Pengantar Analisis Antar Kejadian

Dr. Danardono, MPH

danardono@ugm.ac.id

Program Studi Statistika

Jurusan Matematika UGM

Analisis Antar Kejadian

Data Antar Kejadian (DAK)

- *event-history data*
- *time-to-event data*
- data durasi
- data survival

Analisis Antar Kejadian

Analisis data yang memanfaatkan informasi kronologis dari kejadian-kejadian atau peristiwa (*events*)

Analisis "Buku Harian"

Mahasiswa A			Mahasiswa B	
03:30	tidur		11:00	tidur
08:50	bangun		05:00	bangun
09:13	kuliah, terlambat		08:58	kuliah, *on time*
09:30	ngantuk ... zzz		09:30	aktif ....
11:30	capek, lapar		11:30	belajar, *fresh*
12:00	mau makan		12:20	makan
	dompet ketinggalan		13:00	kuliah
12:20	pulang dulu...		...	dst...
...	dst...			

Analisis "Riwayat Hidup"

Person A	
7 th	masuk SD
10 th	keluar SD
12 th	pengamen jalanan
13 th	terlibat penjambretan
16 th	terlibat curanmor
18 th	terlibat perampokan
21 th	bos mafia gang X
25 th	pertikaian antar gang
...	dst...

Person B	
7 th	masuk SD
13 th	masuk SMP
16 th	masuk SMU
19 th	masuk PT
23 th	mencari pekerjaan
24 th	bekerja di prsh. X
28 th	kepala cabang
...	dst...

Analisis Rekam Medis

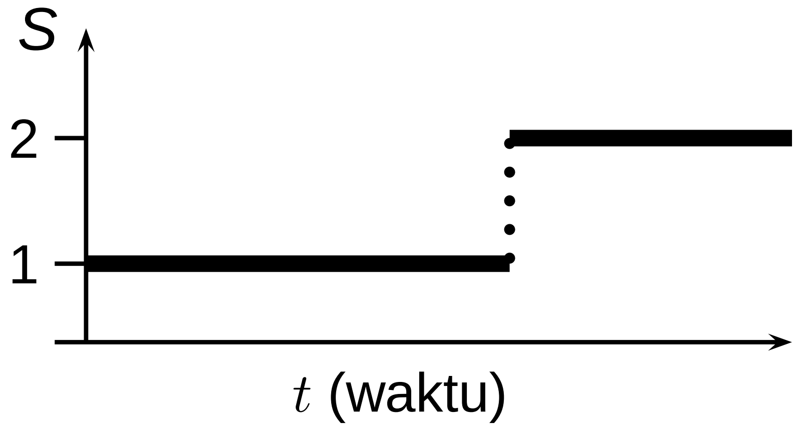
Person A			Person B	
0 th	lahir normal		0 th	lahir normal
1-5 th	sehat		1-5 th	sehat
5-16 th	sehat		5-16 th	sehat
16-40 th	merokok		16-21th	berhenti merokok
40 th	gejala kanker paru		21 th	sehat
...	dst ...		...	dst ...

Aplikasi AAK

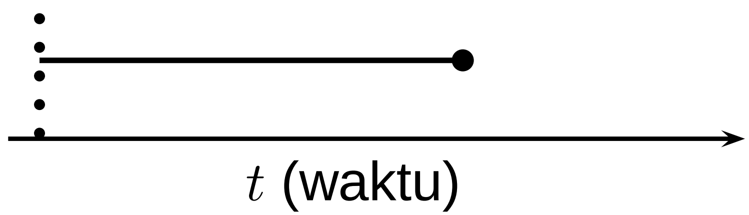
- epidemiologi
- biostatistika
- sosiologi
- psikologi
- demografi
- ekonomi

Contoh DAK

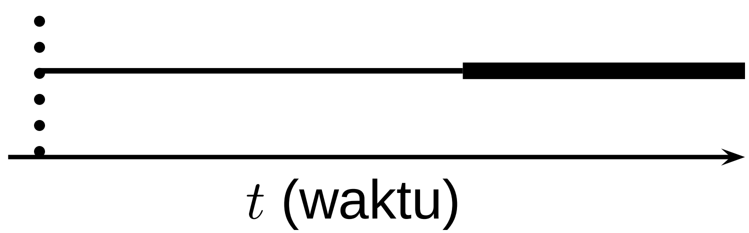
a) Representasi data antar kejadian



b) Alternatif representasi (data survival)



c) Alternatif representasi



S : *state space* (status))

Contoh 1: data survival, kejadian (*event*) yang menjadi perhatian adalah kematian

status 1 :	hidup
2 :	mati

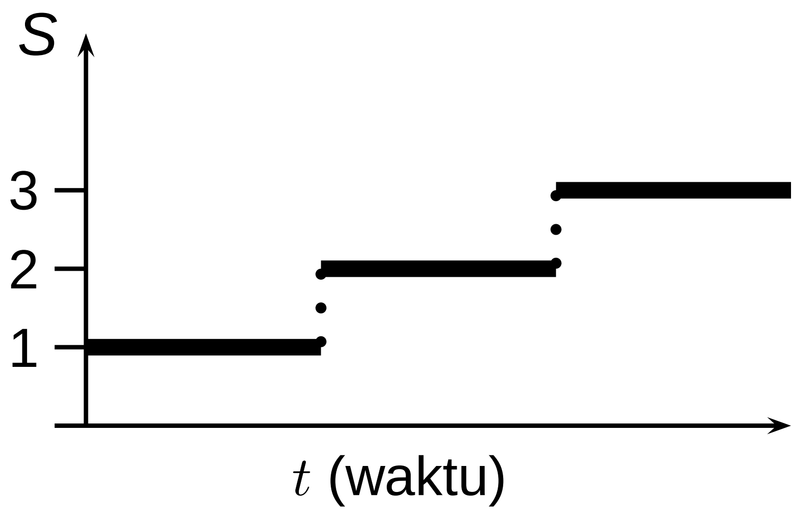
Contoh 2: *event* yang menjadi perhatian adalah saat anak disapih (berhenti disusui oleh ibunya)

status 1 :	disusui
2 :	disapih

Contoh 3: *event* yang menjadi perhatian adalah saat seseorang mulai bisa naik sepeda

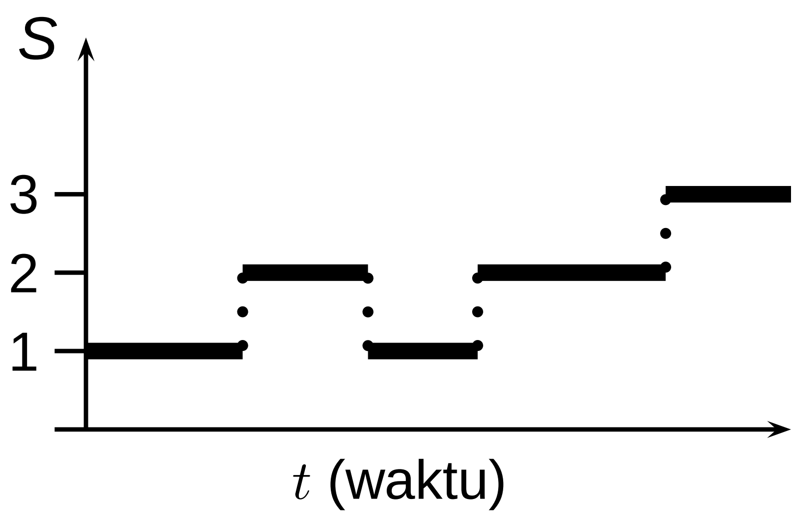
status 1 :	belum bisa naik sepeda
2 :	bisa naik sepeda

Contoh DAK



Contoh 4: data multistatus (*multistate*) dengan tiga macam status yang *irreversible*, misalnya tahapan penyakit yang progresif.

status 1 :	stadium 1
2 :	stadium 2
3 :	stadium 3



Contoh 5: data multistatus dengan kemungkinan beberapa status yang irreversible

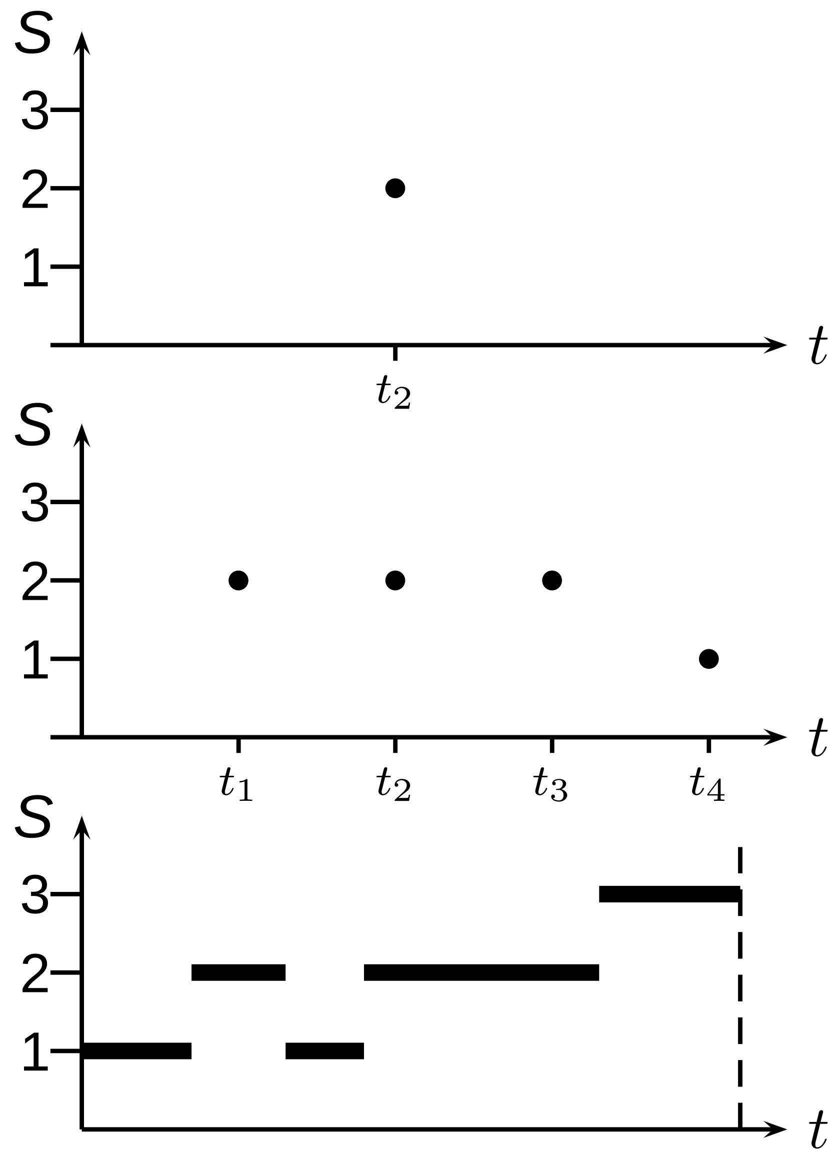
status 1 :	sehat
2 :	sakit
3 :	meninggal

Rancangan Pengumpulan Data

a) *Cross-sectional*

b) *Panel*

c) *Event-oriented* (longitudinal)

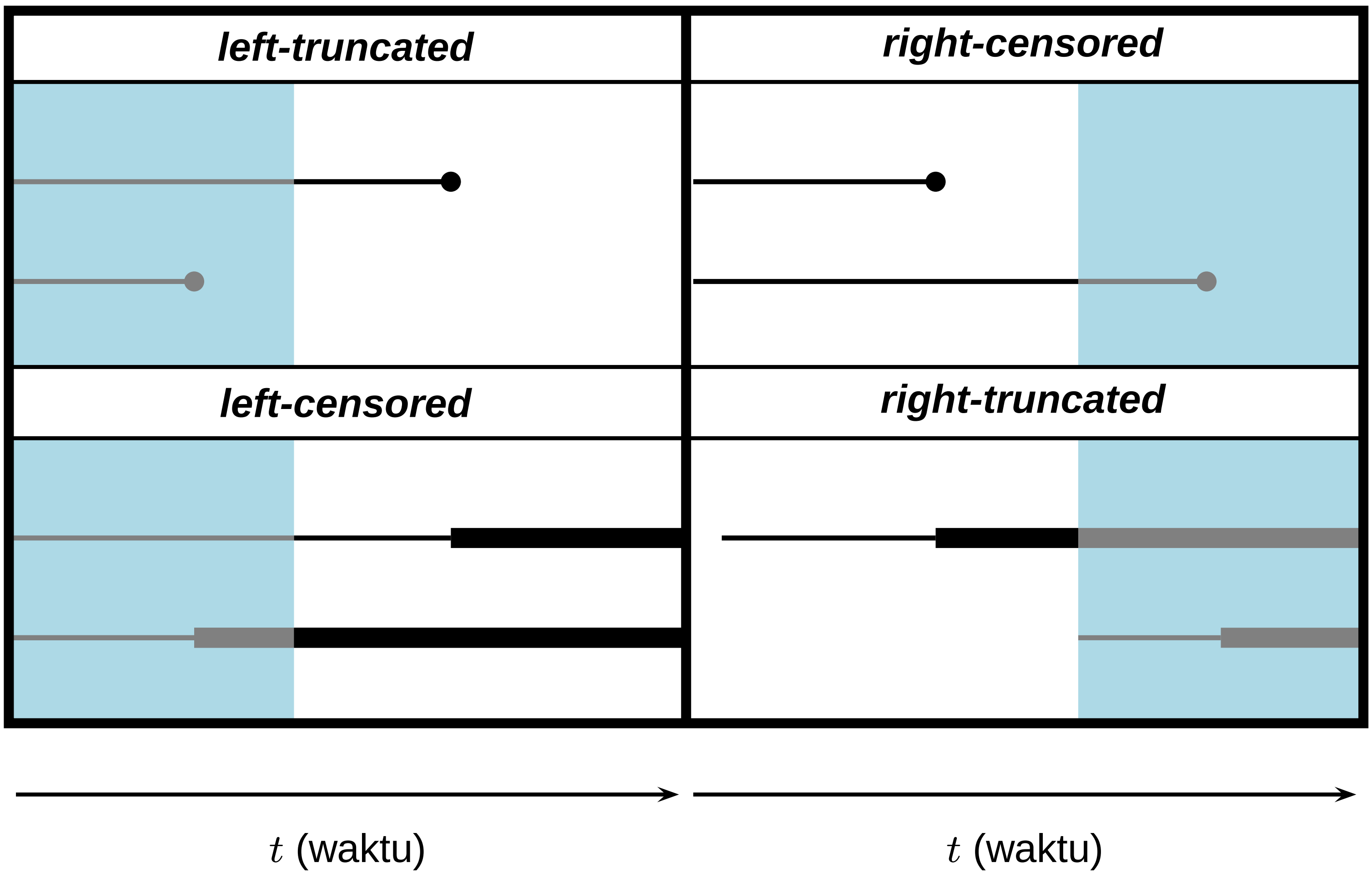


S (state space):

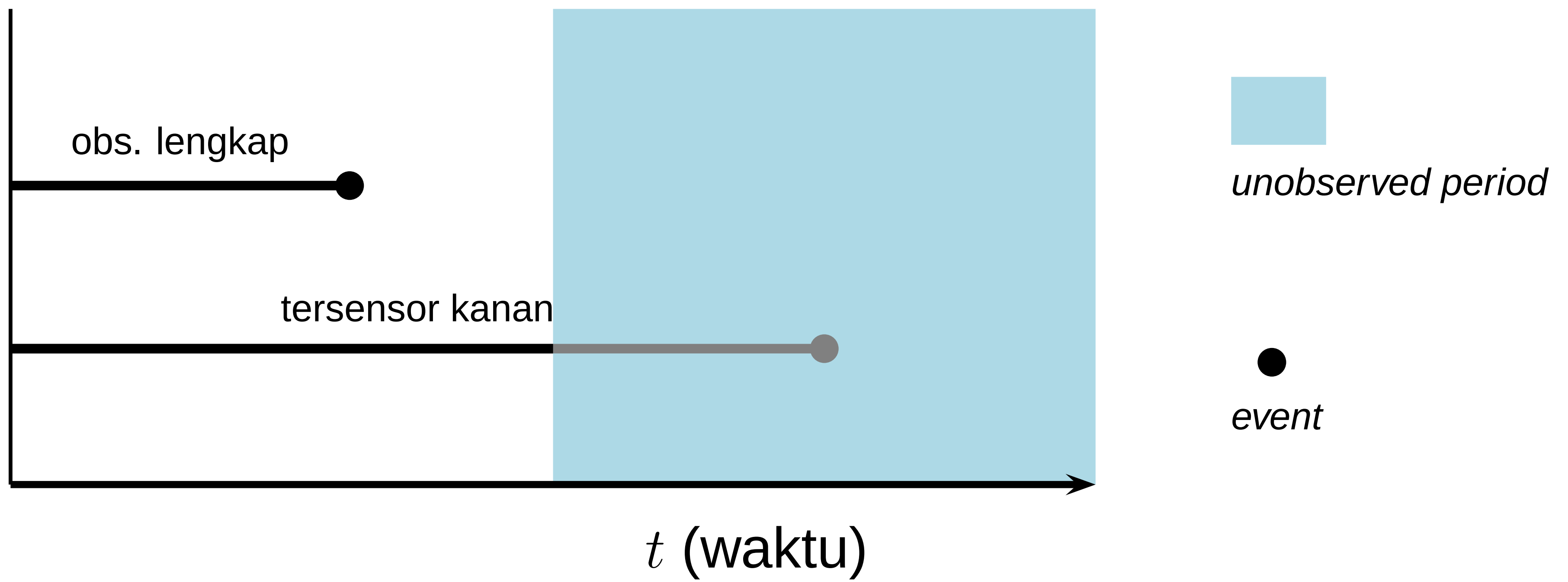
1	:	sehat
2	:	sakit
3	:	meninggal

Tersensor dan Terpotong

Kendala yang sering muncul dalam DAK adalah adanya data tersensor (*censored*) dan terpotong (*truncated*).



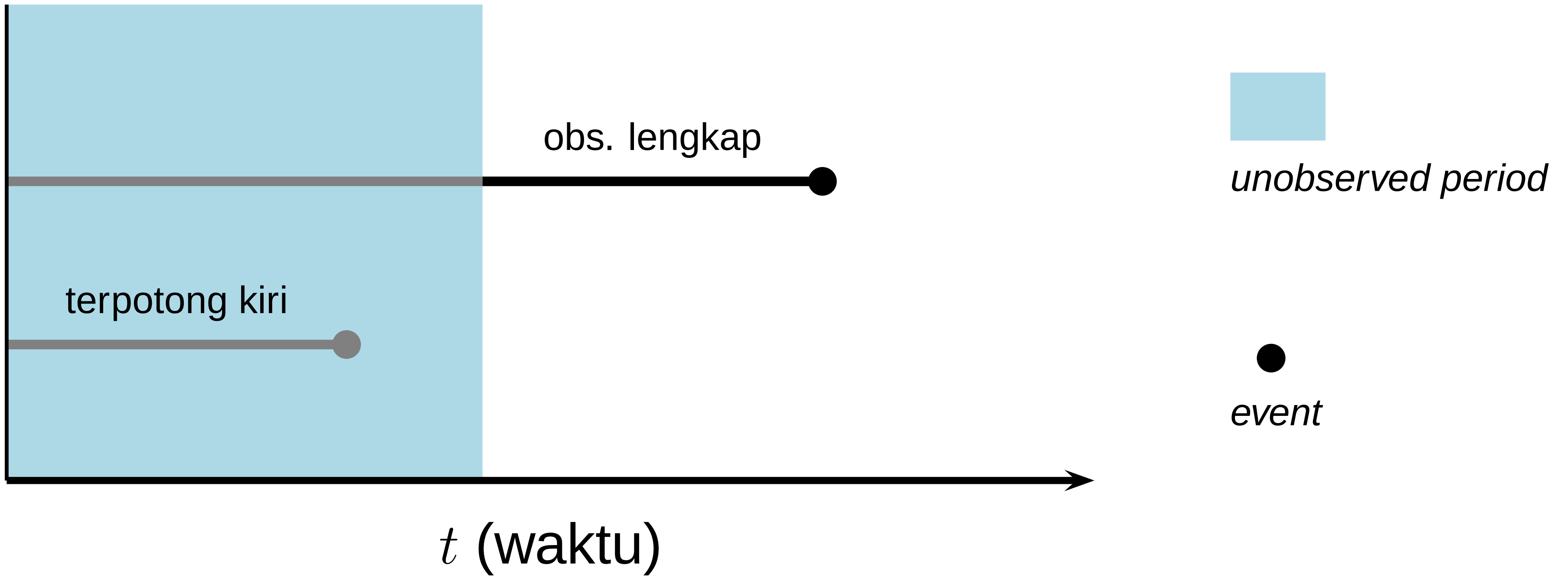
Tersensor Kanan (*Right-censored*)



Contoh:
Suatu eksperimen menggunakan tikus percobaan dilakukan untuk mengetahui seberapa lama tikus dapat hidup setelah pemberian suatu zat yang dapat mengakibatkan kanker.

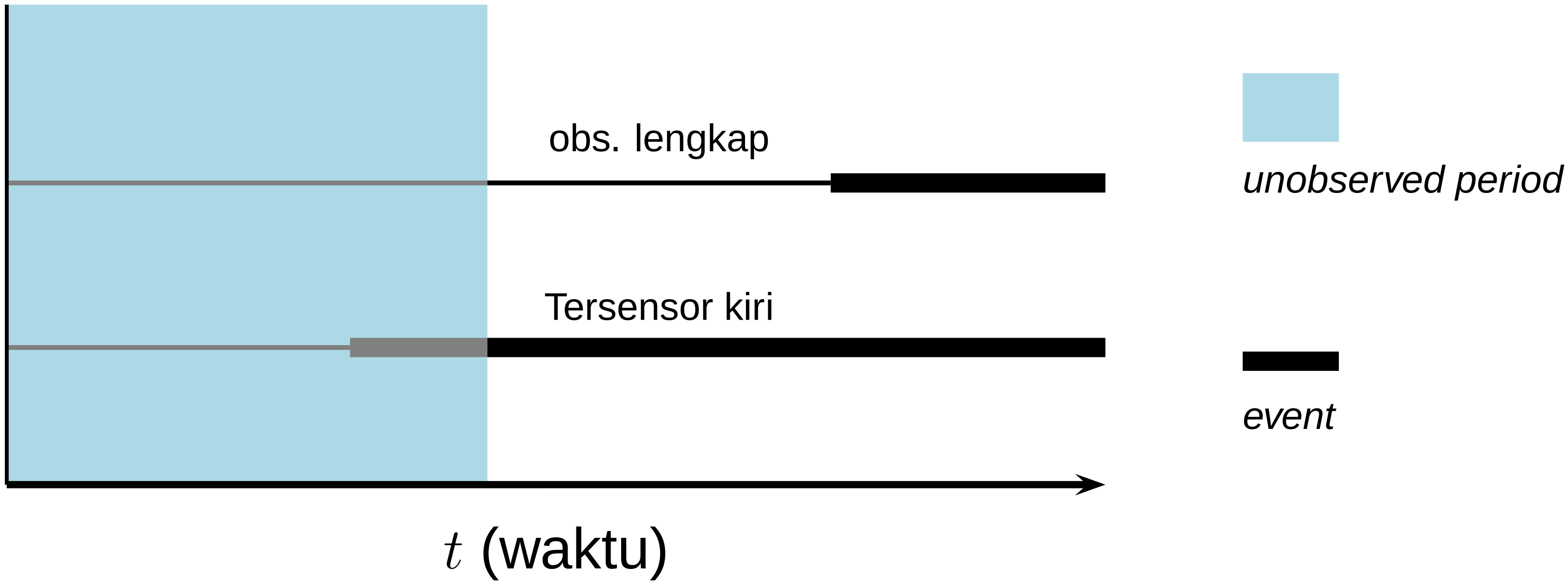
- Tipe I: Jika saat tersensornya ditentukan lebih dahulu
- Tipe II: Jika saat tersensornya ditentukan setelah tercapai persentase atau banyak sampel tertentu yang telah mendapatkan *event*.

Terpotong Kiri (*Left-Truncated*)



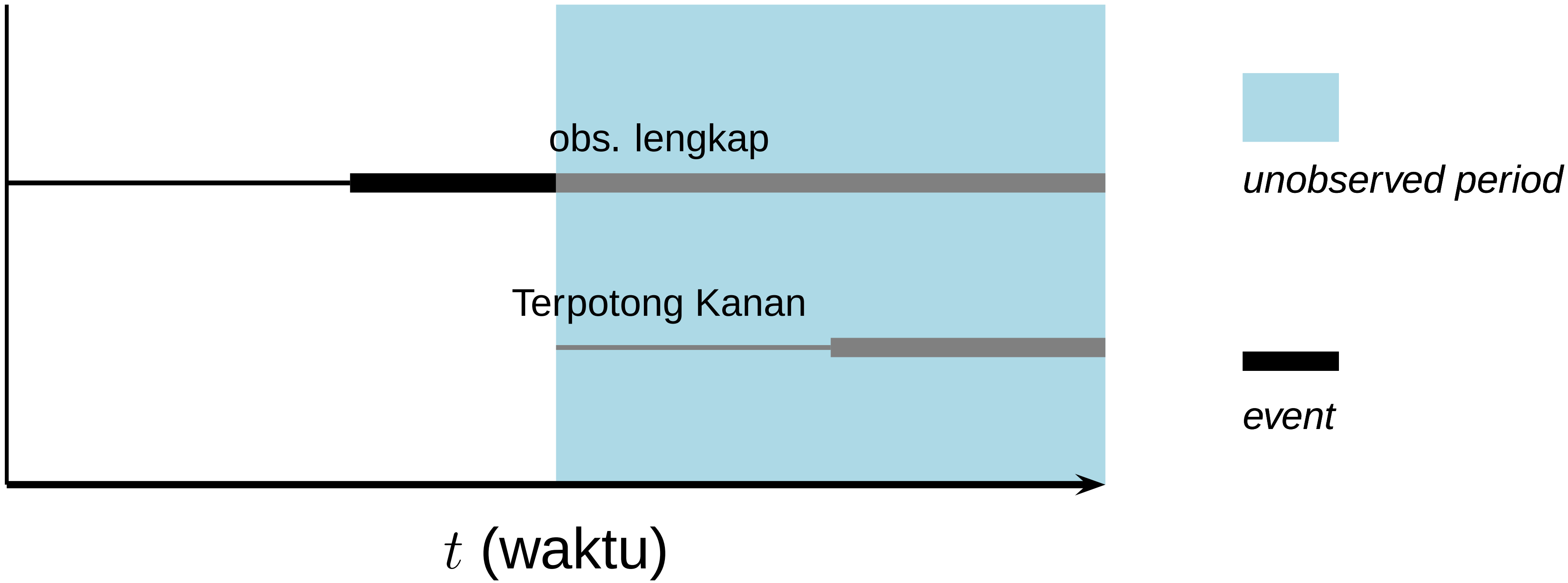
Contoh:
Suatu studi tentang morbiditas dan mortalitas pegawai pada suatu institusi dilakukan ketika pegawai telah berusia 40 tahun ke atas. Apabila seorang pegawai telah meninggal sebelum berusia 40, dia tidak masuk dalam sampel (*left-truncated*).

Tersensor Kiri (*Left-Censored*)



Contoh:
Suatu studi dilakukan untuk mengetahui faktor-faktor yang mempengaruhi usia pertama kali merokok. Apabila responden ingat usia saat dia pertama kali merokok, dikatakan observasi yang diperoleh adalah lengkap. Bila responden tidak ingat kapan dia mulai merokok, tapi hanya ingat mulai merokok sebelum usia tertentu, maka dikatakan observasi tersebut tersensor kiri.

Terpotong Kanan (*Right-Truncated*)



Contoh:
Suatu studi tentang AIDS dilakukan secara retrospektif. Yang menjadi perhatian adalah durasi mulai infeksi HIV sampai terdiagnosis AIDS. Hanya individu yang telah terdiagnosis AIDS sebelum mulai studi saja yang akan masuk dalam studi. Individu yang belum terdiagnosis AIDS tidak masuk dalam studi adalah sampel yang terpotong kanan.

Fungsi Survival

Probabilitas satu individu hidup (tinggal dalam suatu *status*) lebih lama daripada t

$$S(t) = P(T > t)$$

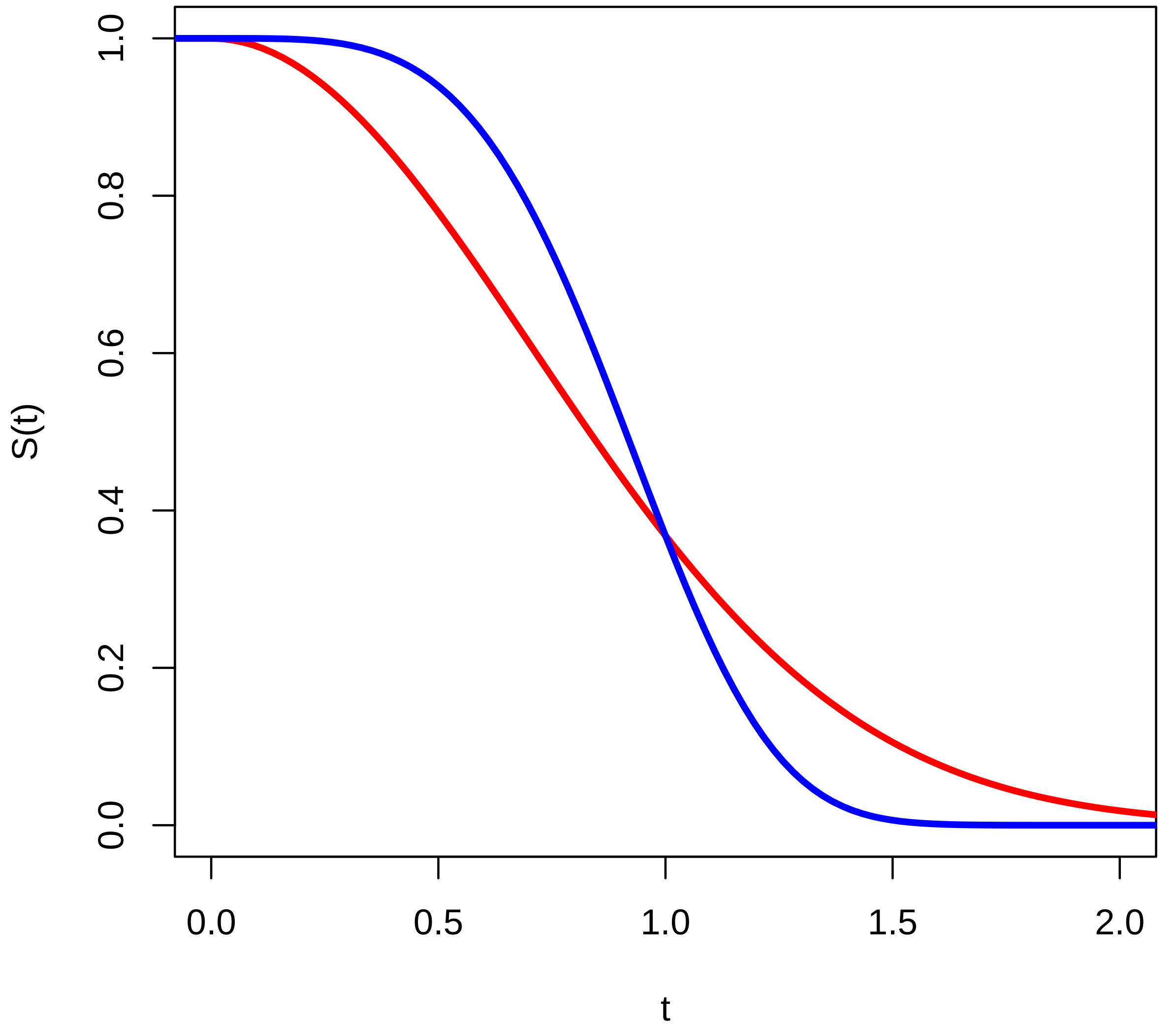
$S(t)$ adalah fungsi *non-increasing* terhadap waktu t dengan sifat

$$S(t) = \begin{cases} 1 & \text{untuk } t = 0 \\ 0 & \text{untuk } t = \infty \end{cases}$$

