

Biostatistika dan Epidemiologi (MMS-4411)

Dr. Danardono, MPH

danardono@ugm.ac.id

Program Studi Statistika
Jurusan Matematika FMIPA UGM

Materi dan Jadwal

Minggu ke-	Pokok Bahasan	Sub Pokok Bahasan
1.	Pendahuluan	<ol style="list-style-type: none">1 Kompetensi dan profesi terkait2 Rencana pembelajaran dan penilaian3 Studi Kasus 1 (tentang kompetensi, profesi, penelitian atau desain)
2.	Desain Penelitian	<ol style="list-style-type: none">1 Penelitian dalam bidang ilmu hayati, kedokteran, kesehatan dan epidemiologi2 Penelitian observasional3 <i>Cross-sectional Study</i>4 <i>Follow-up study</i>5 <i>Case-control study</i>6 Penelitian klinis (<i>clinical trial</i>)

Materi dan Jadwal

3.	Desain Penelitian (lanjutan); Statistik dan ukuran dalam epidemiologi	<ol style="list-style-type: none">1 Pengenalan desain lanjut (nested case-control, case-cohort)2 Kaitan desain penelitian dengan analisis statistik3 Prevalence dan incidence4 Odds ratio, risk ratio dan risk difference5 Studi kasus 2 (tentang penelitian, desain dan analisis)
4.	Alat bantu analisis dan penelitian	<ol style="list-style-type: none">1 Perangkat lunak statistika2 Pencarian informasi
5.	Model linear	<ol style="list-style-type: none">1 Analisis regresi dan ANAVA2 Model linear terumumkan
6.	Analisis Data kategorik	<ol style="list-style-type: none">1 Crosstabulation dan stratifikasi2 Regresi logistik3 Regresi Poisson
7.	Uji Diagnostik	<ol style="list-style-type: none">1 <i>Sensitivity</i> dan <i>specificity</i>2 Kurva ROC

Materi dan Jadwal

8.	Analisis data longitudinal dan survival	<ol style="list-style-type: none">1 Model Regresi untuk data longitudinal2 <i>Kaplan-Meier</i> dan <i>Life Table</i>3 Model Regresi data survival
9.	Studi Kasus 3	Studi kasus tentang penggunaan metode yang tepat, <i>skill</i> olah data, analisis dan interpretasi
10.	Ringkasan Metode	<ol style="list-style-type: none">1 Hubungan antara beberapa metode2 Kelebihan dan kekurangan masing-masing metode3 Dasar teori lebih lanjut
11.	Topik lanjut	Metode yang tidak standar atau metode terkini dari penelitian terakhir atau yang sedang dikembangkan
12.	Praktek Konsultasi	<ol style="list-style-type: none">1 Penulisan laporan2 Komunikasi dan presentasi

Materi dan Jadwal

13.	Studi kasus 4	Studi kasus yang berkaitan dengan praktek konsultasi (individual atau kelompok)
14.	Studi kasus 5	Studi kasus yang berkaitan dengan praktek konsultasi (individual atau kelompok)

Penilaian

No	Unsur Penilaian	Persentase
1.	Ujian Akhir	35
2.	Sisipan	25
3.	Tugas/Presentasi	30
4.	PR/Kuis	10

Pustaka dan Sumber Informasi

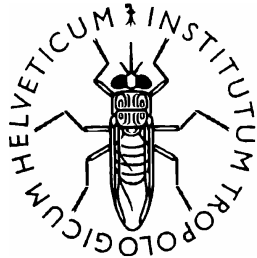
Buku Teks:

1. Le, Chap T. *Introductory Biostatistics*. Wiley, 2003
2. Clayton, D. dan Hills, M. *Statistical Models in Epidemiology*. Oxford University Press, 1993
3. Newman, S. C. *Biostatistical Methods in Epidemiology*. Wiley, 2001
4. Kleinbaum, D. G., Kupper, L. L. dan Morgenstern, H. *Epidemiologic Research: Principles and Quantitative Methods*. Wadsworth, Inc., 1982

Sumber informasi internet:

1. <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed>
2. <http://www.pitt.edu/super1/>

Diskusi



Swiss Tropical Institute
Institut Tropical Suisse
Schweizerisches Tropeninstitut

The Swiss Tropical Institute is seeking

Computing or Biostatistics specialist(s)

to work on projects on the epidemiology of malaria. Candidates should have a post-graduate degree in computer science, biostatistics or a related field. Those from malaria endemic countries are especially encouraged to apply. Salary will be dependent on age and qualifications. Suitable candidates who do not have a doctorate would be encouraged to register for a PhD at the University of Basel.

