UNIVERSITAS INDONESIA

MODEL COX STRATIFIKASI

SKRIPSI

DWI ANJAR FERIANA
0706261612

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

UNIVERSITAS INDONESIA

MODEL COX STRATIFIKASI

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana sains

DWI ANJAR FERIANA
0706261612

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
PROGRAM STUDI SARJANA MATEMATIKA
DEPOK
JULI 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	: Dwi Anjar Feriana
NPM	: 0706261612
Tanda Tangan	:
Tanggal	: 7 Juli 2011

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh	:
Nama	: Dwi Anjar Feriana
NPM	: 0706261612
Program Studi	: Matematika
Judul Skripsi	: Model Cox Stratifikasi

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

DEWAN PENGUJI

Pembimbing	: Sarini Abdullah, S.Si., M.Stats	( )
Penguji	: Dr. Dian Lestari	( )
Penguji	: Dra. Rianti Setiadi, M.Si	( )
Penguji	: Dra. Saskya Mary Soemartojo, M.Si	( )

Ditetapkan di	: Depok
Tanggal	: 9 Juni 2011

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa, 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 Sains Jurusan Matematika pada Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia.

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

(1) Pembimbing tugas akhir penulis, Sarini Abdullah, M.Stats, yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan serta membimbing penulis dalam penyusunan skripsi ini. Terima kasih juga untuk kesabaran, nasehat, doa, dan dukungan yang telah diberikan selama penyusunan skripsi ini.

(2) Dra. Yahma Wisnani, M.Kom selaku pembimbing akademik penulis yang telah memberikan masukan dan dukungan selama 4 tahun masa perkuliahan penulis.

(3) Dr. Yudi Satria dan Rahmi Rusin, M.ScTech, selaku ketua dan sekretaris Departemen Matematika, atas segala bantuan dan dukungan yang telah diberikan.

(4) Seluruh staf Tata Usaha, staf Perpustakaan, serta karyawan Departemen Matematika, terima kasih atas segala bantuannya.

(5) Kedua orang tua penulis, yang telah memberikan bantuan material maupun dukungan selama penulis menjalani masa perkuliahan. Terima kasih atas segala doa, perhatian, kasih sayang, kesabaran, dan berbagai nasehat yang telah diberikan kepada penulis.

(6) Kakak penulis, mba Nunung dan mas Kandar, terima kasih atas doa, dukungan, dan nasehat-nasehatnya selama ini.

(7) Adik penulis, Rinda, terima kasih atas doanya.

(8) Teman-teman matematika angkatan 2007 yang telah memotivasi penulis dalam menyelesaikan tugas akhir ini.

(9) Isna, teman seperjuangan, terima kasih telah menjadi teman curhat sekaligus guru bagi penulis.

(10) Gamar yang telah membantu membuat slide presentasi menjadi lebih bagus dan membantu kesulitan penulis dalam pengetikan tugas akhir ini.

(11) Sahabat-sahabat penulis lainnya, Stefi, Wiwi, Sisca, Putri,Shafa, dan Misda yang telah mendoakan dan mendukung penulis.

(12) Echa, Dila, Ika, Terry, Hellda, Putri, Dinda, dan Lia, terima kasih untuk doa dan dukungan yang telah diberikan kepada penulis. Terima kasih juga atas waktunya untuk mendengarkan keluh kesah penulis.

(13) Naela, Nanda, Rina, Rini, Fitri, Echit, dan seluruh teman-teman Perhimak 2007 yang telah mendoakan penulis.

(11) Teman-teman penulis lainnya baik yang di Matematika, MIPA, maupun UI yang tidak dapat disebutkan satu per satu yang telah mendoakan penulis.

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 pengetahuan.

Penulis

2011

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama	: Dwi Anjar Feriana
NPM	: 0706261612
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 penulis yang berjudul :
Model Cox Stratifikasi

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 penulis selama tetap mencantumkan nama penulis sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini penulis buat dengan sebenarnya.

Dibuat di : Depok
Pada tanggal : 7 Juli 2011
Yang menyatakan

(Dwi Anjar Feriana)

ABSTRAK

Nama : Dwi Anjar Feriana
Program Studi : Matematika
Judul : Model Cox Stratifikasi

Model cox stratifikasi merupakan modifikasi dari model cox *proportional hazard* ketika ada *covariate* yang tidak memenuhi asumsi PH. Modifikasi dilakukan dengan membentuk strata yang berasal dari kombinasi kategori *covariate* yang tidak memenuhi asumsi PH, sehingga akan didapat model *cox proportional hazard* untuk tiap strata. Koefisien regresi pada model cox stratifikasi ditaksir dengan metode maksimum *partial likelihood*. Sebagai contoh penerapan digunakan data berupa waktu sampai meninggal untuk seseorang yang mengidap penyakit kanker paru-paru, dengan awal pengamatan saat pasien diberi suatu perlakuan. Diperoleh hasil bahwa model cox stratifikasi dapat digunakan untuk memperbaiki hasil dari model cox PH. Berdasarkan grafik fungsi *survival*, ada perbedaan "*survival experience*" dari individu-individu pada strata yang berbeda.

Kata Kunci : fungsi *hazard*, model cox *proportional hazard*, model cox stratifikasi, estimasi maksimum *partial likelihood*.
xiv+55 halaman; 4 gambar; 18 tabel
Daftar Pustaka : 5 (1987-2005)

ABSTRACT

Name : Dwi Anjar Feriana
Program Study : Mathematics
Title : Stratified Cox Model

Stratified cox model is a modification of the cox proportional hazard model when there are covariates that violate the PH assumption. The modification is done by forming stratum from combination of covariate category that does not satisfy the PH assumption, so the cox PH model for each stratum will be obtained. Regression coefficients in the stratified cox model are estimated using maximum partial likelihood method. An example, data of the time to event for lung cancer patient, where event is death is used. The study start when a patient is given a treatment and end when the event occurred or censored. It showed that stratified cox model can be used to improve the cox proportional hazard model, by showing different survival experience for patients in different stratum.

Keywords : hazard function, cox proportional hazard model, stratified cox model, maximum partial likelihood estimation
xiv+55 pages ; 4 pictures; 18 tables
Bibliography : 5 (1987-2005)

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB 1
PENDAHULUAN

1.1 Latar Belakang

S*urvival Analysis* adalah sekumpulan prosedur statistika yang digunakan untuk menganalisis data di mana variabel yang diperhatikan adalah waktu sampai terjadinya suatu *event* (Kleinbaum and Klein, 2005). S*urvival analysis* disebut juga sebagai analisis waktu kejadian (*time to event analysis*), di mana waktu sampai terjadinya suatu *event* yang diinginkan disebut *survival time* atau *failure time* (Guolin, 2008). Waktu dapat dinyatakan dalam tahun, bulan, minggu, atau hari dari awal mula dilakukan studi atau pengamatan pada seorang individu sampai suatu peristiwa terjadi pada individu tersebut; selain itu, waktu dapat juga dinyatakan dalam usia individu ketika suatu *event* terjadi. Sedangkan *event* dapat berupa kematian, munculnya suatu penyakit, kambuhnya suatu penyakit setelah dilakukan operasi, atau beberapa hal lain yang dirancang yang bisa diperhatikan dan dapat terjadi pada seorang individu.

Data *survival* dalam analisis semacam ini melibatkan bentuk data *time to event*, misalnya data waktu sampai terjadinya kematian. Sehimpunan data yang digunakan dapat berupa data eksak ataupun data tersensor, dan mungkin juga data terpancung, tetapi dalam tugas akhir ini hanya akan digunakan data eksak dan tersensor. Disebut data eksak apabila waktu tepatnya suatu *event* yang diinginkan terjadi diketahui, sedangkan data tersensor terjadi apabila waktu sampai terjadinya *event* pada individu yang bersangkutan tidak diketahui secara pasti, hanya informasi bahwa sampai saat tertentu kejadiannya belum teramati.

Salah satu tujuan analisis *survival* adalah untuk mengetahui hubungan antara waktu kejadian (*time to failure*) dan variabel prediktor (*covariate*) yang terukur pada saat dilakukan penelitian. Analisis ini dapat dilakukan dengan metode regresi. Salah satunya adalah regresi *cox proportional hazard* (PH).

Model Cox PH menyatakan *hazard rate* dari satu individu pada waktu *t* dengan diketahui variabel-variabel prediktornya. Model ini melibatkan komponen yang disebut fungsi *baseline hazard* yang melibatkan waktu *t* tetapi tidak melibatkan variabel prediktor (*covariate*). Selain itu, model Cox PH juga melibatkan komponen lain yaitu pernyataan eksponensial terhadap jumlah perkalian efek dari masing-masing *covariate* terhadap *covariate* tersebut. Komponen ini menjamin taksiran hazard yang dihasilkan selalu *non-negative*. Hal ini sesuai dengan yang diharapkan karena variabel dependennya (*outcome*) adalah waktu sampai terjadi *event* sehingga tidak mungkin negatif.

Pada model *cox proportional hazard* diasumsikan variabel-variabel prediktornya (*covariate*) memenuhi asumsi *proportional hazard* (PH). Artinya, *hazard rate* untuk satu individu sebanding dengan *hazard rate* individu lain di mana perbandingannya konstan sepanjang waktu. Perbandingan ini disebut juga *hazard ratio* (HR). Dalam kenyataannya, sering kita menemukan kasus dimana tidak semua *covariate* memenuhi asumsi PH. Oleh karena itu, kita memerlukan metode lain untuk mendapatkan hasil yang lebih baik untuk menganalisis data *survival* tersebut.

Pada tugas akhir ini, akan dibahas tentang suatu metode yang dapat mengatasi varibel-variabel yang tidak memenuhi asumsi PH tersebut, yaitu model cox stratifikasi. Model cox stratifikasi ini adalah modifikasi dari model *cox proportional hazard* yang memberikan perhatian atau mengontrol *covariate* yang tidak memenuhi asumsi PH dengan menstratifikasi *covariate* yang tidak memenuhi asumsi PH tersebut. *Covariate* yang memenuhi asumsi PH akan dimasukkan ke dalam model, sedangkan *covariate* yang tidak memenuhi asumsi PH tidak dimasukkan ke dalam model. Karena kuat dugaan bahwa sebenarnya variabel ini berkontribusi, maka variabel ini tidak begitu saja dibuang atau tidak digunakan melainkan tetap diamati efeknya dengan menjadikannya sebagai strata dan tidak masuk ke dalam model. Walaupun variabel ini tidak masuk ke dalam model, tetapi peranannya masih dapat diamati dalam bentuk strata.

1.2 Perumusan Masalah dan Ruang Lingkup Masalah

Perumusan masalah yang diajukan pada tugas akhir ini adalah :

1. Bagaimana bentuk model cox stratifikasi?
2. Bagaimana mencari taksiran model cox stratifikasi tersebut?
3. Bagaimana penerapan model cox stratifikasi pada suatu data?

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. *Covariate* yang digunakan adalah *time independent covariate.*
4. Tidak ada interaksi antar *covariate.*
5. Tidak ada interaksi antar strata.

1.3 Jenis Penelitian dan Metode yang Digunakan

Jenis penelitian yang digunakan dalam pembuatan tugas akhir ini adalah studi literatur. Sedangkan metode yang digunakan untuk menaksir parameter pada model cox stratifikasi adalah *maximum partial likelihood*.

1.4 Tujuan Penulisan

Tujuan penulisan tugas akhir ini adalah :

1. Menentukan bentuk model cox stratifikasi
2. Mencari taksiran dari model cox stratifikasi
3. Memberikan contoh mengenai penerapan model cox stratifikasi pada 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 suatu *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. Waktu akhir yang juga terdefinisi dengan baik.

2.1 *Survival Time*

Dalam analisis *survival*, respon yang diperhatikan adalah waktu sampai terjadinya suatu *event*, sehingga analisis *survival* seringkali disebut sebagai analisis waktu kejadian (*time to event analysis*). 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 fungsi *survival* dan fungsi *hazard*. Kedua 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 $[0, \infty)$.