Hubungan $S(t)$ dengan distribusi kumulatif $F(t)$

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

Fungsi Survival



Fungsi *Hazard*

Tingkat (*rate*) terjadinya suatu *event*

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

Hubungan $h(t)$, $S(t)$ dan $f(t)$

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

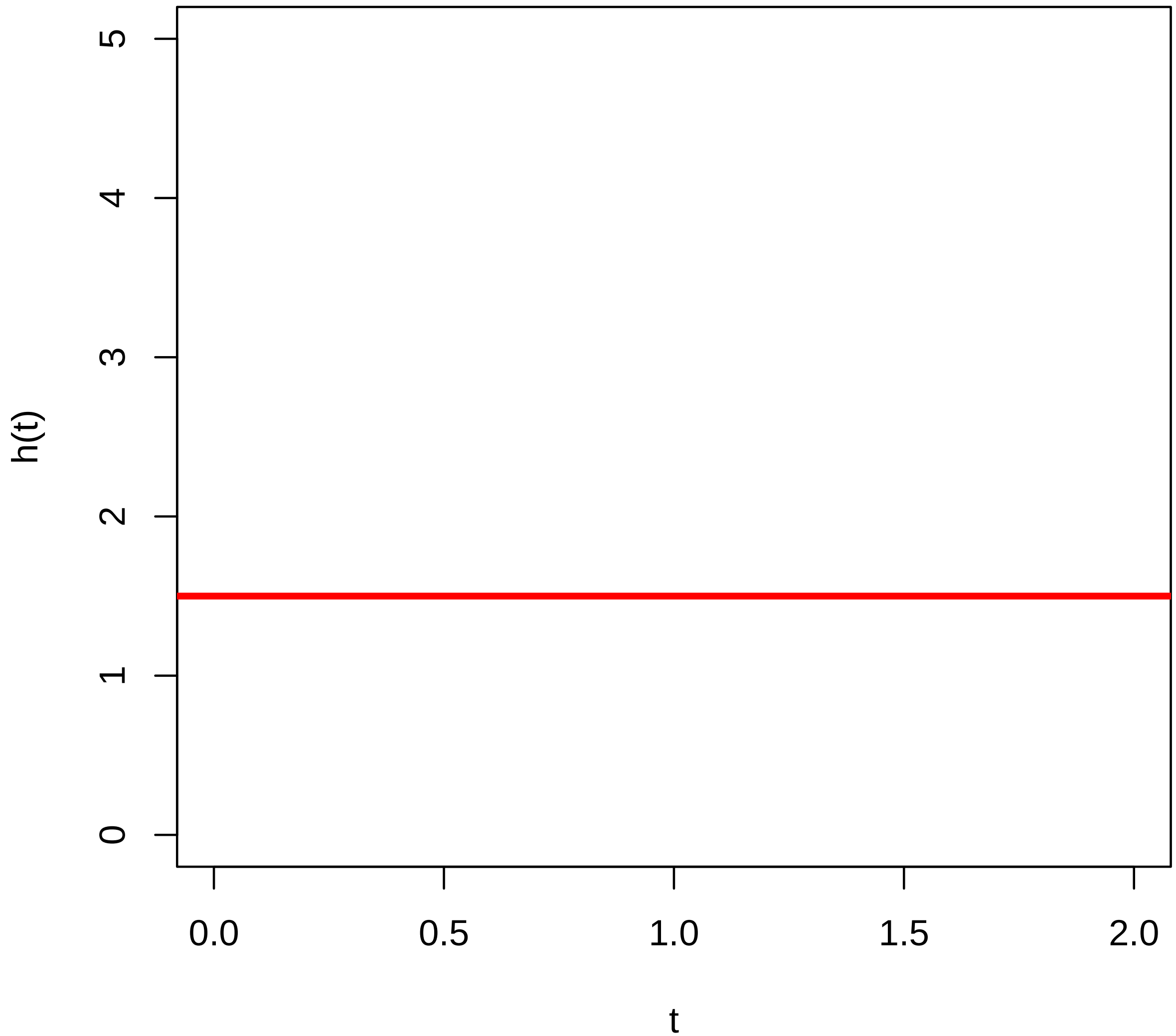
Fungsi *Hazard* Kumulatif

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

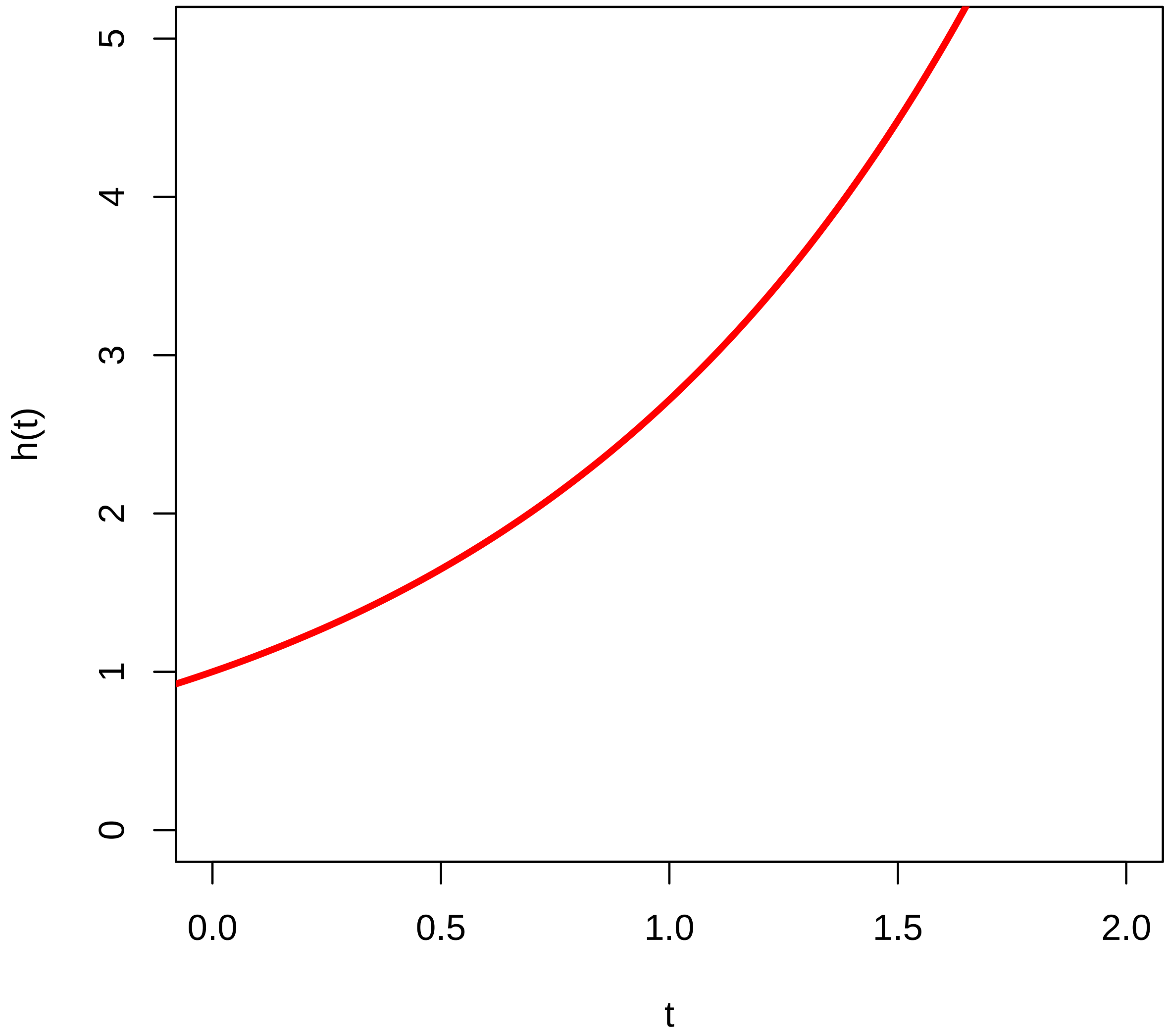
Hubungan $H(t)$ dengan $S(t)$

$$H(t) = -\log S(t)$$

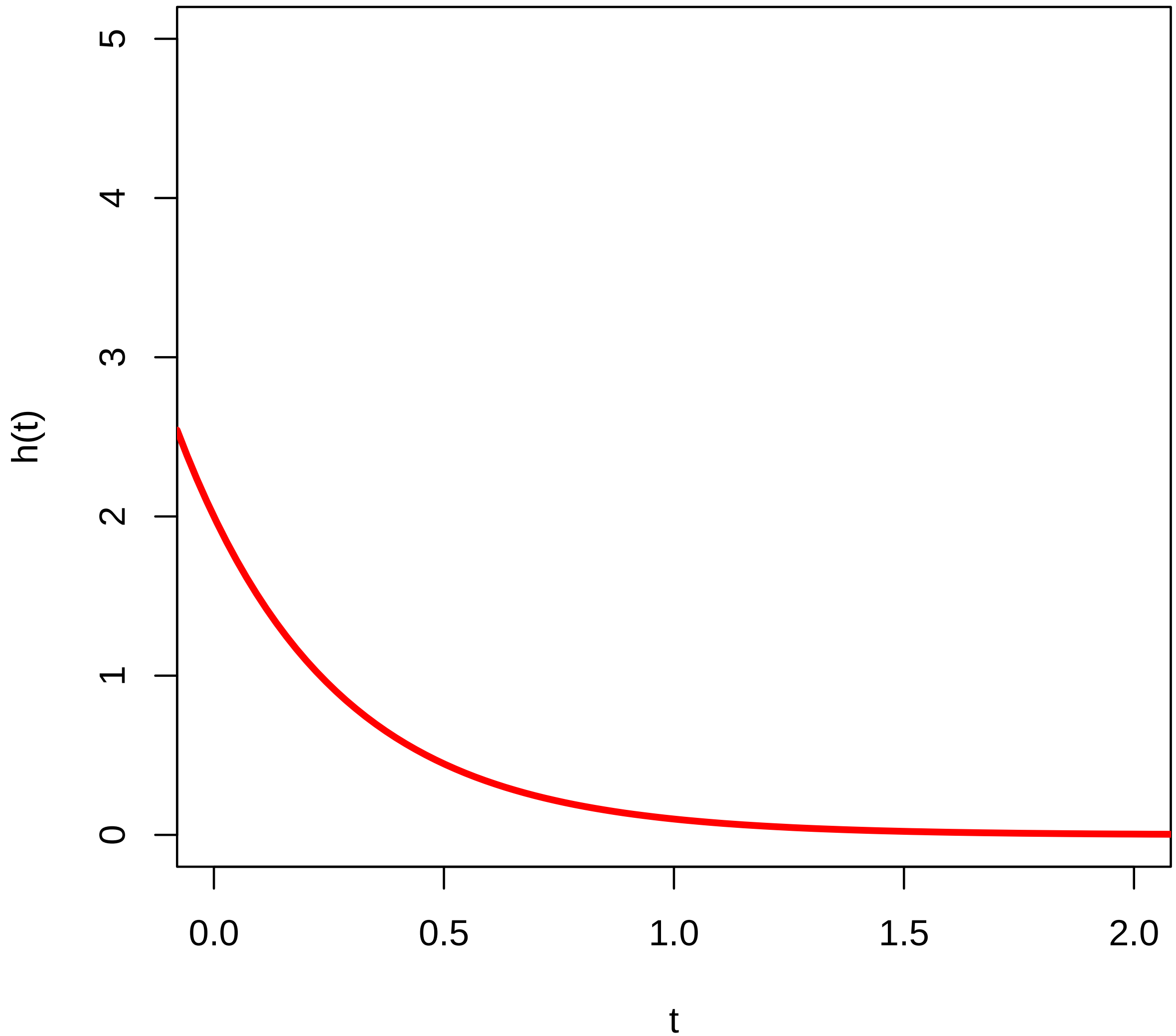
Fungsi *Hazard*



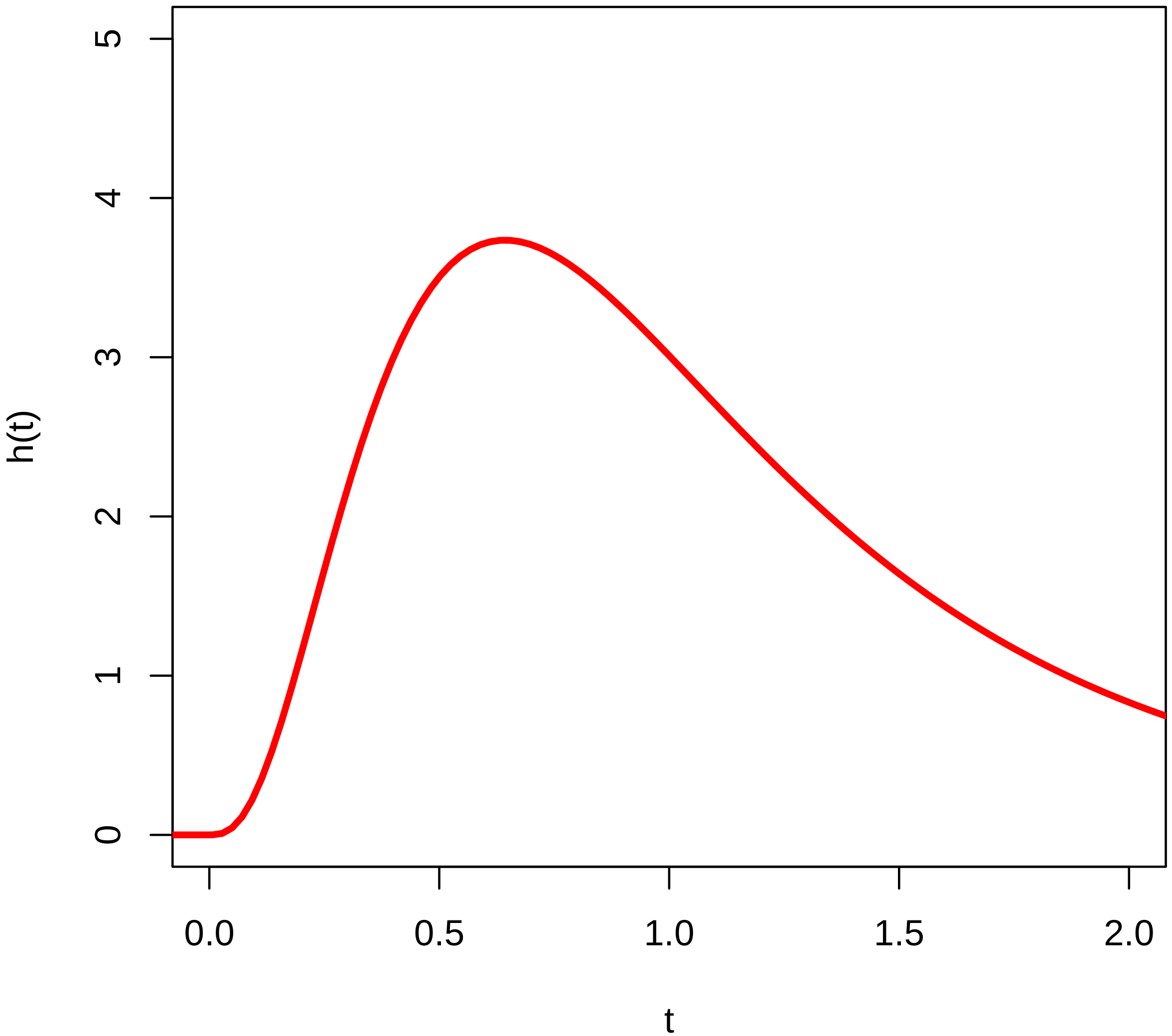
Fungsi *Hazard*



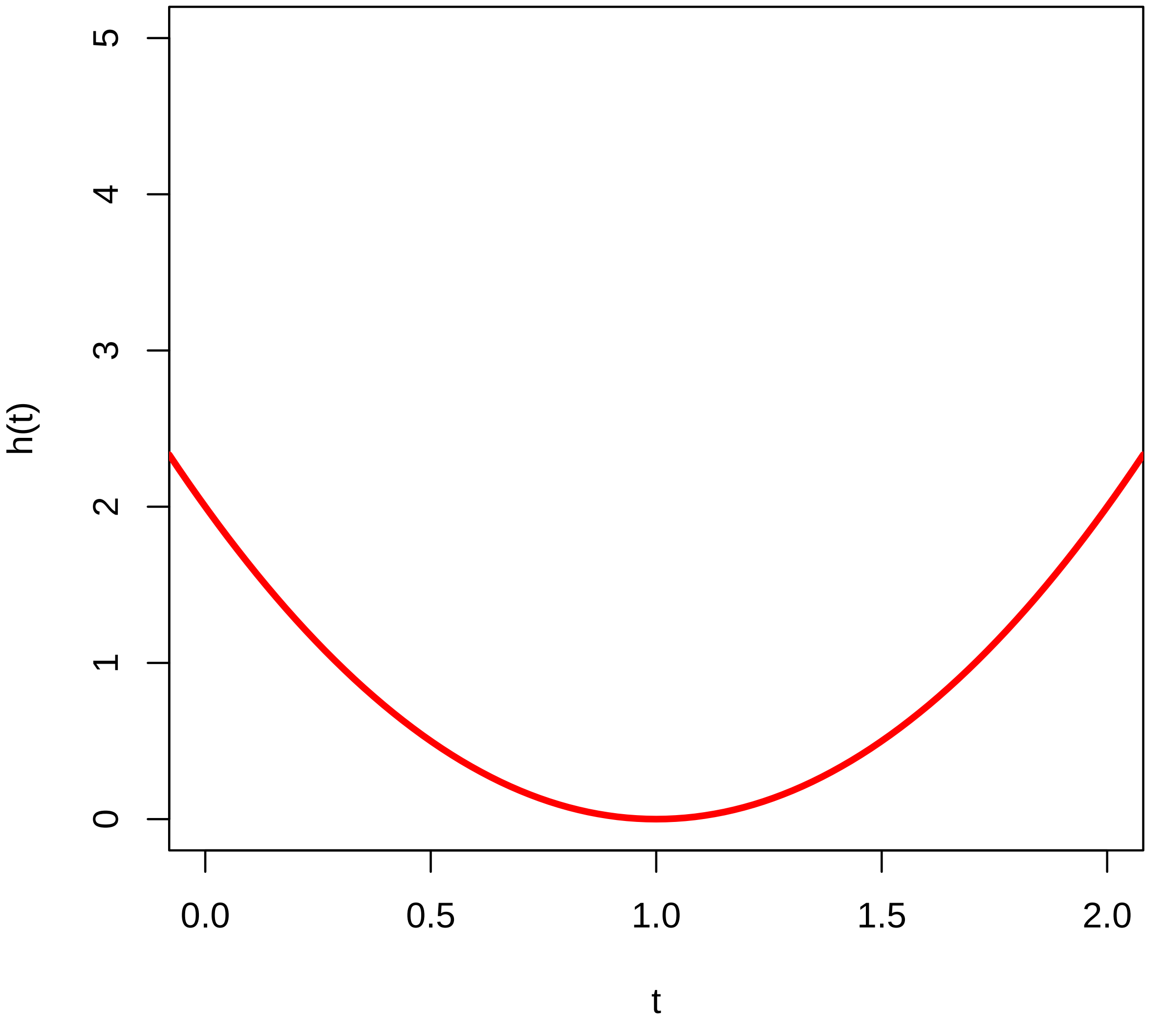
Fungsi *Hazard*



Fungsi *Hazard*



Fungsi *Hazard*



Model Eksponensial

Eksponensial ($\lambda > 0$, $t \geq 0$)

- fungsi densitas

$$f(t) = \lambda \exp(-\lambda t)$$

- fungsi hazard

$$h(t) = \lambda$$

- fungsi survival

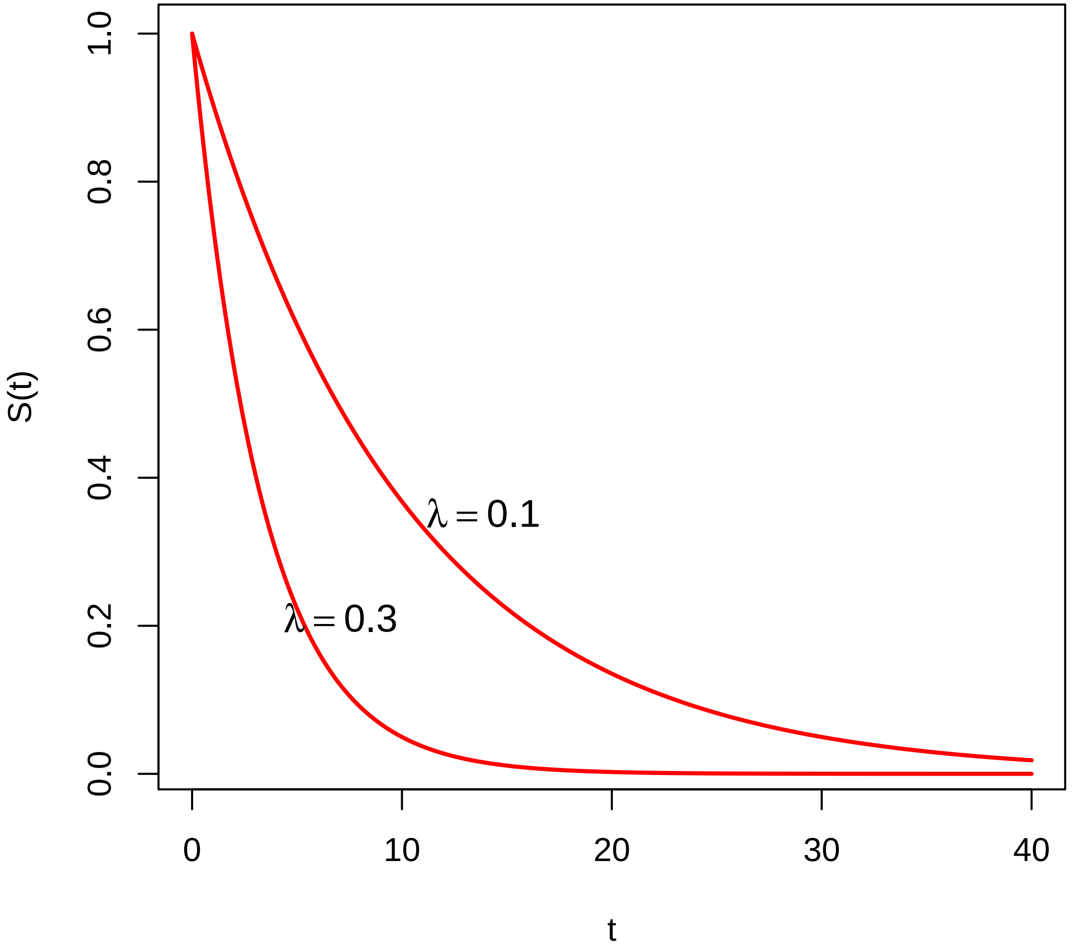
$$S(t) = \exp(-\lambda t)$$

- mean

$$E(t) = \frac{1}{\lambda}$$

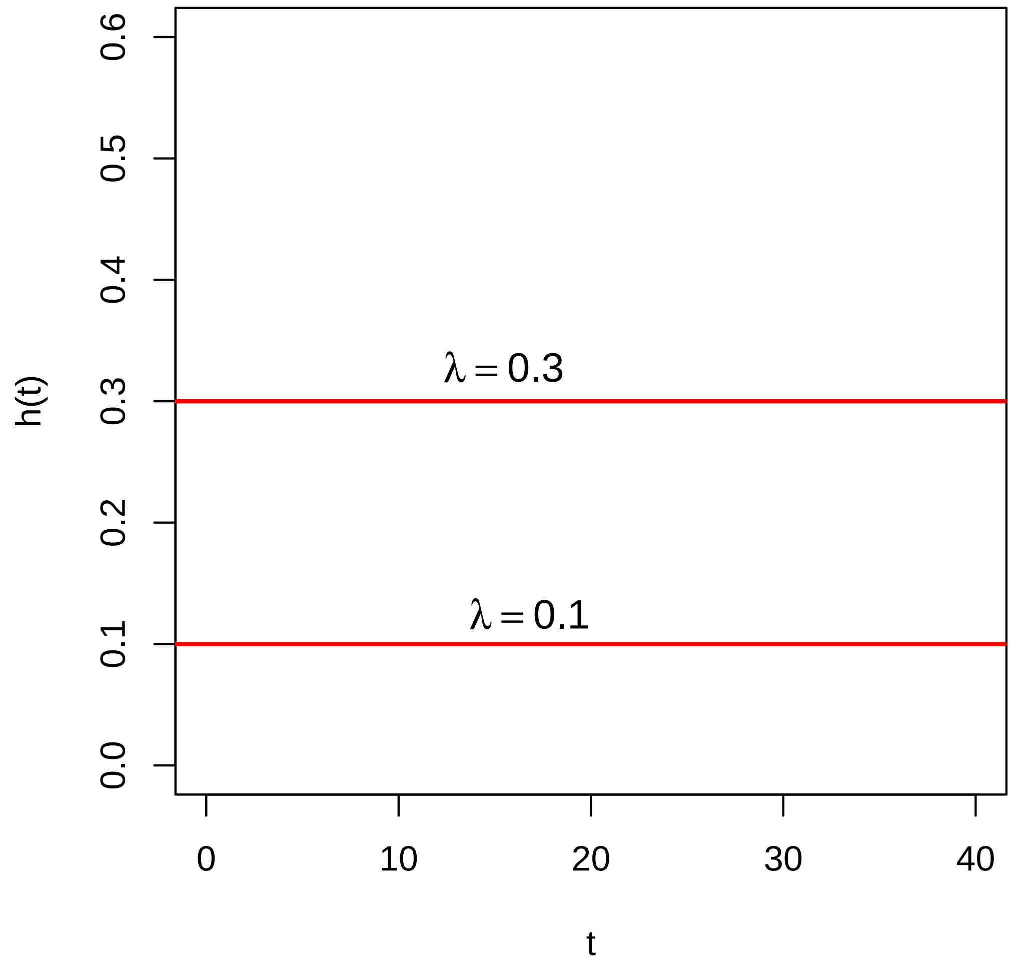
Model Eksponensial

Kurva survival untuk model eksponensial dengan dua nilai λ yang berbeda



Model Eksponensial

Kurva hazard untuk model eksponensial dengan dua nilai λ yang berbeda



Model Weibull

Weibull ($\alpha, \lambda > 0$,$t \geq 0$)
Parameter α dan λ sering disebut sbg. *shape* dan *scale*.

- fungsi densitas

$$f(t) = \alpha\lambda(\lambda t)^{\alpha-1}\exp(-(\lambda t)^{\alpha})$$

- fungsi hazard

$$h(t) = \alpha\lambda(\lambda t)^{\alpha-1}$$

- fungsi survival

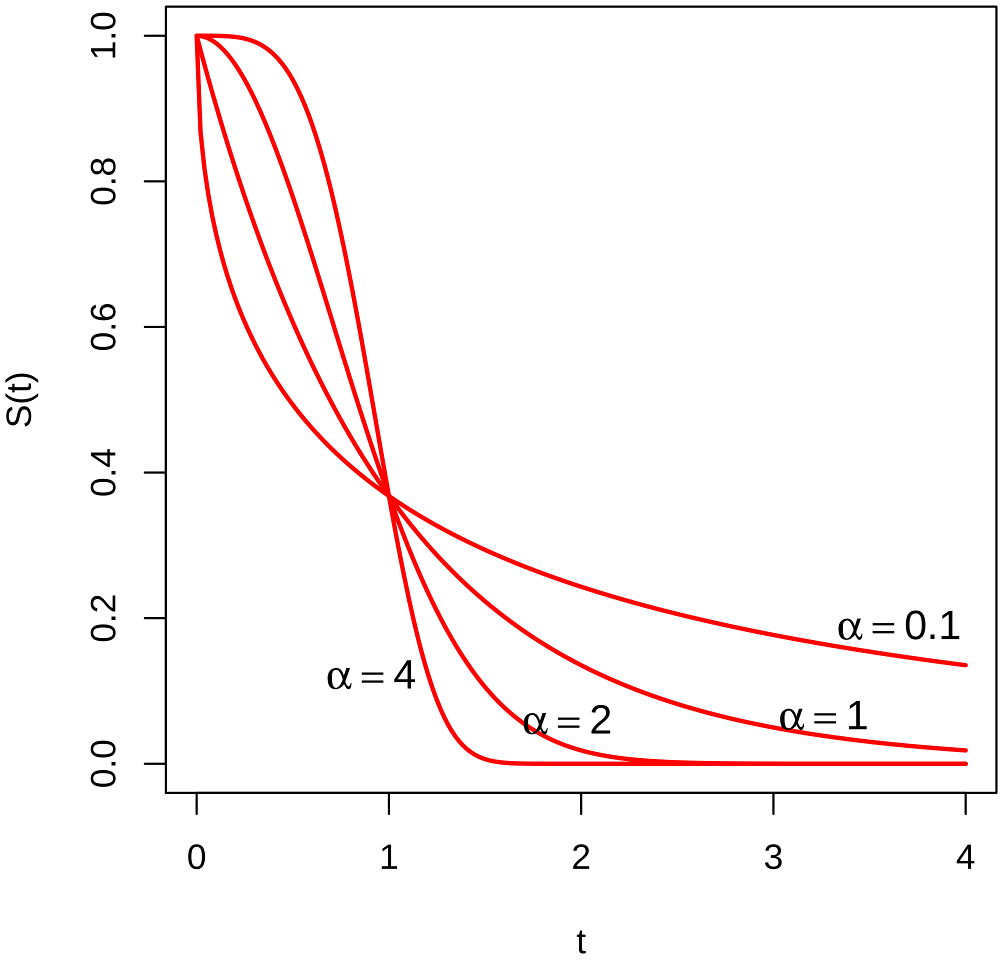
$$S(t) = \exp(-(\lambda t)^{\alpha})$$

- mean

$$E(t) = \frac{\Gamma(1 + 1/\alpha)}{\lambda}$$

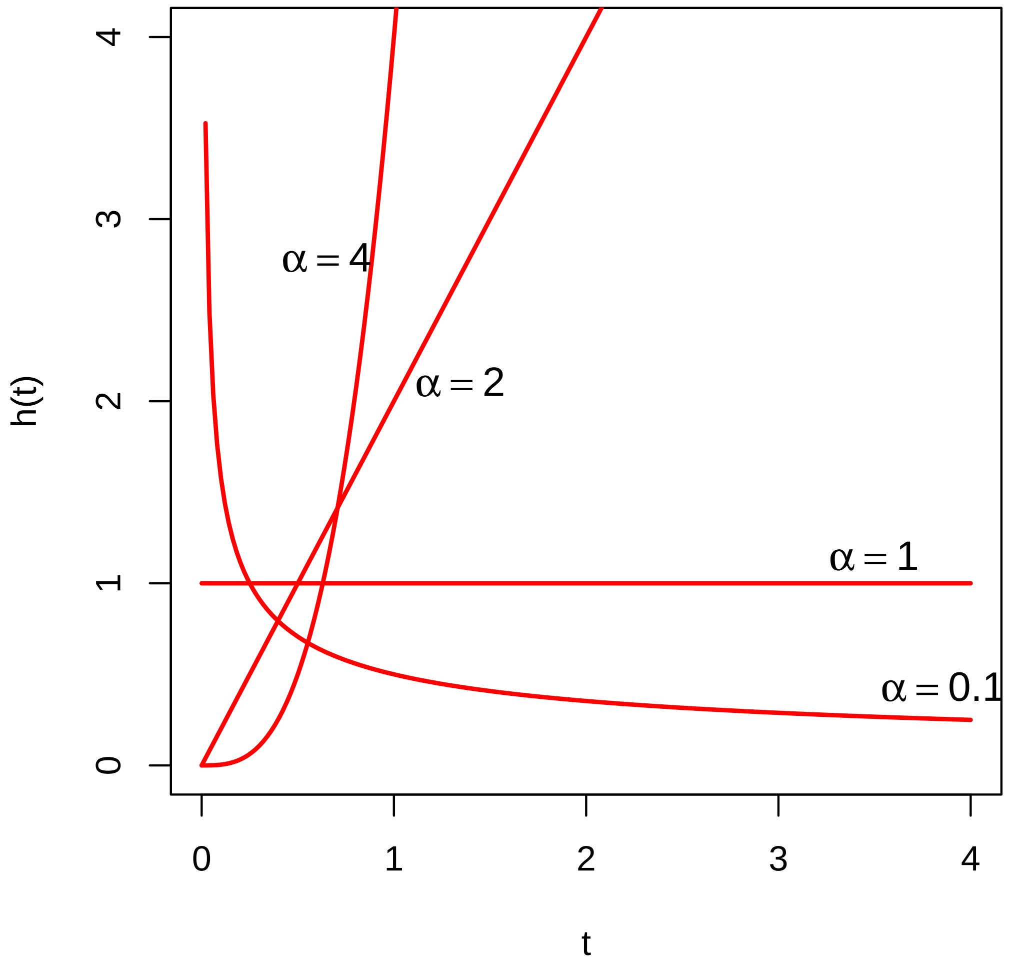
Model Weibull

Kurva survival untuk model Weibull dengan beberapa nilai α yang berbeda



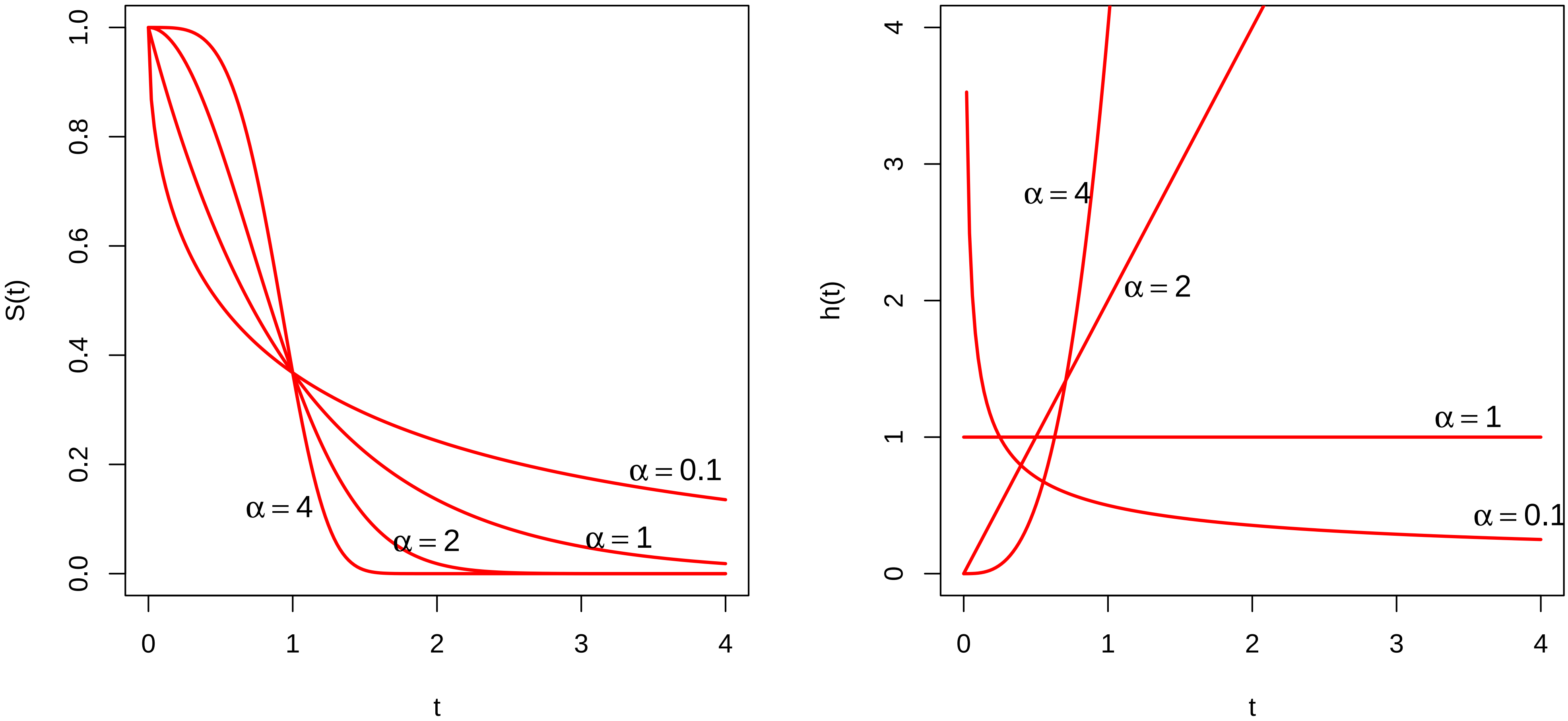
Model Weibull

Kurva hazard untuk model Weibull dengan beberapa nilai α yang berbeda



Model Weibull

Kurva survival dan hazard untuk model Weibull dengan beberapa nilai α yang berbeda



Model Gamma

Gamma ($\beta, \lambda > 0$,$t \geq 0$)

- fungsi densitas

$$f(t) = \frac{\lambda(\lambda t)^{\beta-1}\exp(-\lambda t)}{\Gamma(\beta)}$$

- fungsi hazard

$$h(t) = f(x)/S(x)$$

- fungsi survival

$$S(t) = 1 - I(\lambda t, \beta) = 1 - \frac{1}{\Gamma(\beta)}\int_0^{\lambda t} u^{\beta-1}e^{-u}du$$

- mean

$$E(t) = \beta/\lambda$$

Model Log-normal

log-normal ($\sigma > 0$,$t \geq 0$)

- fungsi densitas

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\log(t) - \mu)^2\right]$$

- fungsi hazard

$$h(t) = f(x)/S(x)$$

- fungsi survival

$$S(t) = 1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)$$

- mean

$$E(t) = \exp(\mu + \sigma^2/2)$$

Estimasi Parameter

Data: $(T_i = t_i, \delta_i)$, $i = 1, 2, \ldots, n$ yang independen satu sama lain

dengan
T_i : durasi atau waktu antar kejadian

$$\delta_i = \begin{cases} 0 & \text{jika } i \text{ tersensor} \\ 1 & \text{jika } i \text{ mendapatkan kejadian (\textit{event})} \end{cases}$$

Estimasi Parameter

Fungsi likelihood untuk data tersensor kanan

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^{n} f(t_i, \boldsymbol{\theta})^{\delta_i} S(t_i, \boldsymbol{\theta})^{1-\delta_i}$$

dengan $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ adalah p parameter yang akan diestimasi; $f(t_i, \boldsymbol{\theta})$ adalah fungsi densitas untuk i yang mendapatkan kejadian dan $S(t_i, \boldsymbol{\theta})$ adalah fungsi survival untuk i yang tidak mendapatkan kejadian.