Applications, including curriculum vitae and names and addresses of two referees should be sent to:

Isabelle Bolliger, Human Resources, Swiss Tropical Institute,
Socinstrasse 57, CH-4002 Basel, Switzerland, email
Isabelle.Bolliger@unibas.ch.

Further information is available from Thomas-A.Smith@unibas.ch.

Diskusi

Biostatistika (Biostatistics)

Epidemiologi (Epidemiology)

Diskusi

Mulailah berpikir dan bertindak bukan hanya sebagai statistisi saja, tapi juga sebagai orang yang mempelajari bidang lain dan dengan sudut pandang yang berbeda dari seorang statistisi

Diskusi

IRON AND ZINC IN INFANCY: RESULTS FROM EXPERIMENTAL TRIALS IN SWEDEN AND INDONESIA

Torbjörn Lind

Epidemiology and Public Health Sciences, Department of Public
Health and Clinical Medicine & Pediatrics Department of Clinical
Sciences Umeå University, 901 87 Umeå, Sweden

Diskusi

ABSTRACT

Background: Iron and zinc are difficult to provide in sufficient amounts in complementary foods to infants world-wide, resulting in high prevalence of both iron and zinc deficiency. These deficiency states cause anemia, delayed neurodevelopment, impaired growth, and increased susceptibility to infections such as diarrhea and respiratory infections.

Diskusi

ABSTRACT

Background: Iron and zinc are difficult to provide in sufficient amounts in complementary foods to infants world-wide, resulting in high **prevalence** of both iron and zinc deficiency. These deficiency states cause anemia, delayed neurodevelopment, impaired growth, and increased **susceptibility** to **infections** such as diarrhea and respiratory infections.

Diskusi

Design: Two different intervention strategies; reduction of a possible inhibitor of iron and zinc absorption, i.e. phytate, or supplementation with iron and zinc, were applied to two different populations in order to improve iron and zinc nutrition:

In a high-income population (Umeå, Sweden), the amount of phytate in commonly consumed infant cereals was reduced. Healthy, term infants (n=300) were at 6 mo of age randomized to phytate-reduced infant cereals, conventional infant cereals, or infant formula and porridge.

In a low income population (Purworejo, Indonesia), daily iron and zinc supplementation was given. Healthy, term infants (n=680) were at 6 mo randomized to supplementation with iron, zinc, a combination of iron and zinc, or placebo.

Blood samples, anthropometrical measurements, and data on infant neurodevelopment and morbidity were collected. Also, in the Swedish study, detailed information on the dietary intake was recorded.

Diskusi

Design: Two different **intervention** strategies; reduction of a possible inhibitor of iron and zinc absorption, i.e. phytate, or supplementation with iron and zinc, were applied to two different **populations** in order to improve iron and zinc nutrition:

In a high-income population (Umeå, Sweden), the amount of phytate in commonly consumed infant cereals was reduced. Healthy, term infants (n=300) were at 6 mo of age **randomized** to phytate-reduced infant cereals, conventional infant cereals, or infant formula and porridge.

In a low income population (Purworejo, Indonesia), daily iron and zinc supplementation was given. Healthy, term infants (n=680) were at 6 mo **randomized** to supplementation with iron, zinc, a combination of iron and zinc, or **placebo**.

Blood samples, anthropometrical measurements, and data on infant neurodevelopment and **morbidity** were collected. Also, in the Swedish study, detailed information on the dietary intake was recorded.