2.2.1 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 :

$$S(t) = P(T > t) \tag{2.1}$$

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 $F(t) = P(T \leq t)$. Fungsi *survival* yang merupakan komplemen dari fungsi distribusi kumulatif yaitu :

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) \tag{2.2}$$

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

$$S(t) = P(T > t) = \int_t^\infty f(t)dt \tag{2.3}$$

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)*

   $$P(t_j) = P(T = t_j),\, j = 1, 2, \ldots, p$$

   dimana $t_1 < t_2 < \cdots < t_p$

2. Fungsi *survival* untuk variabel random diskret T diberikan oleh,

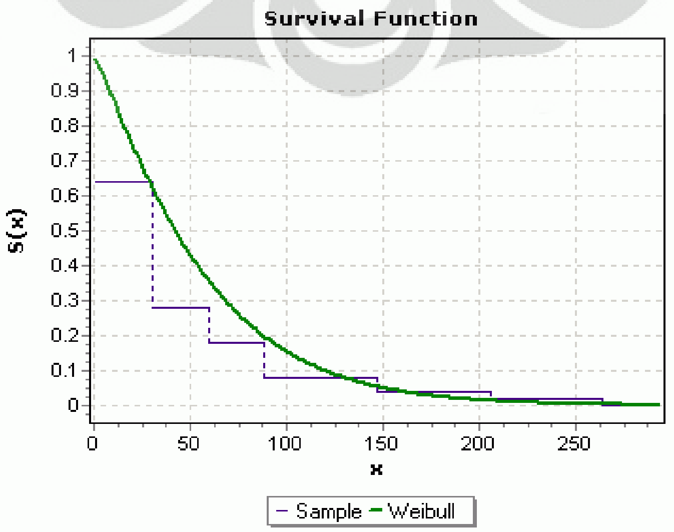
$$S(t) = P(T > t) = \sum_{t_j > t} p(t_j) \tag{2.4}$$

(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 (weibull) dan kurva *survival* dalam praktiknya (sampel):



Gambar 2.1. Kurva fungsi survival secara teori dan secara praktik

2.2.2 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 :

$$h(t) = lim_{\Delta t \to 0} \frac{P(t \leq T < t+\Delta t | T \geq t)}{\Delta t} \tag{2.5}$$

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 :

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

$$h(t) = \frac{f(t)}{S(t)} \tag{2.6}$$

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

$$f(t) = F'(t) = \frac{d(F(t))}{dt} = \frac{d(1-S(t))}{dt} = -\frac{d(S(t))}{dt} = -S'(t) \tag{2.7}$$

Maka,

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -S'(t).\frac{dlnS(t)}{dS(t)} = -\frac{dS(t)}{dt}.\frac{dlnS(t)}{dS(t)} = -\frac{dlnS(t)}{dt}$$

dari persamaan di atas diperoleh,

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

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

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

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

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

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

$$S(t) = \exp\left[-\int_0^t h(x)dx\right] \quad (2.8)$$

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 :

$$h(t_j) = P(T = t_j | T \geq t_j), j = 1, 2, \dots, p$$

$$= \frac{P(T = t_j \cap T \geq t_j)}{P(T \geq t_j)} = \frac{P(T = t_j)}{P(T \geq t_j)}$$

$$= \frac{P(T = t_j)}{P(T > t_{j-1})}$$

$$h(t_j) = \frac{p(t_j)}{S(t_{j-1})} \quad (2.9)$$

dimana $S(t_0) = 1$

Diketahui, $S(t_j) = P(T > t_j)$, dan

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

$$\leftrightarrow p(t_j) = S(t_{j-1}) - S(t_j) \quad (2.10)$$

Berdasarkan persamaan (2.9), diperoleh

$$h(t_j) = \frac{p(t_j)}{S(t_{j-1})} = \frac{S(t_{j-1}) - S(t_j)}{S(t_{j-1})} = 1 - \frac{S(t_j)}{S(t_{j-1})}\ ,\ j = 1, 2, \dots \qquad (2.11)$$

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

$$S(t) = \prod_{t_j \le t} \frac{S(t_j)}{S(t_{j-1})} \qquad (2.12)$$

Jadi, berdasarkan persamaan (2.11) dan (2.12), hubungan fungsi survival dengan fungsi hazard yaitu :

$$S(t) = \prod_{t_j \le t} [1 - h(t_j)] \qquad (2.13)$$

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 ($h(t) \ge 0$) 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, $h(t) \ge 0$
- Tidak memiliki batas atas.

Beberapa bentuk kurva fungsi *hazard* antara lain:

- Dalam kehidupan nyata, untuk kasus di mana *hazard* bernilai konstan jarang ditemui. Contohnya, fungsi hazard untuk orang sehat mengalami kematian.
- 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.
- 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.
- *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 dua 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.

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 merokok. Studi dilakukan selama 12 bulan. Pertanyaan yang diajukan adalah “Kapan pertama kali Anda merokok?” 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 merokok) telah terjadi sebelum usia anak laki-laki tersebut diwawancarai tetapi usia tepatnya ia merokok 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 merokok,
- Waktu akhir (*end of study*): setelah 12 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 umum, data tersensor kanan dapat terjadi karena 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 δ, 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:

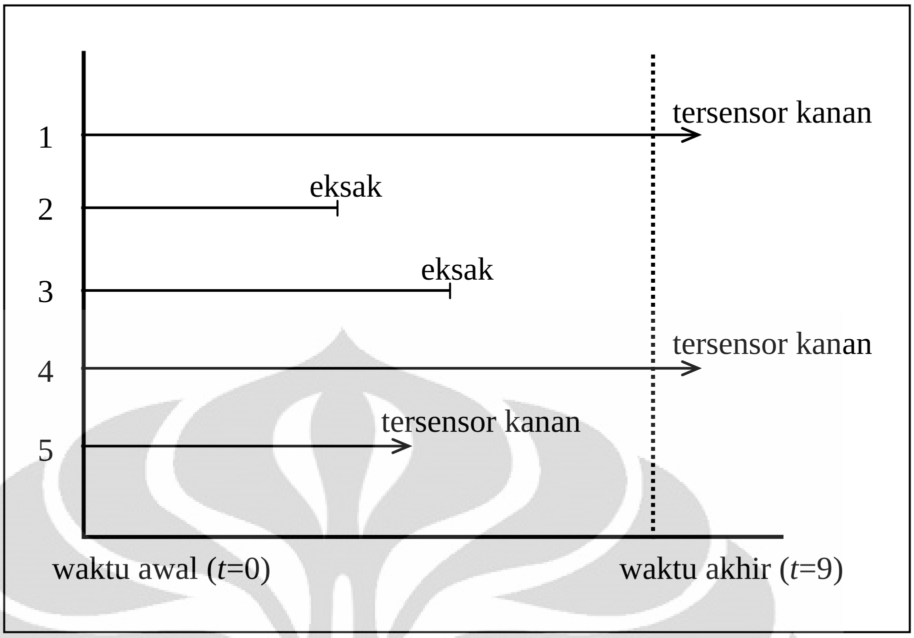
$$9^+, 4, 6, 9^+, 5^+.$$

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 , δ_i) :

$$(9,0), (4,1), (6,1), (9,0), (5,0).$$

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

Gambar 2.2 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 2.2. 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

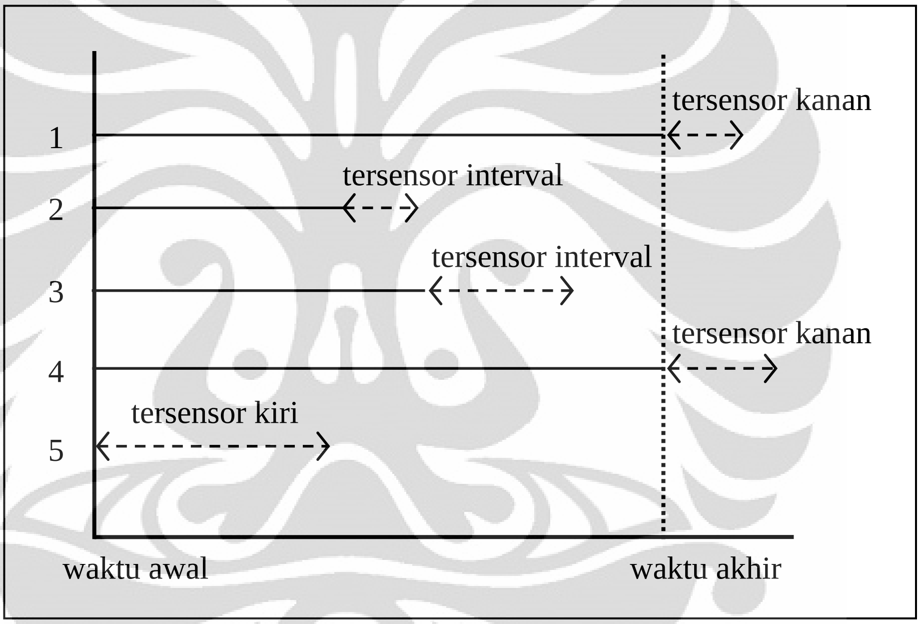
Sensor interval terjadi ketika hanya diketahui bahwa suatu *event* yang diinginkan terjadi dalam suatu periode waktu. Data tersensor kiri dan tersensor kanan merupakan kasus khusus dari data tersensor interval.

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



Gambar 2.3. 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

$$h(t,\boldsymbol{X}) = h_0(t)\exp^{\sum_{y=1}^{m}\beta_y X_y},\ \boldsymbol{X} = (X_1, X_2, \dots, X_m) \qquad (2.14)$$

di mana

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

β_y = koefisien regresi

X_y = *covariate*, y = 1, 2,…, m

Model ini menyatakan *hazard rate* dari satu individu pada waktu *t* dengan diketahui *covariat*e ***X***. Model ini terdiri dari dua kuantitas seperti terlihat pada tabel 2.1.

Tabel 2.1. Dua kuantitas pembentuk *hazard rate*

$\boldsymbol{h_0(t)}$	$\mathbf{exp}\left[\sum_{y=1}^{m}\boldsymbol{\beta_y X_y}\right]$
- 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 ($\mathbf{exp}\left[\sum_{y=1}^{m}\boldsymbol{\beta_y X_y}\right]$)**

  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 $\boldsymbol{X}$ 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 kita miliki tidak diketahui distribusinya, sehingga bentuk $h_0(t)$nya juga tidak dapat diketahui.

Model semiparametrik lebih sering digunakan karena walaupun bentuk fungsional $h_0(t)$ tidak diketahui, 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 rasio dari *hazard rate* satu individu dengan *hazard rate* dari individu lain. Hal ini dapat ditunjukkan sebagai berikut.

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

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

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

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

maka *hazard rasio*-nya adalah

$$\begin{aligned} \mathrm{HR} = \frac{h_A(t, X^*)}{h_B(t, X)} &= \frac{h_0(t)\exp\left[\sum_{y=1}^{m}\beta_y {X_y}^*\right]}{h_0(t)\exp\left[\sum_{y=1}^{m}\beta_y X_y\right]} \\ &= \exp\left(\sum_{y=1}^{m}\beta_y {X_y}^* - \sum_{y=1}^{m}\beta_y X_y\right) \\ &= \exp\left[\sum_{y=1}^{m}\beta_y (X_y^* - X_y)\right] \qquad (2.15) \end{aligned}$$

Apabila nilai *hazard ratio* (HR) konstan sepanjang waktu, maka dapat dikatakan bahwa $X_1, X_2, \dots, X_m$ memenuhi asumsi PH.

Sifat dari model Cox PH adalah meskipun kita tidak mengetahui bentuk $h_0(t)$, akan tetapi kita tetap dapat menaksir koefisien regresi ($\boldsymbol{\beta}$). Seperti yang telah diketahui bahwa kita harus menaksir $\boldsymbol{\beta}$ untuk mengetahui efek dari *covariate*-nya. Besarnya efek ini dapat dihitung tanpa harus menaksir fungsi *baseline hazard*. Jadi, dengan asumsi yang terbatas, kita dapat mengetahui informasi penting yang diperoleh dari data *survival* melalui nilai *hazard ratio* dan *survival experience*.

2.5 *Maximum Likelihood Estimator*

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

$$f(X;\theta) = f(X_1, X_2, \dots, X_m; \boldsymbol{\theta})$$
$$= f(X_1;\boldsymbol{\theta}) \times f(X_2;\boldsymbol{\theta}) \times \dots \times f(X_m;\boldsymbol{\theta}) \qquad (2.16)$$

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

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

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

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

Salah satu penaksiran maksimum likelihood dari θ didapat dengan menyelesaikan persamaan $\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$ atau dengan menggunakan logaritma natural $\frac{\partial \ln\ L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$.