Estimasi Parameter

Fungsi log-likelihood untuk data tersensor kanan

$$\ell(\boldsymbol{\theta}) \propto \sum_{i=1}^{n} (\delta_i) \log(f(t_i, \boldsymbol{\theta})) + \sum_{i=1}^{n} (1 - \delta_i) \log(S(t_i, \boldsymbol{\theta}))$$

dengan $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ adalah p parameter yang akan diestimasi; $f(t_i, \boldsymbol{\theta})$ adalah fungsi densitas untuk i yang mendapatkan kejadian dan $S(t_i, \boldsymbol{\theta})$ adalah fungsi survival untuk i yang tidak mendapatkan kejadian.

Estimasi Parameter

Digunakan metode kemungkinan maksimum (***MLE***: *Maximum Likelihood Estimation*) untuk mengestimasi $\boldsymbol{\theta}$.

MLE dari $\boldsymbol{\theta}$, ditulis $\hat{\boldsymbol{\theta}}$ adalah $(\hat{\theta}_1, \ldots, \hat{\theta}_p)$ yang memaksimumkan $\ell(\boldsymbol{\theta})$

$$\ell(\hat{\boldsymbol{\theta}}) = \max_{\text{semua } \boldsymbol{\theta}} \ell(\boldsymbol{\theta})$$

$\hat{\boldsymbol{\theta}}$ adalah penyelesaian dari

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_j} = 0, \quad j = 1, 2, \ldots, p$$

Eksponensial - data lengkap

Fungsi log-likelihood

$$\ell(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} t_i$$

MLE dari λ

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i}$$

Mean dari Eksponensial: $\mu = E(x) = 1/\lambda$, sehingga $\hat{\mu} = \bar{t}$, dengan $\bar{t} = \sum_{i=1}^{n} t_i / n$

Eksponensial - data lengkap

Interval konfidensi $100(1-\alpha)\%$ untuk λ dibentuk berdasarkan statistik $2n\hat{\mu}/\mu$ yang berdistribusi chi-square dengan derajad bebas $2n$

$$\frac{\hat{\lambda}\chi^2_{2n,\alpha/2}}{2n} < \lambda < \frac{\hat{\lambda}\chi^2_{2n,1-\alpha/2}}{2n}$$

dengan $\chi^2_{2n,p}$ adalah kuantil ke-p dari distribusi chi-square dengan derajad bebas $2n$.

Eksponensial - data lengkap

Diketahui waktu remisi (minggu) dari 21 pasien leukemia akut:
1, 1, 2, 2, 3, 4, 4, 5, 5, 6, 8, 8,
9,10, 10, 12, 14, 16, 20, 24, 34

Interval konfidensi 95% untuk λ dari data di atas:

$$\begin{aligned}
\frac{\hat{\lambda}\chi^2_{2n,\alpha/2}}{2n} &< \lambda < \frac{\hat{\lambda}\chi^2_{2n,1-\alpha/2}}{2n} \\
\frac{0,106 \times 25,999}{42} &< \lambda < \frac{0,106 \times 62,777}{42} \\
0,066 &< \lambda < 0,156
\end{aligned}$$

Eksponensial - data tersensor

Data: $(T_i = t_i, \delta_i)$, $i = 1, 2, \ldots, n$ yang independen satu sama lain demikian juga dengan T_i dan δ_i

Fungsi likelihood

$$L(\lambda) = \prod_{i=1}^{n} \lambda^{\delta_i} \exp(-\lambda t_i)$$

Fungsi log-likelihood

$$\ell(\lambda) = \sum_{i=1}^{n} \left[\delta_i \log \lambda - \lambda \sum_{i=1}^{n} t_i \right]$$

Eksponensial - data tersensor

MLE dari λ

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} t_i}$$

Bila banyaknya data yang lengkap adalah k

$$\hat{\lambda} = \frac{k}{\sum_{i=1}^{n} t_i}$$

Eksponensial - data tersensor

Dalam suatu penelitian 10 tikus percobaan terpapar (*exposed*) ke suatu jenis penyakit kanker. Setelah 5 tikus mati percobaan dihentikan diperoleh data lama hidup tikus sbb: 4, 5, 8, 9, 10, 10+, 10+, 10+, 10+, 10+. (tanda + menunjukkan tersensor)

Metode Non-Parametrik untuk Survival

Penduga untuk $S(t)$ bila data tidak tersensor:

$$\hat{S}(t) = \frac{s}{N}$$

dimana s adalah banyaknya individu yang masih hidup lebih lama dari t ; N adalah total banyaknya individu

Untuk Data yang tersensor:

- Kaplan-Meier
- Nelson-Aalen

Kaplan-Meier

Estimator untuk $S(t)$ (sering disebut juga sebagai Product-Limit estimator)

$$\hat{S}(t) = \begin{cases} 1 & \text{jika } t < t_1 \\ \prod_{t_i \leq t}(1 - \frac{d_i}{Y_i}) & \text{jika } t_i \leq t \end{cases}$$

dimana d_i adalah banyaknya *event* dan Y_i adalah banyaknya individu yang beresiko (*number at risk*)

Kaplan-Meier

Variansi dari KM estimator (Greenwood's formula)

$$\mathrm{var}[\hat{S}(t)] = \hat{S}(t)^2 \sum_{t_i \le t} \frac{d_i}{Y_i(Y_i - d_i)}$$

Alternatif:

$$\mathrm{var}[\hat{S}(t)] = \hat{S}(t)^2 \frac{[1 - \hat{S}(t)]}{Y(t)}$$

Nelson-Aalen

Estimator untuk fungsi hazard kumulatif:

$$\hat{H}(t) = \begin{cases} 0 & \text{jika } t < t_1 \\ \sum_{t_i \le t} \frac{d_i}{Y_i} & \text{jika } t_i \le t \end{cases}$$

dengan variansi

$$\hat{\mathrm{Var}}(\hat{H}(t)) = \sum_{t_i \le t} \frac{d_i}{Y_i^2}$$

Kaplan-Meier

t	d_i	Y_i	$\hat{S}(t)$
4	1	10	$(1 - \frac{1}{10}) = 0,9$
5	1	9	$0,9(1 - \frac{1}{9}) = 0,8$
6	1	8	$0,8(1 - \frac{1}{8}) = 0,7$
8	3	7	$0,7(1 - \frac{3}{7}) = 0,4$
10	2	4	$0,4(1 - \frac{2}{5}) = 0,2$
11	1	2	$0,2(1 - \frac{1}{2}) = 0,1$
12	1	1	$0,1(1 - \frac{1}{1}) = 0,0$

Kaplan-Meier di SPSS

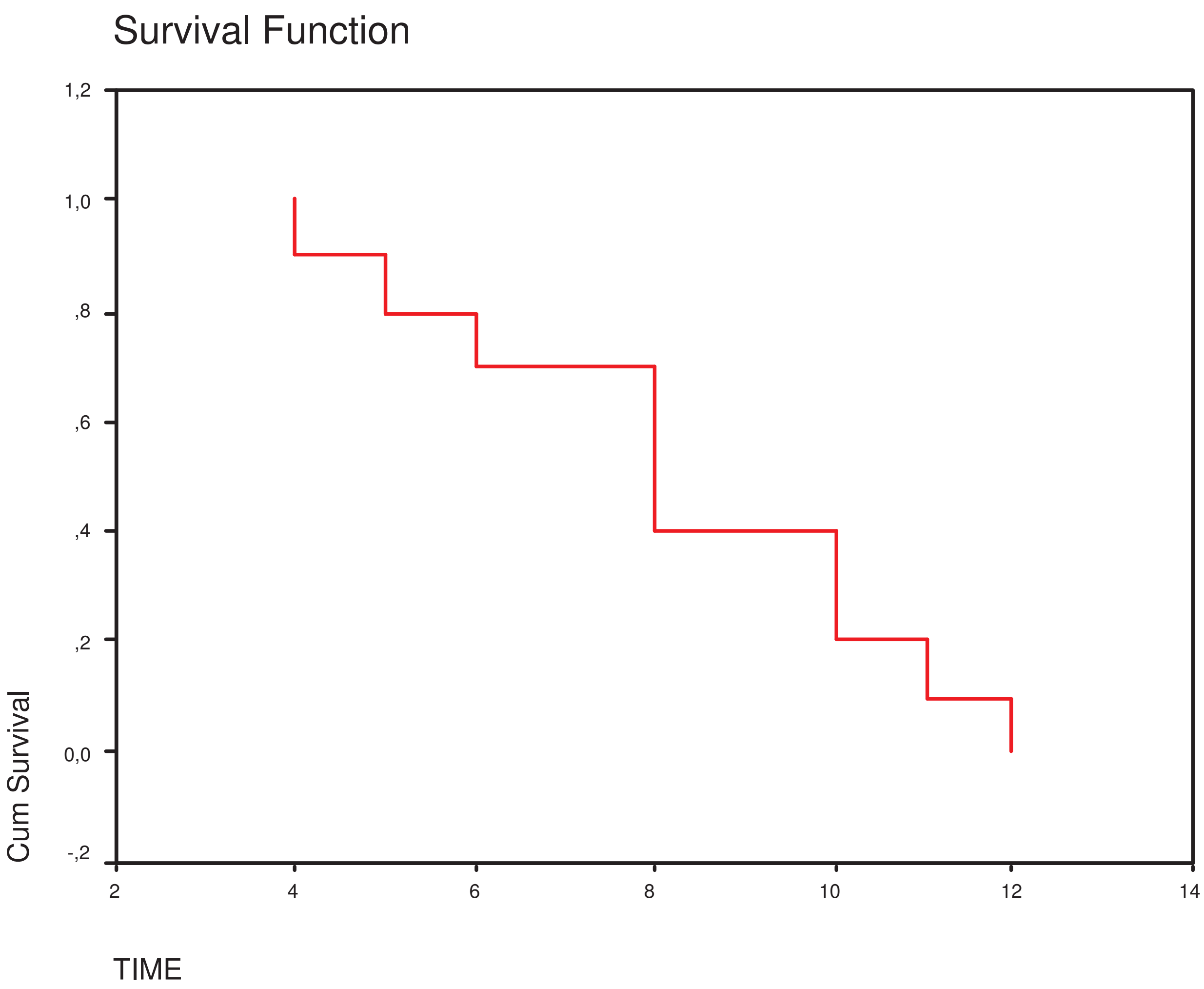
```
Survival Analysis for TIME

Time     Status   Cumulative Standard  Cumulative  Number
                   Survival   Error      Events   Remaining

  4,00      1,00    ,9000      ,0949          1         9
  5,00      1,00    ,8000      ,1265          2         8
  6,00      1,00    ,7000      ,1449          3         7
  8,00      1,00                              4         6
  8,00      1,00                              5         5
  8,00      1,00    ,4000      ,1549          6         4
 10,00      1,00                              7         3
 10,00      1,00    ,2000      ,1265          8         2
 11,00      1,00    ,1000      ,0949          9         1
 12,00      1,00    ,0000      ,0000         10         0


 Number of Cases:  10  Censored:   0   (   ,00%)   Events: 10
```

Kaplan-Meier di SPSS

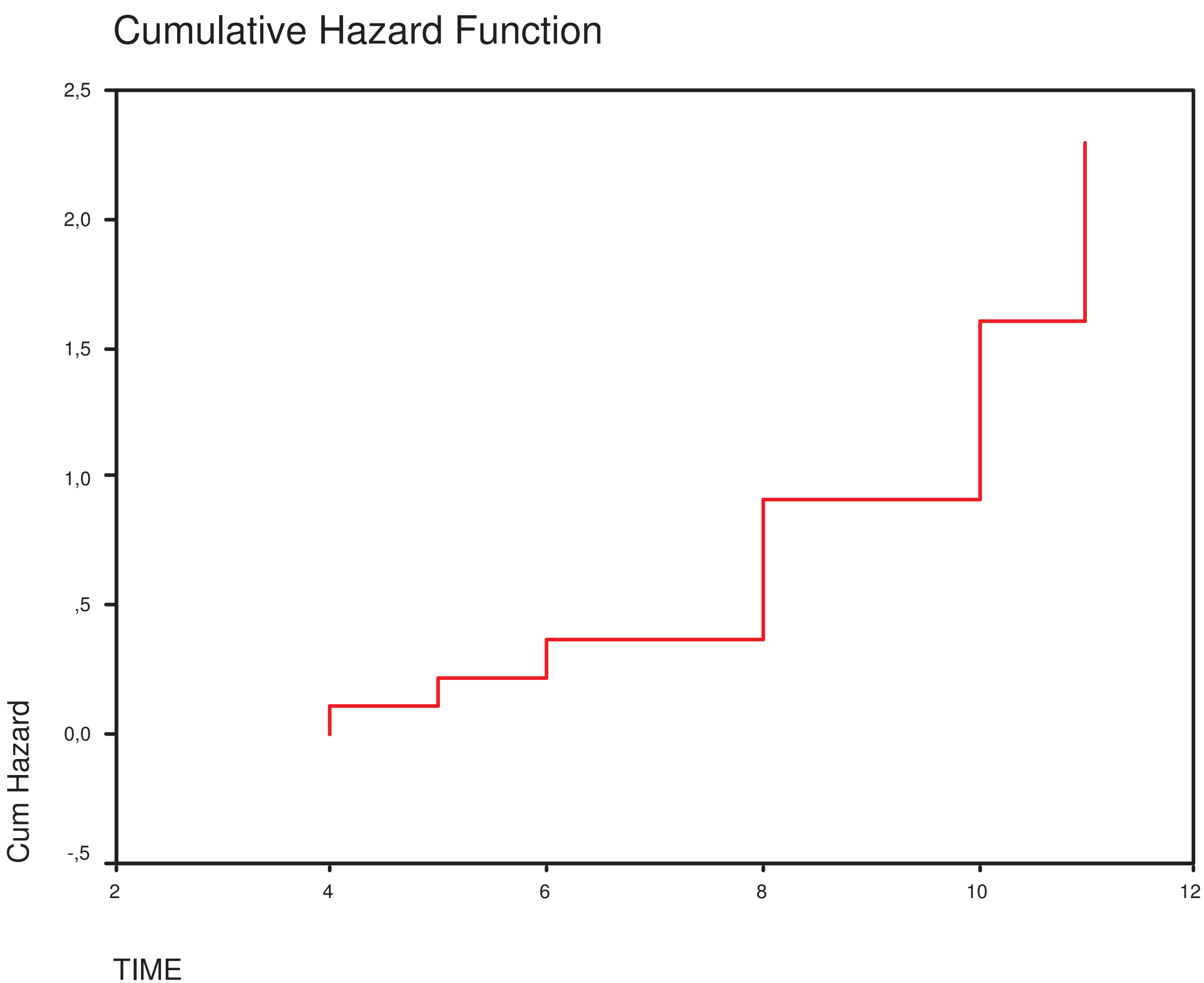
```
KM
  time  /STATUS=status(1)
  /PRINT TABLE MEAN
  /PLOT SURVIVAL HAZARD .
```

Menu: Analyze – Survival – Kaplan-Meier

Kaplan-Meier di SPSS



Fungsi Hazard kumulatif di SPSS

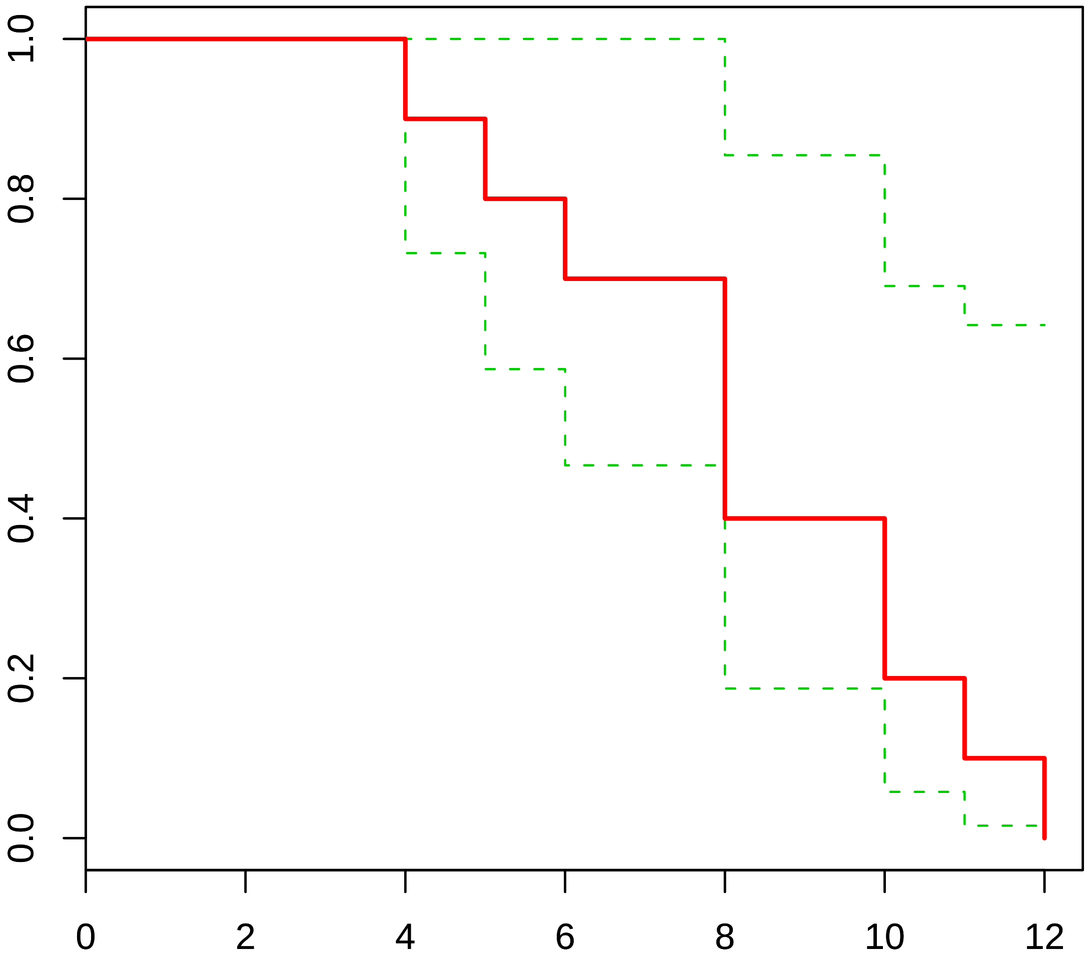


Kaplan-Meier di R

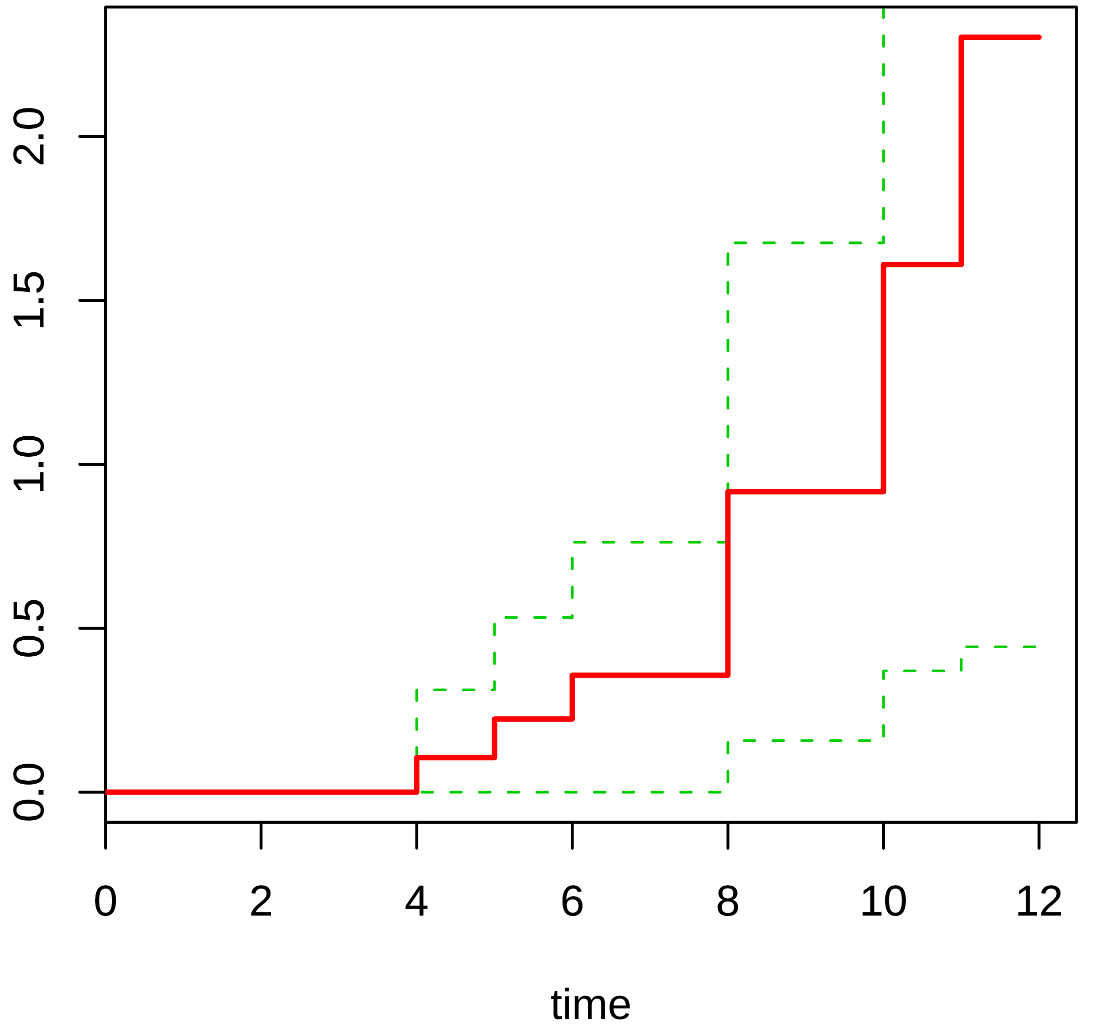
```
library(survival)
c1<-survfit(Surv(TIME,STATUS)~1,data=contohKM)

windows(width=5,height=5)
plot(c1,col=3)
par(new=T)
plot(c1,col=2,lwd=2,conf.int=F)

plot(c1,xlab="time",col=3,fun="cumhaz")
par(new=T)
plot(c1,col=2,lwd=2,conf.int=F,fun="cumhaz")
```

(Lebih rumit dari SPSS, tapi lebih *powerful* dan fleksibel)