Diskusi

Results: In the Swedish study, the reduction of phytate had little effect on iron and zinc status, growth, development or incidence of diarrhea or respiratory infections, possibly due to the presence of high contents of ascorbic acid, which may counteract the negative effects of phytate. In the Indonesian study, significant negative interaction between iron and zinc was evident for several of the outcomes; Hb and serum ferritin improved more in the iron only group compared to placebo or the combined iron and zinc group. Further, supplementation with iron alone improved infant psychomotor development and knee-heel length, whereas supplementation with zinc alone improved weight and knee-heel length compared to placebo. Combined iron and zinc supplementation did decrease the prevalence of iron deficiency anemia and low serum zinc, but had no other positive effects. Vomiting was more common in the combined group.

Diskusi

Results: In the Swedish [study](#), the reduction of phytate had little [effect](#) on iron and zinc status, growth, development or [incidence](#) of diarrhea or respiratory infections, possibly due to the presence of high contents of ascorbic acid, which may counteract the negative effects of phytate. In the Indonesian study, significant negative [interaction](#) between iron and zinc was [evident](#) for several of the [outcomes](#); Hb and serum ferritin improved more in the iron only group [compared](#) to placebo or the combined iron and zinc group. Further, supplementation with iron alone improved infant psychomotor development and knee-heel length, whereas supplementation with zinc alone improved weight and knee-heel length compared to placebo. Combined iron and zinc supplementation did decrease the prevalence of iron deficiency anemia and low serum zinc, but had no other positive effects. Vomiting was more common in the combined group.

Diskusi

Results (continued): Analyses of dietary intake from the Swedish study showed that dietary iron intake in the 6-11 mo period was significantly associated with Hb, but not serum ferritin at 9 and 12 mo, whereas the opposite was true in the 12-17 mo period, i.e. dietary iron intake was significantly associated with serum ferritin, but not Hb at 18 mo.

Diskusi

Results (continued): *Analyses* of dietary intake from the Swedish study showed that dietary iron intake in the 6-11 mo period was *significantly associated* with Hb, but not serum ferritin at 9 and 12 mo, whereas the opposite was true in the 12-17 mo period, i.e. dietary iron intake was significantly associated with serum ferritin, but not Hb at 18 mo.

Diskusi

Conclusions: The phytate content of commercial infant cereals does not seem to contribute to poor iron and zinc status of Swedish infants as feared. However, the current definitions of iron and zinc deficiency in infancy may overestimate the problem, and a change in the recommended cutoffs is suggested. These studies also indicate that dietary iron is preferably channeled towards erythropoiesis during infancy, but to an increasing amount channeled towards storage in early childhood. This suggests that in evaluating dietary programs, Hb may be superior in monitoring response to dietary iron in infancy, whereas S-Ft may respond better later in childhood. However, as shown in this study, increasing Hb may not necessarily be an indicator of iron deficiency, as more dietary iron increased Hb regardless of iron status.

Diskusi

Conclusions (continued): In the low-income setting combined supplementation with iron and zinc resulted in significant negative interaction. Thus, it is not possible to recommend routine iron-zinc supplementation at the molar concentration and mode used in this study. It is imperative that further research efforts are focused at finding cost-effective strategies to prevent iron and zinc deficiency in low-income populations.

Tujuan Penelitian Epidemiologi

1. Mendeskripsikan status kesehatan populasi dengan cara melakukan enumerasi kejadian sakit, menghitung frekuensi relatif dan mendapatkan kecenderungan atau *trend* penyakit;
2. Menjelaskan penyebab penyakit dengan cara menentukan faktor yang menjadi sebab dari suatu penyakit tertentu dan cara transmisinya;
3. Melakukan prediksi kejadian sakit dan distribusi status kesehatan dalam populasi;
4. Melakukan pengendalian penyebaran penyakit dalam populasi dengan pencegahan kejadian sakit, penyembuhan kasus sakit, menambah lama hidup bersama dengan suatu penyakit, atau meningkatkan status kesehatannya