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

$$\frac{\partial \ln\ L(\theta_1, \theta_2, \dots, \theta_n)}{\partial \theta_i} = 0, dengan\ i = 1, 2, .., m \qquad (2.17)$$

BAB 3
MODEL COX STRATIFIKASI

3.1 Model Cox Stratifikasi

Salah satu model pada metode semiparametrik yang sering digunakan untuk mengetahui pengaruh dari *covariate* yang diperhatikan terhadap *survival time* adalah model *cox proportional hazard* (PH). Misalnya, kita memiliki sebanyak *m* *covariate*, maka model Cox PH yang terbentuk adalah sebagai berikut:

$$\boldsymbol{h(t,X) = h_0(t)\,\mathrm{exp}\left[\sum_{y=1}^{m} \beta_y X_y\right],}$$

di mana $\boldsymbol{X} = (X_1, X_2, \dots., X_m)$ adalah variabel prediktor (*covariate*). Berdasarkan model di atas, bagian fungsi *baseline hazard*, $h_0(t)$, tidak dapat ditentukan sehingga *hazard rate* dari masing-masing objek tidak dapat diketahui.

Salah satu pengukuran yang biasa digunakan dalam interpretasi model *cox proportional hazard* adalah *Hazard Ratio* (HR). Misalnya, individu A memiliki *hazard rate* $h_A(t, X^*)$ di mana $X^* = (X_1^*, X_2^*, \dots, X_m^*)$ dan individu B memiliki *hazard rate* $h_B(t, X)$ di mana $X = (X_1, X_2, \dots, X_m)$, maka rasio hazardnya sebagai berikut :

$$\begin{aligned} HR = \frac{h_A(t,X^*)}{h_B(t,X)} &= \frac{h_0(t)\,\exp\left[\sum_{y=1}^{m} \beta_y X_y{}^*\right]}{h_0(t)\,\exp\left[\sum_{y=1}^{m} \beta_y X_y\right]} \\ &= \exp\left(\sum_{y=1}^{m} \beta_y X_y{}^* - \sum_{y=1}^{m} \beta_y X_y\right) \\ &= \exp\left[\sum_{y=1}^{m} \beta_y (X_y^* - X_y)\right] \end{aligned}$$

Dari *hazard ratio* di atas, terlihat bahwa nilainya tidak bergantung pada waktu, artinya perbandingan *hazard* kedua objek tetap sepanjang waktu. Hal ini disebut dengan asumsi *Proportional Hazard* (PH).

Dalam beberapa kasus, *hazard ratio* dari suatu *covariate* dapat berbeda pada waktu yang berbeda. Ini berarti bahwa *hazard ratio* dipengaruhi oleh waktu. Pada kondisi ini, maka asumsi *proportional hazard* tidak terpenuhi. Jika asumsi ini tidak terpenuhi, maka model Cox PH tidak tepat jika digunakan untuk melihat pengaruh dari *covariate* tersebut. Jika hal ini terjadi, maka dianjurkan untuk menggunakan model lain yang sesuai dengan keadaan tersebut dan salah satunya adalah menggunakan model Cox Stratifikasi.

3.1.1 Pemodelan

Model cox stratifikasi adalah model yang didapatkan dengan memodifikasi model *cox proportional hazard*. Modifikasi dilakukan dengan mengontrol *covariate* yang tidak memenuhi asumsi PH. Pengontrolan ini dilakukan dengan cara menstratifikasi *covariate* tersebut.

Misalkan, ada sebanyak *m covariate*. Misalkan pula tidak ada interaksi antar *covariate*. Model Cox PH yang terbentuk adalah

$$h(t,\boldsymbol{X}) = h_0(t)\exp[\beta_1 X_1 + \cdots + \beta_k X_k + \beta_{k+1} X_{k+1} + \cdots + \beta_m X_m].$$

Dari *m* covariate tersebut, misalkan ada sebanyak *k covariate* yang memenuhi asumsi PH dan ada *p covariate* yang tidak memenuhi asumsi PH, di mana *p* = *m* – *k*. Tanpa menghilangkan sifat keumuman, misalkan *covariate* yang tidak memenuhi asumsi PH tersebut adalah $X_{k+1}, X_{k+2}, \dots, X_m$ dan sesudah dilakukan stratifikasi variabel terbentuk k* strata (proses stratifikasi akan dijelaskan lanjut pada subbab 3.1.2).

a. Model dengan interaksi

Jika terdapat k* strata dan *k covariate* yang terdapat pada model, maka model cox stratifikasi dengan interaksi adalah sebagai berikut:

$$\begin{aligned} h_g(t,X) = h_{0g}(t)\exp[&\beta_1 X_1 + \beta_2 X_2 \dots + \beta_k X_k \\ &+\beta_{11}(Z_1{}^*.X_1) + \beta_{21}(Z_1{}^*.X_2) + \cdots + \beta_{k1}(Z_1{}^*.X_k) \\ &+\beta_{12}(Z_2{}^*.X_1) + \beta_{22}(Z_2{}^*.X_2) + \cdots + \beta_{k2}(Z_2{}^*.X_k) + \cdots \\ &+\beta_{1,k^*-1}(Z^*{}_{k^*-1}.X_1) + \beta_{2,k^*-1}(Z^*{}_{k^*-1}.X_2) + \cdots + \\ &\beta_{k,k^*-1}(Z^*{}_{k^*-1}.X_k)] \end{aligned} \tag{3.1}$$

di mana

$$Z_l^* = \begin{cases} 1, strata\ l+1 \\ 0, bukan \end{cases}, l = 1, 2, \dots, k^* - 1$$

Berdasarkan model 3.1, terlihat bahwa jika menggunakan model dengan interaksi, maka akan ada tambahan parameter sebanyak $k \times (k^* - 1)$ yang harus ditaksir. Keuntungan apabila menggunakan model dengan interaksi adalah didapatkan model yang lebih akurat, yaitu dalam hal efek *covariate* yang mungkin berbeda pada strata yang berbeda dapat diakomodir. Akan tetapi, menimbang banyaknya parameter yang harus ditaksir, maka model tanpa interaksi lebih sederhana, seperti yang akan dibahas berikut.

b. Model tanpa interaksi

Model cox stratifikasi tanpa interaksi adalah sebagai berikut:

$$h_g(t, X) = h_{0g}(t)\exp[\beta_1 X_1 + \beta_2 X_2 \dots + \beta_k X_k],\ g = 1,2, \dots, k^* \quad (3.2)$$

dimana:

g = jumlah strata

$h_{0g}(t)$ = fungsi *baseline hazard* untuk tiap strata

$\beta_1, \beta_2, \dots, \beta_k$ = koefisien regresi.

Pada model 3.2, efek dari *covariate* terhadap $h(t)$ adalah sama pada strata yang berbeda. Akan tetapi, nilai $h_{0g}(t)$ berbeda untuk strata yang berbeda. Jadi, efek strata tetap dapat terlihat melalui komponen $h_{0g}(t)$. Dalam tugas akhir ini, akan digunakan model tanpa interaksi.

Secara umum, jika ada data berukuran *n*, dengan *covariate* X_1, X_2,…, X_k, maka skema data pada model cox stratifikasi adalah seperti pada tabel 3.1.

Tabel 3.1 Skema Data pada Model Cox Stratifikasi

	i	X_1	X_2	…	X_k	T
Strata 1	1					
	2					
	:					
	n_1					
Strata 2	1					
	2					
	:					
	n_2					
:	:					
	:					
Strata k*	1					
	2					
	:					
	n_{k*}					

dengan $n_1 + n_2 + \cdots + n_{k^*} = n$.

3.1.2 Stratifikasi Variabel

Pada model Cox Stratifikasi, *covariate* pada model Cox PH yang tidak memenuhi asumsi PH akan dikeluarkan dari model. Akan tetapi, walaupun tidak termasuk dalam model, peran dari *covariate* ini tetap ada. Hal ini dapat dilakukan melalui stratifikasi, yang pada akhirnya terlihat kontribusi masing-masing variabel tersebut pada strata yang berbeda.

Langkah awal yang harus dilakukan adalah identifikasi *covariate* yang tidak memenuhi asumsi PH. Mengacu pada model (3.2), sebut variabel-variabel yang tidak memenuhi asumsi PH adalah

$$X_{k+1} \rightarrow Z_1, X_{k+2} \rightarrow Z_2, \dots, X_m \rightarrow Z_p.$$

Langkah selanjutnya adalah stratifikasi variabel Z_i, $i = 1, 2, \dots, p$.

Secara umum, variabel terbagi atas dua macam pengukuran, yaitu variabel kategorik dan variabel numerik. Stratifikasi variabel ini dilakukan dengan ketentuan sebagai berikut:

- Z_i kategorik

  Tidak perlu dilakukan kategorisasi karena jumlah kategori sudah diketahui.
- Z_i numerik

  Tidak ada ketentuan khusus dalam menentukan banyaknya kategori. Jumlah kategori biasanya disesuaikan dengan kebutuhan dan tujuan penelitian.

Beberapa cara yang dapat digunakan untuk mengkategorikan variabel numerik adalah sebagai berikut :

1. *Judgment* peneliti
2. Persentase yang sama di tiap kelompok
3. Interval yang sama
4. *Clustering*

Selanjutnya, definisikan sebuah variabel baru yang nilainya merupakan kombinasi dari level pada $Z_1, Z_2, \dots, Z_p$; yang kemudian kita sebut sebagai variabel Z^*. Cara pendefinisian variabel baru ini mengikuti langkah-langkah sebagai berikut:

1. Kategorisasi masing-masing *covariate* Z_i, $i = 1, 2, \ldots, p$
2. Kombinasikan seluruh kategori tersebut
3. Hasil kombinasi tersebut akan digunakan sebagai nilai-nilai dari Z^*, disebut strata.

Sebagai contoh, misal ada 2 *covariate* yang tidak memenuhi asumsi PH, yaitu variabel Z_1 dan Z_2, dimana masing-masing variabel memiliki sebanyak 2 dan 3 kategori yaitu:

$$Z_1 = \begin{cases} 1, kategori\ 1 \\ 2, kategori\ 2 \end{cases} \quad dan \quad Z_2 = \begin{cases} 1, kategori\ 1 \\ 2, kategori\ 2 \\ 3, kategori\ 3 \end{cases}$$

Dari kedua variabel tersebut akan dikombinasikan sehingga menghasilkan kategori baru.

Tabel 3.2 Hasil Kombinasi variabel Z_1 dan Z_2

Z_2 \ Z_1	Kategori 1	Kategori 2
Kategori 1	11	12
Kategori 2	21	22
Kategori 3	31	32

Dari tabel tersebut, dapat dilihat bahwa kombinasi yang terbentuk ada enam. Keenam kombinasi tersebut adalah kategori dari variabel baru yang kita stratifikasi dalam model cox stratifikasi ($k^* = 6$).

Secara umum, misal masing-masing variabel $Z_1, Z_2, \ldots, Z_p$ mempunyai kategori sebanyak $k_1, k_2, \ldots, k_p$. Kombinasi yang dihasilkan adalah sebanyak

$$k^* = k_1 \times k_2 \times \ldots \times k_p.$$

Jadi, variabel baru (Z^*) memiliki k^* kategori, dimana k^* adalah jumlah kombinasi (strata) hasil dari pengkategorian variabel-variabel Z_i. Sehingga diperoleh model sebagai berikut:

$$h_g(t, X) = h_{0g}(t) exp[\beta_1 X_1 + \cdots + \beta_k X_k], g = 1, 2 \ldots., k^*.$$

Pada model tersebut, g menyatakan strata ke-g. Sebagai catatan, variabel Z^* tidak dimasukkan ke dalam model, tetapi variabel-variabel X yang memenuhi asumsi PH, yang dimasukkan ke dalam model. Selain itu, fungsi *baseline hazard* $h_{0g}(t)$ dapat berbeda untuk tiap-tiap strata, sehingga fungsi hazard pun dapat berbeda untuk strata yang berbeda. Akan tetapi, koefisien $\beta_1, \beta_2 \ldots, \beta_k$ adalah sama untuk semua strata. Oleh karena itu, taksiran dari *hazard ratio* (HR) juga akan sama untuk tiap strata.

Untuk lebih jelasnya, perhatikan contoh berikut.