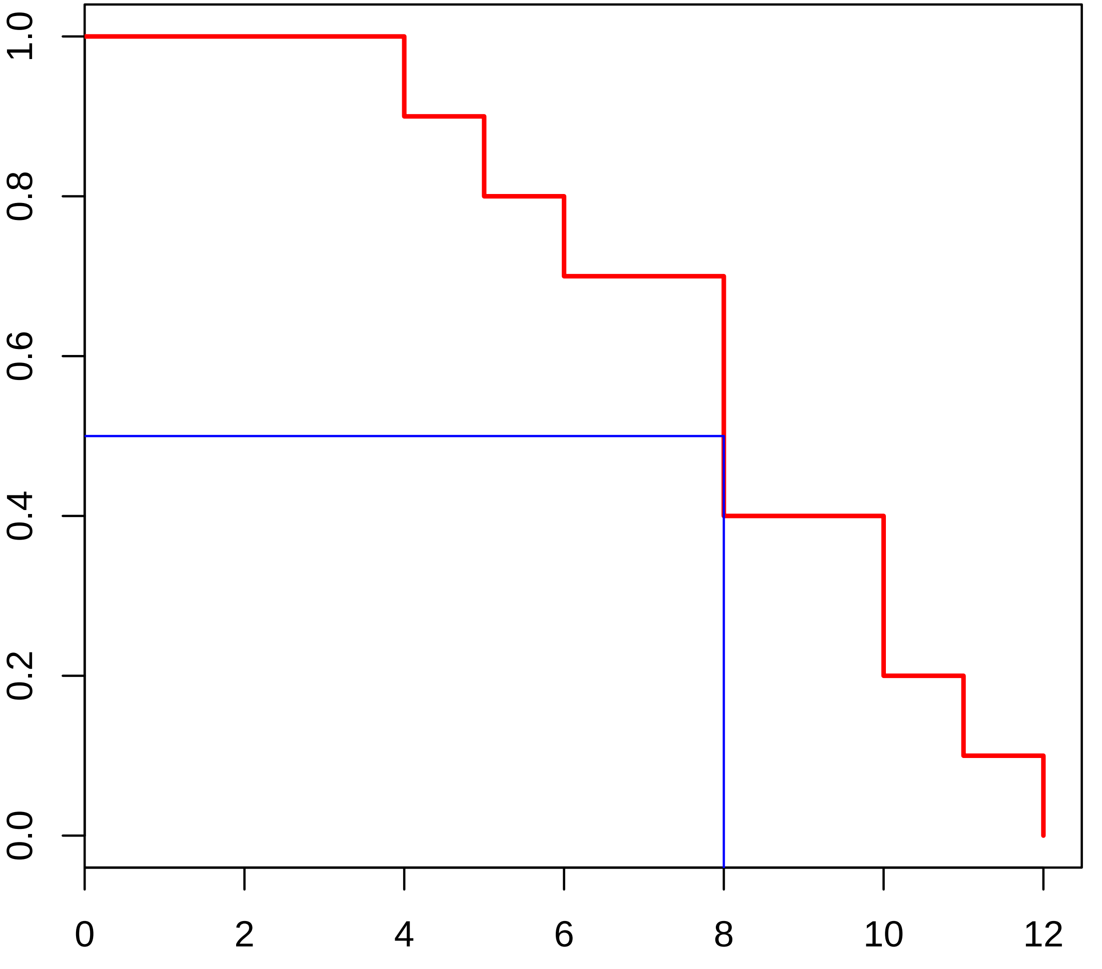
Kaplan-Meier di R



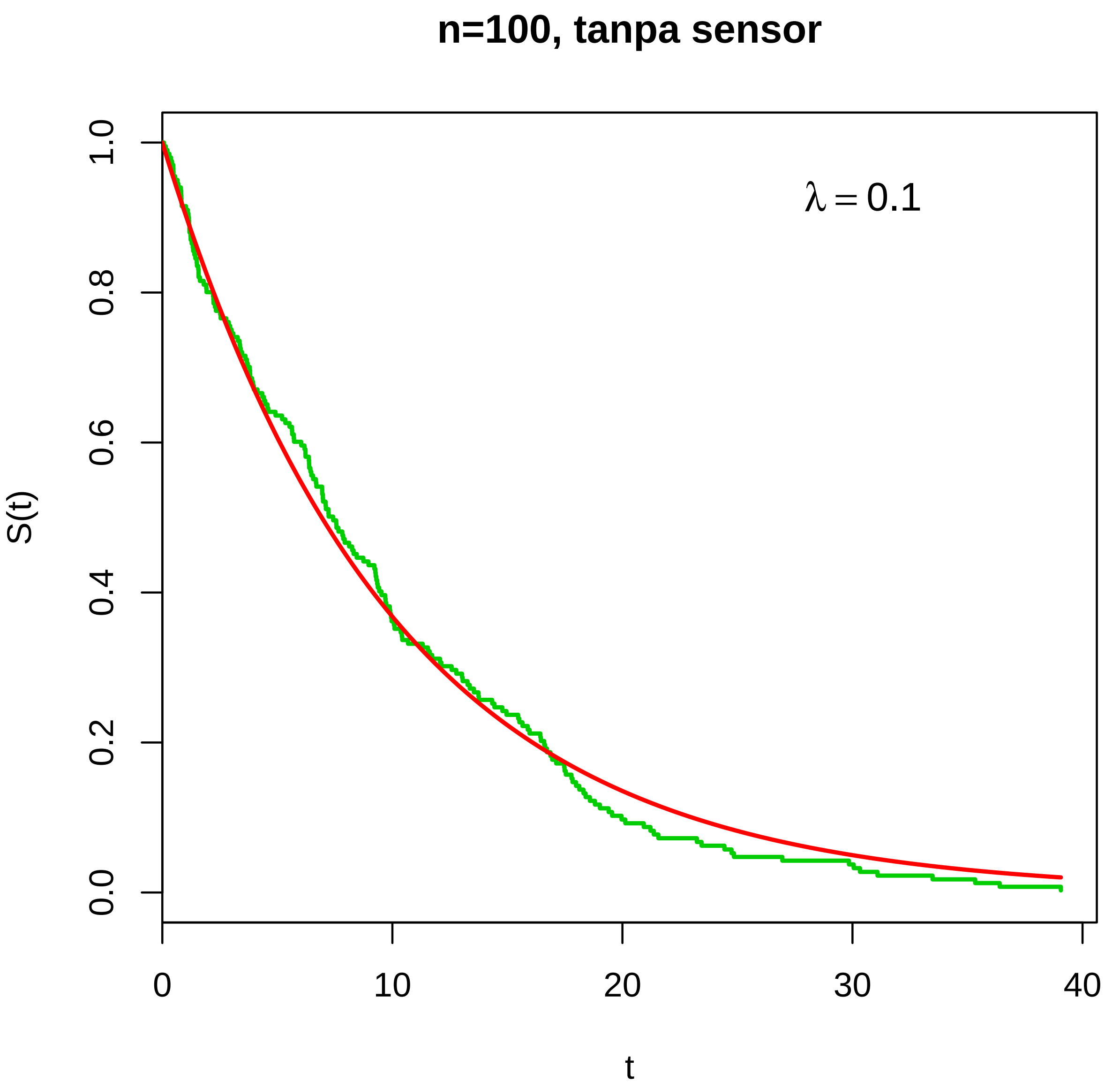
Fungsi Hazard Kumulatif di R



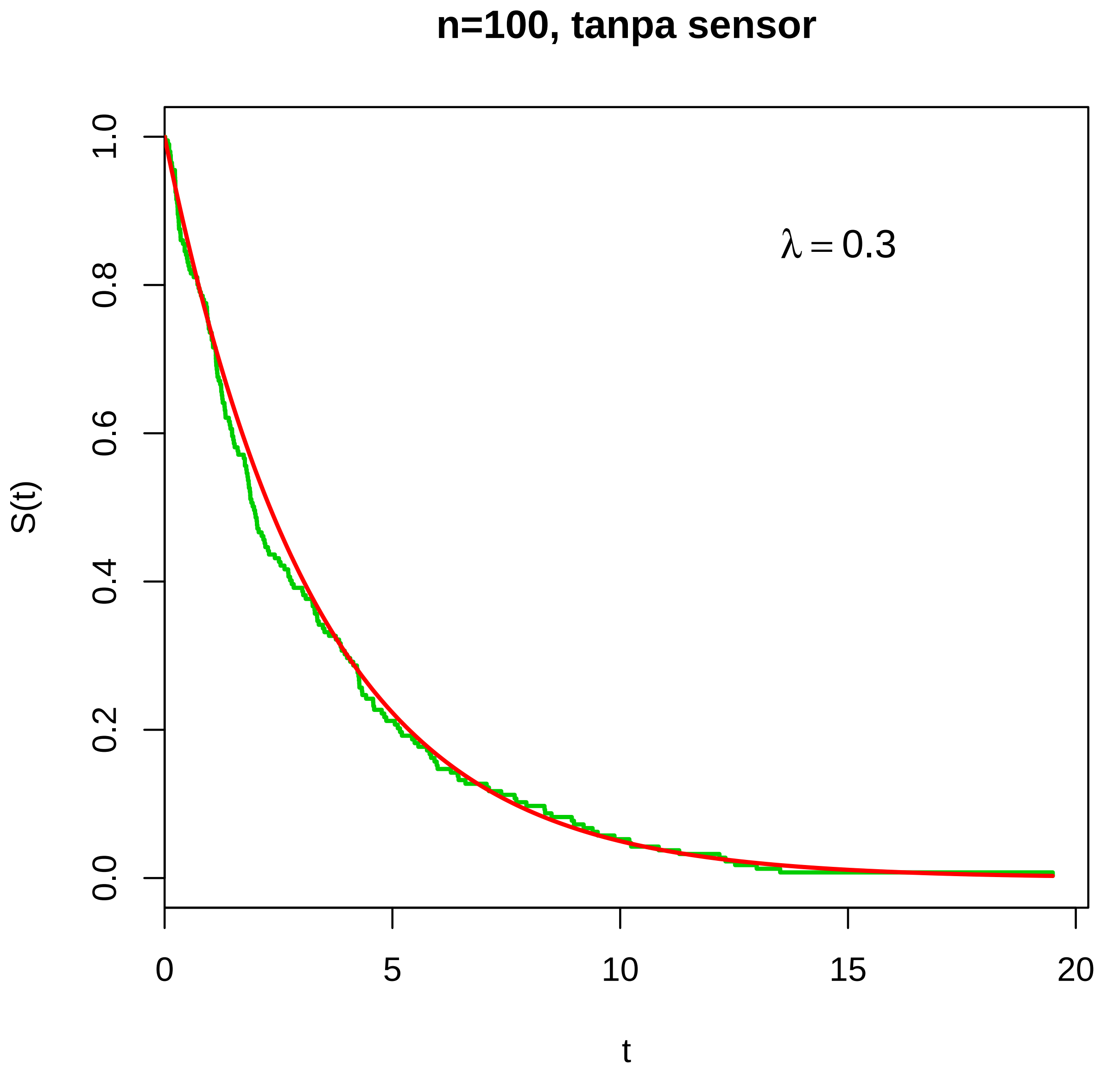
Median Survival Time



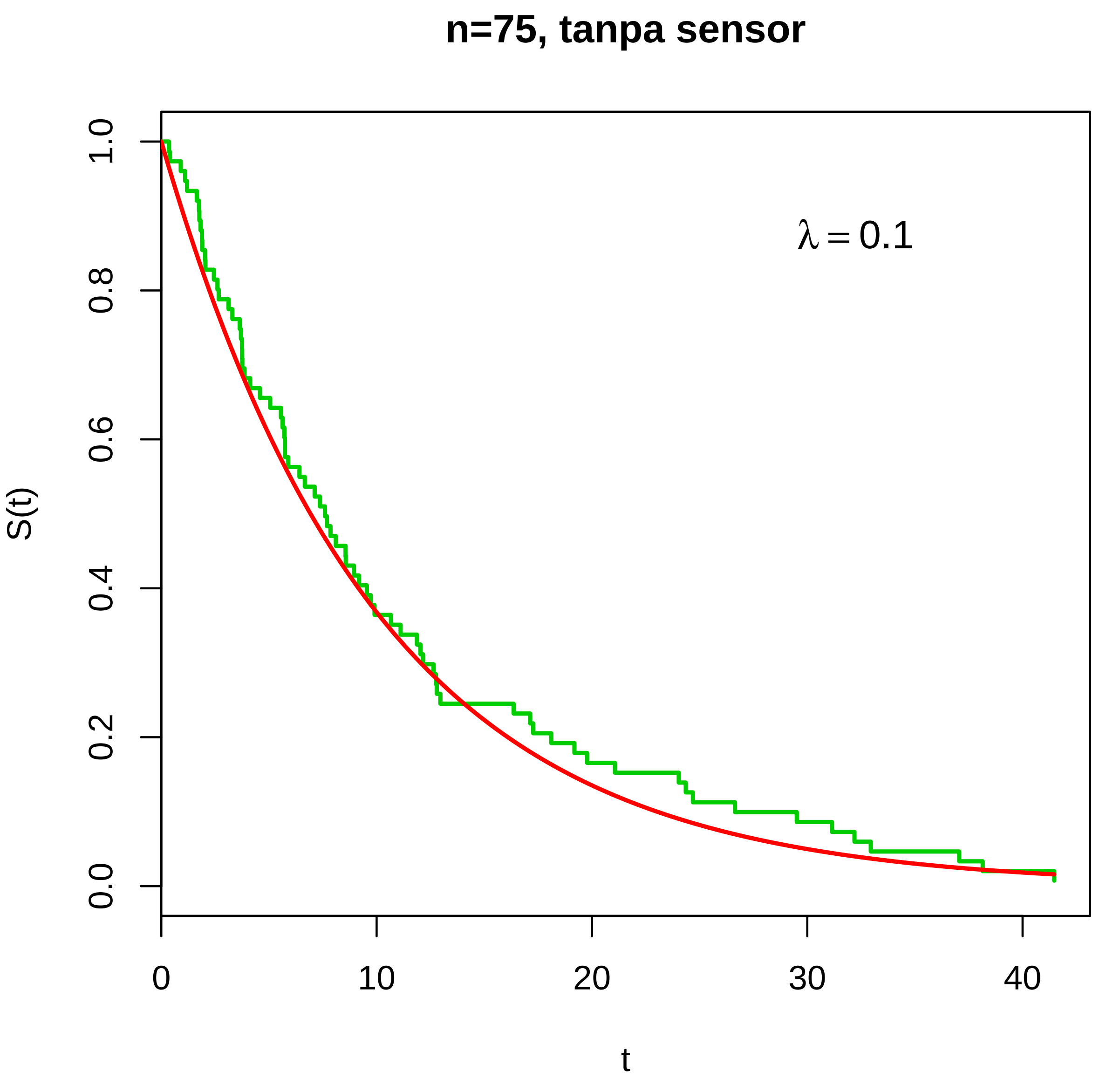
Model Eksponensial - KM



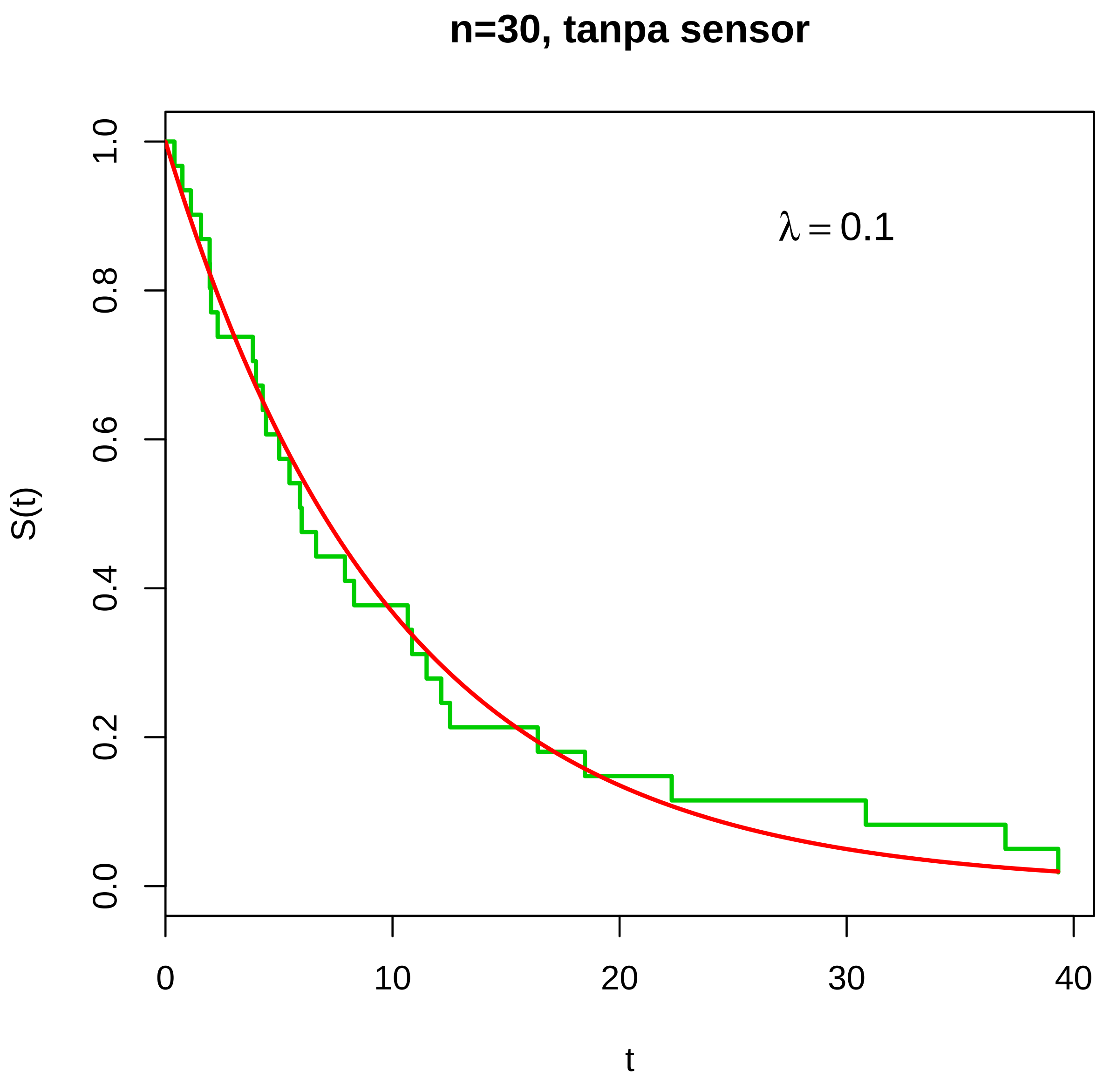
Model Eksponensial - KM



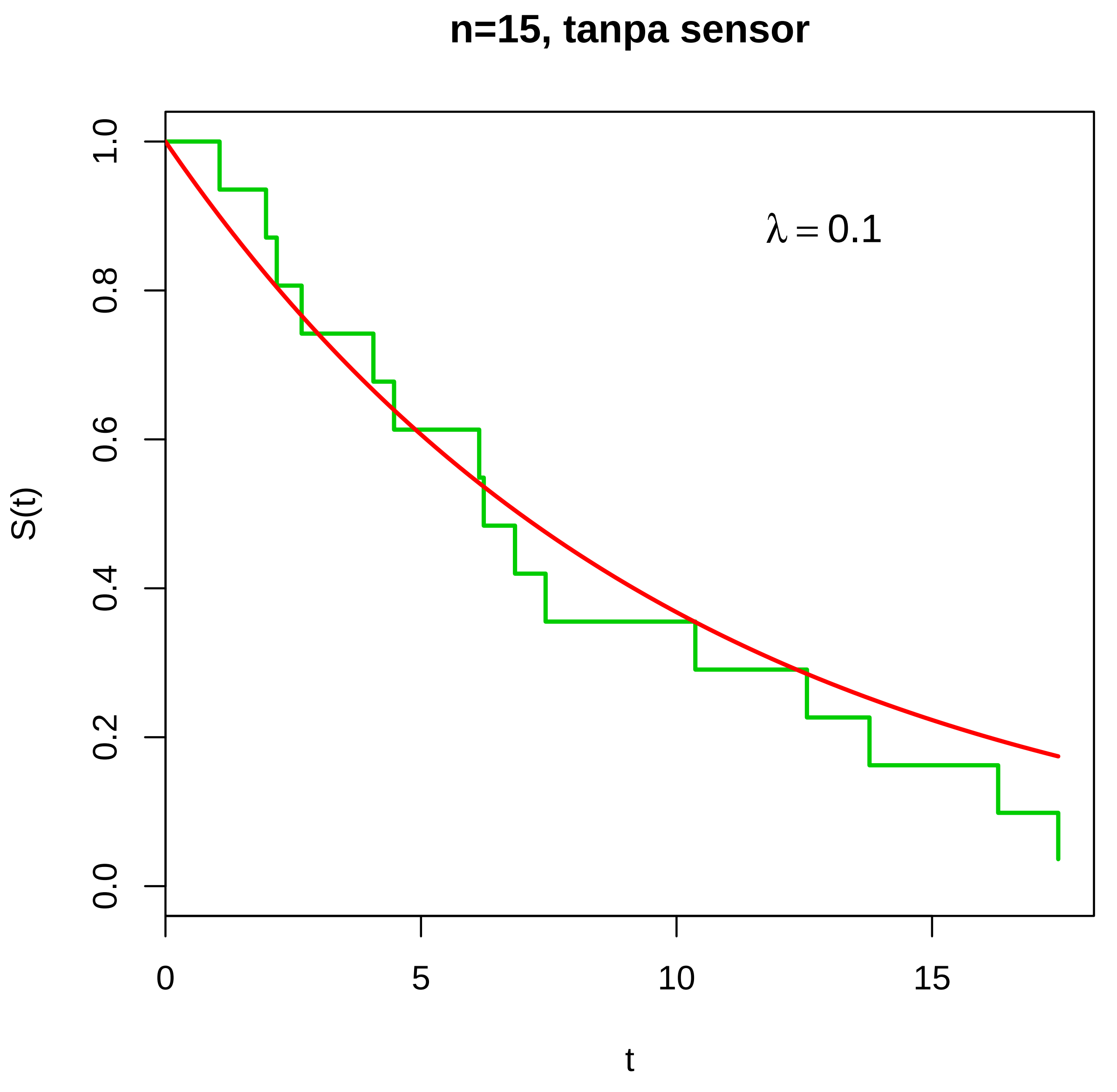
Model Eksponensial - KM



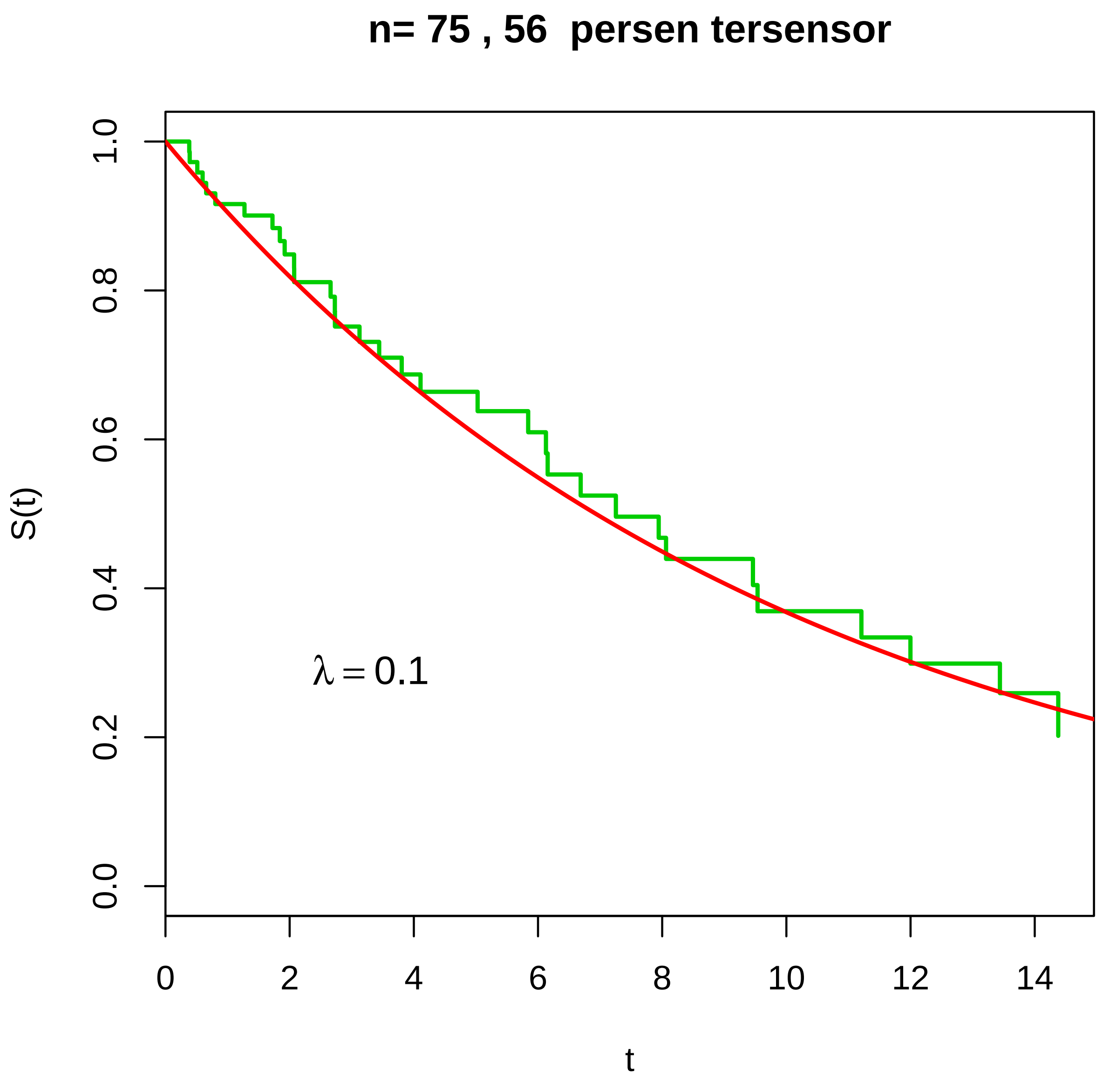
Model Eksponensial - KM



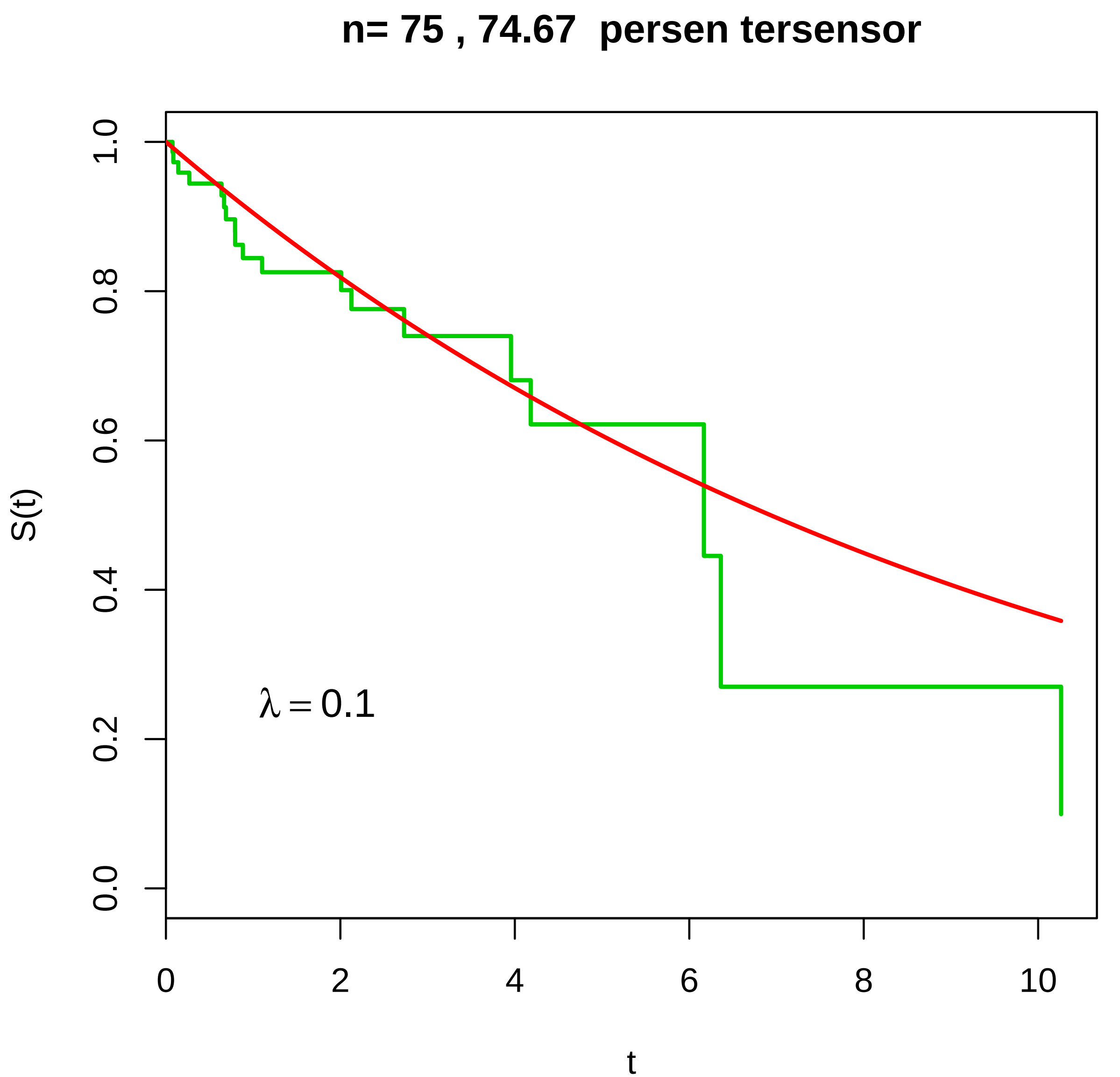
Model Eksponensial - KM



Model Eksponensial - KM



Model Eksponensial - KM

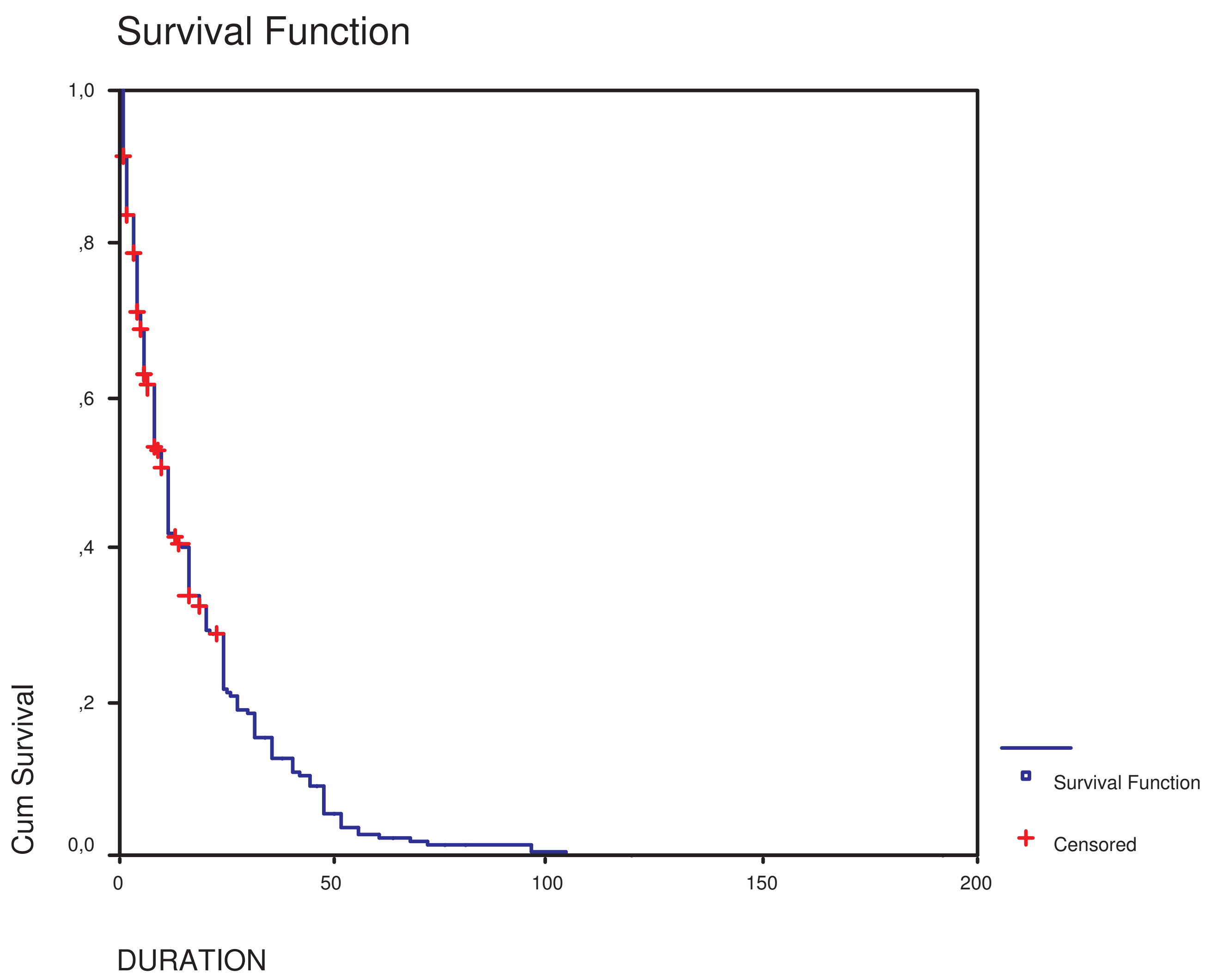


Data ASI

	dur	d	race	pvty	smk	alco	agmth	ybirth	yschool	pc3
1	16	1	1	0	0	1	24	82	14	0
2	1	1	1	0	1	0	26	85	12	0
3	4	0	1	0	0	0	25	85	12	0
4	3	1	1	0	1	1	21	85	9	0
5	36	1	1	0	1	0	22	82	12	0

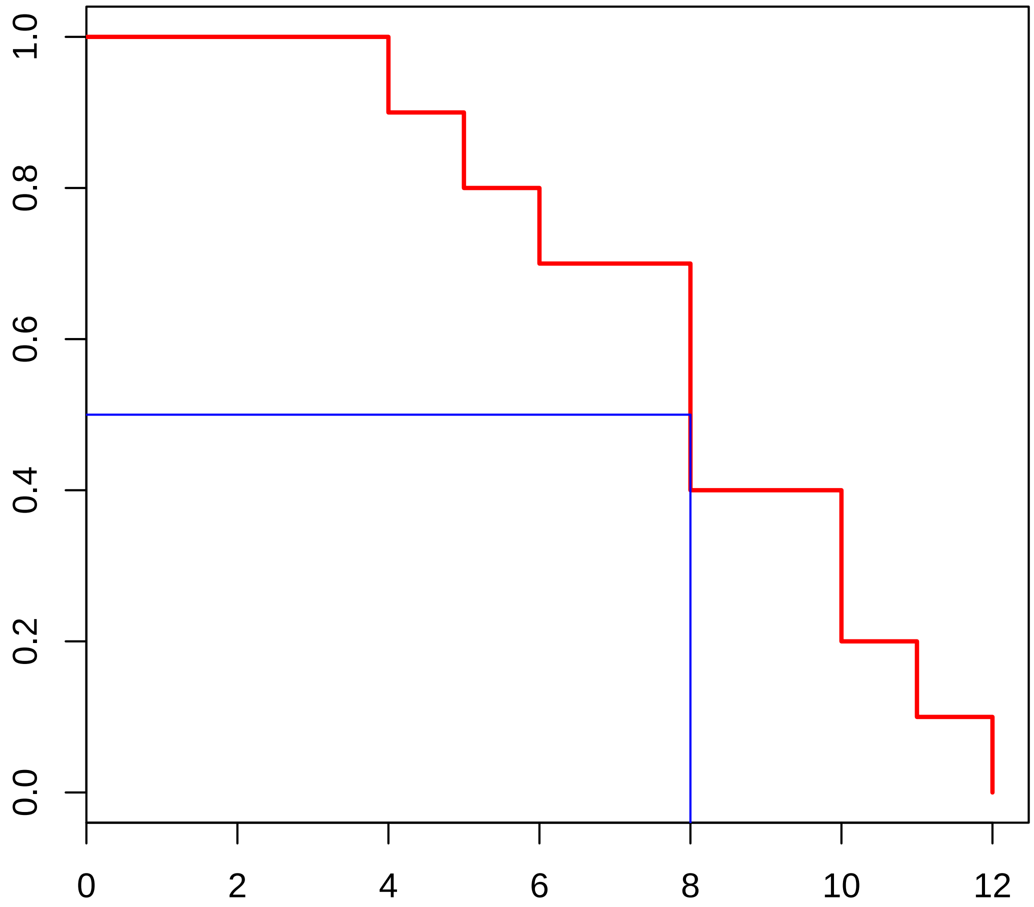
```
KM
  dur  /STATUS=d(1)
  /PRINT TABLE MEAN
  /PLOT SURVIVAL HAZARD .
```

KM Data ASI

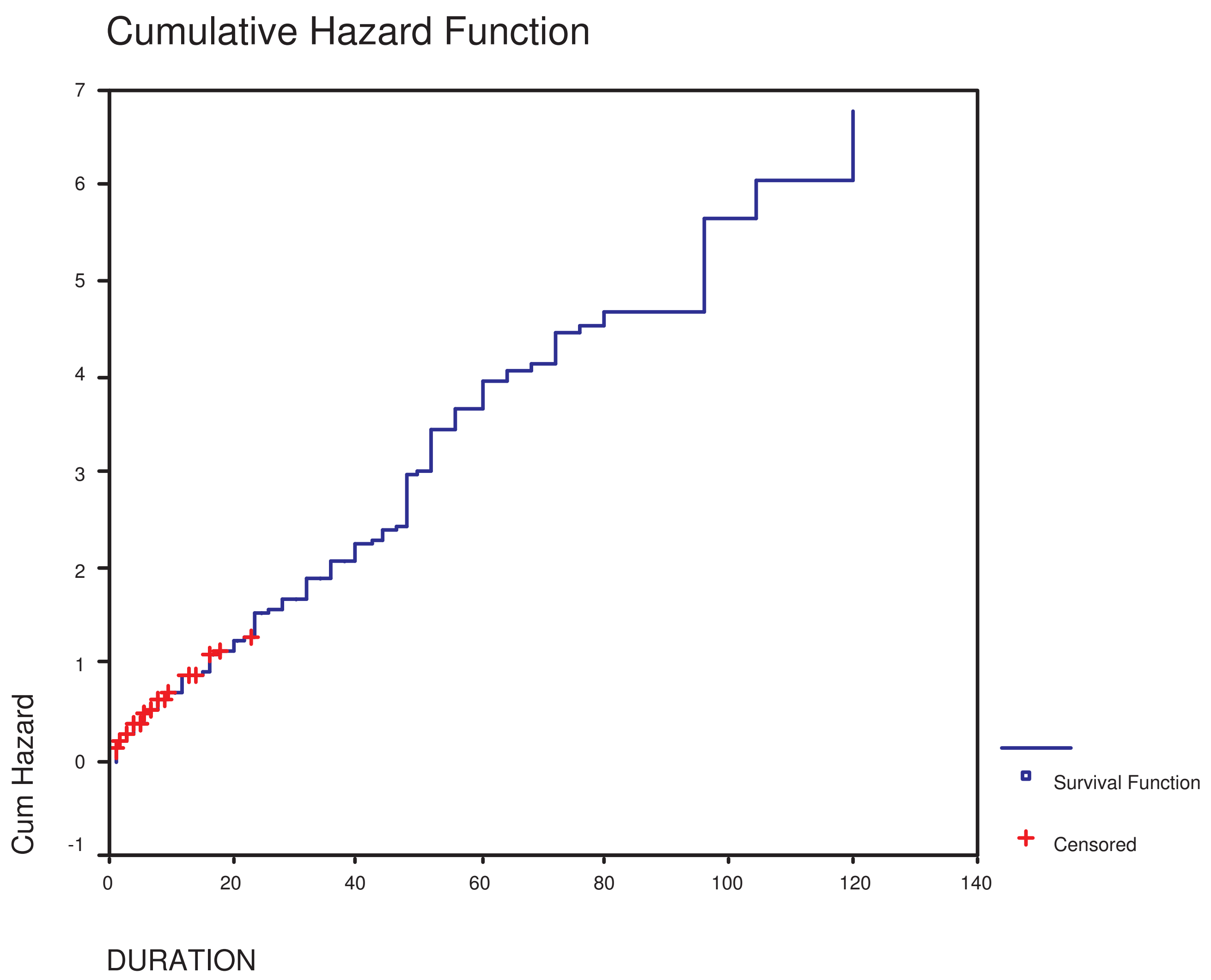


Survival Function
1,0
,8
,6
,4
,2
0,0
Cum Survival
0
50
100
150
200
DURATION
Survival Function
Censored

Median Survival Time ASI



Hazard Kumulatif ASI



Membandingkan Distribusi Survival

Membandingkan dua populasi yang masing-masing mempunyai fungsi survival $S_1(t)$ dan $S_2(t)$

Hipotesis null: $H_0 : S_1(t) = S_2(t)$

Hipotesis alternatif:
$H_1 : S_1(t) > S_2(t)$
$H_1 : S_1(t) < S_2(t)$
$H_1 : S_1(t) \neq S_2(t)$

Membandingkan Distribusi Survival

Metode Non-parametrik

Untuk data tidak tersensor

- Wilcoxon (1945)
- Mann-Whitney (1947)
- Sign test (1977)

Membandingkan Distribusi Survival

Metode Non-parametrik

Untuk data tersensor

- Gehan's generalized Wilcoxon test (1965)
- the Cox-Mantel test (Cox 1959, 1972; Mantel, 1966)
- the logrank test (1972)
- Peto and Peto's generalized Wilcoxon test (1972)
- Cox's F-test (1964)

Logrank Test

Berdasarkan *observed* dan *expected* event pada setiap *event-time*

Untuk 2 grup
Statistik penguji:

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

dengan χ^2 $\sim$Chi-square(df=1)

Logrank Test

Contoh:
grup 1: 23, 16+, 18+, 20+, 24+
grup 2: 15, 18, 19, 19, 20

$$H_0 : S_1(t) = S_2(t)$$
$$H_1 : S_1(t) \neq S_2(t)$$

Logrank Test

Contoh:
grup 1: 23, 16+, 18+, 20+, 24+
grup 2: 15, 18, 19, 19, 20

$$H_0 : S_1(t) = S_2(t)$$
$$H_1 : S_1(t) \neq S_2(t)$$

t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}

t : *event-time*

d_t: banyaknya event

n_1, n_2: *number at risk*

e_{1t}, e_{2t}: *expected event*

Logrank Test

Contoh:
grup 1: 23, 16+, 18+, 20+, 24+
grup 2: 15, 18, 19, 19, 20

$H_0 : S_1(t) = S_2(t)$
$H_1 : S_1(t) \neq S_2(t)$

t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15					
18					
19					
20					
23					

