analisis perubahan pada titik tertentu
> event.times <- as.numeric(unlist(dimnames(table(smonths[event==1]))))
> n.etimes <- length(event.times)
> logliks <- rep(0,n.etimes)
> for (i in 1:n.etimes)
+ {
+ t.cathtype2 <- cathtype1
+ t.cathtype2[stop.int1<=event.times[i]] <- 0
+ ph.cp <- coxph(Surv(start.int1,stop.int1,event1)~cathtype1+t.cathtype2,
method="efron")
+ logliks[i] <- ph.cp$loglik[2]
+ }
> t.cathtype2<-cathtype1
 > t.cathtype2[stop.int1<=event.times[4]]<-0
 > ph.cp<-
coxph(Surv(start.int1,stop.int1,event1)~cathtype1+t.cathtype2,method="efron")
>  cbind(event.times,logliks)
     event.times    logliks
 [1,]         0.5  -97.56271
 [2,]         1.5  -99.93807
```

```
 [3,]        2.5  -97.31866
 [4,]        3.5  -97.20138
 [5,]        4.5  -99.42334
 [6,]        5.5 -100.24255
 [7,]        6.5  -98.59629
 [8,]        8.5 -100.20066
 [9,]        9.5 -100.86155
[10,]       10.5 -101.44951
[11,]       11.5 -101.95374
[12,]       15.5 -100.61801
[13,]       16.5 -101.26864
[14,]       18.5 -101.85368
[15,]       23.5 -102.41635
[16,]       26.5 -103.02781
> summary(ph.cp)
Call:
coxph(formula = Surv(start.int1, stop.int1, event1) ~ cathtype1 +
    t.cathtype2, method = "efron")
  n= 1132
                coef exp(coef) se(coef)      z Pr(>|z|)
cathtype1    1.10023   3.00485  0.78294  1.405  0.15995
t.cathtype2 -3.19436   0.04099  1.09095 -2.928  0.00341 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
            exp(coef) exp(-coef) lower .95 upper .95
cathtype1     3.00485     0.3328  0.647709   13.9401
t.cathtype2   0.04099    24.3946  0.004832    0.3478
Rsquare= 0.012   (max possible= 0.168 )
Likelihood ratio test= 14.06  on 2 df,   p=0.0008845
Wald test            = 9.57  on 2 df,   p=0.008343
Score (logrank) test = 12.99  on 2 df,   p=0.001513
```