Misal kita memiliki dua *covariate*, yaitu X_1 dan X_2. Misalkan pula, X_1 adalah variabel numerik dan X_2 variabel kategorik, yaitu :

$$X_2 = \begin{cases} 0, & pria \\ 1, & wanita \end{cases},$$

maka model Cox PHnya seperti berikut:

$$h(t, X) = h_0(t) e^{X_1\beta_1 + X_2\beta_2}.$$

Misalkan setelah dilakukan pengecekan asumsi, ternyata didapatkan hasil bahwa variabel X_2 tidak memenuhi asumsi PH. Berdasarkan metode di atas, maka perlu dilakukan modifikasi model. Modifikasi dilakukan dengan mengeluarkan X_2 dari model untuk dijadikan strata. Dalam kasus ini, karena X_2 adalah variabel kategorik dengan 2 kategori, maka kategori tersebut adalah strata sehingga model cox stratifikasinya adalah sebagai berikut:

$$h_g(t, X) = h_{0g}(t)\exp[\beta_1 X_1], \; g = 1,2.$$

Pada masing-masing strata ambil individu yang memiliki karakteristik sama berdasarkan X_1, yaitu :

- Pada strata 1

  A dengan $X_1 = x$ dan B dengan $X_1 = x + \Delta x$

  $$HR = \frac{h_A}{h_B} = \frac{h_{01}(t)\exp[\beta(x)]}{h_{01}(t)\exp[\beta(x + \Delta x)]} = e^{-\beta\Delta x}$$

- Pada strata 2

  C dengan $X_1 = x$ dan D dengan $X_1 = x + \Delta x$

  $$HR = \frac{h_C}{h_D} = \frac{h_{02}(t)\exp[x]}{h_{02}(t)\exp[x + \Delta x]} = e^{-\beta\Delta x}$$

Dari ilustrasi di atas, terlihat bahwa *hazard ratio* (HR) pada strata 1 dan strata 2 adalah sama.

Meskipun *hazard ratio* bernilai sama untuk setiap strata, kita akan tetap dapat melihat efek dari X_2 berdasarkan stratanya. Hal ini karena mengacu pada model (3.2) bahwa fungsi hazard terdiri dari dua komponen, yaitu $h_{0g}(t)$ dan

$\exp\left[\sum_{y=1}^{k} \beta_y X_y\right]$. Pada bagian $\exp\left[\sum_{y=1}^{k} \beta_y X_y\right]$, yang berperan adalah $\boldsymbol{X}$ yang terdapat di dalam model. Sedangkan $\boldsymbol{X}$ yang tidak terdapat dalam model, digunakan untuk membentuk strata. Efek strata terlihat pada komponen $h_{0g}(t)$. Secara singkat dapat disajikan pada tabel 3.3.

Tabel 3.3 Komponen $h_g(t, X)$

$\boldsymbol{h_{0g}(t)}$	$\mathbf{exp}\left[\sum_{y=1}^{k} \boldsymbol{\beta_y X_y}\right]$
- bergantung pada strata → berdasarkan $\boldsymbol{X}$ yang tidak ada di model (pada contoh, bergantung pada X_2)	- bergantung pada $\boldsymbol{X}$ yang ada di model (pada contoh, bergantung pada X_1)

3.2 Metode Penaksiran Parameter Model Cox Stratifikasi

Pada model regresi, untuk dapat mengetahui berapa besar pengaruh atau efek dari suatu variabel prediktor terhadap variabel responnya, maka harus dicari taksiran masing-masing parameter dari model regresi tersebut.

Tabel 3.4 Fungsi Likelihood

Full Likelihood	Partial Likelihood
$L = \prod_{t \epsilon D} f(t) \prod_{t \epsilon C} S(t)$	$L_P = \prod_{j \epsilon D} \frac{h(t_j, \boldsymbol{X})}{\sum_{i \epsilon R(t_j)} h(t_i, \boldsymbol{X})}$

di mana :

t = *time to event*

j = objek yang mengalami *event*

D = himpunan objek yang mengalami *event*

C = himpunan objek yang tersensor

$f(t)$ = *pdf*

$S(t)$ = *survival function*

$R(t_j)$ = himpunan objek yang beresiko mengalami *event* pada waktu t_j

$h(t_j, \boldsymbol{X})$ = *hazard rate* objek ke-*j*

$h(t_i, \boldsymbol{X})$ = *hazard rate* objek ke-*i*

Pada model regresi Cox PH digunakan metode *Maximum Partial Likelihood Estimation* (MPLE) dalam penaksiran parameter model. Begitu juga pada model Cox Stratifikasi. Pada penaksiran parameter model Cox Stratifikasi, digunakan metode yang sama dengan model Cox PH yaitu *Maximum Partial Likelihood Estimation*. Taksiran maksimum partial likelihood dari parameter model stratifikasi diturunkan dengan memaksimumkan fungsi partial likelihood, sebut L_P. L_P sering dinotasikan dengan $L_P(\boldsymbol{\beta})$ dimana $\boldsymbol{\beta}$ menyatakan kumpulan dari parameter yang belum diketahui, yaitu $\boldsymbol{\beta} = \beta_1, \beta_2, \dots, \beta_k$. Adapan fungsi $L_P(\boldsymbol{\beta})$ digunakan karena :

1. Pada data survival, data dapat berupa data tersensor ataupun teramati. Untuk data yang tersensor, tidak diketahui pasti kapan terjadinya *event*, sehingga tidak dapat diperhitungkan kontribusinya secara langsung terhadap fungsi likelihood. Jadi, hanya objek yang teramati saja yang berkontribusi langsung terhadap pembentukan likelihood.

   Contoh :

   Misalnya, kita memiliki sampel berukuran *n*, dimana ada sebanyak *r* objek yang mengalami *event* dan sebanyak *n-r* objek yang tersensor. Pada fungsi partial likelihood, hanya *r* objek yang mengalami *event* yang dapat dilihat kontribusinya secara langsung. Sedangkan *n-r* objek yang tersensor tetap berkontribusi tetapi tidak dapat dilihat secara langsung.
2. Pada fungsi partial likelihood, tidak diketahui distribusi dari variabel responnya. Oleh karena itu, kontribusi yang diperhitungkan pada $L_P(\boldsymbol{\beta})$ hanya bergantung pada statistik terurut dari data sampel.Contoh,
   - Data 1

Tabel 3.5 (a) Data *survival* I

i	1	2	3
t_i	3	4	5

dimana i = objek, t_i = *time to event* ke-i.

Pr(objek mengalami event pada t=5) = Pr(objek mengalami event pada t=5|T≥5)

= 1/1

- Data 2

Tabel 3.5 (b) Data *survival* II

i	1	2	3
t_i	2	5	8

dimana i = objek, t_i = *time to event* ke-i.

Pr(objek mengalami event pada t=5) = Pr(objek mengalami event pada t=5|T≥5)

= 1/2

Terlihat bahwa nilai probabilitas saat t = 5 dari data I dan data II berbeda. Sementara, jika distribusi dari T diketahui, maka probabilitasnya akan sama untuk t=5, yaitu f(5). Akan tetapi, Robin (1989) dan Tsiatis (1990) telah menunjukkan bahwa taksirannya berdistribusi *asymptotically* normal.

Fungsi partial likelihood dapat ditulis sebagai perkalian dari likelihood masing-masing objek yang *event*-nya **teramati**.

Misalkan, $t_1 < t_2 < \cdots < t_n$ menyatakan *survival time* yang telah diurutkan dan hasil pengamatannya seperti pada tabel 3.6.

Tabel 3.6 Hasil Pengamatan

$\boldsymbol{i}$	$\boldsymbol{T_i}$	$\boldsymbol{\delta_i}$	$\boldsymbol{L_i}$
1	t_1	1	L_1
2	t_2	0	-
⋮	⋮	⋮	⋮
N	t_n	1	L_n

dimana

- i = data ke-i
- t_i = *survival time* ke-i

- $\delta_i = \begin{cases} 1, & event\ teramati \\ 0, & event\ tersensor \end{cases}$
- L_i = kontribusi likelihood dari data yang **teramati** pada waktu t_i.

Jika himpunan dari objek yang teramati dinotasikan dengan *D*, maka fungsi *partial* likelihoodnya adalah:

$$L_P(\boldsymbol{\beta}) = \prod_{j \in D} L_j$$

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* terhadap setiap parameter di dalam model, yaitu :

$$\frac{\partial \ln L_P}{\partial \beta_s} = 0, \qquad s = 1, 2, \ldots, k$$

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

3.2.1 Fungsi Partial Likelihood pada Model Cox Stratifikasi

Secara umum, distribusi dari variabel random terkait perlu diketahui untuk pembentukan fungsi likelihood. Namun, pada model Cox Stratifikasi, distribusi dari variabel respon (*survival time, T*) tidak diketahui. Inilah yang membedakannya dengan model parametrik.

Pembentukan fungsi partial likelihood dalam model Cox Stratifikasi berdasarkan pada urutan kejadian yang teramati. Setelah diurutkan, bentuk likelihood dari masing-masing data (objek) yang ***teramati***. Likelihood dari masing-masing data yang teramati menyatakan probabilitas objek tersebut mengalami *event* pada waktu t, bersyarat bahwa $T_i \geq t$. L_i bergantung pada himpunan objek-objek yang masih beresiko untuk mengalami *event* sampai waktu t_i yang disebut himpunan resiko ($R(t_{(i)})$). Banyaknya anggota dari himpunan ini akan berkurang dengan meningkatnya *survival time* (t_i).

Meskipun partial likelihood di atas tidak melibatkan objek yang tersensor, informasi dari objek yang tersensor tetap berkontribusi dalam pembentukan partial likelihoodnya. Objek tersebut berkontribusi dalam penentuan himpunan resiko pada waktu t_i dengan ketentuan sebagai berikut:

- Data yang tersensor pada dan setelah waktu t_i ; data tersebut masih menjadi anggota dari himpunan resiko pada waktu t_i
- Data yang tersensor sebelum waktu t_i ; data tersebut tidak lagi menjadi anggota dari himpunan resiko pada waktu t_i

Untuk menggambarkan fungsi partial likelihood dari model Cox Stratifikasi, perhatikan contoh berikut.

Misalkan, sebuah bank sedang mengadakan undian berhadiah. A, B, dan C masing-masing memiliki satu kupon undian. Dari ketiga individu tersebut akan dipilih pemenang pada waktu t_i (i = 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 individu menjadi pemenang bila urutan terpilihnya adalah A, kemudian B, dan terakhir C?

Berdasarkan urutannya, dapat dilihat dalam tabel 3.7.

Tabel 3.7 Probabilitas masing-masing individu

i	t_1	t_2	t_3
Individu yang terpilih	A	B	C
Jumlah kupon yang dimiliki	1	1	1
Total jumlah kupon yang tersedia	3	2	1
Probabilitas	1/3	1/2	1/1

1. Probabilitas A terpilih sebelum B dan C terpilih yaitu 1/3 (karena anggota himpunan resikonya masih ada 3 individu). Setelah A terpilih, kupon A tidak dapat dipilih lagi sehingga keluar dari himpunan resiko.
2. Probabilitas B terpilih sebelum C terpilih yaitu 1/2 (karena anggota himpunan resikonya masih ada 2 individu). Setelah B terpilih, kupon B tidak dapat dipilih lagi sehingga keluar dari himpunan resiko.

3. Probabilitas C terpilih terakhir yaitu 1/1 (karena anggota himpunan resiko tinggal C sendiri). Setelah C terpilih, tidak ada lagi anggota di dalam himpunan resiko.

Jadi, probabilitas terpilihnya pemenang dengan diberikan urutan kejadian seperti itu adalah :

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

Selanjutnya, contoh di atas akan dikembangkan. Misalkan A mempunyai 3 kupon, B mempunyai 2 kupon, dan C mempunyai 4 kupon. Berapa probabilitas terpilihnya pemenang undian jika urutan terpilihnya adalah A, lalu B, dan terakhir C?

Berdasarkan urutannya, dapat dilihat dalam tabel 3.8.

Tabel 3.8 Probabilitas masing-masing individu

i	t_1	t_2	t_3
Pemenang yang terpilih	A	B	C
Jumlah kupon yang dimiliki	3	2	4
Total jumlah kupon yang tersedia	9	6	4
Probabilitas	3/9	2/6	4/4

Jumlah kupon seluruhnya adalah 9 sehingga:

1. Bersyarat bahwa pada saat t_1 akan diberikan hadiah, maka :
   - Pr(A menang) = 3/9
   - Pr(B menang) = 2/9
   - Pr(C menang) = 4/9

   Pada saat t_1, salah satu dari A, B , dan C akan menjadi pemenang.