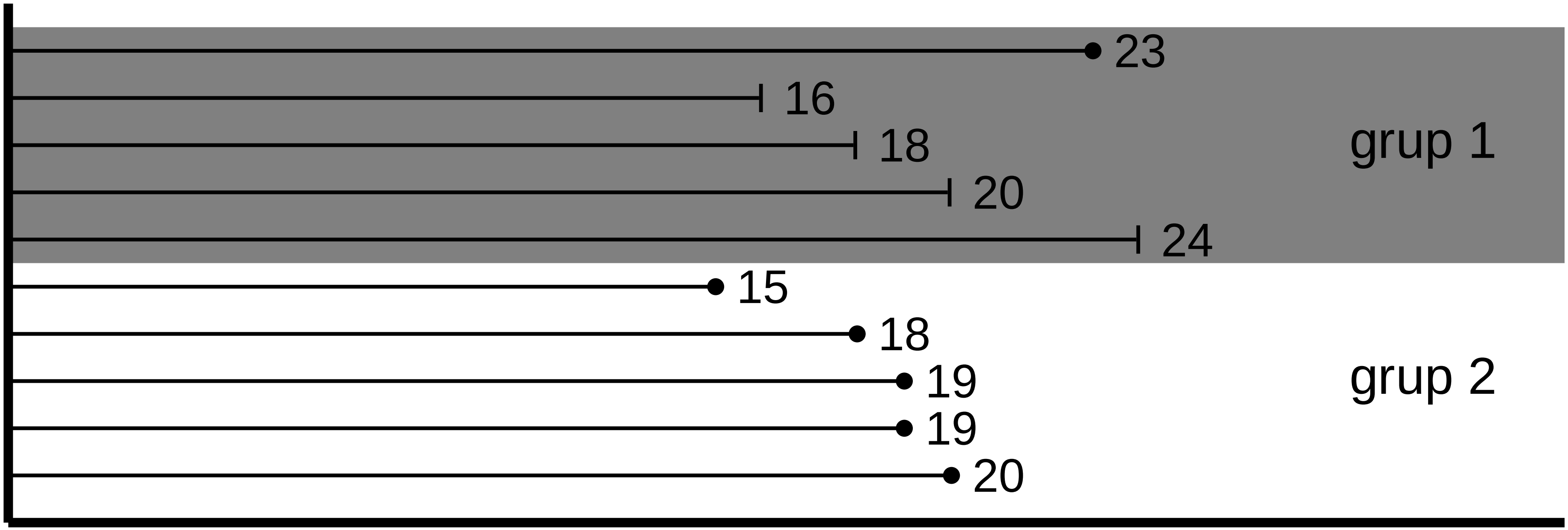
t : *event-time*

d_t: banyaknya event

n_1, n_2: *number at risk*

e_{1t}, e_{2t}: *expected event*

Logrank Test



t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15					
18					
19					
20					
23					

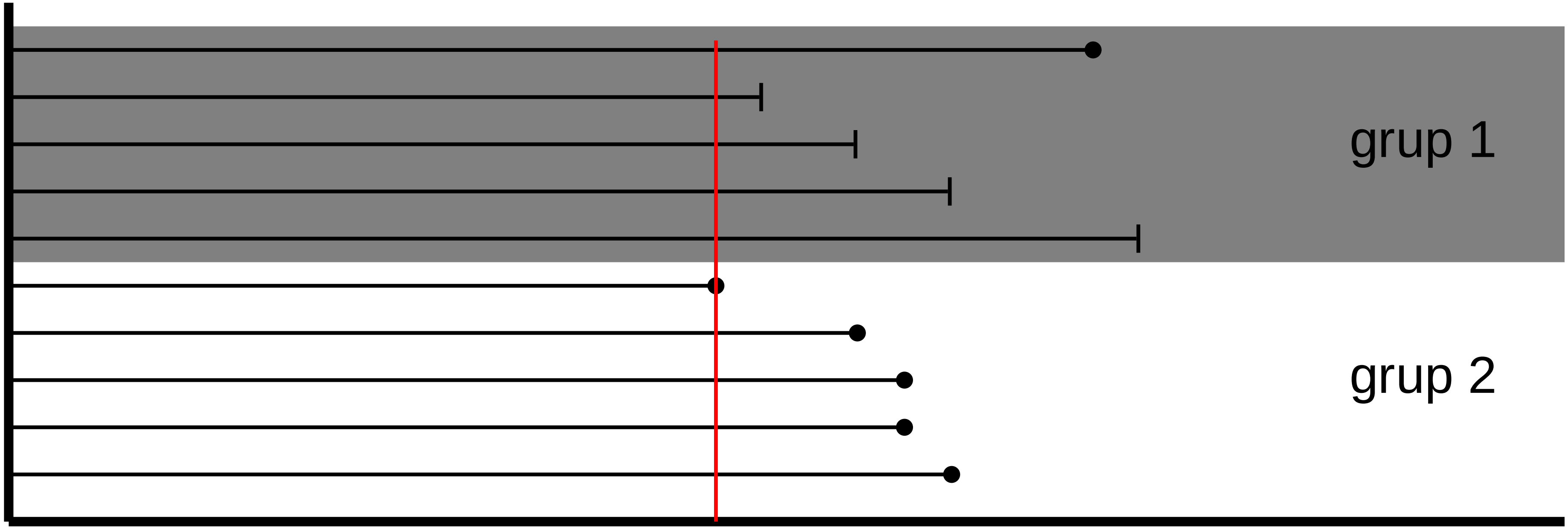
t : *event-time*

d_t: banyaknya event

n_1, n_2: *number at risk*

e_{1t}, e_{2t}: *expected event*

Logrank Test



t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5		
18					
19					
20					
23					

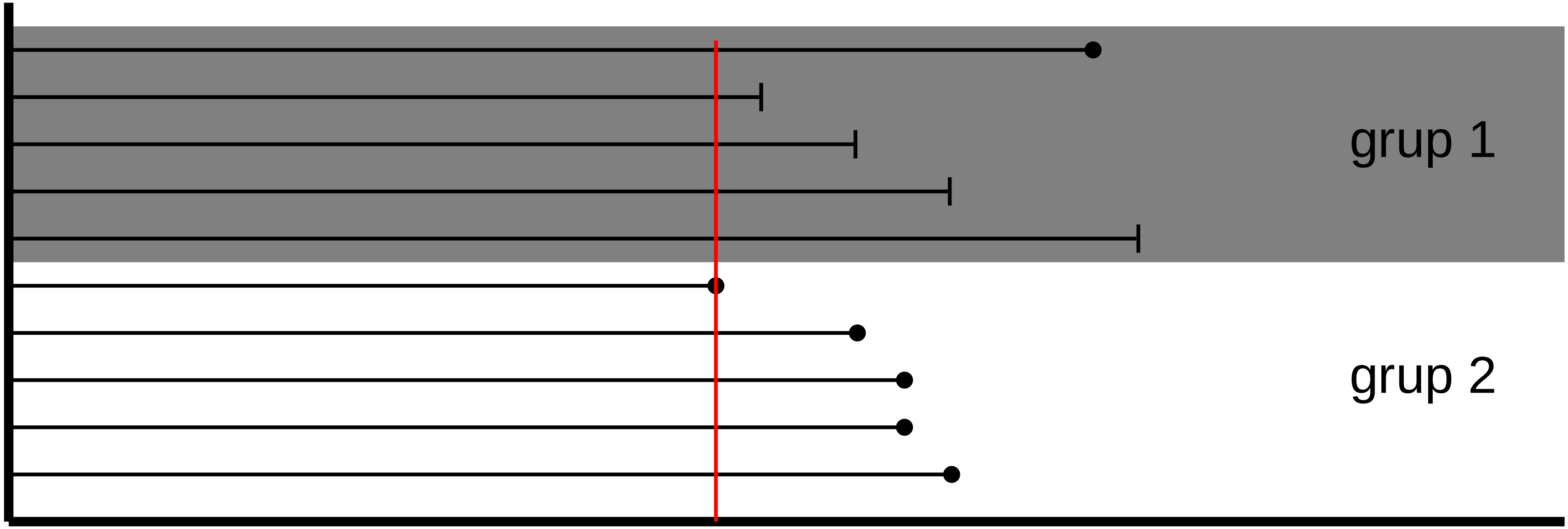
t : *event-time*

d_t: banyaknya event

n_1, n_2: *number at risk*

e_{1t}, e_{2t}: *expected event*

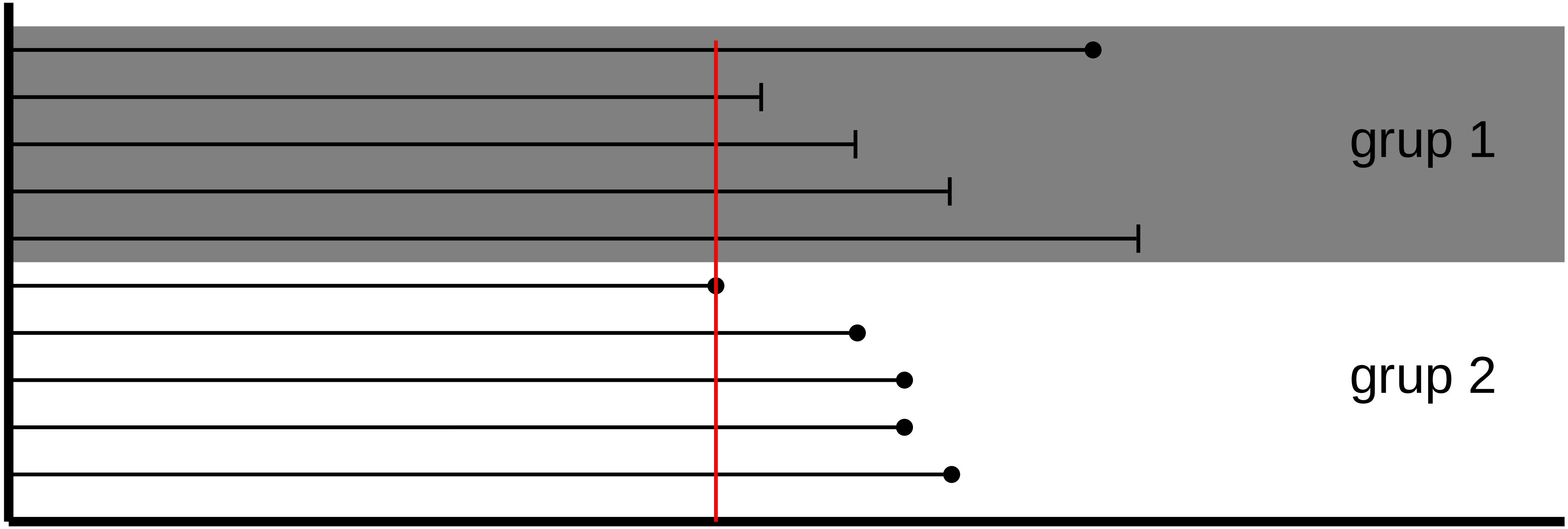
Logrank Test



t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5		
18					
19					
20					
23					

$$e_{1t} = \frac{n_{1t}}{n_{1t}+n_{2t}} \times d_t$$
$$e_{2t} = \frac{n_{2t}}{n_{1t}+n_{2t}} \times d_t$$

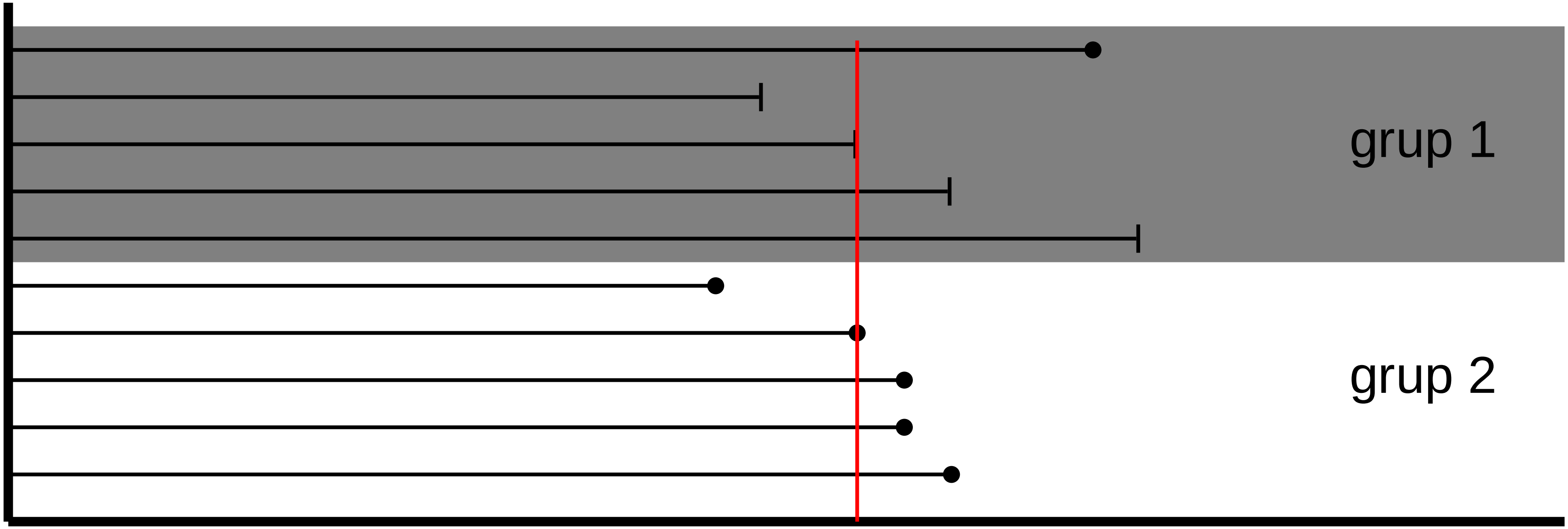
Logrank Test



t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5	0,5	0,5
18					
19					
20					
23					

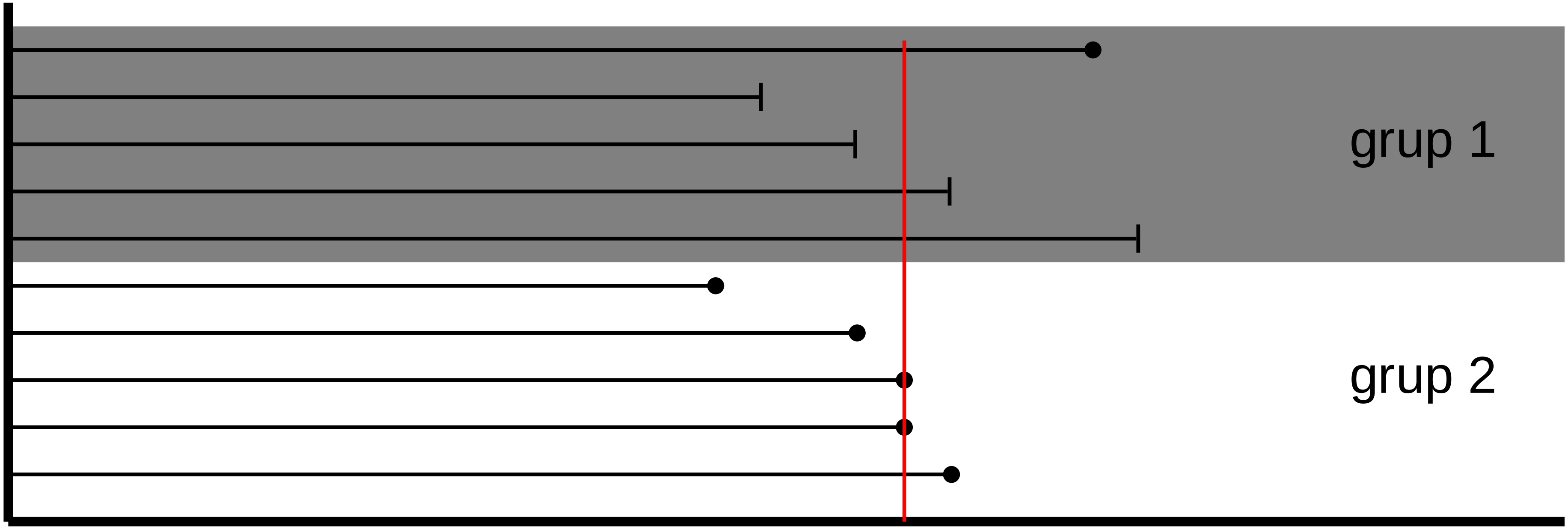
$$e_{1t} = \frac{n_{1t}}{n_{1t}+n_{2t}} \times d_t$$
$$e_{2t} = \frac{n_{2t}}{n_{1t}+n_{2t}} \times d_t$$

Logrank Test



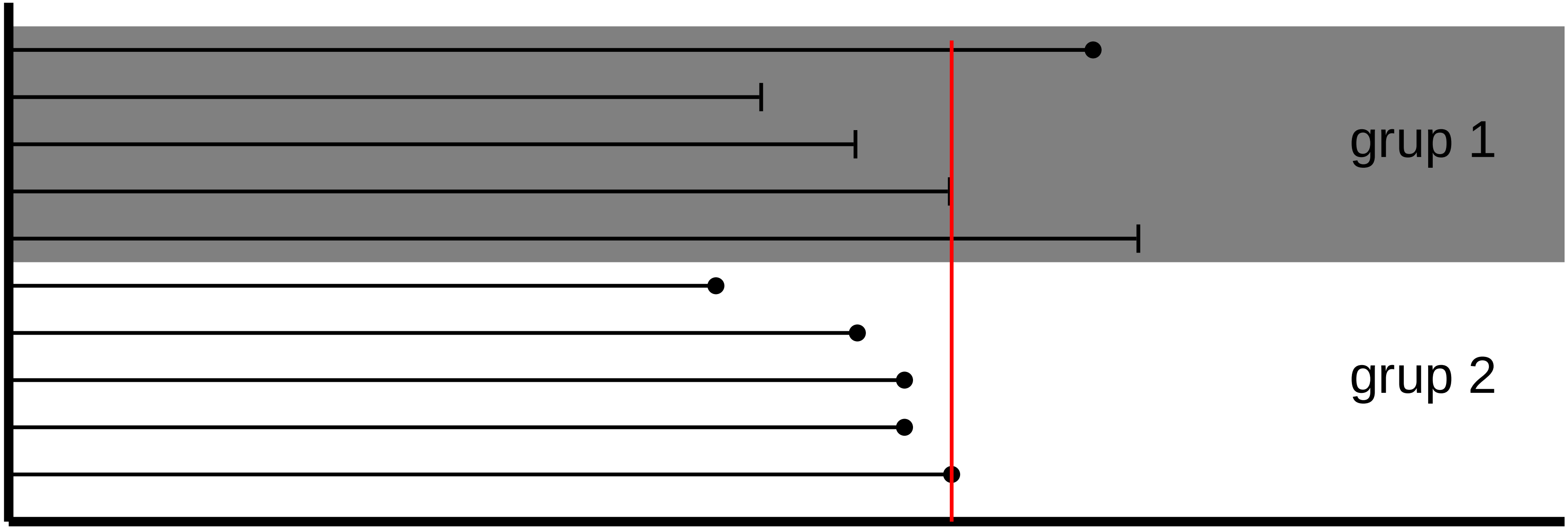
t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5	0,5	0,5
18	1	4	4	0,5	0,5
19					
20					
23					

Logrank Test



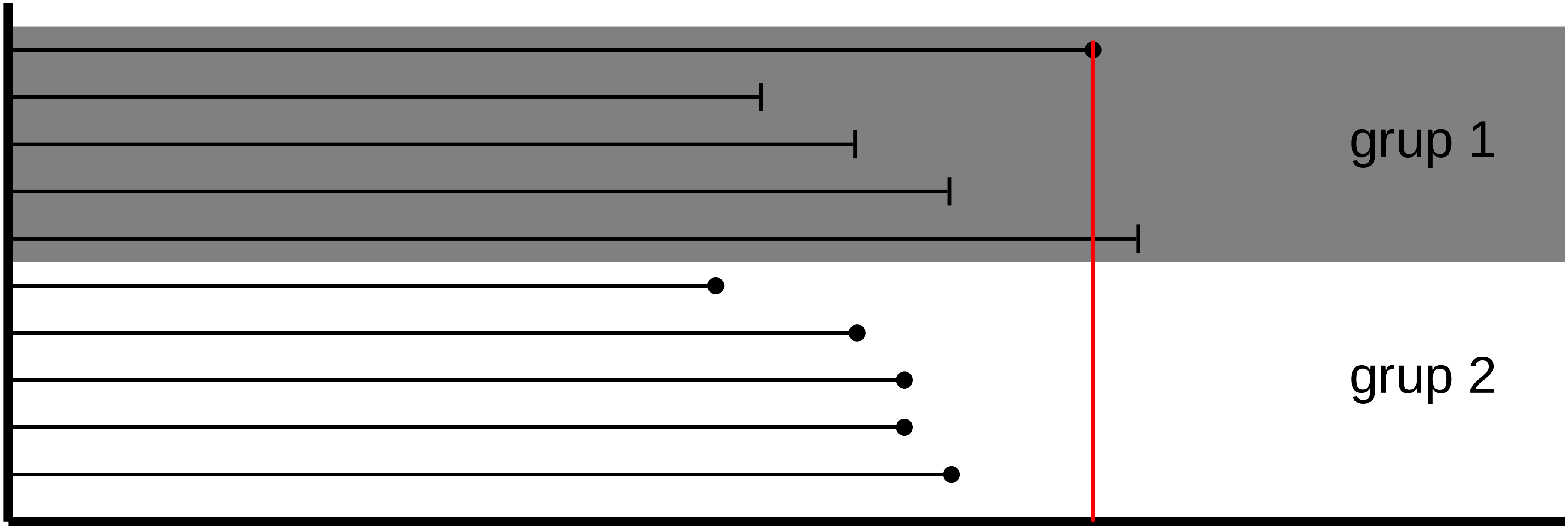
t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5	0,5	0,5
18	1	4	4	0,5	0,5
19	2	3	3	1,0	1,0
20					
23					

Logrank Test



t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5	0,5	0,5
18	1	4	4	0,5	0,5
19	2	3	3	1,0	1,0
20	1	3	1	0,75	0,25
23					

Logrank Test



t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5	0,5	0,5
18	1	4	4	0,5	0,5
19	2	3	3	1,0	1,0
20	1	3	1	0,75	0,25
23	1	2	0	1,0	0

Logrank Test

Contoh:
grup 1: 23, 16+, 18+, 20+, 24+
grup 2: 15, 18, 19, 19, 20

$$H_0 : S_1(t) = S_2(t)$$
$$H_1 : S_1(t) \neq S_2(t)$$

t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5	0,5	0,5
18	1	4	4	0,5	0,5
19	2	3	3	1,0	1,0
20	1	3	1	0,75	0,25
23	1	2	0	1,0	0
				3,75	2,25

Logrank Test

Contoh:
grup 1: 23, 16+, 18+, 20+, 24+
grup 2: 15, 18, 19, 19, 20

$H_0 : S_1(t) = S_2(t)$
$H_1 : S_1(t) \neq S_2(t)$

t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5	0,5	0,5
18	1	4	4	0,5	0,5
19	2	3	3	1,0	1,0
20	1	3	1	0,75	0,25
23	1	2	0	1,0	0
				3,75	2,25

$E_1 = 3,75$

$E_2 = 2,25$

Logrank Test

Contoh:

grup 1: 23, 16+, 18+, 20+, 24+ $O_1 = 1$

grup 2: 15, 18, 19, 19, 20 $O_2 = 5$

$$H_0 : S_1(t) = S_2(t)$$
$$H_1 : S_1(t) \neq S_2(t)$$

t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5	0,5	0,5
18	1	4	4	0,5	0,5
19	2	3	3	1,0	1,0
20	1	3	1	0,75	0,25
23	1	2	0	1,0	0
				3,75	2,25

$$E_1 = 3,75$$
$$E_2 = 2,25$$

Logrank Test

Contoh:
grup 1: 23, 16+, 18+, 20+, 24+
grup 2: 15, 18, 19, 19, 20

$$H_0 : S_1(t) = S_2(t)$$
$$H_1 : S_1(t) \neq S_2(t)$$

t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5	0,5	0,5
18	1	4	4	0,5	0,5
19	2	3	3	1,0	1,0
20	1	3	1	0,75	0,25
23	1	2	0	1,0	0
				3,75	2,25

$$E_1 = 3,75$$
$$E_2 = 2,25$$
$$O_1 = 1$$
$$O_2 = 5$$

Logrank Test

Contoh:
grup 1: 23, 16+, 18+, 20+, 24+
grup 2: 15, 18, 19, 19, 20

$H_0 : S_1(t) = S_2(t)$
$H_1 : S_1(t) \neq S_2(t)$

t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5	0,5	0,5
18	1	4	4	0,5	0,5
19	2	3	3	1,0	1,0
20	1	3	1	0,75	0,25
23	1	2	0	1,0	0
				3,75	2,25

$$\begin{aligned} \chi^2 &= \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \\ &= \frac{(1 - 3,75)^2}{3,75} + \frac{(5 - 2,25)^2}{2,25} \\ &= 5,378 \end{aligned}$$

Logrank Test

Contoh:
grup 1: 23, 16+, 18+, 20+, 24+
grup 2: 15, 18, 19, 19, 20

$H_0 : S_1(t) = S_2(t)$
$H_1 : S_1(t) \neq S_2(t)$

t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5	0,5	0,5
18	1	4	4	0,5	0,5
19	2	3	3	1,0	1,0
20	1	3	1	0,75	0,25
23	1	2	0	1,0	0
				3,75	2,25

$$\begin{aligned}\chi^2 &= \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \\ &= \frac{(1 - 3,75)^2}{3,75} + \frac{(5 - 2,25)^2}{2,25} \\ &= 5,378\end{aligned}$$

p-value$= 0,0204 < 0,05$

Model Regresi

Data ASI (penyapihan)

```
> data(bfeed)
> bfeed
   duration delta race poverty smoke alcohol agemth ybirth yschool pc3mth
1        16     1    1       0     0       1     24     82      14      0
2         1     1    1       0     1       0     26     85      12      0
3         4     0    1       0     0       0     25     85      12      0
4         3     1    1       0     1       1     21     85       9      0
5        36     1    1       0     1       0     22     82      12      0
6        36     1    1       0     0       0     18     82      11      0
7        16     1    1       1     1       0     20     81       9      0
8         8     0    1       0     1       0     24     85      12      0
9        20     1    1       1     0       0     24     85      12      0
10       44     1    1       0     0       0     24     82      14      0
11       20     1    1       0     1       0     26     84      12      0
12       30     1    1       0     1       0     22     84      12      1
13       24     1    1       0     0       0     19     83      12      0
14       13     1    1       0     0       0     22     80      14      0
15        6     1    1       0     0       0     27     84      16      0
16        2     1    1       0     0       0     22     81      12      1
```

Model Regresi

Data ASI (penyapihan)

variabel respon : Lama periode menyusui (minggu) dan status menyusui (disapih atau belum)

variabel penjelas: ras (kulit putih, hitam, yang lain); tingkat kemiskinan, perokok atau tidak, peminum atau tidak, usia ibu (saat melahirkan), pendidikan ibu, pemeriksaan kehamilan

Model Regresi

Data ASI (penyapihan)

variabel respon : Lama periode menyusui (minggu) dan status menyusui (disapih atau belum)

variabel penjelas: ras (kulit putih, hitam, yang lain); tingkat kemiskinan, perokok atau tidak, peminum atau tidak, usia ibu (saat melahirkan), pendidikan ibu, pemeriksaan kehamilan

Bagaimana pengaruh variabel penjelas terhadap variabel respon?

Model Regresi

Model Regresi untuk data antar kejadian:

- Model Regresi Parametrik
- Regresi Cox
- Model Hazard Aditif

Model Regresi Parametrik

- AFT (*accelerated failure-time model*)
- model linear dalam log durasi (lama antar kejadian)
- model hazard proporsional

Model Regresi Parametrik

- Representasi fungsi hazard AFT

$$h(t \mid \boldsymbol{X}) = h_0(\exp(\boldsymbol{X\beta})t)\exp(\boldsymbol{X\beta})$$

dengan $\boldsymbol{X}$ adalah matriks $(n \times p)$ dari variabel penjelas; $\boldsymbol{\beta}^{\mathrm{T}} = (\beta_1 \ldots \beta_p)$ adalah vektor $(p \times 1)$ parameter regresi.

- Representasi $\log T$

$$\log T = \mu + \boldsymbol{X\alpha} + \sigma\epsilon$$

dengan $\boldsymbol{\alpha}^{\mathrm{T}} = (\alpha_1 \ldots \alpha_p)$ dan μ adalah parameter regresi; ϵ adalah suku *error* berdistribusi tertentu dan $\sigma > 0$ adalah suatu parameter skala.

Model AFT

Model AFT dapat ditulis sebagai fungsi hazard atau survival

$$H(t \mid \boldsymbol{x}) = H_0(\exp(\boldsymbol{x\beta})t), \qquad \text{untuk semua } t$$

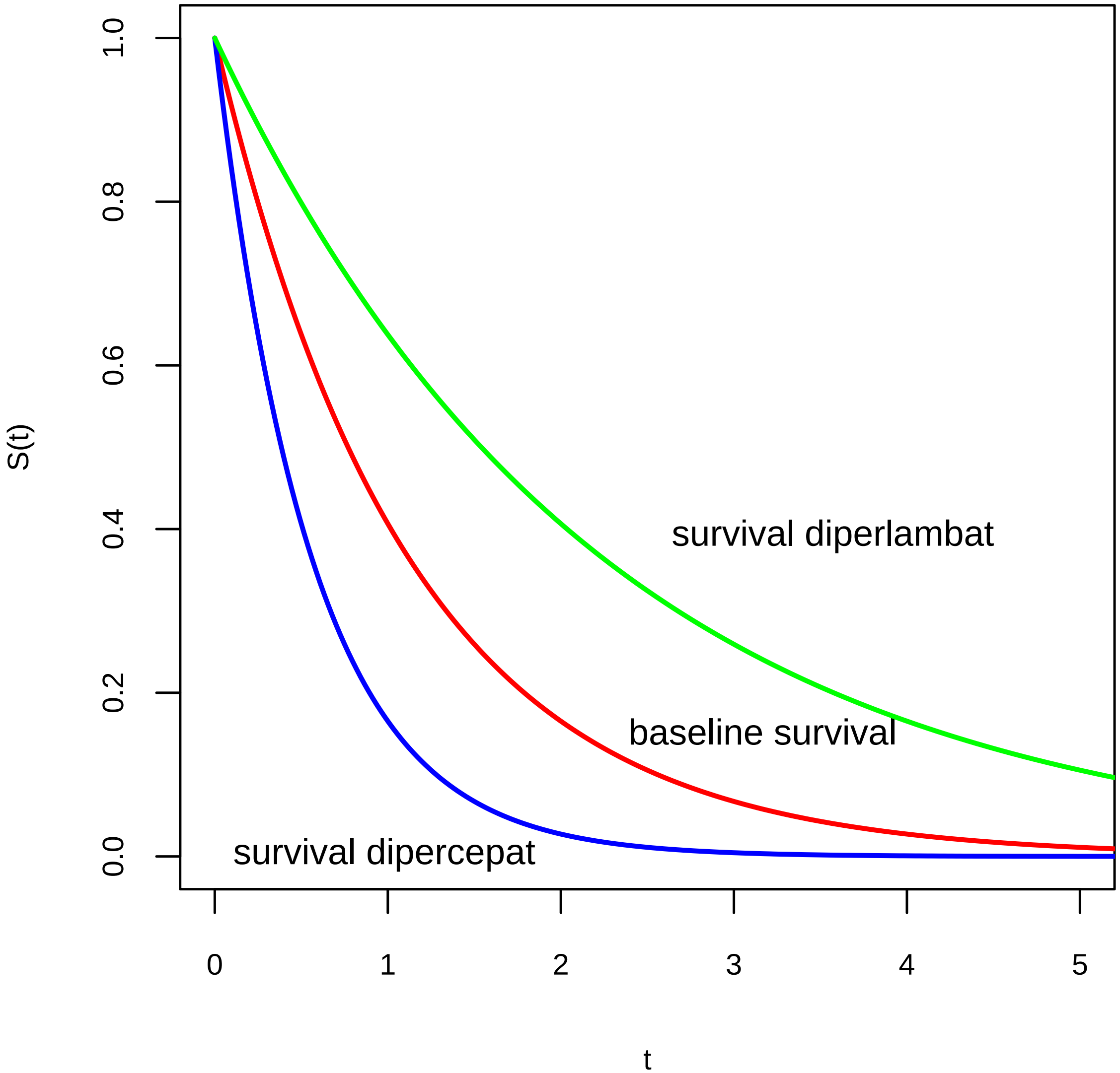
atau

$$S(t \mid \boldsymbol{x}) = S_0(\exp(\boldsymbol{x\beta})t), \qquad \text{untuk semua } t$$

dengan H_0 adalah baseline fungsi hazard kumulatif dan S_0 baseline fungsi survival