Tujuan Penelitian Epidemiologi

1. Mendeskripsikan status kesehatan populasi dengan cara melakukan enumerasi kejadian sakit, menghitung frekuensi relatif dan mendapatkan kecenderungan atau *trend* penyakit; **describe**
2. Menjelaskan penyebab penyakit dengan cara menentukan faktor yang menjadi sebab dari suatu penyakit tertentu dan cara transmisinya; **explain**
3. Melakukan prediksi kejadian sakit dan distribusi status kesehatan dalam populasi; **predict**
4. Melakukan pengendalian penyebaran penyakit dalam populasi dengan pencegahan kejadian sakit, penyembuhan kasus sakit, menambah lama hidup bersama dengan suatu penyakit, atau meningkatkan status kesehatannya **control**

Type Penelitian

- Observasional (tanpa ada manipulasi atau perlakuan)
 - Deskriptif
 - Analitik
- Eksperimental (ada manipulasi atau perlakuan dengan randomisasi)
 - Laboratorium
 - Penelitian klinis
 - Intervensi komunitas
- Quasi-eksperimental (ada manipulasi atau perlakuan tanpa randomisasi)
 - Laboratorium/Klinis
 - Program/Kebijakan

Type Penelitian

?

populasi

Type Penelitian



populasi

Variabel

Type Penelitian

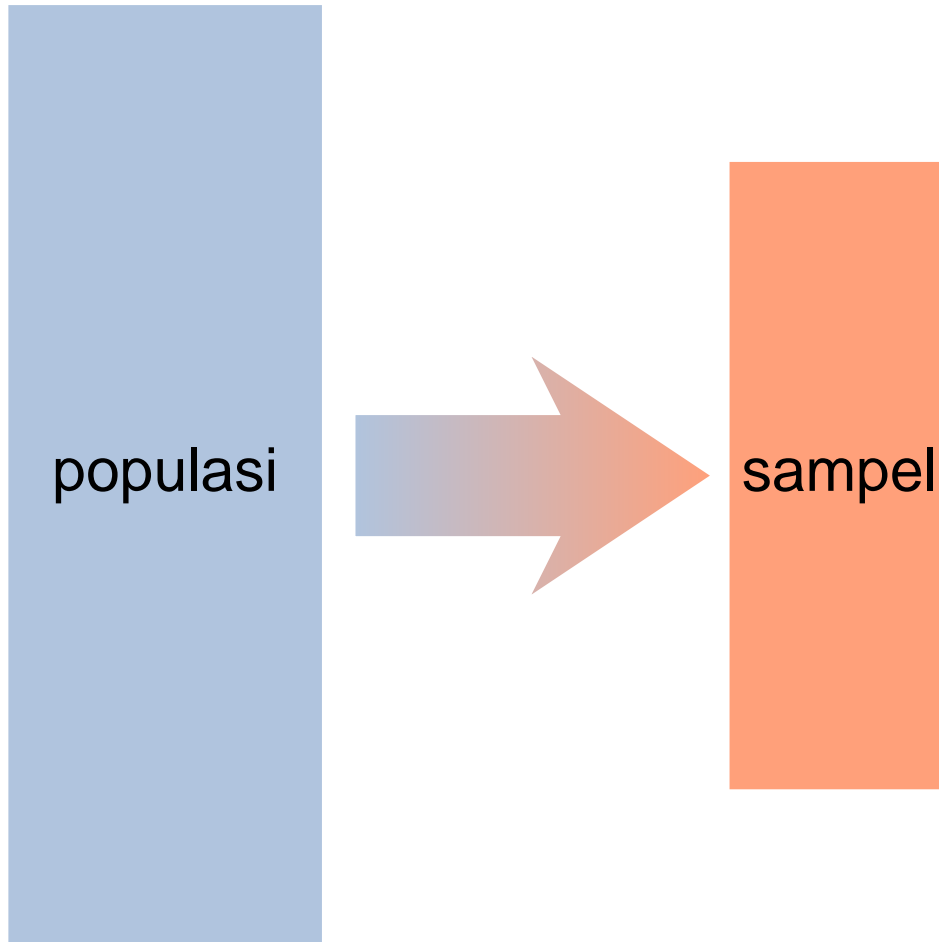


populasi

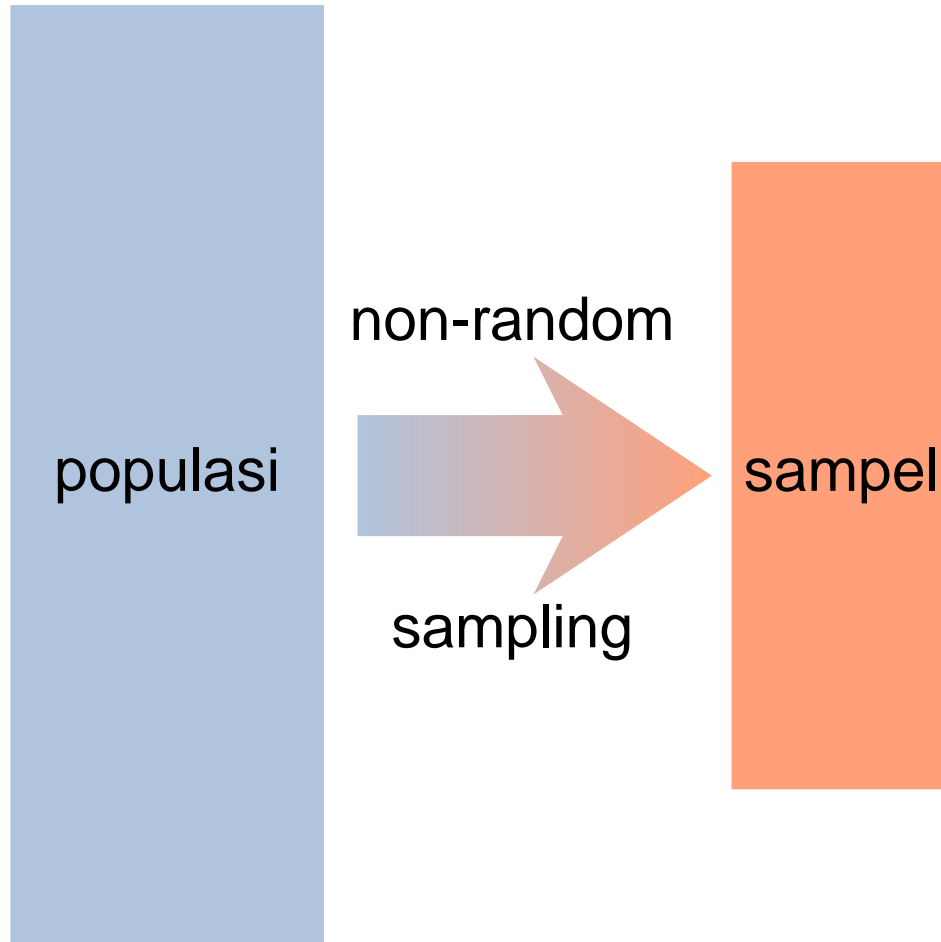
X_1, X_2, \dots v. independen, v. penjelas, faktor, *exposure*, *confounder*