   Misal, M1 = kejadian terpilih 1 pemenang pada waktu t_1.

   A = kejadian terpilih A

   B = kejadian terpilih B

   C = kejadian terpilih C

   Maka :

$$P(A|M_1) = \frac{P(A \cap M_1)}{P(M_1)} = \frac{\frac{3}{9}}{\frac{3}{9} + \frac{2}{9} + \frac{4}{9}} = \frac{3}{9}$$

2. Bersyarat bahwa pada saat t_2 akan diberikan hadiah, maka :
   - Pr(B menang) = 2/6
   - Pr(C menang) = 4/6

   Pada saat t_2, salah satu dari B dan C akan menjadi pemenang.

   Misal, M2 = kejadian terpilih 1 pemenang pada waktu t_2

   B = kejadian terpilih B

   C = kejadian terpilih C

   Maka :

$$P(B|M_2) = \frac{P(B \cap M_2)}{P(M_2)} = \frac{\frac{2}{6}}{\frac{2}{6} + \frac{4}{6}} = \frac{2}{6}$$

3. Dengan cara yang sama, maka :

$$P(C|M_3) = \frac{P(C \cap M_3)}{P(M_3)} = \frac{\frac{4}{4}}{\frac{4}{4}} = \frac{4}{4} = 1$$

Jadi, probabilitas terpilihnya pemenang undian dengan diberikan urutan kejadian A, B, lalu C adalah :

$$\frac{3}{9} \times \frac{2}{6} \times \frac{4}{4} = \frac{1}{9}$$

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

Jika contoh-contoh tersebut dikaitkan dengan likelihood pada model Cox Stratifikasi, maka perhatikan tabel 3.9.

Tabel 3.9 Likelihood pada Model Cox Stratifikasi

				Notasi			Model Cox Stratifikasi
Urutan waktu kejadian	t_1	t_2	t_3	t_1	t_2	t_3	*time to event*
Probabilitas	A	B	C	A	B	C	Hazard rate dari objek yang teramati
	3/9	2/6	4/4	$h_A(t_1)$	$h_B(t_2)$	$h_C(t_3)$	
Total Jumlah Kupon	9	6	4	$R(t_1)$	$R(t_2)$	$R(t_3)$	Himpunan Resiko
Probabilitas bersyarat	$P(A \mid M_1)$	$P(B \mid M_2)$	$P(C \mid M_3)$	$\frac{h_A(t_1)}{h(t_1)}$	$\frac{h_B(t_2)}{h(t_2)}$	$\frac{h_C(t_3)}{h(t_3)}$	Kontribusi likelihood

Keterangan:

- $h_A(t_1)$ = hazard rate A pada waktu t_1
- $h_B(t_2)$ = hazard rate B pada waktu t_2
- $h_C(t_3)$ = hazard rate C pada waktu t_3
- $h(t_1) = h_A(t_1) + h_B(t_2) + h_C(t_3)$
- $h(t_2) = h_B(t_2) + h_C(t_3)$
- $h(t_3) = h_C(t_3)$

Berdasarkan tabel di atas, dapat disimpulkan bahwa *hazard rate* seseorang berperan penting dalam pembentukan fungsi partial likelihood.

Untuk lebih memahami bagaimana pembentukan likelihood pada model Cox Stratifikasi, maka perhatikan contoh berikut. Misal, kita memiliki data *survival* seperti pada tabel 3.10.

Tabel 3.10 Data *Survival*

i	***t_i***	$\boldsymbol{\delta}$	***X***
1	3	1	0
2	5	1	0
3	5	0	1
4	8	0	1
5	10	1	1

di mana $X = \begin{cases} 1, & untuk\ kategori\ A \\ 0, & untuk\ kategori\ B \end{cases}$ dan $\delta = \begin{cases} 1, & jika\ data\ teramati \\ 0, & jika\ data\ tersensor \end{cases}$.

Dari data tersebut diketahui bahwa ada sebanyak 3 individu yang mengalami *event* dan sebanyak 2 individu yang tersensor sehingga bentuk likelihoodnya sebagai berikut :

$$L_P = \prod_{j \epsilon D} L_j = L_1 \times L_2 \times L_3$$

di mana $L_j = \frac{h(t_j)}{\sum_{i \epsilon R(t_j)} h(t_i)} = \frac{h_0(t) \exp\left(X_{t_j}^T \beta\right)}{h_0(t) \sum_{i \in R(t_j)} \exp\left(X_{t_i}^T \beta\right)} = \frac{\exp\left(X_{t_j}^T \beta\right)}{\sum_{i \in R(t_j)} \exp\left(X_{t_i}^T \beta\right)}.$

Berdasarkan data tersebut, maka didapatkan

$$L_1 = \frac{\exp(X\beta)}{\sum_{i=1}^{5} \exp(X_i \beta)} = \frac{1}{1+1+e^{\beta}}$$

$$L_2 = \frac{\exp(X\beta)}{\sum_{i=1}^{4} \exp(X_i \beta)} = \frac{1}{1+e^{\beta}}$$

$$L_3 = \frac{\exp(X\beta)}{\sum_{i=1}^{1} \exp(X_i \beta)} = \frac{e^{\beta}}{e^{\beta}} = 1$$

sehingga

$$L_P = \prod_{j \epsilon D} L_j = L_1 \times L_2 \times L_3 = \frac{1}{1+1+e^{\beta}} . \frac{1}{1+e^{\beta}} . 1 = \frac{1}{(2+e^{\beta})(1+e^{\beta})}.$$

Misalkan ada k^* strata. Misalkan pula pada strata ke-g terdapat n_g *time to event* dan *Dg* adalah himpunan *time to event* yang berada pada strata ke-*g*. Misalkan pula dari n_g time to event tersebut, sebanyak r_g teramati dan n_g - r_g yang tersensor, maka pada model cox stratifikasi, fungsi partial likelihood didapatkan dari perkalian fungsi partial likelihood pada tiap strata seperti berikut:

$$L_P = \prod_g L_{Sg} = L_{S1} \times L_{S2} \times \ldots \times L_{Sk^*}, g = 1, 2, \ldots, k^*$$

di mana :

k^*= jumlah strata

L_g= likelihood pada strata ke-g, dengan

$$L_{Sg} = \prod_{j\epsilon\, D_g} L_j$$

Tabel 3.11 (*Partial) Likelihood Function*

Strata	1	2	…	k*
Likelihood	$L_{S1} = \prod_{j\epsilon\, D_1} \frac{h(t_j, \boldsymbol{X})}{\sum_{i\epsilon R(t_j)} h(t_i)} = \frac{h(t_1, \boldsymbol{X})}{\sum_{i\epsilon R(t_1)} h(t_i)} \times \frac{h(t_2, \boldsymbol{X})}{\sum_{i\epsilon R(t_2)} h(t_i)} \times \ldots \times \frac{h(t_{r_1}, \boldsymbol{X})}{\sum_{i\epsilon R(t_{r_1})} h(t_i)}$	$L_{S2} = \prod_{j\epsilon\, D_2} \frac{h(t_j, \boldsymbol{X})}{\sum_{i\epsilon R(t_j)} h(t_i)} = \frac{h(t_1, \boldsymbol{X})}{\sum_{i\epsilon R(t_1)} h(t_i)} \times \frac{h(t_2, \boldsymbol{X})}{\sum_{i\epsilon R(t_2)} h(t_i)} \times \ldots \times \frac{h(t_{r_2}, \boldsymbol{X})}{\sum_{i\epsilon R(t_{r_2})} h(t_i)}$	…	$L_{Sk^*} = \prod_{j\epsilon\, D_{k^*}} \frac{h(t_j, \boldsymbol{X})}{\sum_{i\epsilon R(t_j)} h(t_i)} = \frac{h(t_1, \boldsymbol{X})}{\sum_{i\epsilon R(t_1)} h(t_i)} \times \frac{h(t_2, \boldsymbol{X})}{\sum_{i\epsilon R(t_2)} h(t_i)} \times \ldots \times \frac{h(t_{r_{k^*}}, \boldsymbol{X})}{\sum_{i\epsilon R(t_{r_{k^*}})} h(t_i)}$

L_{Sg} adalah fungsi likelihood untuk masing-masing strata, di mana g = 1, 2,…, k*.

Untuk lebih memudahkan pemahaman, perhatikan contoh berikut.

Misal, ada sebanyak 4 strata di mana tiap-tiap strata memiliki ukuran sampel (n) masing-masing 5, 6, 5, dan 6 dengan rincian data seperti pada tabel 3.12.

Tabel 3.12 Data *Survival*

i	**Strata**							
	1		**2**		**3**		**4**	
	t	***X***	***t***	***X***	***t***	***X***	***t***	***X***
1	3	1	2+	0	5	1	3	1
2	4	0	3	0	6	1	4	1
3	5	1	3+	1	7	0	5	1
4	5+	0	4	1	8	1	6	0
5	6	1	6	0	9	0	7	0
6	-	-	8	0	-	-	8	1

Fungsi likelihood untuk tiap strata adalah:

$$L_{Sg}(\beta) = \prod_{j\epsilon\, D_g} \frac{\exp\left(X_{t_j}^T\beta\right)}{\sum_{i\in R(t_j)} \exp\left(X_{t_i}^T\beta\right)}, g = 1,2,3,4$$

- Untuk strata 1 (g = 1)

- i = 1 (t = 3, $X = 1$)

$$\frac{e^\beta}{e^\beta + e^\beta + e^\beta + e^0 + e^0} = \frac{e^\beta}{3e^\beta + 2}$$

- i = 2 (t = 4, $X = 0$)

$$\frac{e^0}{e^\beta + e^\beta + e^0 + e^0} = \frac{1}{2e^\beta + 2}$$

- i = 3 (t = 5, $X = 1$)

$$\frac{e^\beta}{e^\beta + e^\beta + e^0} = \frac{e^\beta}{2e^\beta + 1}$$

- i = 4 (t = 5+, $X = 0$)

  Karena data tersebut merupakan data tersensor, maka tidak dimasukkan dalam fungsi likelihoohnya.

- i = 5 (t = 6, $X = 1$)

$$\frac{e^\beta}{e^\beta} = 1$$

Sehingga fungsi likelihood untuk strata 1 adalah sebagai berikut:

$$L_{S1}(\beta) = \prod_{j\epsilon\, D_1} \frac{\exp\left(X_{t_j}^T \beta\right)}{\sum_{i\in R(t_j)} \exp\left(X_{t_i}^T \beta\right)} = \frac{e^{\beta}}{3e^{\beta}+2} \cdot \frac{1}{2e^{\beta}+2} \cdot \frac{e^{\beta}}{2e^{\beta}+1} \cdot 1$$

$$= \frac{e^{2\beta}}{(3e^{\beta}+2)(2e^{\beta}+2)(2e^{\beta}+1)}$$

- Untuk strata 2 (g = 2)

Dengan cara yang sama pada strata 1, maka fungsi likelihood pada strata ke 2 adalah sebagai berikut:

$$L_{S2}(\beta) = \prod_{j\epsilon\, D_2} \frac{\exp\left(X_{t_j}^T \beta\right)}{\sum_{i\in R(t_j)} \exp\left(X_{t_i}^T \beta\right)}$$

$$= \frac{1}{2e^{\beta}+3} \cdot \frac{e^{\beta}}{e^{\beta}+2} \cdot \frac{1}{2} \cdot 1$$

$$= \frac{e^{\beta}}{2(2e^{\beta}+3)(e^{\beta}+2)}$$

- Untuk strata 3 (g = 3)

$$L_{S3}(\beta) = \prod_{j\epsilon\, D_3} \frac{\exp\left(X_{t_j}^T \beta\right)}{\sum_{i\in R(t_j)} \exp\left(X_{t_i}^T \beta\right)}$$

$$= \frac{e^{\beta}}{3e^{\beta}+2} \cdot \frac{e^{\beta}}{2e^{\beta}+2} \cdot \frac{1}{e^{\beta}+2} \cdot \frac{e^{\beta}}{e^{\beta}+1} \cdot 1$$

$$= \frac{e^{3\beta}}{(3e^{\beta}+2)(2e^{\beta}+2)(e^{\beta}+2)(e^{\beta}+1)}$$

- Untuk strata 4 (g = 4)

$$L_{S4}(\beta) = \prod_{j\epsilon\, D_4} \frac{\exp\left(X_{t_j}^T \beta\right)}{\sum_{i\in R(t_j)} \exp\left(X_{t_i}^T \beta\right)}$$

$$= \frac{e^{\beta}}{4e^{\beta}+2} \cdot \frac{e^{\beta}}{3e^{\beta}+2} \cdot \frac{e^{\beta}}{2e^{\beta}+2} \cdot \frac{1}{e^{\beta}+2} \cdot \frac{1}{e^{\beta}+1} \cdot 1$$

$$= \frac{e^{3\beta}}{(4e^{\beta}+2)(3e^{\beta}+2)(2e^{\beta}+2)(e^{\beta}+2)(e^{\beta}+1)}$$