Model AFT

Fungsi Survival Model AFT, Eksponensial ($\lambda = 0{,}9$):



$$S(t \mid x) = S_0(xt)$$
$$= \exp(-x\lambda t)$$

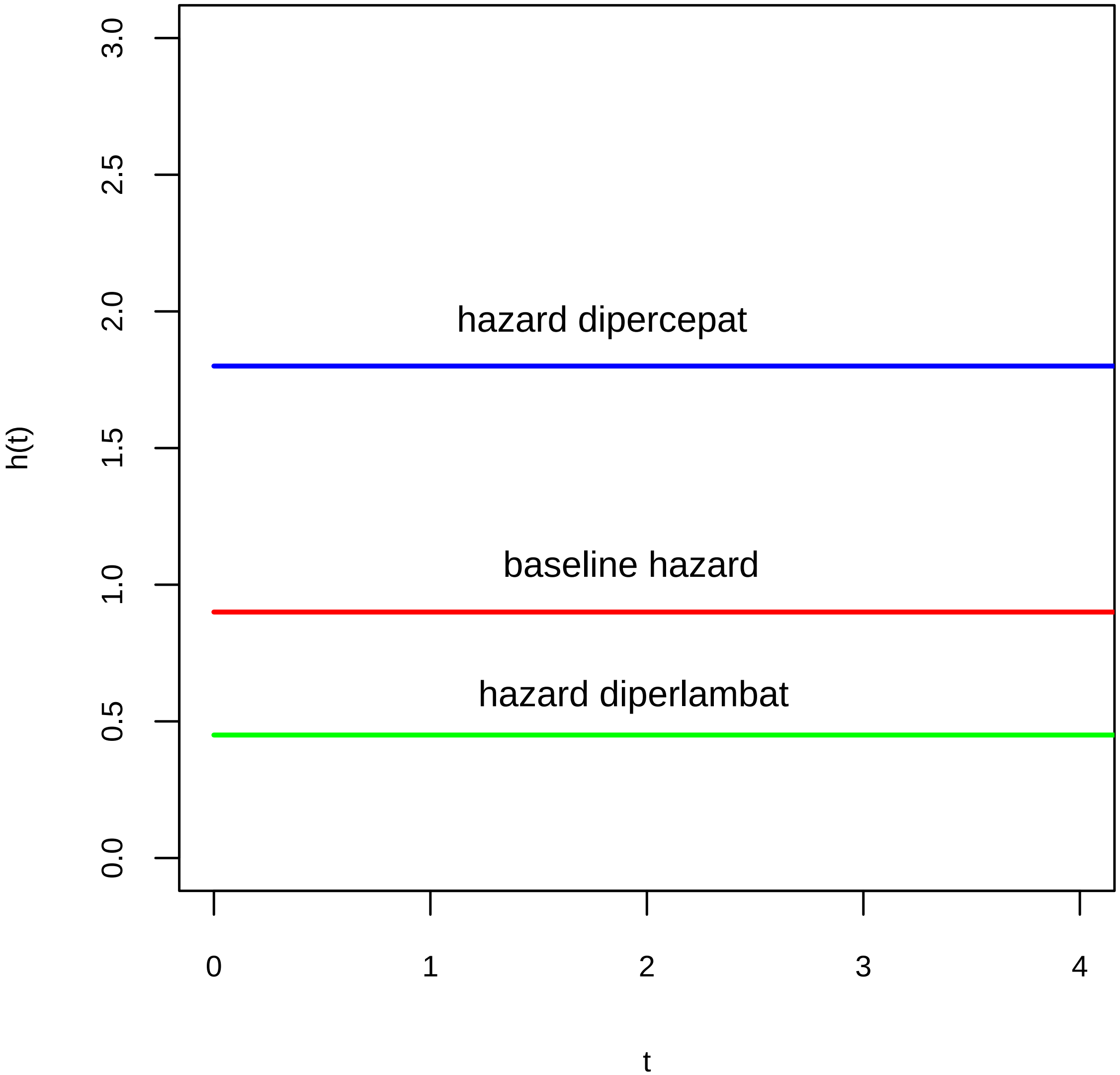
Baseline survival:
$$S_0(t) = \exp(-\lambda t)$$

Survival diperlambat:
$$S(t \mid 0{,}5) = \exp(-0{,}5\lambda t)$$

Survival dipercepat:
$$S(t \mid 2) = \exp(-2\lambda t)$$

Model AFT

Fungsi hazard Model AFT, Eksponensial ($\lambda = 0{,}9$):



$$h(t \mid x) = h_0(t)x$$
$$= x\lambda$$

Baseline hazard:

$h_0(t) = \lambda$

hazard diperlambat:

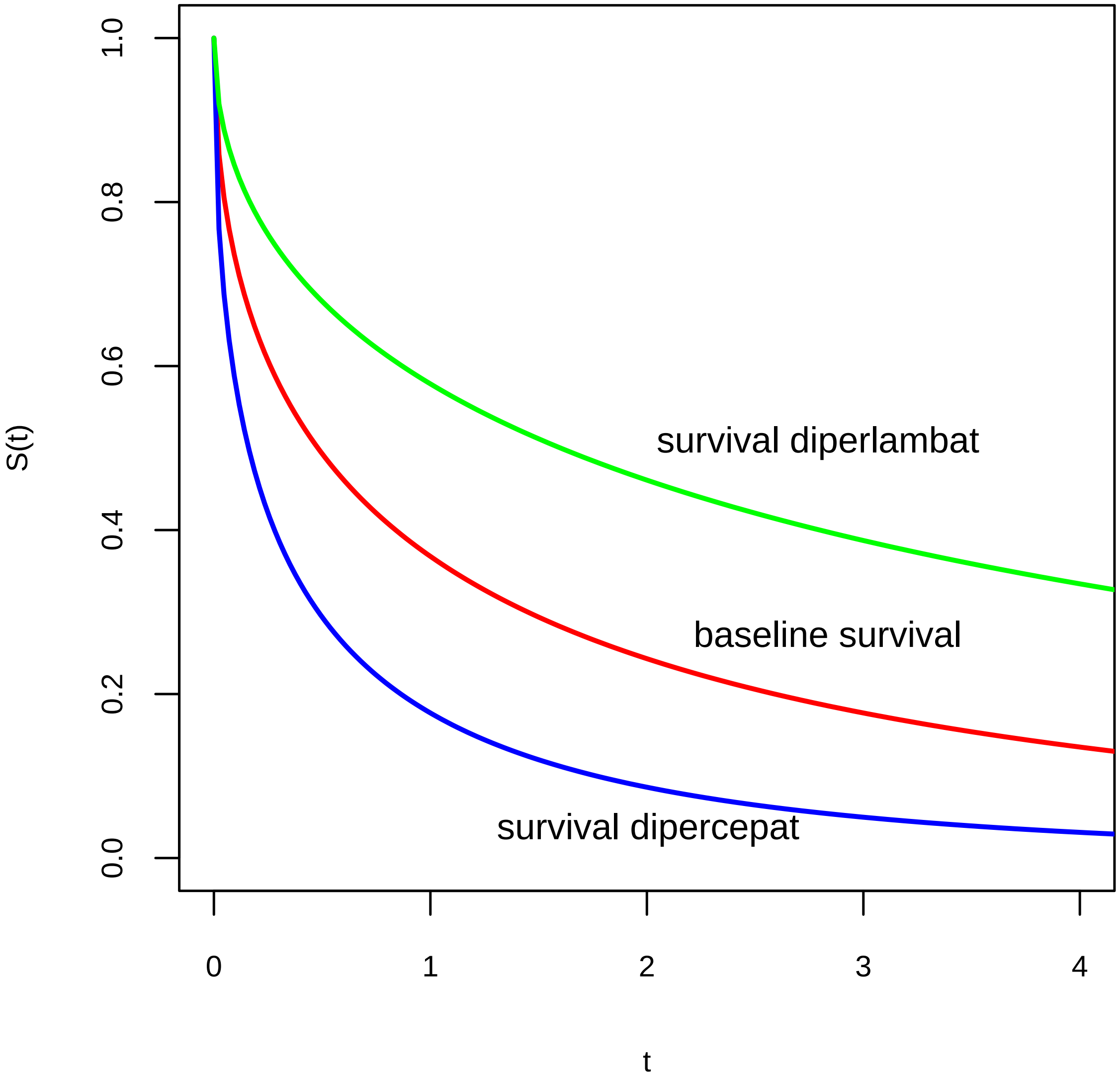
$h(t \mid 0{,}5) = 0{,}5\lambda$

hazard dipercepat:

$h(t \mid 2) = 2\lambda$

Model AFT

Fungsi Survival Model AFT, Weibull($\lambda = 1, \alpha = 0{,}5$):



$$S(t \mid x) = S_0(xt)$$
$$= \exp(-(x\lambda t)^\alpha)$$

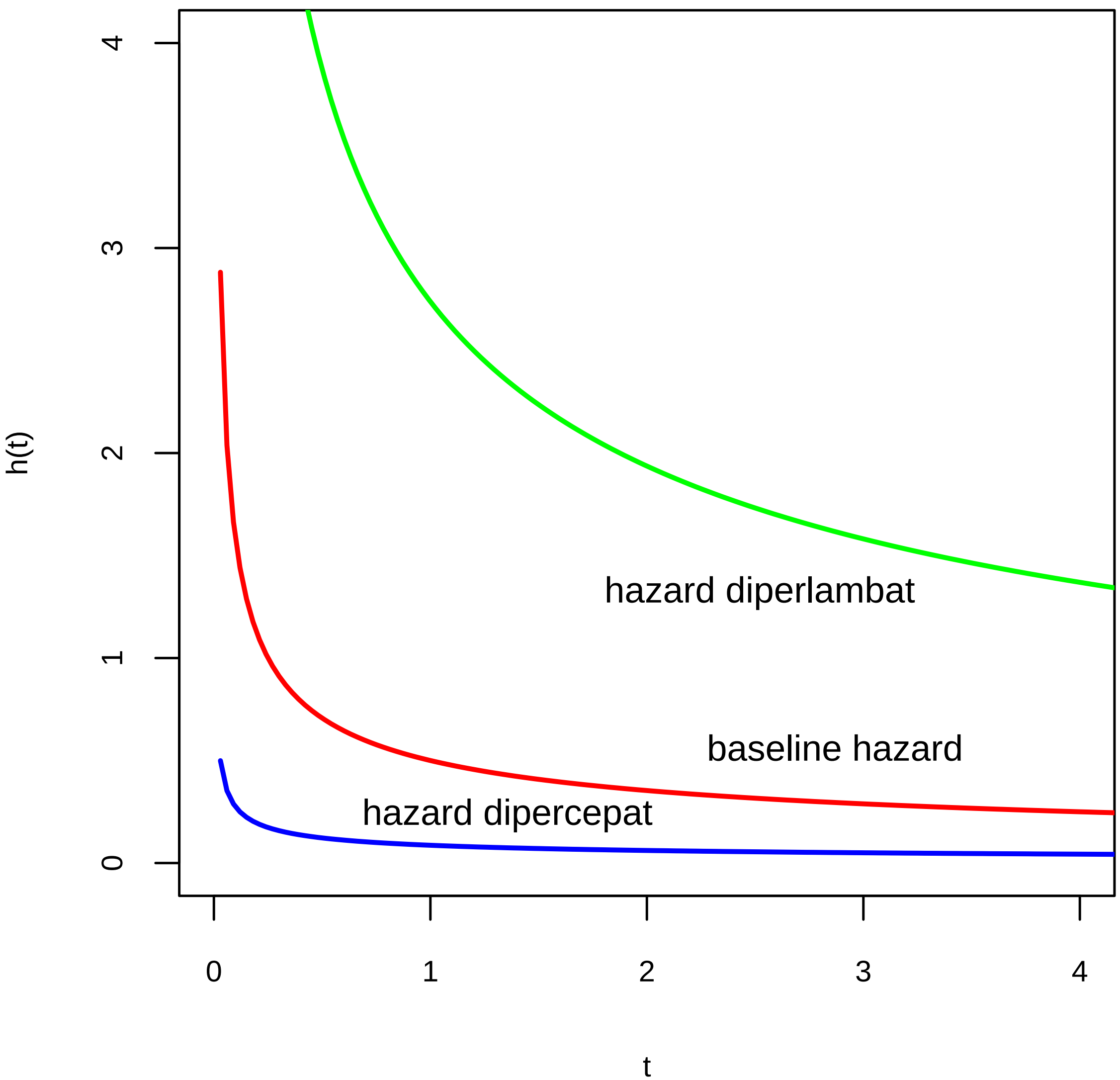
Baseline survival:
$$S_0(t) = \exp(-(\lambda t)^\alpha)$$

Survival diperlambat:
$$S(t \mid 0{,}3) = \exp(-(0{,}3\lambda t)^\alpha)$$

Survival dipercepat:
$$S(t \mid 3) = \exp(-(3\lambda t)^\alpha)$$

Model AFT

Fungsi hazard Model AFT, Weibull($\lambda = 1, \alpha = 0{,}5$):



$$h(t \mid x) = h_0(xt)x$$
$$= \alpha\lambda x(x\lambda t)^{\alpha-1}$$

Baseline hazard:

$$h_0(t) = \alpha\lambda(\lambda t)^{\alpha-1}$$

hazard diperlambat:

$$h(t \mid 0{,}3) = \alpha\lambda 0{,}3(0{,}3\lambda t)^{\alpha-1}$$

hazard dipercepat:

$$h(t \mid 3) = \alpha\lambda 3(3\lambda t)^{\alpha-1}$$

Estimasi Parameter

Data: $(t_i, \delta_i, \boldsymbol{x}_i)$, $i = 1, 2, \ldots, n$ yang independen satu sama lain

dengan

t_i adalah durasi atau waktu antar kejadian

$$\delta_i = \begin{cases} 0 & \text{jika } i \text{ tersensor} \\ 1 & \text{jika } i \text{ mendapatkan kejadian (\textit{event})} \end{cases}$$

$\boldsymbol{x}_i = \begin{pmatrix} x_{1i} & \ldots & x_{pi} \end{pmatrix}$ adalah vektor variabel penjelas untuk subyek (individu) i

Estimasi Parameter

Fungsi likelihood untuk data tersensor kanan

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^{n} f(t_i, \boldsymbol{\theta} \mid \boldsymbol{x}_i)^{\delta_i} S(t_i, \boldsymbol{\theta} \mid \boldsymbol{x}_i)^{1-\delta_i}$$

dengan $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ adalah p parameter yang akan diestimasi; $f(t_i, \boldsymbol{\theta} \mid \boldsymbol{x}_i)$ adalah fungsi densitas untuk i yang mendapatkan kejadian dan mempunyai variabel penjelas $\boldsymbol{x}_i$; $S(t_i, \boldsymbol{\theta}) \mid \boldsymbol{x}_i$ adalah fungsi survival untuk i yang tidak mendapatkan kejadian (tersensor kanan) dan mempunyai variabel penjelas $\boldsymbol{x}_i$.

Estimasi Parameter

Contoh Data:
Data 90 laki-laki yang terdiagnosis kanker larynx (library KMsurv dalam **R**).

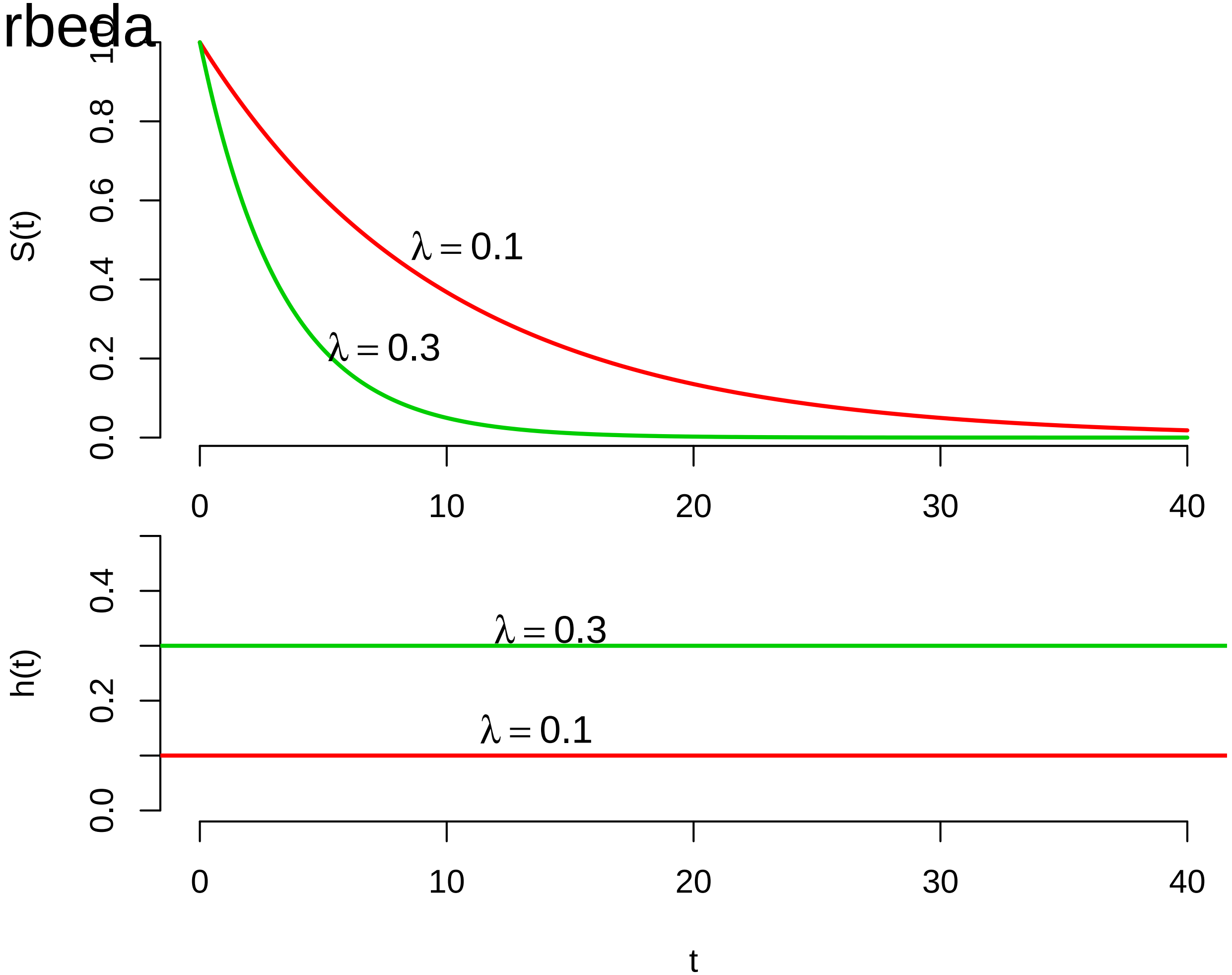
stage	:	Stage of disease (1=stage 1, 2=stage 2, 3=stage 3, 4=stage 4)
time	:	Time to death or on-study time, months
age	:	Age at diagnosis of larynx cancer
diagyr	:	Year of diagnosis of larynx cancer
delta	:	Death indicator (0=alive, 1=dead)

Estimasi Parameter

```
> m<-survreg(Surv(time,delta)~factor(stage)+age,data=larynx)
> summary(m)

Call:
survreg(formula = Surv(time, delta) ~ factor(stage) + age, data = larynx)
                  Value Std. Error      z        p
(Intercept)      3.5288     0.9041  3.903 9.50e-05
factor(stage)2 -0.1477     0.4076 -0.362 7.17e-01
factor(stage)3 -0.5866     0.3199 -1.833 6.68e-02
factor(stage)4 -1.5441     0.3633 -4.251 2.13e-05
age            -0.0175     0.0128 -1.367 1.72e-01
Log(scale)     -0.1223     0.1225 -0.999 3.18e-01

Scale= 0.885

Weibull distribution
Loglik(model)= -141.4   Loglik(intercept only)= -151.1
        Chisq= 19.37 on 4 degrees of freedom, p= 0.00066
Number of Newton-Raphson Iterations: 5
n= 90
```

Hazard Proporsional

Kurva survival untuk model eksponensial dengan dua nilai λ yang berbeda



Hazard Proporsional

Misalkan ada dua orang yang masing-masing mempunyai hazard $\lambda_1 = 0,1$ dan $\lambda_2 = 0,3$

hazard ratio:
$\frac{\lambda_2}{\lambda_1} = \frac{0{,}3}{0{,}1} = 3$

Hazard Proporsional

Misalkan ada dua orang yang masing-masing mempunyai hazard $\lambda_1 = 0,1$ dan $\lambda_2 = 0,3$

hazard ratio:
$\frac{\lambda_2}{\lambda_1} = \frac{0{,}3}{0{,}1} = 3$

konstant, independen terhadap waktu

Cox's Regression Model

Cox's regression model atau Cox's proportional hazards (Cox;1972,1975):

$$h(t \mid \boldsymbol{x}) \quad = \quad h_0(t)\psi(\boldsymbol{x}, \boldsymbol{\beta})$$

dengan $\boldsymbol{x} = (x_1, \ldots, x_p)$ adalah vektor kovariat (variabel independen) dan $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_p)$ adalah parameter dari model regresi

Cox's Regression Model

Cox's regression model atau Cox's proportional hazards (Cox;1972,1975):

$$
\begin{aligned}
h(t \mid \boldsymbol{x}) &= h_0(t)\psi(\boldsymbol{x}, \boldsymbol{\beta}) \\
\begin{array}{r}\text{fungsi hazard} \\ \text{bergantung pada } \boldsymbol{x}\end{array} &= \begin{array}{l}\text{baseline hazard} \\ \text{tdk bergantung pd } \boldsymbol{x}\end{array} \times \text{fungsi kovariat}
\end{aligned}
$$

Cox's Regression Model

Cox's regression model atau Cox's proportional hazards (Cox;1972,1975):

$$
\begin{array}{rcl}
h(t \mid \boldsymbol{x}) & = & h_0(t)\psi(\boldsymbol{x},\boldsymbol{\beta}) \\
\begin{array}{r}\text{fungsi hazard}\\ \text{bergantung pada } \boldsymbol{x}\end{array} & = & \begin{array}{l}\text{baseline hazard}\\ \text{tdk bergantung pd } \boldsymbol{x}\end{array} \times \text{fungsi kovariat}
\end{array}
$$

Bentuk fungsional dari $\psi(\boldsymbol{x},\boldsymbol{\beta})$

- $\psi(\boldsymbol{x},\boldsymbol{\beta}) = \exp(\boldsymbol{x}\boldsymbol{\beta})$
- $\psi(\boldsymbol{x},\boldsymbol{\beta}) = \exp(1 + \boldsymbol{x}\boldsymbol{\beta})$
- $\psi(\boldsymbol{x},\boldsymbol{\beta}) = \log(1 + \exp(\boldsymbol{x}\boldsymbol{\beta}))$

Cox's Regression Model

Cox's regression model atau Cox's proportional hazards (Cox;1972,1975):

$$
\begin{array}{rcl}
h(t \mid \boldsymbol{x}) & = & h_0(t)\psi(\boldsymbol{x}, \boldsymbol{\beta}) \\
\begin{array}{r}\text{fungsi hazard} \\ \text{bergantung pada } \boldsymbol{x}\end{array} & = & \begin{array}{l}\text{baseline hazard} \\ \text{tdk bergantung pd } \boldsymbol{x}\end{array} \times \text{fungsi kovariat}
\end{array}
$$

Bentuk fungsional dari $\psi(\boldsymbol{x}, \boldsymbol{\beta})$

- $\psi(\boldsymbol{x}, \boldsymbol{\beta}) = \exp(\boldsymbol{x}\boldsymbol{\beta})$
- $\psi(\boldsymbol{x}, \boldsymbol{\beta}) = \exp(1 + \boldsymbol{x}\boldsymbol{\beta})$
- $\psi(\boldsymbol{x}, \boldsymbol{\beta}) = \log(1 + \exp(\boldsymbol{x}\boldsymbol{\beta}))$

Cox's Regression Model

Model:

$$h(t \mid \boldsymbol{x}) = h_0(t)\exp(\boldsymbol{x}\boldsymbol{\beta})$$

Misalkan:

$$x = \begin{cases} 0 & \text{placebo} \\ 1 & \text{obat baru} \end{cases}$$

Cox's Regression Model

Model:

$$h(t \mid \boldsymbol{x}) \quad = \quad h_0(t) \exp(\boldsymbol{x\beta})$$

Hazard ratio:

$$\frac{h(t \mid x = 1)}{h(t \mid x = 0)} \quad = \quad \frac{h_0(t) \exp(1 \times \beta)}{h_0(t) \exp(0 \times \beta)}$$

Cox's Regression Model

Model:

$$h(t \mid \boldsymbol{x}) \quad = \quad h_0(t) \exp(\boldsymbol{x}\boldsymbol{\beta})$$

Hazard ratio:

$$\begin{aligned} \frac{h(t \mid x = 1)}{h(t \mid x = 0)} &= \frac{h_0(t) \exp(1 \times \beta)}{h_0(t) \exp(0 \times \beta)} \\ &= \exp(\beta) \end{aligned}$$

Cox's Regression Model

Model:

$$h(t \mid \boldsymbol{x}) = h_0(t)\exp(\boldsymbol{x}\boldsymbol{\beta})$$

Hazard ratio:

$$\begin{aligned}\frac{h(t \mid x=1)}{h(t \mid x=0)} &= \frac{h_0(t)\exp(1 \times \beta)}{h_0(t)\exp(0 \times \beta)} \\ &= \exp(\beta)\end{aligned}$$

jika $\beta = 0 \Rightarrow$ obat baru dan placebo sama efeknya

Cox's Regression Model

Model:

$$h(t \mid \boldsymbol{x}) = h_0(t) \exp(\boldsymbol{x\beta})$$

Hazard ratio:

$$\frac{h(t \mid x = 1)}{h(t \mid x = 0)} = \frac{h_0(t) \exp(1 \times \beta)}{h_0(t) \exp(0 \times \beta)} = \exp(\beta)$$

jika $\beta < 0$ $\Rightarrow$ obat baru memberikan efek yang lebih baik daripada placebo (resiko kematian lebih rendah)

Cox's Regression Model

Model:

$$h(t \mid \boldsymbol{x}) = h_0(t)\exp(\boldsymbol{x}\boldsymbol{\beta})$$

Hazard ratio:

$$\begin{aligned}\frac{h(t \mid x=1)}{h(t \mid x=0)} &= \frac{h_0(t)\exp(1\times\beta)}{h_0(t)\exp(0\times\beta)}\\ &= \exp(\beta)\end{aligned}$$

jika $\beta > 0 \Rightarrow$ obat baru memberikan efek yang lebih buruk daripada placebo (resiko kematian lebih tinggi)

Cox's Regression Model

Model:

$$h(t \mid \boldsymbol{x}) \;=\; h_0(t)\exp(\boldsymbol{x}\boldsymbol{\beta})$$

Secara umum nilai estimasi $\boldsymbol{\beta}$ dapat digunakan untuk mengidentifikasi faktor resiko (*risk factors*, *prognostic factors*) yang berkaitan dengan variabel dependen *time-to-event* T.

Cox's Regression Model

Model:

$$h(t \mid \boldsymbol{x}) = h_0(t)\exp(\boldsymbol{x}\boldsymbol{\beta})$$

Dapat dituliskan dalam $H(t \mid \boldsymbol{x})$ atau $S(t \mid \boldsymbol{x})$

$$H(t \mid \boldsymbol{x}) = H_0(t)\exp(\boldsymbol{x}\boldsymbol{\beta})$$

$$S(t \mid \boldsymbol{x}) = S_0(t)^{\exp(\boldsymbol{x}\boldsymbol{\beta})}$$

dengan H_0 adalah *baseline* hazard kumulatif dan S_0 adalah *baseline* survival

Estimasi untuk β

- Parametrik: $h_0(t)$ ditentukan dari distribusi probabilitas tertentu
- Semi-Parametrik: *Partial-likelihood*
- Non-Parametrik: *Smoothing, GAM*

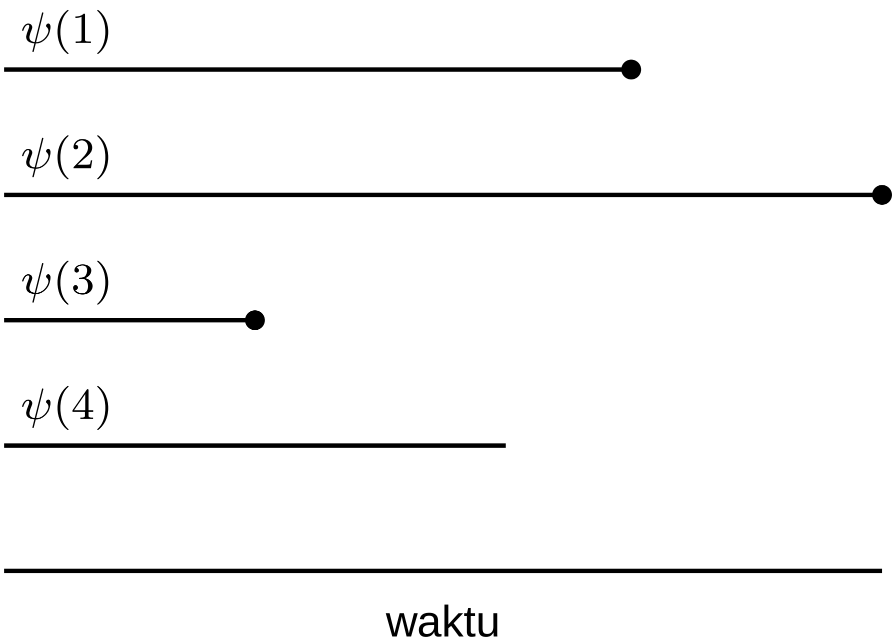
Partial likelihood

Cox (1972,1975):

$$L(\boldsymbol{\beta}) = \prod_{k \in \mathcal{D}} \frac{\exp(\boldsymbol{x}_k \boldsymbol{\beta})}{\sum_{j \in R_k} \exp(\boldsymbol{x}_j \boldsymbol{\beta})}$$

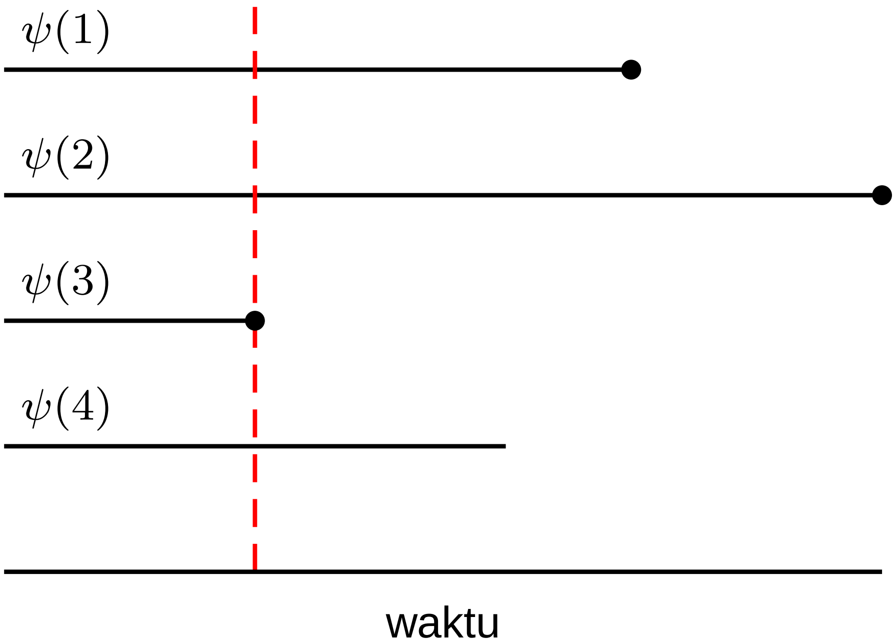
- $\boldsymbol{x}$ adalah vektor kovariat (variabel penjelas)
- $\boldsymbol{\beta}$ adalah parameter regresi yang akan diestimasi
- $\mathcal{D}$ adalah himpunan indeks j dari semua waktu kejadian (semua t_j yang mendapatkan kejadian)
- R_k adalah himpunan resiko (*risk set*) , semua individu (subyek) yang belum mendapatkan kejadian pada saat tertentu

Partial likelihood



$$\mathcal{D} = \{1, 2, 3\}$$

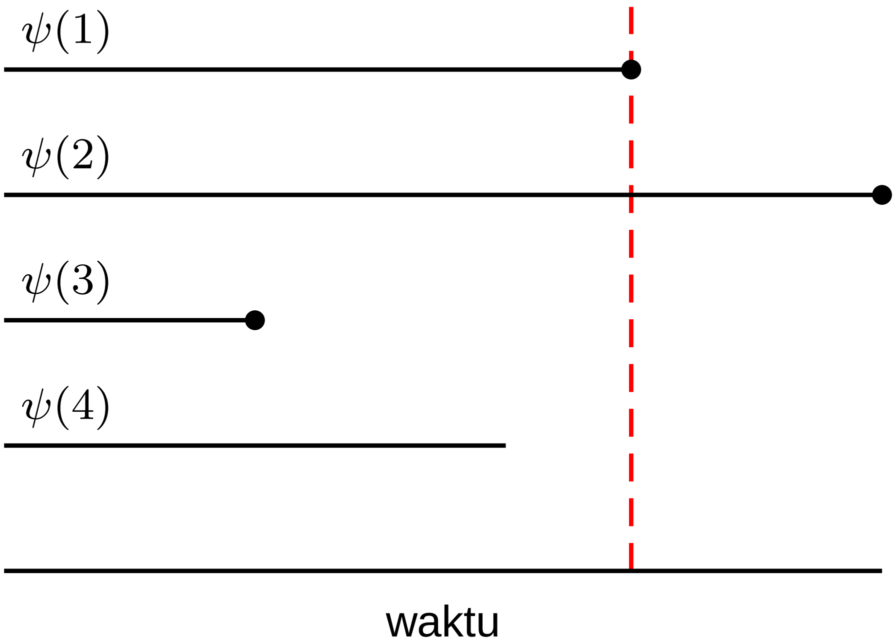
Partial likelihood



$$\frac{\psi(3)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)}$$

$$\mathcal{D} = \{1, 2, 3\}$$
$$R_3 = \{1, 2, 3, 4\}$$

Partial likelihood

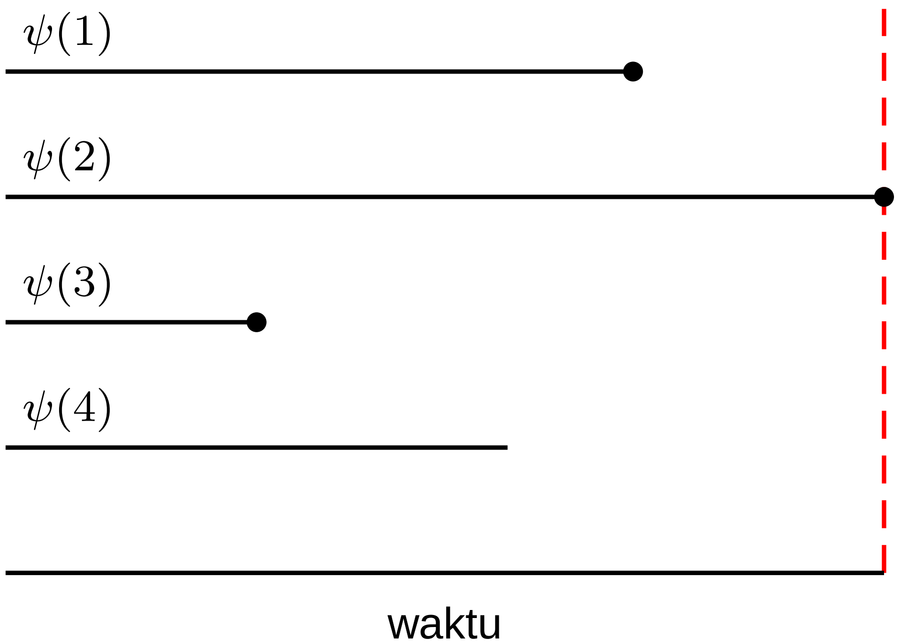


$$\frac{\psi(1)}{\psi(1)+\psi(2)}$$

$$\frac{\psi(3)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)}$$

$$\mathcal{D} = \{1, 2, 3\}$$
$$R_3 = \{1, 2, 3, 4\}$$
$$R_1 = \{1, 2, \}$$

Partial likelihood



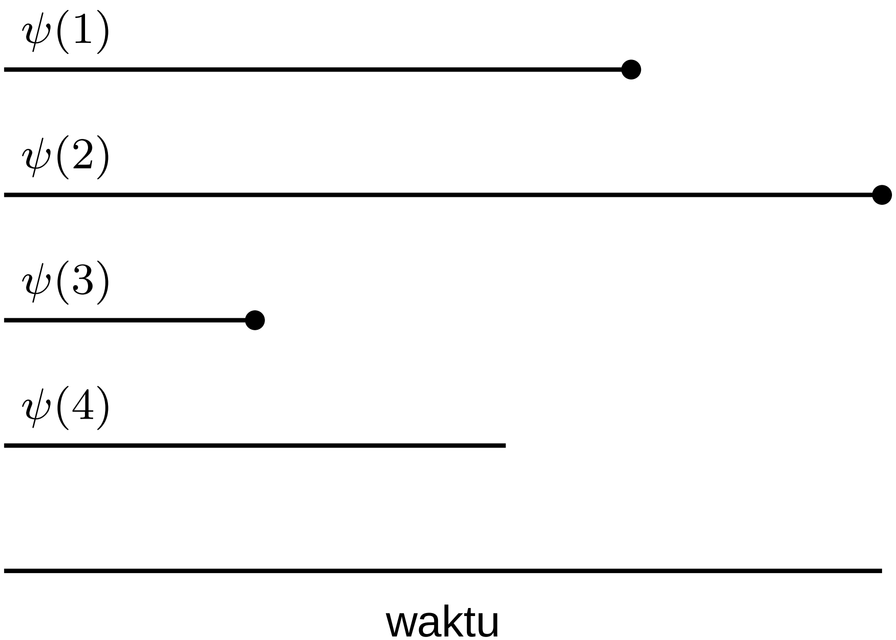
$$\frac{\psi(1)}{\psi(1)+\psi(2)}$$

$$\frac{\psi(2)}{\psi(2)}$$

$$\frac{\psi(3)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)}$$

$$\mathcal{D} = \{1, 2, 3\}$$
$$R_3 = \{1, 2, 3, 4\}$$
$$R_1 = \{1, 2, \}$$
$$R_2 = \{2\}$$