Y_1, Y_2, \dots v. dependen, *outcome*, *response*

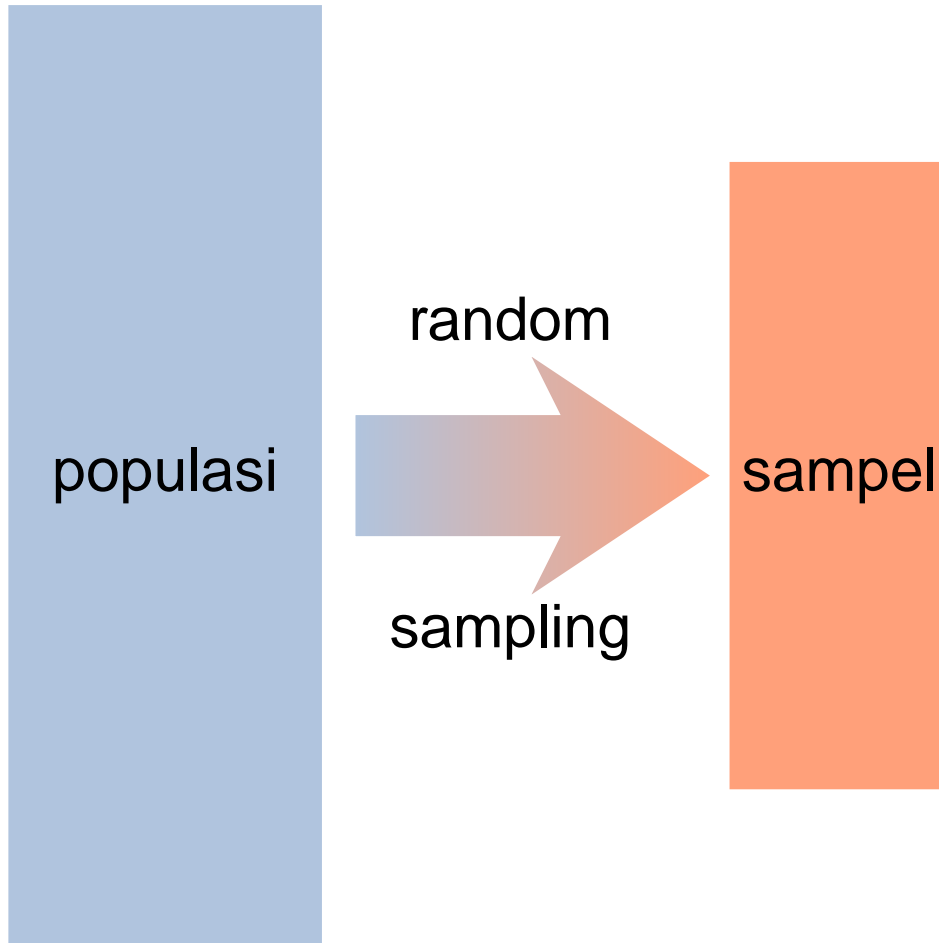
Type Penelitian