Sehingga didapatkan fungsi parsial likelihoodnya adalah sebagai berikut :

$$L_P = L_{S1} \times L_{S2} \times L_{S3} \times L_{S4}$$

$$= \frac{e^{2\beta}}{(3e^{\beta} + 2)(2e^{\beta} + 2)(2e^{\beta} + 1)} \times \frac{e^{\beta}}{2(2e^{\beta} + 3)(e^{\beta} + 2)}$$

$$\times \frac{e^{3\beta}}{(3e^{\beta} + 2)(2e^{\beta} + 2)(e^{\beta} + 2)(e^{\beta} + 1)}$$

$$\times \frac{e^{3\beta}}{(4e^{\beta} + 2)(3e^{\beta} + 2)(2e^{\beta} + 2)(e^{\beta} + 2)(e^{\beta} + 1)}$$

$$= \frac{e^{9\beta}}{32(3e^{\beta} + 2)^3(e^{\beta} + 1)^6(2e^{\beta} + 1)^2(2e^{\beta} + 3)(e^{\beta} + 2)^2}$$

$$\ln L_P = 9\beta - 3\ln\left(3e^{\beta} + 2\right) - \ln 32 - 6\ln\left(e^{\beta} + 1\right) - 2\ln\left(2e^{\beta} + 1\right)$$
$$- \ln(2e^{\beta} + 3) - 2\ln\left(e^{\beta} + 2\right)$$

Untuk menaksir $\boldsymbol{\beta}$, maka selesaikan persamaan $\frac{\partial \ln L_P}{\partial \beta} = 0.$

Berdasarkan contoh di atas, secara umum langkah-langkah untuk menentukan partial likelihood dari model Cox Stratifikasi adalah sebagai berikut:

1. Bentuk model Cox Stratifikasi
2. Masukkan data ke dalam masing-masing strata
3. Urutkan waktu kejadian pada masing-masing strata
4. Bentuk fungsi likelihood untuk tiap strata

- Konstruksi kontribusi masing-masing objek yang diamati pada himpunan *D* terhadap fungsi partial likelihood. Misal, individu A pada strata 1 di waktu t_j dengan fungsi hazard $h_A\left(t_j, \boldsymbol{X}(t_j)\right)$, maka

$$L_A = \frac{h_A\left(t_j, \boldsymbol{X}(t_j)\right)}{\sum_{i \in R(t_{(j)})} h_i(t_i, \boldsymbol{X}(t_i))}$$

dimana

L_A = Probabilitas bahwa individu A mengalami *event* di waktu t_j bersyarat bahwa terdapat suatu *event* di waktu t_j

$h_A\left(t_j, X(t_j)\right)$ = *hazard rate* A dengan *covariate* $\boldsymbol{X}$ mengalami *event* di waktu t_j

$\sum_{i\in R(t_{(j)})} h_i(t_i, X(t_i))$ = jumlah *hazard rate* dari individu-individu yang beresiko mengalami *event* di t_j

Mengacu pada persamaan (3.2),

$$L_A = \frac{h_{0g}(t)\exp[\beta_1 X_1 + \beta_2 X_2 \ldots + \beta_k X_k]}{\sum_{\mathrm{i\in R(t_{(j)})}} h_{0g}(t)\exp[\beta_1 X_1 + \beta_2 X_2 \ldots + \beta_k X_k]}$$

$$= \frac{\exp[\beta_1 X_1 + \beta_2 X_2 \ldots + \beta_k X_k]}{\sum_{\mathrm{i\in R(t_{(j)})}} \exp[\beta_1 X_1 + \beta_2 X_2 \ldots + \beta_k X_k]}$$

Misal, $\psi_A = \exp[\beta_1 X_1 + \beta_2 X_2 \ldots + \beta_k X_k]$, maka

$$L_A = \frac{\psi_A}{\sum_{i\in R(t_{(j)})} \psi_i}$$

Proses yang sama dilakukan untuk objek lain yang menjadi anggota himpunan *D* pada strata 1 ($j\epsilon D_1$). Selanjutnya, lakukan hal yang sama pada strata 2, 3,…, k*.

- Fungsi likelihood untuk tiap strata

$$L_{S_g}(\boldsymbol{\beta}) = \prod_{j\epsilon\, D_g} L_j = \prod_{j\epsilon\, D_g} \frac{\psi_j}{\sum_{i\epsilon R(t_j)} \psi_i}, \qquad g = 1,2, \ldots, k^*$$

5. Kalikan fungsi likelihood pada seluruh strata

$$L_P(\boldsymbol{\beta}) = \prod_{g=1}^{k^*} L_{S_g} = \prod_{g=1}^{k^*}\prod_{j\epsilon\, D_g} L_j = \prod_{g=1}^{k^*}\prod_{j\epsilon\, D_g} \frac{\psi_j}{\sum_{i\epsilon R(t_j)} \psi_i}$$

6. Bentuk $\ln L_P(\boldsymbol{\beta})$

$$\ln L_P(\boldsymbol{\beta}) = \ln\left[\prod_{g=1}^{k^*} \prod_{j\epsilon\, D_g} \frac{\psi_j}{\sum_{i\epsilon R(t_j)} \psi_i}\right]$$

$$= \sum_{g=1}^{k^*} \left[\ln \prod_{j\epsilon\, D_g} \frac{\psi_j}{\sum_{i\epsilon R(t_j)} \psi_i}\right]$$

$$= \sum_{g=1}^{k^*} \left[\sum_{j\epsilon\, D_g} \ln \psi_j - \ln \sum_{i\epsilon R(t_j)} \psi_i\right]$$

$$= \sum_{g=1}^{k^*} \left[\sum_{j\epsilon\, D_g} X_{t_j}\boldsymbol{\beta} - \ln \sum_{i\epsilon R(t_j)} e^{X_{t_i}\boldsymbol{\beta}}\right]$$

$$= \sum_{g=1}^{k^*} \sum_{j\epsilon\, D_g} X_{t_j}\boldsymbol{\beta} - \ln \sum_{i\epsilon R(t_j)} e^{X_{t_i}\boldsymbol{\beta}}$$

7. Selanjutnya, cari taksiran dari parameter-parameter model dengan metode *Maximum Partial Likelihood Estimator*.

- $\frac{\partial \ln L_p}{\partial \beta_1} = 0$

$$\frac{\partial}{\partial \beta_1}\left(\ln \prod_{g=1}^{k^*} \prod_{j\epsilon\, D_g(t_j)} \frac{\psi_j}{\sum_{i\epsilon R(t_j)} \psi_i}\right) = \frac{\partial}{\partial \beta_1}\left(\sum_{g=1}^{k^*}\left[\ln \prod_{j\epsilon\, D_g(t_j)} \frac{\psi_j}{\sum_{i\epsilon R(t_j)} \psi_i}\right]\right)$$

$$= \frac{\partial}{\partial \beta_1}\left(\sum_{g=1}^{k^*}\left[\sum_{j\epsilon\, D_g(t_j)} \ln \psi_j - \ln \sum_{i\epsilon R(t_j)} \psi_i\right]\right)$$

$$= \frac{\partial}{\partial \beta_1}\sum_{g=1}^{k^*}\sum_{j\epsilon\, D_g(t_j)} X_{t_j}\boldsymbol{\beta} - \frac{\partial}{\partial \beta_1}\sum_{g=1}^{k^*}\ln \sum_{i\epsilon R(t_j)} e^{X_{t_i}\boldsymbol{\beta}}$$

$$= 0$$

- $\frac{\partial \ln L_p}{\partial \beta_2} = 0$

$$\frac{\partial}{\partial \beta_2}\left(\ln \prod_{g=1}^{k^*} \prod_{j\epsilon\, D_g(t_j)} \frac{\psi_j}{\sum_{i\epsilon R(t_j)} \psi_i}\right) = \frac{\partial}{\partial \beta_2}\left(\sum_{g=1}^{k^*}\left[\ln \prod_{j\epsilon\, D_g(t_j)} \frac{\psi_j}{\sum_{i\epsilon R(t_j)} \psi_i}\right]\right)$$

$$= \frac{\partial}{\partial \beta_2}\left(\sum_{g=1}^{k^*}\left[\sum_{j\epsilon\, D_g(t_j)} \ln \psi_j - \ln \sum_{i\epsilon R(t_j)} \psi_i\right]\right)$$

$$= \frac{\partial}{\partial \beta_2}\sum_{g=1}^{k^*}\sum_{j\epsilon\, D_g(t_j)} X_{t_j}\boldsymbol{\beta} - \frac{\partial}{\partial \beta_2}\sum_{g=1}^{k^*}\ln \sum_{i\epsilon R(t_j)} e^{X_{t_i}\boldsymbol{\beta}}$$

$$= 0$$

Secara umum,

$$\frac{\partial \ln L_P}{\partial \beta_s} = \frac{\partial}{\partial \beta_2}\sum_{g=1}^{k^*}\sum_{j\epsilon\, D_g(t_j)} X_{t_j}\boldsymbol{\beta} - \frac{\partial}{\partial \beta_2}\sum_{g=1}^{k^*}\ln \sum_{i\epsilon R(t_j)} e^{X_{t_i}\boldsymbol{\beta}} = 0,$$

$$s = 1, 2, \ldots, k$$

Persamaan tersebut dapat diselesaikan dengan iterasi.

Robbin (1989) & Tsiatis (1990) telah membuktikan bahwa taksiran yang didapat berdistribusi *asymptotically* normal:

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

dimana

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

dengan

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

3.2.2 Partial Likelihood untuk Model Cox Stratifikasi jika Terdapat Ties

Partial likelihood pada subbab 3.2.1 di atas adalah partial likelihood yang digunakan pada data yang tidak mengandung *ties*. Sedangkan pada kenyataannya, seringkali data survival yang kita punya mengandung *ties*. *Ties* adalah keadaan di mana terdapat dua individu atau lebih yang mengalami *event* pada waktu yang sama. Pada subbab ini, akan dibahas partial likelihood untuk data yang mengandung *ties*.

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

Tabel 3.13 Data *Survival* dengan terdapat *ties*

i	1	2	3	4	5
t_i	1+	2	2	3	5+

Pada waktu t=2, terdapat dua objek yang mengalami event. Tidak diketahui objek mana yang mengalami event terlebih dahulu. Untuk lebih memudahkan dalam penulisan, fungsi hazard dari objek ke-j dinotasikan dengan ψ_j .

1. Misalkan dianggap objek ke-2 yang mengalami event terlebih dahulu, maka partial likelihoodnya adalah:

$$L_{P1}(\boldsymbol{\beta}) = \frac{\psi_2}{(\psi_2 + \psi_3 + \psi_4 + \psi_5)} \frac{\psi_3}{(\psi_3 + \psi_4 + \psi_5)} \frac{\psi_4}{(\psi_4 + \psi_5)}$$

2. Bila dianggap objek ke-3 yang mengalami event terlebih dahulu, maka partial likelihoodnya adalah:

$$L_{P2}(\boldsymbol{\beta}) = \frac{\psi_3}{(\psi_2 + \psi_3 + \psi_4 + \psi_5)} \frac{\psi_2}{(\psi_2 + \psi_4 + \psi_5)} \frac{\psi_4}{(\psi_4 + \psi_5)}$$

Nilai $\psi_2 \neq \psi_3$ yang menyebabkan perbedaan pada himpunan resikonya sehingga

$$L_{P1}(\boldsymbol{\beta}) \neq L_{P2}(\boldsymbol{\beta}).$$

Partial likelihood yang diberikan ketika menganggap objek ke-2 yang terjadi lebih dahulu akan berbeda nilainya bila menganggap objek ke-3 yang terjadi lebih dahulu. Seharusnya, untuk data yang sama maka likelihoodnya juga harus sama. Inilah alasan mengapa urutan sangat penting dalam pembentukan partial likelihood.