Partial likelihood



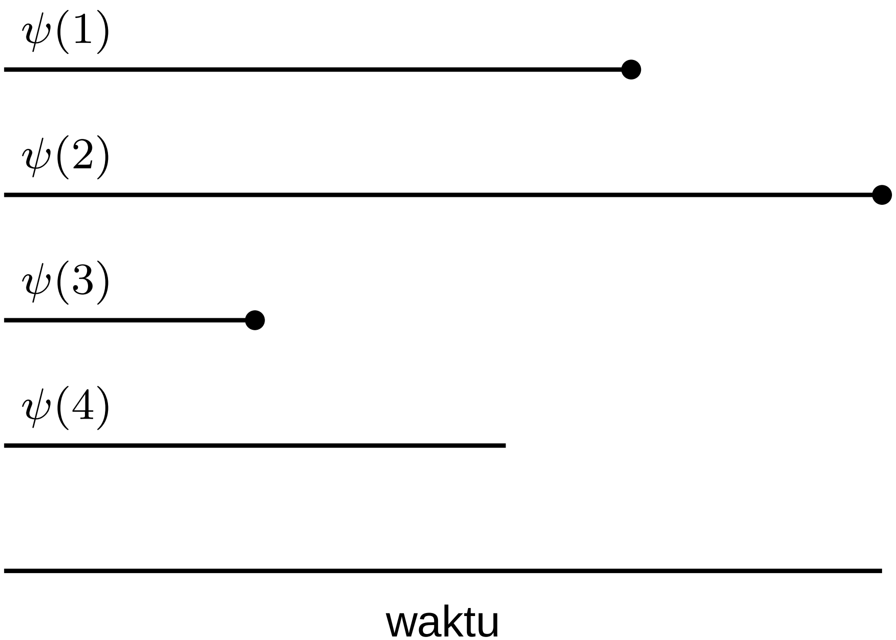
$$\frac{\psi(1)}{\psi(1)+\psi(2)}$$

$$\frac{\psi(2)}{\psi(2)}$$

$$\frac{\psi(3)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)}$$

$$\mathcal{D} = \{1, 2, 3\}$$
$$R_3 = \{1, 2, 3, 4\}$$
$$R_1 = \{1, 2, \}$$
$$R_2 = \{2\}$$

Partial likelihood



$$\frac{\psi(1)}{\psi(1)+\psi(2)}$$

$$\frac{\psi(2)}{\psi(2)}$$

$$\frac{\psi(3)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)}$$

$$\begin{aligned} \mathcal{D} &= \{1,2,3\} \\ R_3 &= \{1,2,3,4\} \\ R_1 &= \{1,2,\} \\ R_2 &= \{2\} \end{aligned}$$

$$\begin{aligned} L(\boldsymbol{\beta}) &= \left(\frac{\psi(1)}{\psi(1)+\psi(2)}\right)\left(\frac{\psi(2)}{\psi(2)}\right) \\ &\quad \left(\frac{\psi(3)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)}\right) \end{aligned}$$

Contoh Partial likelihood

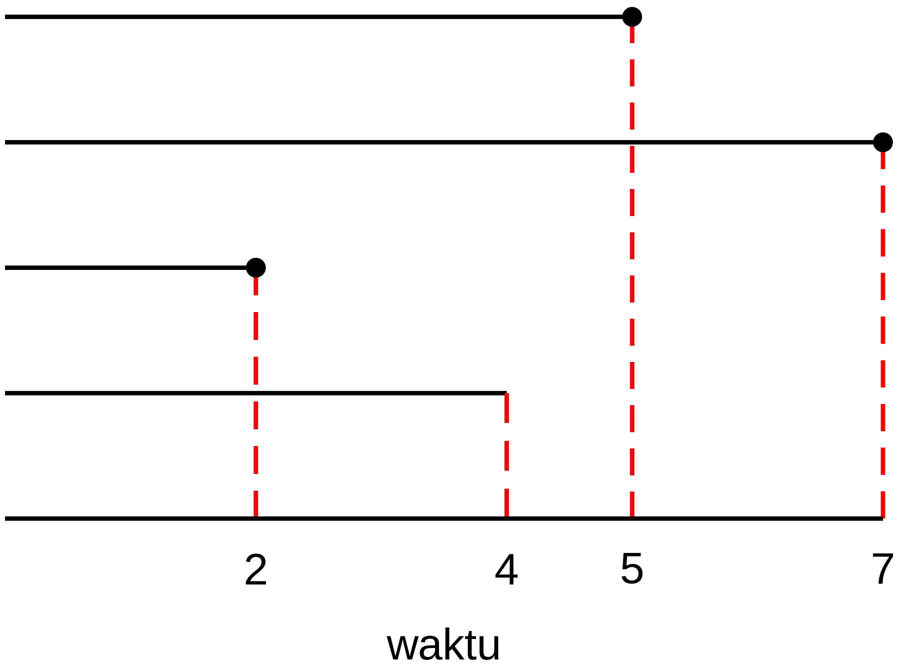
Diketahui data sebagai berikut:

t	δ	x
5	1	2,58
7	1	1,36
2	1	-0,54
4	0	3,30

Contoh Partial likelihood

Diketahui data sebagai berikut:

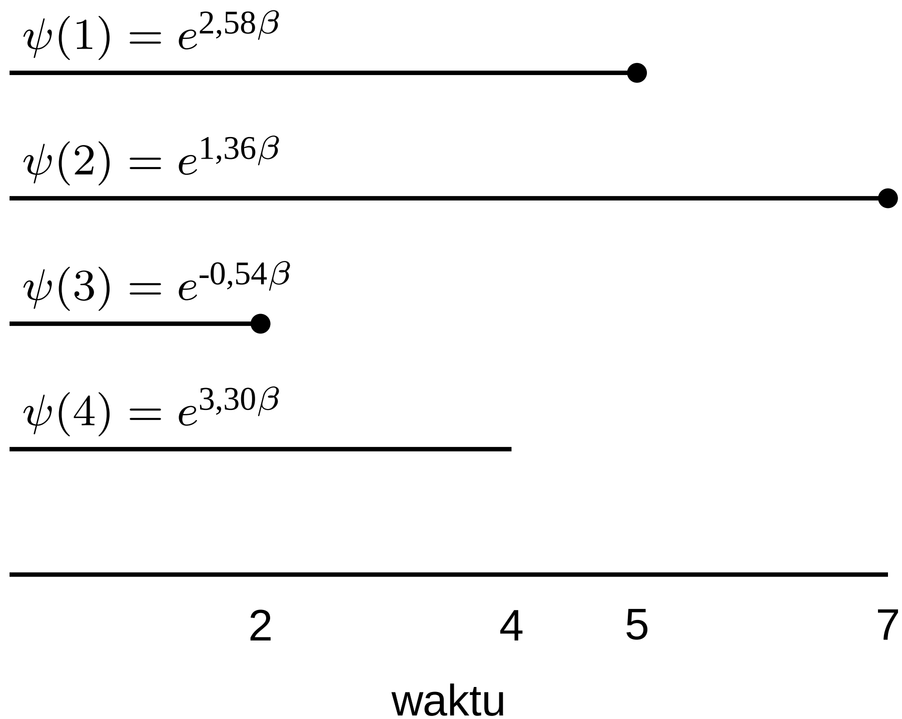
t	δ	x
5	1	2,58
7	1	1,36
2	1	-0,54
4	0	3,30



Contoh Partial likelihood

Diketahui data sebagai berikut:

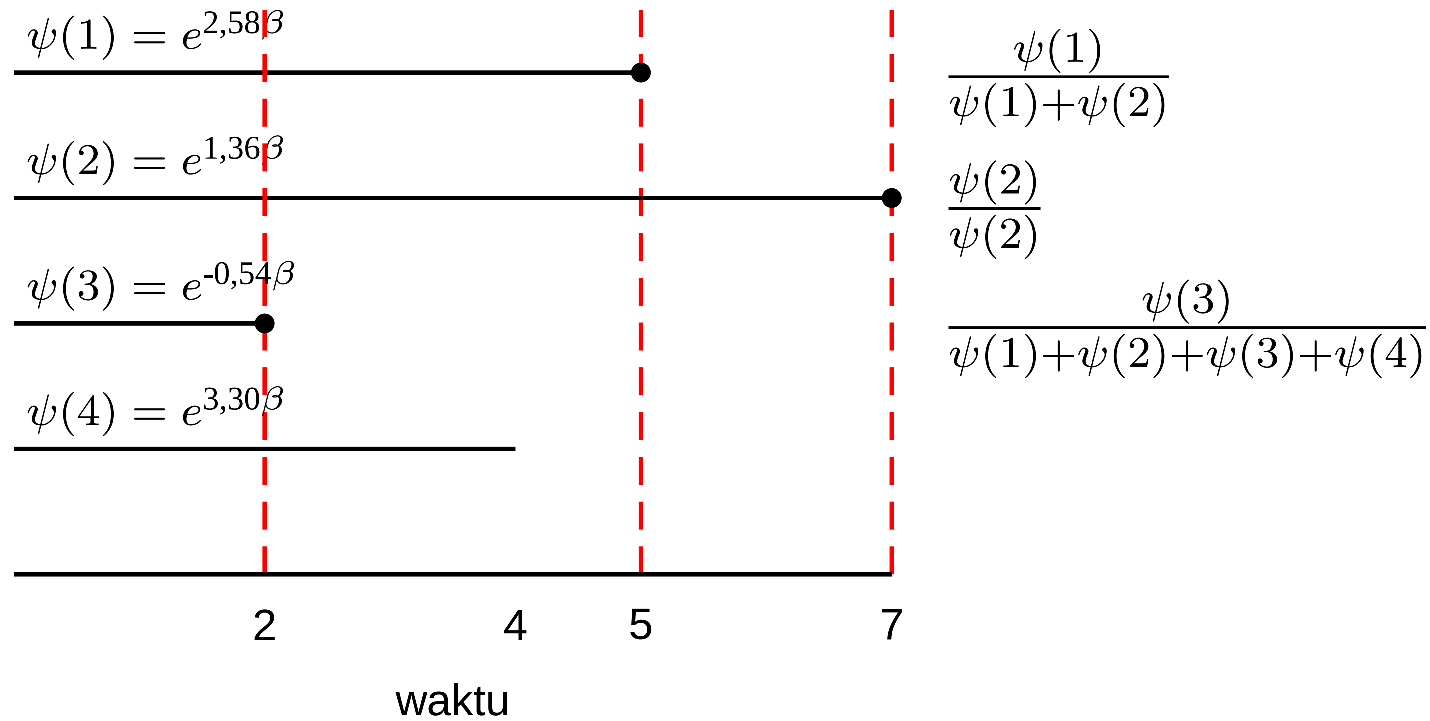
t	δ	x
5	1	2,58
7	1	1,36
2	1	-0,54
4	0	3,30



Contoh Partial likelihood

Diketahui data sebagai berikut:

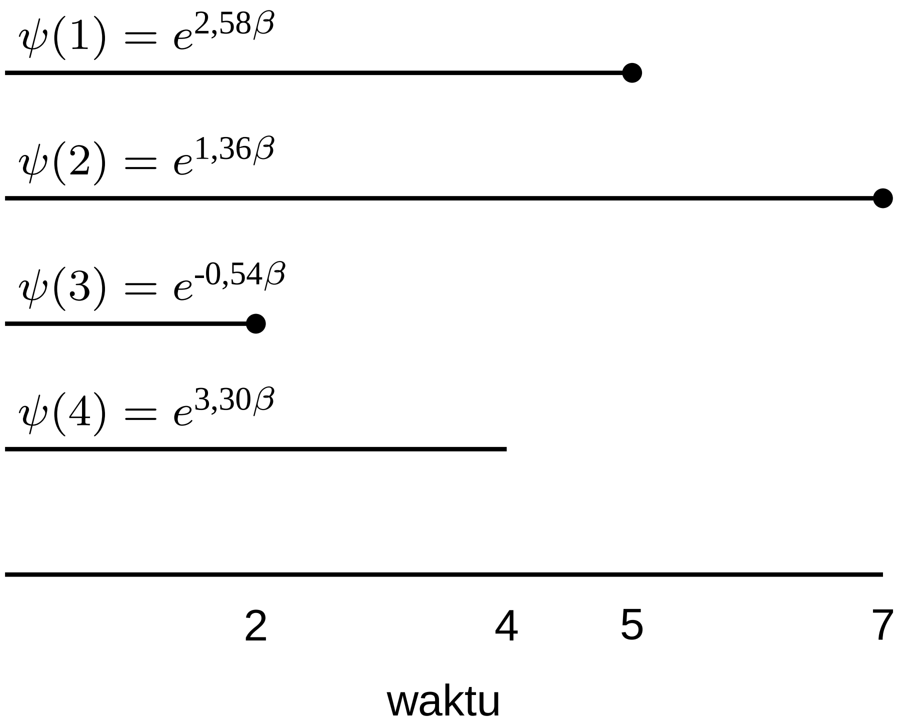
t	δ	x
5	1	2,58
7	1	1,36
2	1	-0,54
4	0	3,30



Contoh Partial likelihood

Diketahui data sebagai berikut:

t	δ	x
5	1	2,58
7	1	1,36
2	1	-0,54
4	0	3,30



$$\frac{e^{2,58\beta}}{e^{2,58\beta}+e^{1,36\beta}}$$

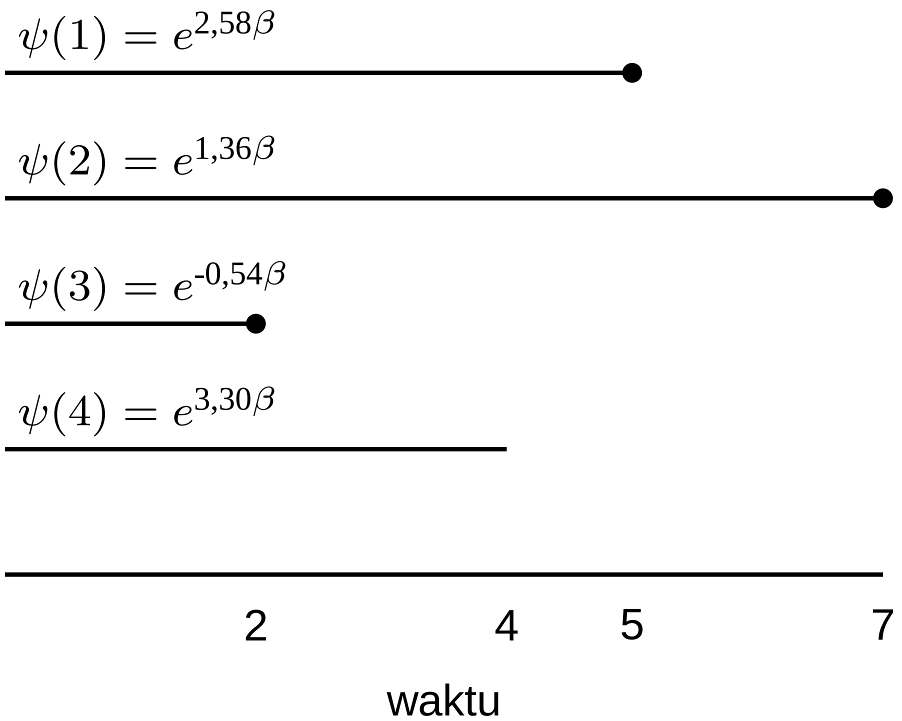
$$\frac{e^{1,36\beta}}{e^{1,36\beta}}$$

$$\frac{e^{-0,54\beta}}{e^{2,58\beta}+e^{1,36\beta}+e^{-0,54\beta}+e^{3,30\beta}}$$

Contoh Partial likelihood

Diketahui data sebagai berikut:

t	δ	x
5	1	2,58
7	1	1,36
2	1	-0,54
4	0	3,30



$$\frac{e^{2{,}58\beta}}{e^{2{,}58\beta}+e^{1{,}36\beta}}$$

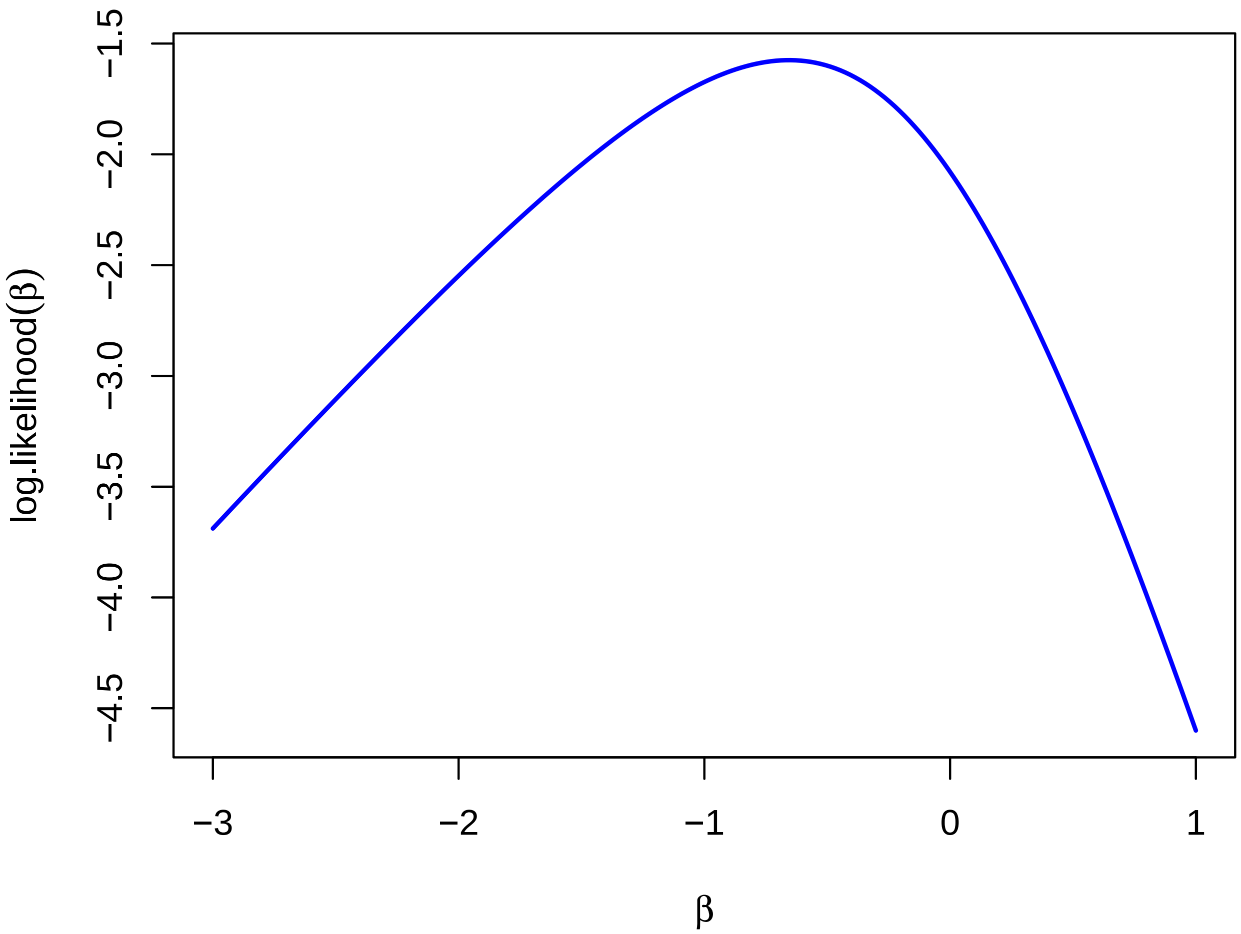
$$\frac{e^{1{,}36\beta}}{e^{1{,}36\beta}}$$

$$\frac{e^{-0{,}54\beta}}{e^{2{,}58\beta}+e^{1{,}36\beta}+e^{-0{,}54\beta}+e^{3{,}30\beta}}$$

Mencari penduga β yang memaksimalkan fungsi partial likelihood

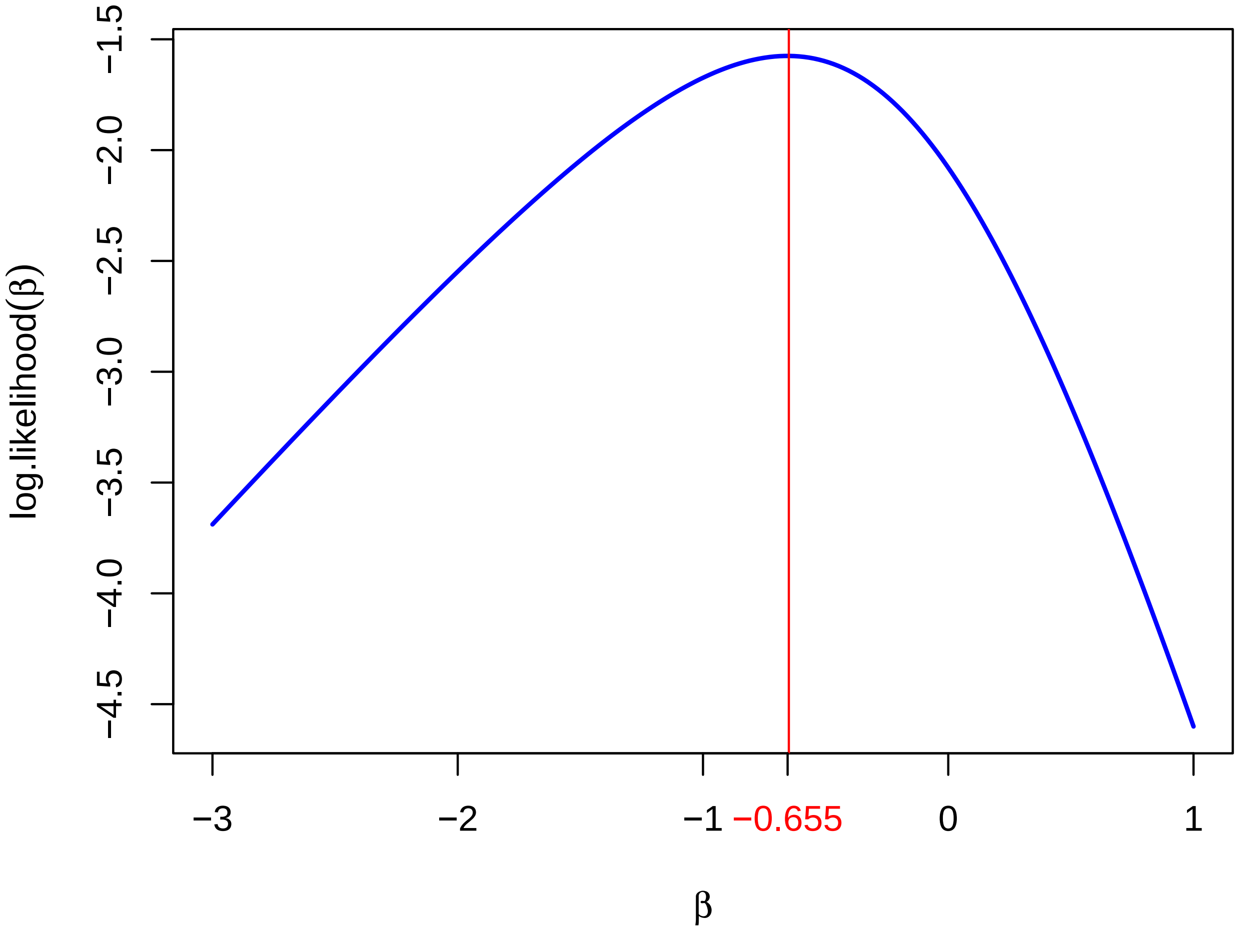
$$L(\beta) = \left(\frac{\psi(1)}{\psi(1)+\psi(2)}\right)\left(\frac{\psi(2)}{\psi(2)}\right)\left(\frac{\psi(3)}{\psi(1)+\psi(2)+\psi(3)+\psi(4)}\right)$$

Contoh Partial likelihood



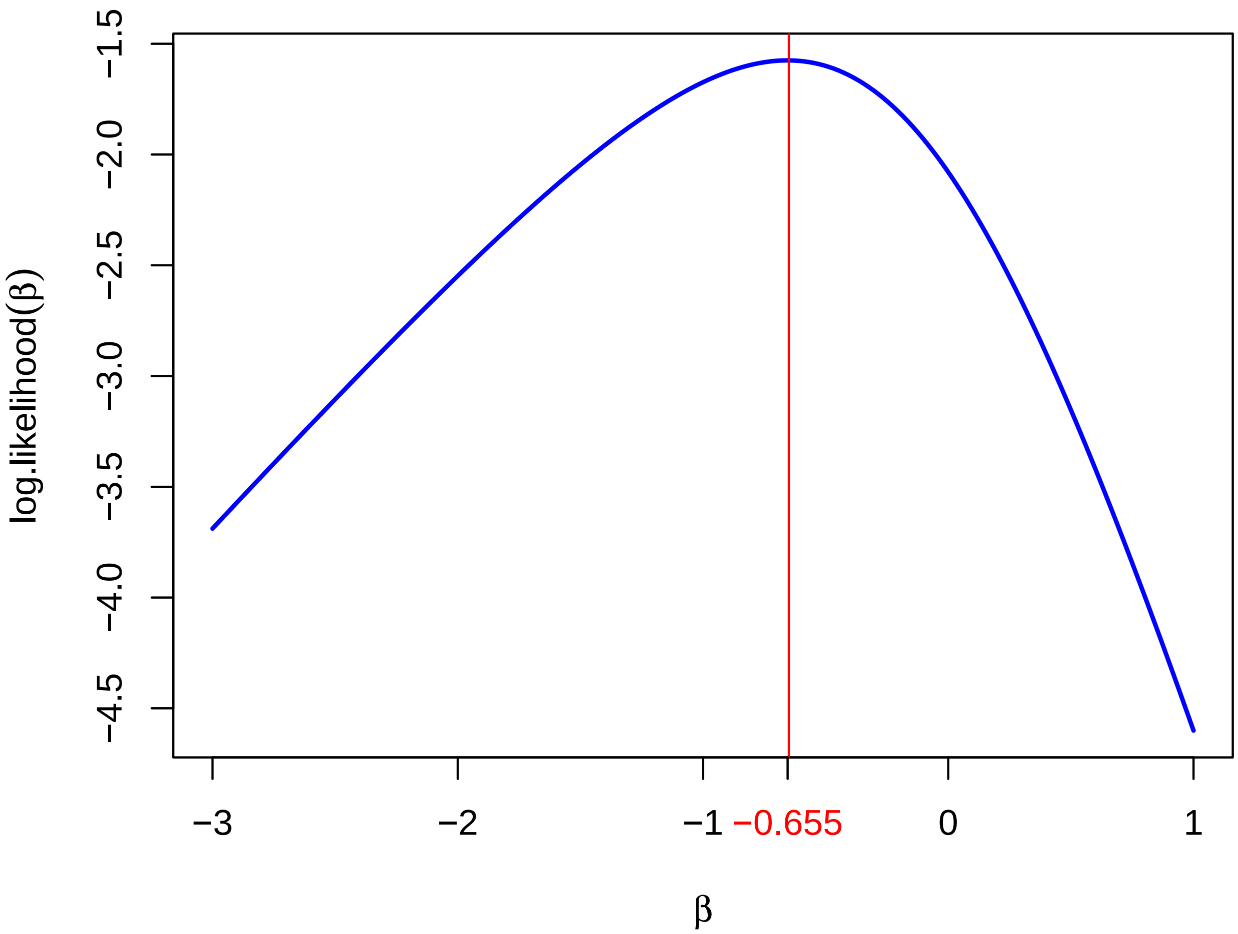
$$L(\beta) \quad = \quad \left(\frac{e^{2{,}58\beta}}{e^{2{,}58\beta} + e^{1{,}36\beta}}\right)\left(\frac{e^{-0{,}54\beta}}{e^{2{,}58\beta} + e^{1{,}36\beta} + e^{-0{,}54\beta} + e^{3{,}30\beta}}\right)$$

Contoh Partial likelihood



$$L(\beta) = \left(\frac{e^{2{,}58\beta}}{e^{2{,}58\beta} + e^{1{,}36\beta}}\right)\left(\frac{e^{-0{,}54\beta}}{e^{2{,}58\beta} + e^{1{,}36\beta} + e^{-0{,}54\beta} + e^{3{,}30\beta}}\right)$$

Contoh Partial likelihood



Estimasi β yang memaksimalkan $L(\beta)$ adalah $\hat{(}\beta) =$ -0,655 dengan nilai partial likelihood $\log(L($-0,655$)) =$ -1,575, atau $L($-0,655$) = 0{,}207$

Contoh Partial likelihood

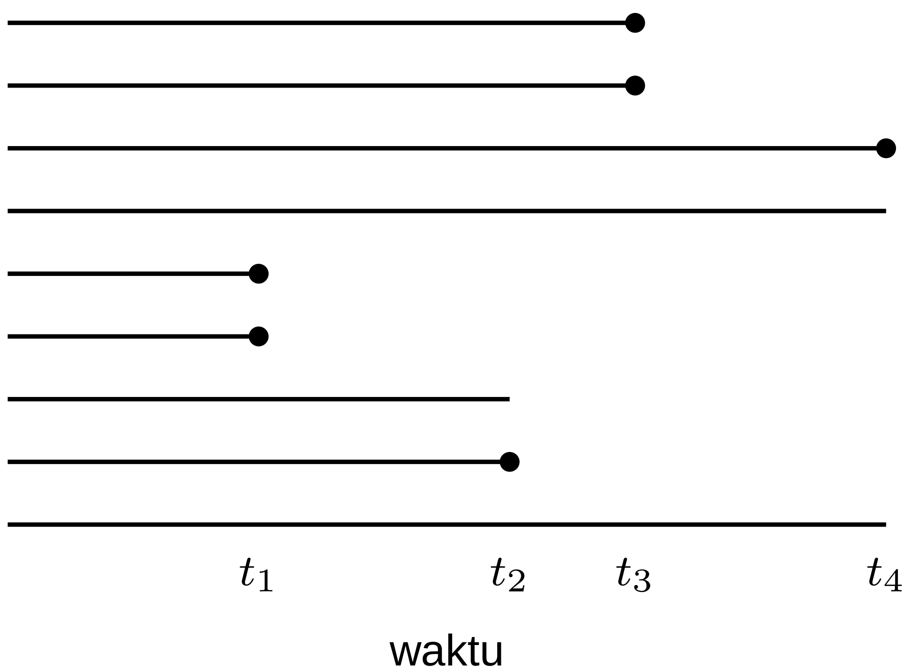
```
> X
  tt d     x
1  5 1  2.58
2  7 1  1.36
3  2 1 -0.54
4  4 0  3.30
> coxph(Surv(tt,d)~x,data=X)
Call:
coxph(formula = Surv(tt, d) ~ x, data = X)

    coef exp(coef) se(coef)      z    p
x -0.655     0.519    0.718 -0.913 0.36

Likelihood ratio test=1.01  on 1 df, p=0.315  n= 4
```

Contoh Partial likelihood

```
> m<-coxph(Surv(tt,d)~x,data=X)
> m$loglik
[1] -2.079442 -1.574940
```

Partial Likelihood dengan *ties*

Data: $t_1 < t_2 < \ldots < t_{n(\mathcal{D})}$ dengan $n(\mathcal{D})$ adalah banyaknya waktu t yang mendapatkan kejadian; d_k adalah banyaknya kejadian saat t_k (jika d_k>1 dinamakan *ties*); $\mathfrak{D}_k$ adalah himpunan individu yang mendapatkan kejadian saat t_k; $S_k = \sum_{j \in \mathfrak{D}} \boldsymbol{x}_j$ adalah jumlahan nilai variabel $\boldsymbol{x}$ pada saat t_k.



Partial Likelihood dengan *ties*

Digunakan 3 metode:

- Breslow

$$L(\boldsymbol{\beta}) = \prod_{k \in \mathcal{D}} \frac{\exp(S_k \boldsymbol{\beta})}{\left[\sum_{j \in R_k} \exp(\boldsymbol{x}_j \boldsymbol{\beta})\right]^{d_k}}$$

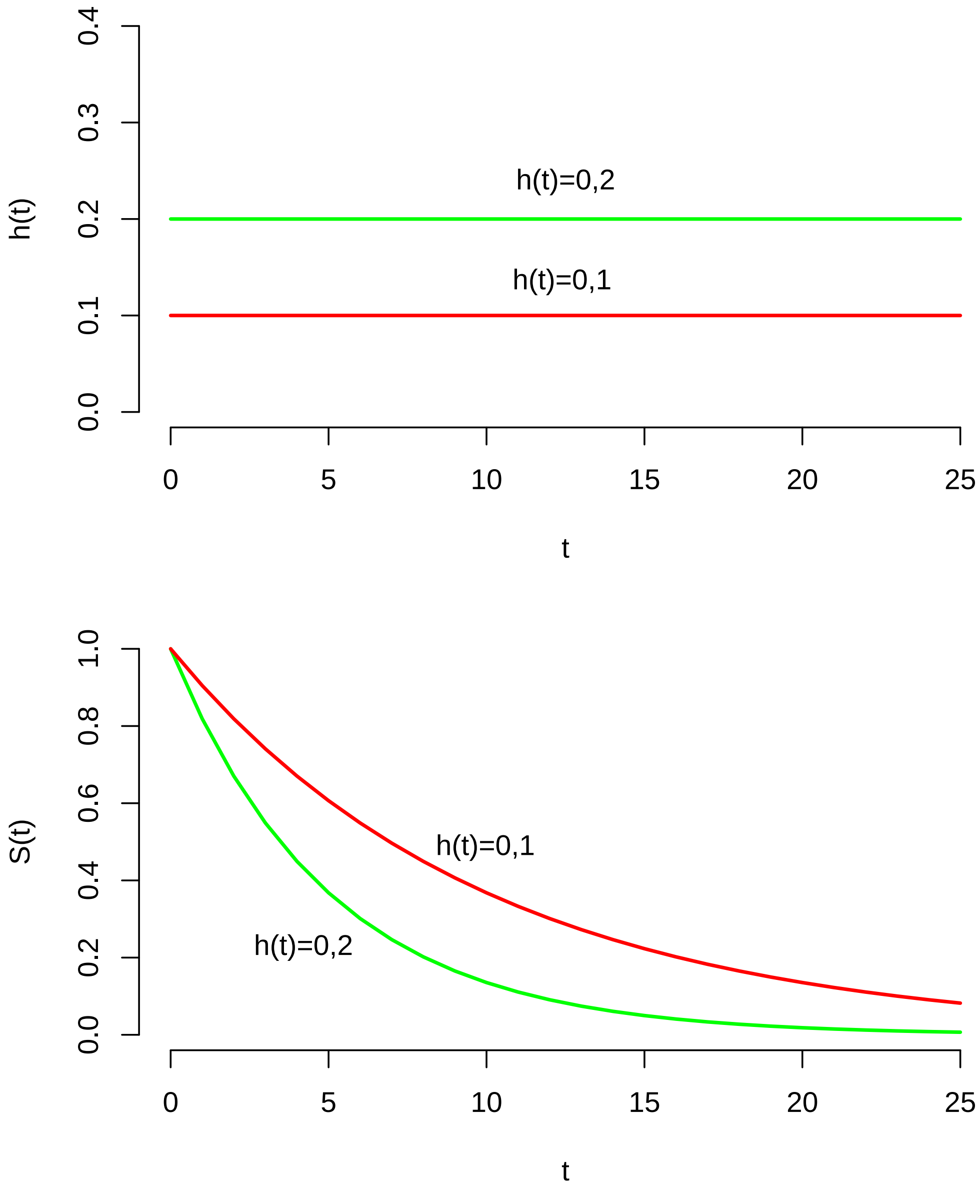
- Efron

$$L(\boldsymbol{\beta}) = \prod_{k \in \mathcal{D}} \frac{\exp(S_k \boldsymbol{\beta})}{\prod_{j=1}^{d_k} \left[\sum_{i \in R_k} \exp(\boldsymbol{x}_i \boldsymbol{\beta}) - \frac{j-1}{d_k} \sum_{i \in \mathfrak{D}_k} \exp(\boldsymbol{x}_i \boldsymbol{\beta})\right]}$$

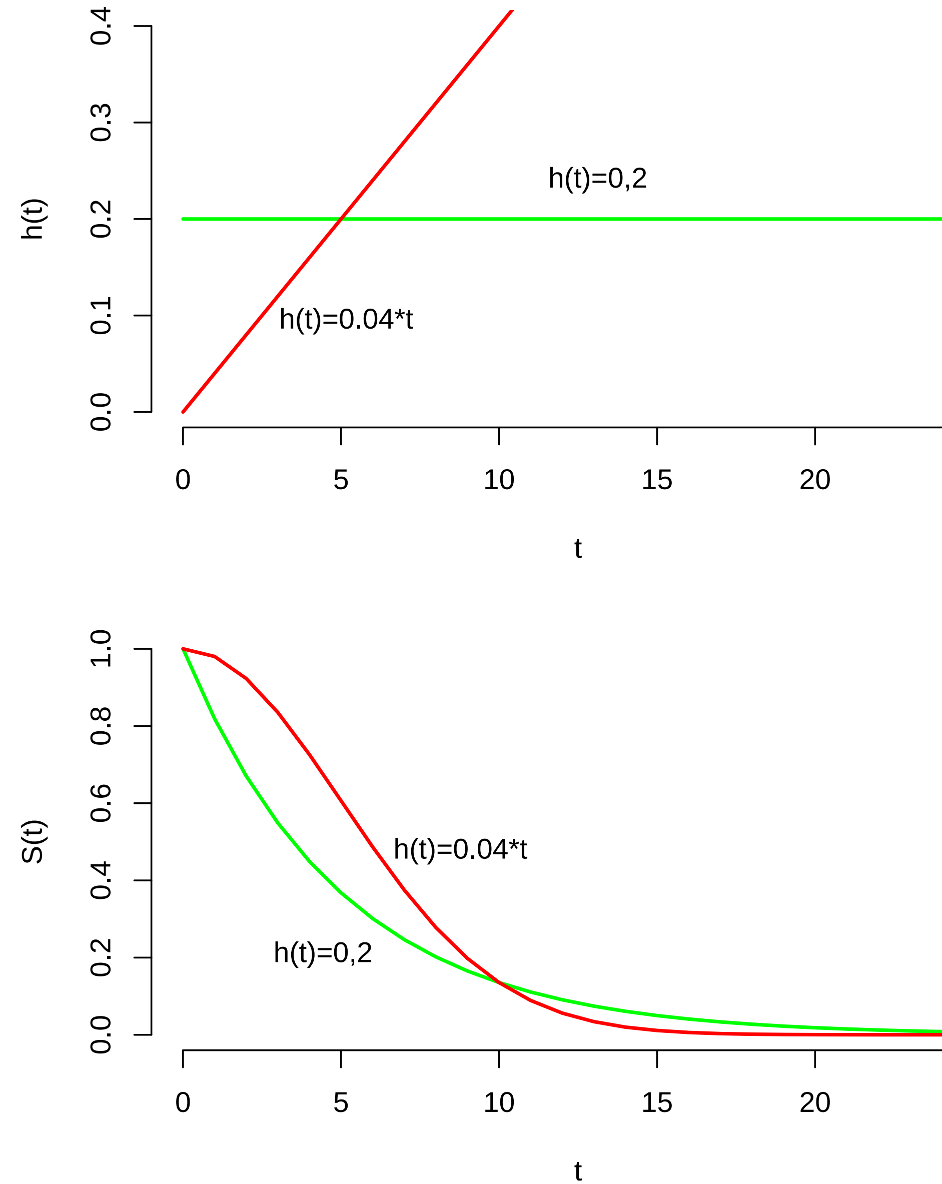
- Diskret

Non-proporsionalitas

Hazard proporsional

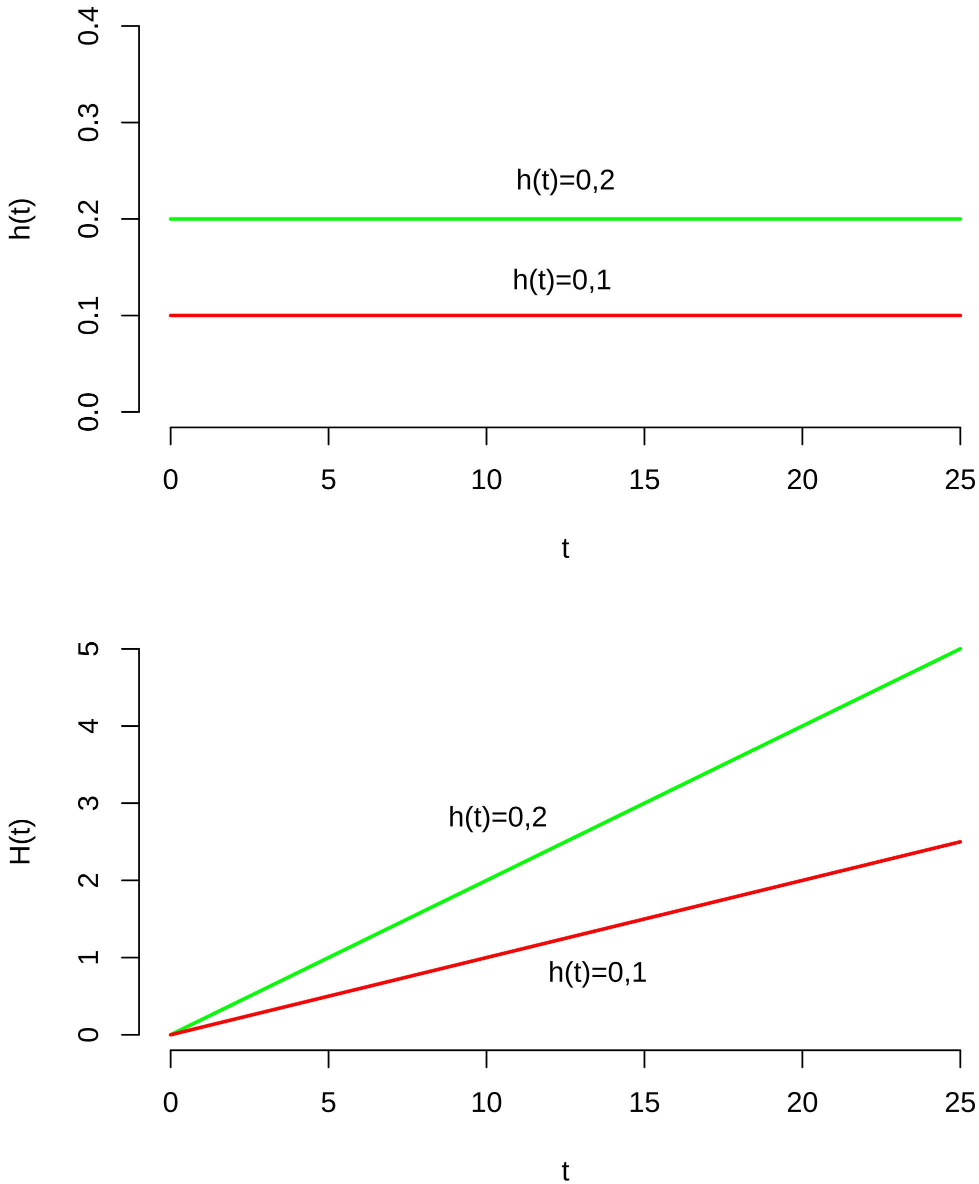


Hazard non-proporsional

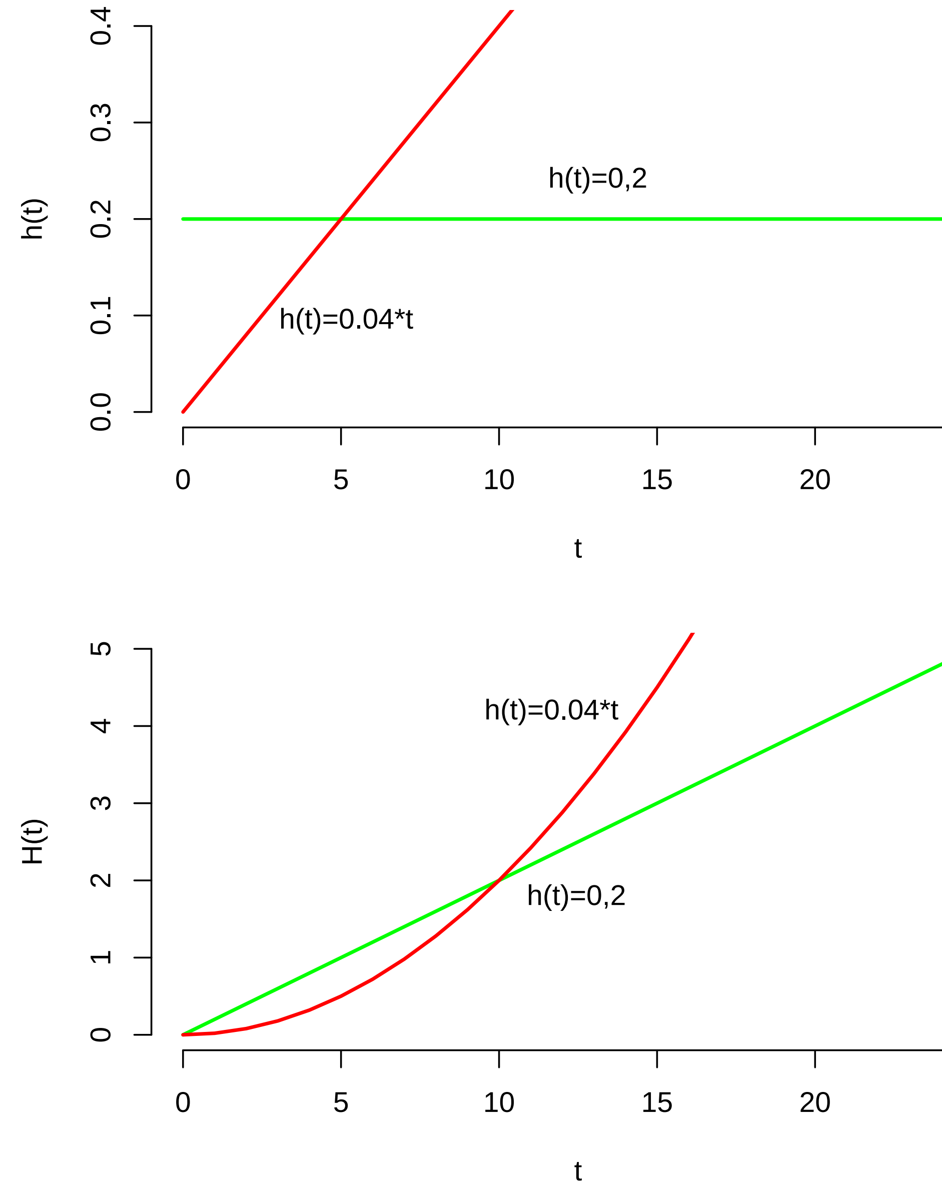


Non-proporsionalitas

Hazard proporsional



Hazard non-proporsional



Stratifikasi

Baseline hazard berbeda antar strata namun parameter $\boldsymbol{\beta}$ sama untuk tiap strata

$$h_j(t \mid \boldsymbol{x}) = h_{0j} \exp(\boldsymbol{x}\boldsymbol{\beta})$$

dengan $j = 1, \ldots, s$ adalah banyaknya strata.
Estimasi untuk $\boldsymbol{\beta}$ menggunakan partial likelihood

$$\ell(\boldsymbol{\beta}) = \ell_1(\boldsymbol{\beta}) + \ell_2(\boldsymbol{\beta}) + \ldots + \ell_s(\boldsymbol{\beta})$$

dengan $\ell_j(\boldsymbol{\beta})$, $j = 1, \ldots, s$ adalah partial likelihood yang dihitung hanya pada subset data dalam strata ke-j.

Cox's Regression Model: R

Data ASI:

```
m1<-coxph(Surv(DUR,D)~SMK+ALCO+race+PVTY,
      data=bfeed)
summary(m1)
```

Cox's Regression Model: R

Data ASI:

```
Call:
coxph(formula = Surv(DUR, D) ~ SMK + ALCO + race + PVTY, data = bfeed)
  n= 927
              coef exp(coef) se(coef)     z       p
SMK          0.288      1.33   0.0768  3.75 0.00018
ALCO         0.141      1.15   0.1218  1.16 0.25000
raceblack    0.178      1.19   0.1041  1.71 0.08700
raceother    0.345      1.41   0.0950  3.63 0.00029
PVTY        -0.162      0.85   0.0882 -1.84 0.06600

            exp(coef) exp(-coef) lower .95 upper .95
SMK              1.33      0.750     1.147      1.55
ALCO             1.15      0.868     0.907      1.46
raceblack        1.19      0.837     0.974      1.47
raceother        1.41      0.708     1.172      1.70
PVTY             0.85      1.176     0.715      1.01
```

Cox’s Regression Model: SPSS

Data ASI:

```
COXREG
  dur  /STATUS=d(1)
   /CONTRAST (race)=Indicator(1)
  /METHOD=ENTER smk alco race pvty
  /PRINT=CI(95) .
```

Cox's Regression Model: SPSS

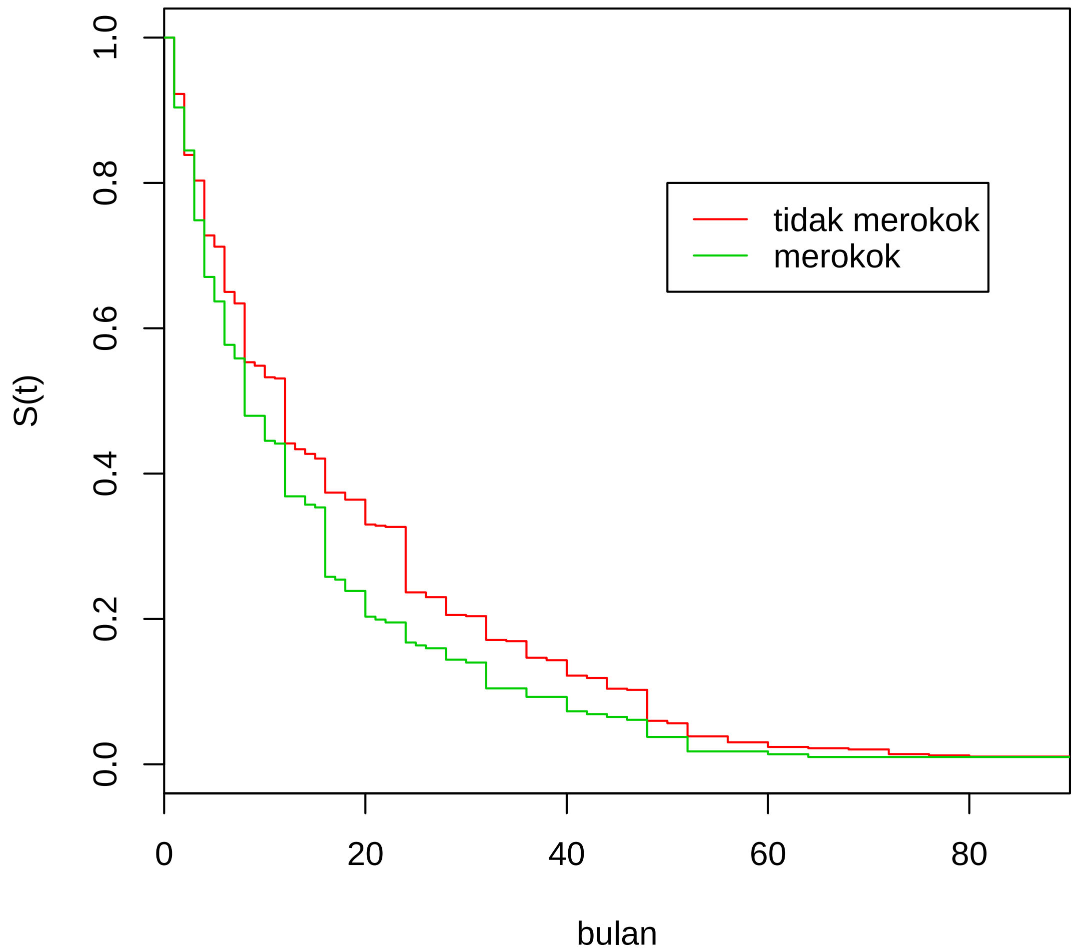
Data ASI:

Variable	B	S.E.	Wald	df	Sig	R
SMK	,2756	,0768	12,8651	1	,0003	,0322
ALCO	,1354	,1217	1,2384	1	,2658	,0000
RACE			12,5149	2	,0019	,0285
RACE(1)	,1578	,1041	2,2981	1	,1295	,0053
RACE(2)	,3264	,0950	11,7917	1	,0006	,0305
PVTY	-,1480	,0882	2,8191	1	,0932	-,0088

		95% CI for Exp(B)	
Variable	Exp(B)	Lower	Upper
SMK	1,3173	1,1331	1,5313
ALCO	1,1450	,9021	1,4534
RACE(1)	1,1709	,9548	1,4359
RACE(2)	1,3859	1,1504	1,6697
PVTY	,8624	,7256	1,0251

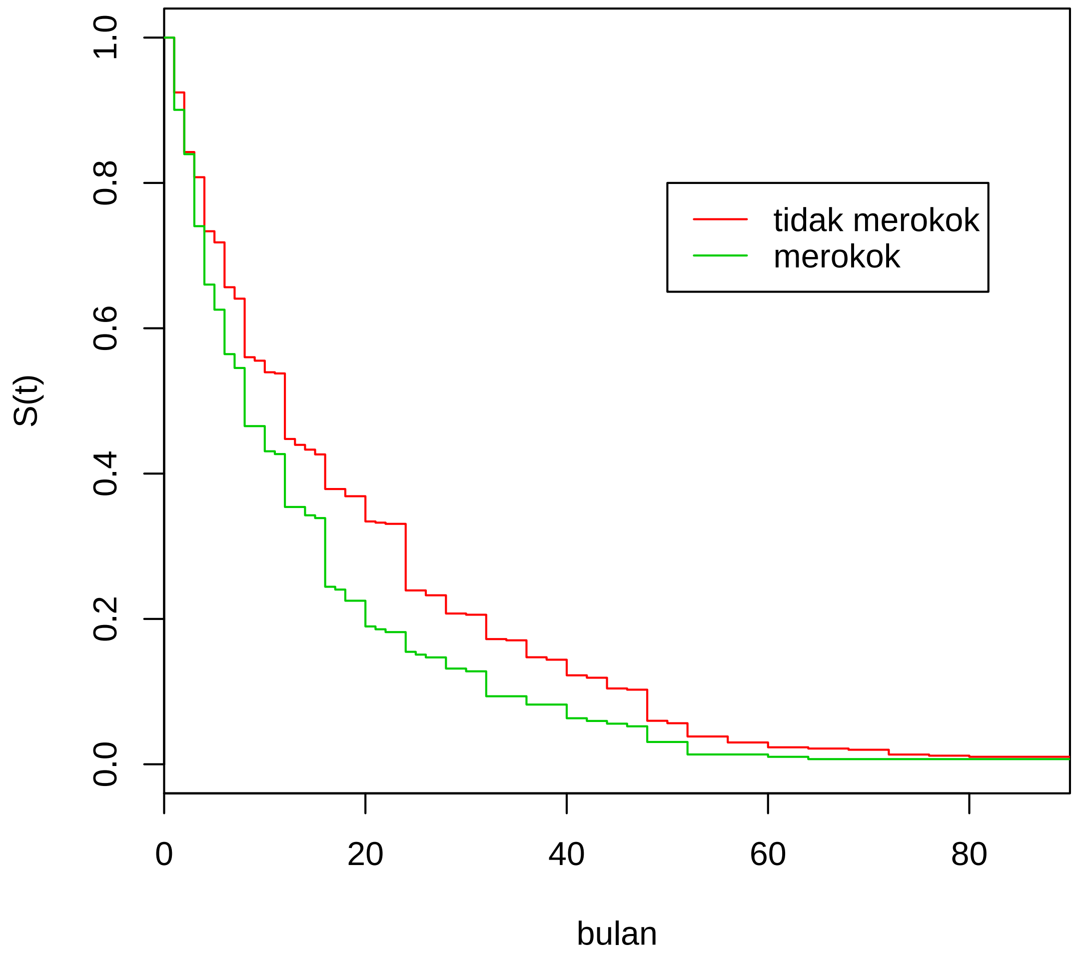
Cox's Regression Model

Kurva survival: status merokok



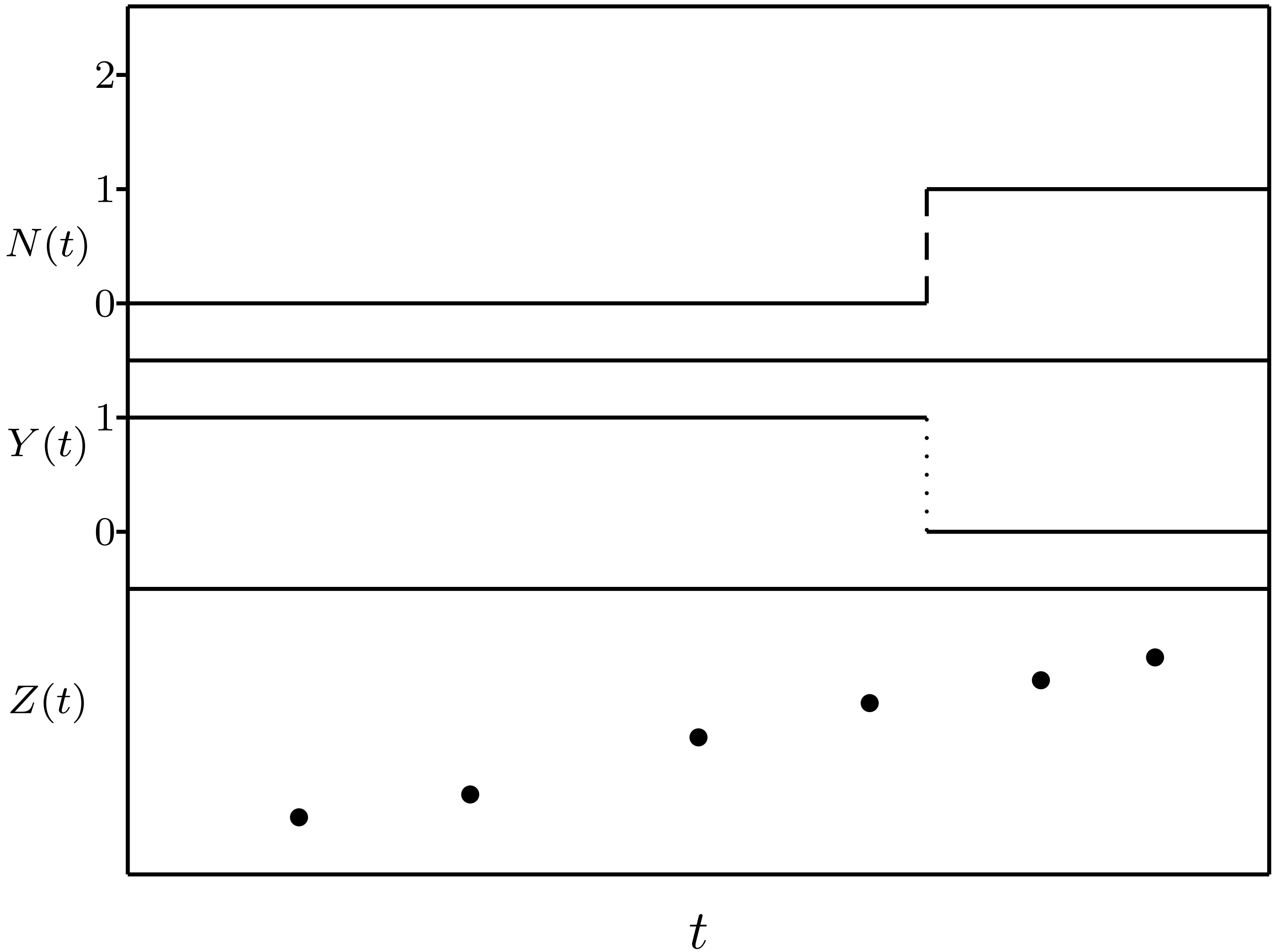
Cox’s Regression Model

Kurva survival: status merokok, dengan memasukkan variabel lain



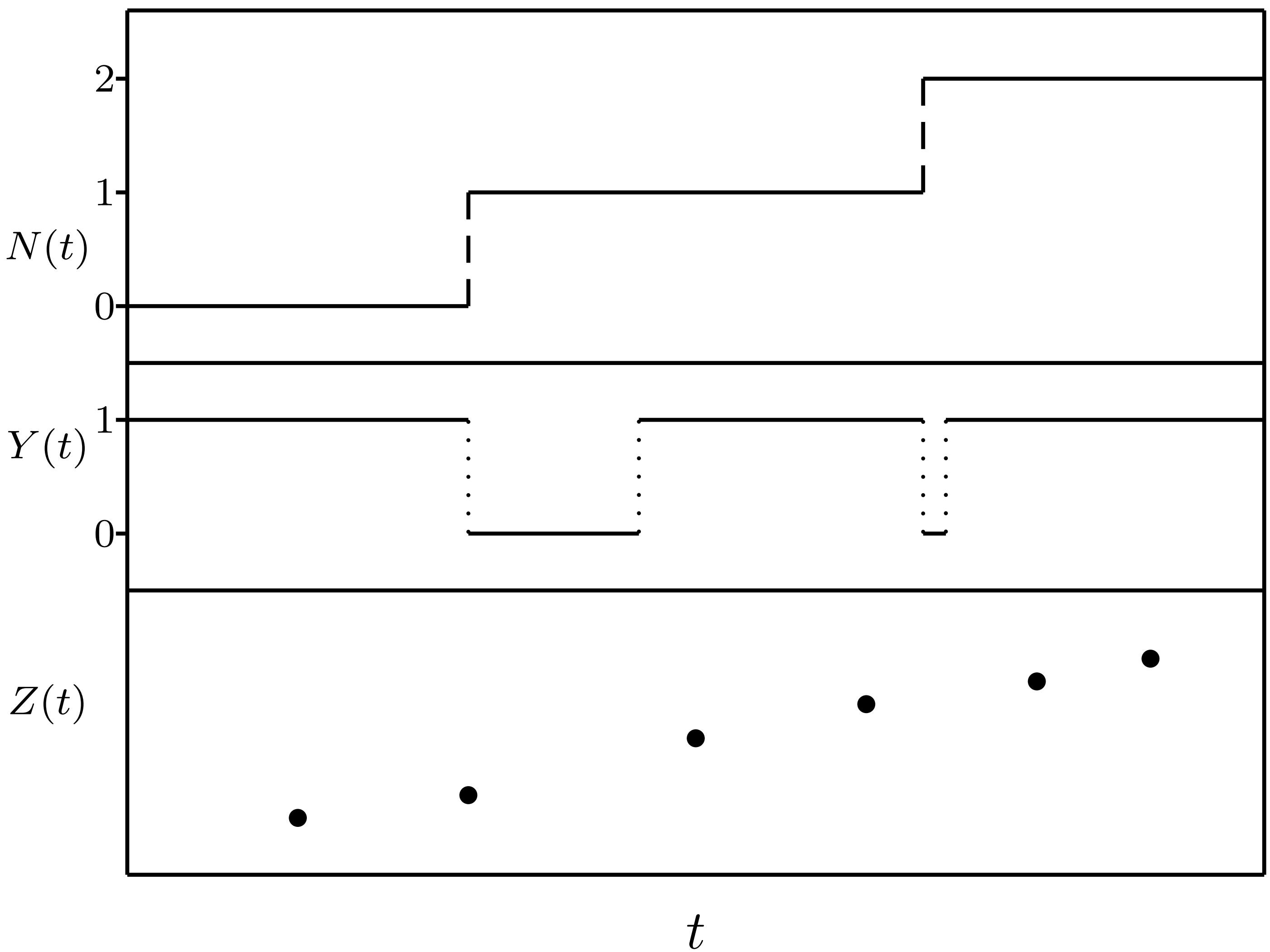
Proses Cacah

Proses cacah (*counting process*) dalam AAK: $\{N(t), Y(t), \boldsymbol{Z}(t)\}$



Proses Cacah

Proses cacah (*counting process*) dalam AAK: $\{N(t), Y(t), \boldsymbol{Z}(t)\}$



Proses Cacah

Model hazard multiplikatif
Data: $\{N_i(t), Y_i(t), \boldsymbol{Z}_i(t)\}, t \geq 0$ untuk individu ke-i,
$i = 1, 2, \ldots, n$

$$Y_i(t)h(t \mid \boldsymbol{Z}_i(t)) \quad = \quad Y_i(t)h_0(t)\exp(\boldsymbol{Z}_i(t)\boldsymbol{\beta})$$

dengan $Y_i(t)$ adalah proses resiko saat t

$$Y_i(t) = \begin{cases} 1 & \text{jika } i \text{ beresiko untuk mendapat kejadian} \\ 0 & \text{jika } i \text{ tidak beresiko untuk mendapat kejadian} \end{cases}$$

dan $\boldsymbol{Z}_i(t)$ adalah nilai variabel penjelas individu i saat t

Proses Cacah

Partial likelihood untuk n individu

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \prod_{t \geq 0} \left\{ \frac{Y_i(t) \exp(\boldsymbol{Z}_i(t)\boldsymbol{\beta})}{\sum_{j=1}^{n} \exp(\boldsymbol{Z}_j(t)\boldsymbol{\beta})} \right\}^{\Delta N_i(t)}$$

dengan

$$\Delta N_i(t) = \begin{cases} 1 & \text{jika } N_i(t) - N_i(t-) = 1 \\ 0 & \text{yang lain} \end{cases}$$

Dalam praktek $\Delta N_i(t)$ adalah indikator δ_i dalam data survival (indikator apakah individu mendapatkan kejadian atau tidak)