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

Pada pendekatan Efron, himpunan resikonya diselesaikan dengan pengurangan terhadap rata-rata dari nilai $\psi_2\ dan\ \psi_3$ karena tidak diketahui objek mana yang mengalami *event* terlebih dahulu. Bentuk partial likelihoodnya adalah sebagai berikut:

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

Secara umum, partial likelihood untuk pendekatan Efron yaitu :

$$L_P(\boldsymbol{\beta}) \approx \prod_{j \epsilon D} \frac{\prod_{i \epsilon D} \psi_i}{\prod_{q=1}^{d_j} \left[\sum_{i \epsilon R(t_{(j)})} \psi_i - {(q-1)}/{d_q} \left(\sum_{i \epsilon D} \psi_i\right)\right]}$$

dimana

D = himpunan dari objek yang *event*-nya teramati

$R(t_j)$ = himpunan resiko

q = objek yang teramati

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

BAB IV
CONTOH PENERAPAN

Dalam bab ini, akan dibahas penerapan model Cox Stratifikasi pada data. Kemudian akan dijelaskan interpretasi dari taksiran yang telah diperoleh. Selain itu, akan dibentuk juga model cox *proportional hazard*. Perangkat lunak yang akan digunakan dalam penaksiran parameter model, baik model *Cox Proportional Hazard* maupun model Cox Stratifikasi adalah perangkat lunak R versi 2.9.1.

4.1 Data

Data yang digunakan dalam tugas akhir ini adalah data yang diteliti oleh Kalbfleisch dan Prentice pada tahun 1980. Data ini menyatakan waktu sampai meninggal (dalam hari) seseorang yang mengidap penyakit kanker paru-paru.

- Awal : awal dilakukan treatment
- Akhir : meninggal, akhir masa study
- Event : meninggal
- Skala waktu : hari

Variabel penting yang akan diamati pada data ini adalah *treatment status*. Ada dua jenis treatment status, yaitu:

$$X_1 = \begin{cases} 1; standar \\ 2; \quad test \end{cases}$$

Selain itu, ada variabel-variabel lain yang berguna sebagai variabel control adalah *cell type*, *performance status*, dan *age*, di mana:

$$Cell\ type = \begin{cases} large \\ adeno \\ small \\ squamous \end{cases}$$

Data ini terdiri dari 137 observasi, di mana :

- Ada 69 pasien yang mendapat perlakuan standar $(X_1 = 1)$ dan 68 pasien yang mendapat perlakuan test $(X_1 = 2)$
- Ada 9 pasien yang tersensor dan sebanyak 128 pasien yang mengalami *event*

Untuk lebih jelasnya, data dapat dilihat pada lampiran 1.

4.1.1 Model *Cox Proportional Hazard* (PH)

Dengan menggunakan software R, akan didapatkan taksiran parameter pada model cox PH seperti pada tabel 4.1.

Tabel 4.1 Output R untuk model cox PH

	coef	exp(coef)	se(coef)	z	p
Treat	0.3030	1.354	0.20566	1.474	0.14
Cel1	0.4023	1.495	0.28254	1.424	0.15
Cel2	1.1788	3.250	0.29644	3.977	0.00007
Cel3	0.8563	2.355	0.27132	3.156	0.0016
Cel4	NA	NA	NA	NA	NA
Perform	-0.0327	0.968	0.00541	-6.043	0.0000000015
Age	-0.0089	0.991	0.00922	-0.965	0.33

Model Cox PH :

$$h(t,X) = h_0(t)e^{0.3030X_1+0.4023X_2+1.1788X_3+0.8563X_4-0.0327X_5-0.0089X_6}$$

di mana

X_1 = treatment ($X_1 = 1$ → standar, $X_1 = 2$ → test)

X_2 = cel1 (X_2 = 1 → large, X_2 = 0 → bukan)

X_3 = cel2 (X_3 = 1 → adeno, X_3 = 0 → bukan)

X_4 = cel3 (X_4 = 1 → small, X_4 = 0 → bukan)

X_5 = perform

X_6 = age

Dengan asumsi sudah dilakukan pengecekan asumsi, hasilnya ada 2 variabel yang tidak memenuhi asumsi PH, yaitu *performance status* dan *cell type*. Oleh karena itu, akan dilakukan modifikasi model Cox PH menjadi model Cox Stratifikasi.

4.1.2 Model Cox Stratifikasi

Seperti yang telah dijelaskan pada bab 3, tahap-tahap yang harus dilakukan dalam pembentukan model Cox Stratifikasi antara lain :

1. Identifikasi variabel yang tidak memenuhi asumsi PH

   Berdasarkan hasil dari pengecekan asumsi PH, variabel yang tidak memenuhi asumsi PH adalah *performance status* dan *cell type*.

2. Definisikan variabel baru (Z^*)

   Dalam pendefinisian variabel baru, ada beberapa hal yang harus dilakukan, yaitu :

   2.1 Kategorisasi variabel yang tidak memenuhi asumsi

   - *performance status*

     Berdasarkan data, variabel ini adalah variabel numerik dengan interval antara 0 (worst) - 100 (best), sehingga perlu dikategorikan. Variabel ini akan dikategorikan menjadi 2 kategori dengan titik batasnya 60. Sehingga hasilnya adalah :

$$perform = \begin{cases} 1, < 60 \\ 0, \geq 60 \end{cases}$$

   - *cell type*

     Berdasarkan data, telah diketahui bahwa ada 4 kategori dari *cell type*, yaitu *large*, *adeno*, *small*, dan *squamous*.

   2.2 Kombinasi seluruh kategori

   Hasil kombinasi dari 2 variabel tersebut terlihat pada tabel 4.2.

   Tabel 4.2 Hasil kombinasi variabel *Performance status* dan *cell type*

Performance status	Cell Type			
	Large (1)	Adeno (2)	Small (3)	Squamous (4)
1	Strata 1	Strata 3	Strata 5	Strata 7
0	Strata 2	Strata 4	Strata 6	Strata 8

Berdasarkan tabel di atas, terlihat bahwa ada sebanyak 8 kombinasi dari kedua variabel yang tidak memenuhi asumsi PH. Oleh karena itu, jumlah strata yang terbentuk adalah sebanyak 8 strata.

3. Buat model Cox Stratifikasi

Model Cox Stratifikasi yang terbentuk adalah sebagai berikut :

$$h_g(t) = h_{0g}(t)e^{X_1\beta_1 + X_2\beta_2}$$

di mana :

- g menyatakan strata, $g = 1,2,\ldots,8$
- X_1 = treatment
- X_2 = age

Dengan menggunakan software R, akan didapatkan taksiran parameter pada model cox stratifikasi seperti pada tabel 4.3.

Tabel 4.3 Output R untuk model Cox Stratifikasi

	coef	**exp(coef)**	**se(coef)**	**Z**	**p**
Treat	0.12778	1.136	0.2085	0.613	0.54
age	-0.00127	0.999	0.0101	-0.126	0.90

Jadi, model Cox Stratifikasinya adalah :

$$h_g(t) = h_{0g}(t)e^{0.12778X_1 - 0.00127X_2}$$

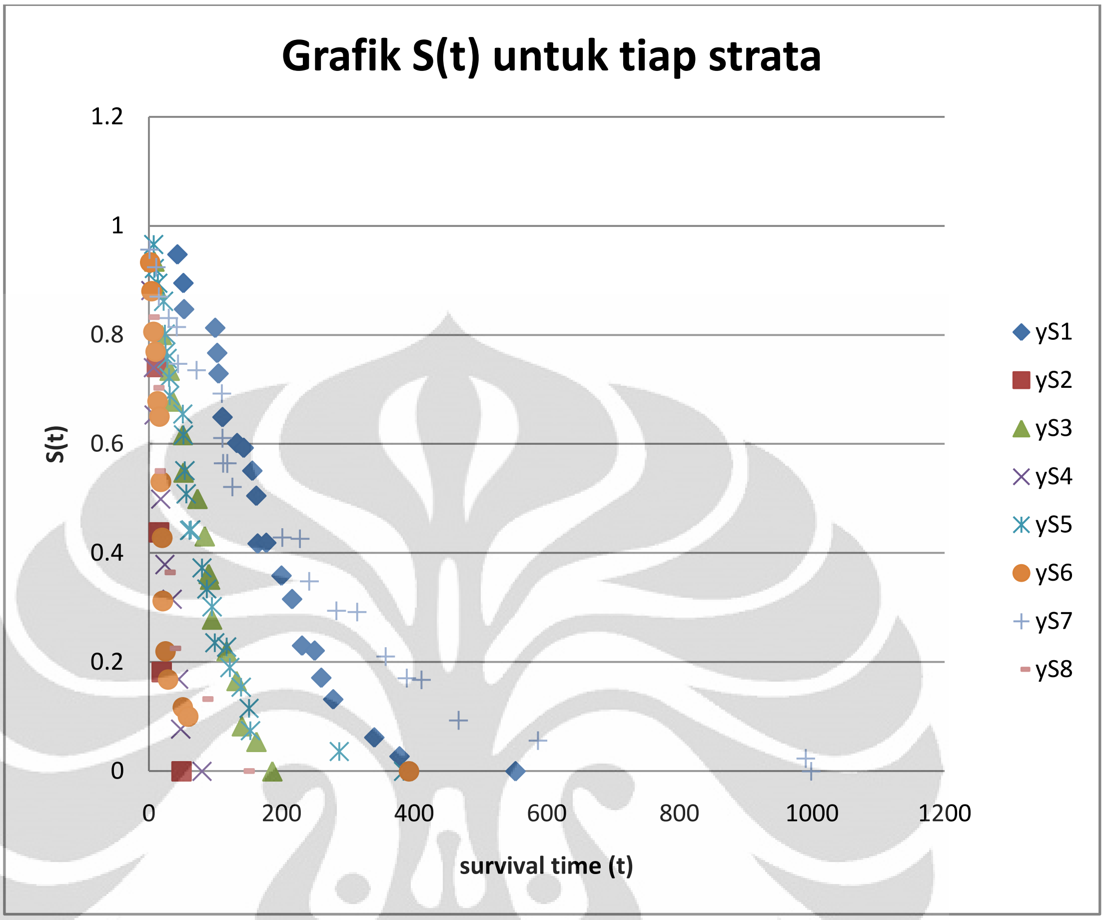
di mana :

X_1 = treatment

X_2 = age

g = 1, 2,…, 8

Dari kedelapan model di atas, dapat dilihat perbedaannya melalui grafik fungsi hazard dari masing-masing strata. Untuk lebih memudahkan, kita bisa menggunakan grafik $S(t)$, dengan nilai $S_0g(t)$nya dapat kita taksir menggunakan taksiran Kaplan-Meier.



Gambar 5.1 Grafik *S*(*t*) untuk tiap strata

Berdasarkan grafik di atas, secara umum menunjukkan *survival experience* yang berbeda untuk setiap strata. Meskipun demikian, ada pengelompokan pada masing-masing strata, yaitu:

- yang menujukkan *survival* baik adalah strata pertama dan ketujuh. Dari kedua strata tersebut, yang paling *survive* adalah strata ketujuh. Berdasarkan stratifikasi variabel, strata ketujuh adalah individu-individu dengan karakteristik *performance status "worst"* dan memiliki tipe sel *squamous*.
- yang menunjukkan *survival* sedang adalah strata ketiga, kelima, dan kedelapan. Dari ketiga strata tersebut, yang lebih *survive* adalah strata kelima. Berdasarkan stratifikasi variabel, strata lima berasal dari *performance status "worst*" dan memiliki tipe sel *small*.
- yang menunjukkan *survival* buruk adalah strata kedua, keempat, dan keenam. Dari ketiga strata tersebut, yang paling baik adalah strata empat.

Berdasarkan stratifikasi variabel, strata empat berasal dari *performance status "best"* dan memiliki tipe sel *adeno*.

Dari kedelapan grafik, dapat diketahui bahwa strata dengan *survival experience* terbaik adalah strata ketujuh, yaitu individu-individu dengan karakteristik *performance status "worst"* dan tipe sel *squamous*. Sedangkan strata dengan *survival experience* terburuk adalah strata kedua, yaitu individu-individu dengan karakteristik *performance status "best"* dan tipe sel *large*.

Pada model cox PH, berdasarkan output pada tabel 4.1, individu-individu dengan *performance status "worst"* akan lebih *survive* daripada individu dengan *performance status "best"*. Sedangkan individu-individu dengan tipe sel *large* akan lebih *survive* daripada individu-individu yang bertipe sel *adeno*, *small* atau *squamous*.

Pada kenyataannya, individu dengan *performance status "worst"* belum tentu lebih *survive* daripada individu dengan *performance status "best"*. Hal ini dapat dilihat pada grafik strata keempat. Individu pada strata empat, yaitu individu *dengan performance status "best"*, lebih *survive* daripada individu pada strata lima, *dengan performance status "worst"*. Selain itu, individu dengan tipe sel *large* tidak selalu lebih *survive* daripada individu dengan tipe sel lain. Misalnya, untuk strata satu dan tujuh. Pada saat t = 200, individu pada strata satu lebih *survive* daripada individu pada strata tujuh. Sedangkan pada saat t > 200, berlaku sebaliknya.

Dengan membandingkan hasil dari kedua model, dapat disimpulkan bahwa model cox stratifikasi dapat digunakan untuk memperbaiki model cox PH. Hal ini disebabkan karena pada model cox PH, variabel *performance status* dan *cell type* dianggap proportional terhadap waktu. Padahal, pada kenyataannya tidak.

BAB 5
PENUTUP

5.1 Kesimpulan

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

1. Jika terdapat covariate yang tidak memenuhi asumsi PH pada model cox proportional hazard, maka dapat diatasi dengan menggunakan model cox stratifiikasi. Hal ini dapat dilakukan dengan membagi data atas beberapa strata, yang ditentukan berdasarkan kategorisasi *covariate* yang tidak memenuhi asumsi PH. Kemudian, dibuat model cox PH pada masing-masing strata.
2. Penaksiran parameter model cox stratifikasi dilakukan dengan metode *maximum partial likelihood*, di mana 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-variabel *treatment, cell type, performance status*, dan *age* terhadap waktu sampai pasien penderita kanker paru-paru mengalami kematian setelah dilakukan suatu *treatment*. Dari hasil analisis diperoleh bahwa model cox stratifikasi dapat menjelaskan dengan lebih baik daripada model cox PH.

5.2 Saran

Saran untuk pengembangan skripsi ini adalah

1. Dapat dijelaskan mengenai bentuk model cox stratifikasi dengan interaksi.
2. Untuk kasus data yang terdapat *ties*, maka dapat diterapkan pendekatan lain, misalnya Breslow atau Diskrit.
3. Dapat dijelaskan mengenai model cox stratifikasi dengan tipe data tersensor kiri ataupun tersensor interval.

DAFTAR PUSTAKA

Daowen, Zhang. *Modelling Survival Data with Paranetric Regression Model.* *http://www.ncbi.nlm.nih.gov/pubmed/1480879*

Dorota M, Dabrowska. 1987. *Non-parametric Regression with Censored Survival Time Data.http://www.jstor.org/*

Klein, J.P. and Moeschberger, M.L. 1997. *Survival Analysis – Techniques for Censored and Truncated Data.* New York: Springer-Verlag

Kleinbaum, D.G. and Klein, M. 2005. *Survival Analysis – A Self Learning Text, Second Edition.* New York: Springer

Rober V. Hogg and Allen T. Craig. 1995. *Introduction to Mathematical Statistics*. New Jersey: Prentice-Hall

LAMPIRAN

Lampiran 1

Data yang digunakan pada contoh penerapan.

No.	treatment	cel1	cel2	cel3	cel4	survival	perform	age	status
1	2	0	0	0	1	1	20	65	1
2	2	0	0	0	1	1	50	35	1
3	2	0	0	1	0	2	40	44	1
4	1	0	1	0	0	3	30	43	1
5	1	0	0	1	0	4	40	35	1
6	1	0	0	1	0	7	50	72	1
7	2	0	0	1	0	7	20	66	1
8	2	0	1	0	0	7	40	58	1
9	1	0	0	0	1	8	40	63	1
10	1	0	1	0	0	8	20	61	1
11	2	0	0	1	0	8	80	68	1
12	2	0	1	0	0	8	50	66	1
13	1	0	0	0	1	10	20	49	1
14	1	0	0	1	0	10	40	67	1
15	1	0	0	0	1	11	70	48	1
16	1	0	1	0	0	12	50	63	1
17	1	1	0	0	0	12	40	68	1
18	1	0	0	1	0	13	60	56	1
19	2	0	0	1	0	13	30	62	1
20	2	0	0	0	1	15	50	40	1
21	2	1	0	0	0	15	30	63	1
22	1	0	0	1	0	16	30	53	1
23	1	0	0	1	0	18	20	42	1
24	1	0	0	1	0	18	30	60	1
25	2	0	1	0	0	18	40	69	1
26	2	0	1	0	0	19	50	42	1
27	2	1	0	0	0	19	30	39	1
28	1	0	0	1	0	20	30	65	1
29	2	0	0	1	0	20	30	54	1
30	1	0	0	1	0	21	40	55	1
31	2	0	0	1	0	21	20	71	1
32	1	0	0	1	0	22	60	68	1
33	2	0	0	1	0	24	60	49	1

34	2	0	1	0	0	24	40	60	1
35	1	0	0	0	1	25	80	52	0
36	2	0	0	0	1	25	20	63	1
37	2	0	0	1	0	25	30	69	1
38	2	0	0	1	0	25	70	70	1
39	1	0	0	1	0	27	60	62	1
40	2	0	0	1	0	29	40	67	1
41	1	0	0	1	0	30	60	61	1
42	2	0	0	0	1	30	70	63	1
43	1	0	0	1	0	31	75	65	1
44	2	0	1	0	0	31	80	39	1
45	2	0	0	0	1	33	30	64	1
46	1	0	1	0	0	35	40	62	1
47	2	0	1	0	0	36	70	61	1
48	1	0	0	0	1	42	60	81	1
49	2	1	0	0	0	43	60	49	1
50	2	0	0	0	1	44	60	70	1
51	2	0	1	0	0	45	40	69	1
52	2	0	1	0	0	48	10	81	1
53	2	1	0	0	0	49	30	37	1
54	1	0	0	1	0	51	60	67	1
55	2	0	0	1	0	51	30	59	1
56	2	0	1	0	0	51	60	62	1
57	1	0	0	1	0	52	70	55	1
58	2	0	1	0	0	52	60	43	1
59	2	1	0	0	0	52	60	45	1
60	2	1	0	0	0	53	60	66	1
61	1	0	0	1	0	54	80	63	1
62	1	0	0	1	0	54	70	67	1
63	1	0	0	1	0	56	80	43	1
64	1	0	0	1	0	59	30	65	1
65	2	0	0	1	0	61	70	71	1
66	1	0	0	1	0	63	50	48	1
67	1	0	0	0	1	72	60	69	1
68	2	0	1	0	0	73	60	70	1
69	2	0	0	1	0	80	50	71	1
70	2	0	1	0	0	80	40	63	1
71	1	0	0	0	1	82	40	69	1
72	2	0	1	0	0	83	99	57	0
73	2	0	1	0	0	84	80	62	1
74	2	0	0	0	1	87	80	48	0
75	2	0	0	1	0	87	60	60	1

76	2	0	1	0	0	90	60	50	1
77	1	0	1	0	0	92	70	60	1
78	1	0	1	0	0	95	80	34	1
79	2	0	0	1	0	95	70	61	1
80	1	0	0	1	0	97	60	67	0
81	2	0	0	1	0	99	70	72	1
82	2	0	0	1	0	99	85	62	1
83	1	0	0	0	1	100	70	70	0
84	1	1	0	0	0	100	60	37	1
85	1	1	0	0	0	103	80	38	1
86	2	0	0	1	0	103	70	36	0
87	1	1	0	0	0	105	80	66	1
88	1	0	0	0	1	110	80	68	1
89	2	0	0	0	1	111	70	62	1
90	2	1	0	0	0	111	60	64	1
91	2	0	0	0	1	112	80	60	1
92	1	0	0	1	0	117	80	46	1
93	1	0	1	0	0	117	80	38	1
94	1	0	0	0	1	118	70	65	1
95	1	0	0	1	0	122	80	53	1
96	1	0	0	1	0	123	40	55	0
97	1	0	0	0	1	126	60	63	1
98	1	0	1	0	0	132	80	50	1
99	2	1	0	0	0	133	75	65	1
100	1	0	0	1	0	139	80	64	1
101	2	0	1	0	0	140	70	63	1
102	1	1	0	0	0	143	90	60	1
103	1	0	0	0	1	144	30	63	1
104	1	0	0	1	0	151	50	69	1
105	1	0	0	1	0	153	60	63	1
106	1	1	0	0	0	156	70	66	1
107	1	0	1	0	0	162	80	64	1
108	1	1	0	0	0	162	80	62	1
109	2	1	0	0	0	164	70	68	1
110	1	1	0	0	0	177	50	66	1
111	1	1	0	0	0	182	90	62	0
112	2	0	1	0	0	186	90	60	1
113	1	1	0	0	0	200	80	41	1
114	2	0	0	0	1	201	80	52	1
115	1	1	0	0	0	216	50	52	1
116	1	0	0	0	1	228	60	38	1
117	2	0	0	0	1	231	50	52	0

118	2	1	0	0	0	231	70	67	1
119	2	0	0	0	1	242	50	70	1
120	1	1	0	0	0	250	70	53	1
121	1	1	0	0	0	260	80	45	1
122	1	1	0	0	0	278	60	63	1
123	2	0	0	0	1	283	90	51	1
124	1	0	0	1	0	287	60	66	1
125	1	0	0	0	1	314	50	43	1
126	2	1	0	0	0	340	80	64	1
127	2	0	0	0	1	357	70	58	1
128	2	1	0	0	0	378	80	65	1
129	1	0	0	1	0	384	60	42	1
130	2	0	0	0	1	389	90	62	1
131	1	0	0	1	0	392	40	68	1
132	1	0	0	0	1	411	70	64	1
133	2	0	0	0	1	467	90	64	1
134	1	1	0	0	0	553	70	47	1
135	2	0	0	0	1	587	60	58	1
136	2	0	0	0	1	991	70	50	1
137	2	0	0	0	1	999	90	54	1

Lampiran 2

Output

R-versi 2.9.1

Sebelumnya pilih menu “Package” → “Load package” → “Survival” → OK

```
> data<-read.table("skripsibeneran.txt",header=T)
> data1<-data.frame(data)
> attach(data1)
> local({pkg <- select.list(sort(.packages(all.available = TRUE)))
+ if(nchar(pkg)) library(pkg, character.only=TRUE)})
Loading required package: splines
> coxph<-coxph(Surv(survival,status)~treat+age,method="efron")
> coxph
Call:
coxph(formula = Surv(survival, status) ~ treat + age, method = "efron")
          coef exp(coef) se(coef)       z    p
treat -0.00365     0.996  0.18251 -0.0200 0.98
age    0.00753     1.008  0.00966  0.7791 0.44
Likelihood ratio test=0.63  on 2 df, p=0.731  n= 137
> coxph2<-coxph(Surv(survival,status)~treat+age+cel1+cel2+cel3+cel4+perform,method="efron")
Warning message:
In coxph(Surv(survival, status) ~ treat + age + cel1 + cel2 + cel3 +  :
  X matrix deemed to be singular; variable 6
> coxph2
Call:
coxph(formula = Surv(survival, status) ~ treat + age + cel1 +
    cel2 + cel3 + cel4 + perform, method = "efron")
           coef exp(coef) se(coef)      z       p
treat    0.3030     1.354  0.20566  1.474 1.4e-01
age     -0.0089     0.991  0.00922 -0.965 3.3e-01
cel1     0.4023     1.495  0.28254  1.424 1.5e-01
cel2     1.1788     3.250  0.29644  3.977 7.0e-05
cel3     0.8563     2.355  0.27132  3.156 1.6e-03
cel4         NA        NA  0.00000     NA      NA
perform -0.0327     0.968  0.00541 -6.043 1.5e-09
```

```
Likelihood ratio test=62  on 6 df, p=1.78e-11  n= 137

> coxstratifikasi1<-
coxph(Surv(survival,status)~treat+age+strata(cell)+strata(perform<60),method="efron")
> coxstratifikasi1
Call:
coxph(formula = Surv(survival, status) ~ treat + age + strata(cell) +
    strata(perform < 60), method = "efron")

          coef exp(coef) se(coef)      z    p
treat  0.12778     1.136   0.2085  0.613 0.54
age   -0.00127     0.999   0.0101 -0.126 0.90
Likelihood ratio test=0.38  on 2 df, p=0.829  n= 137
```