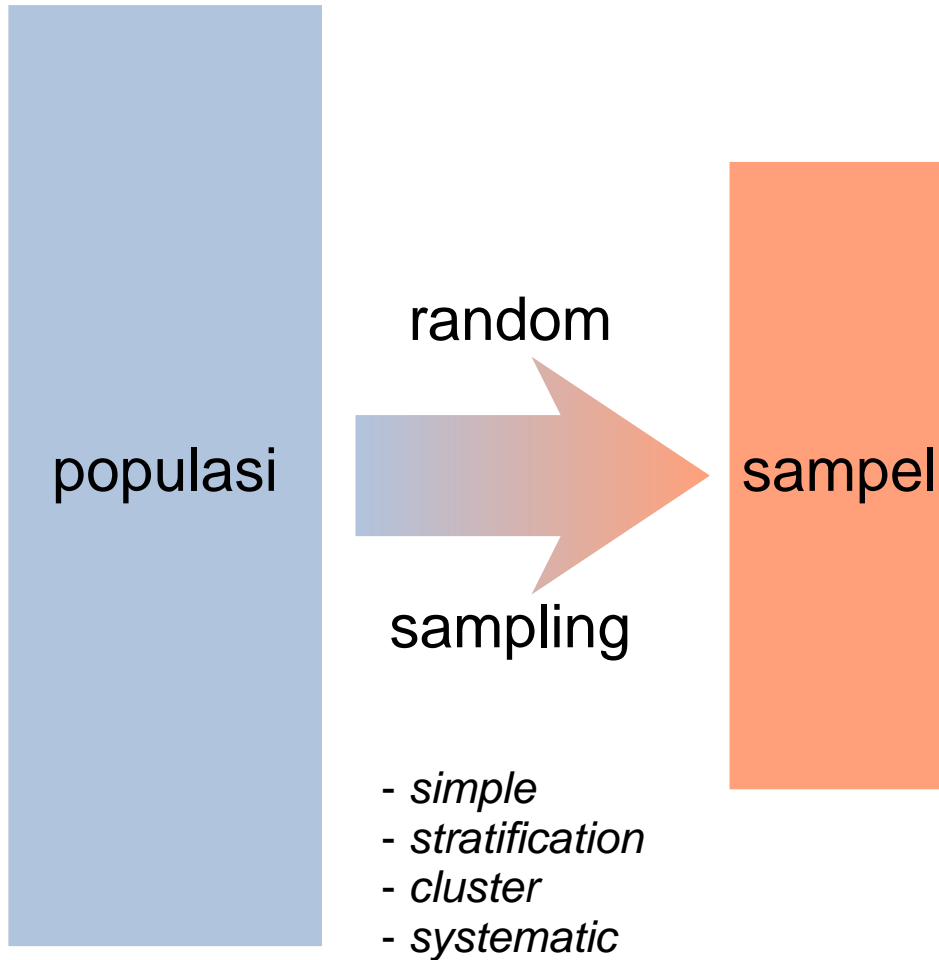
Type Penelitian



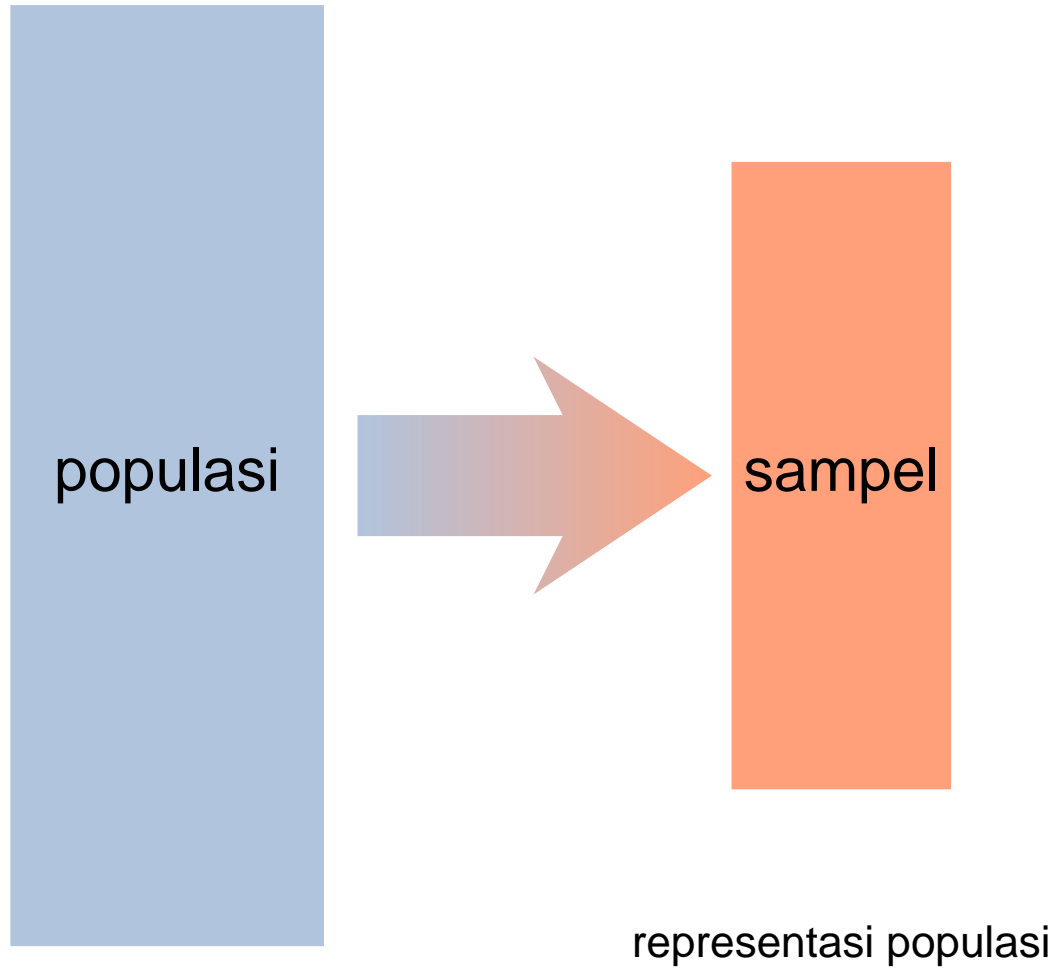
Type Penelitian



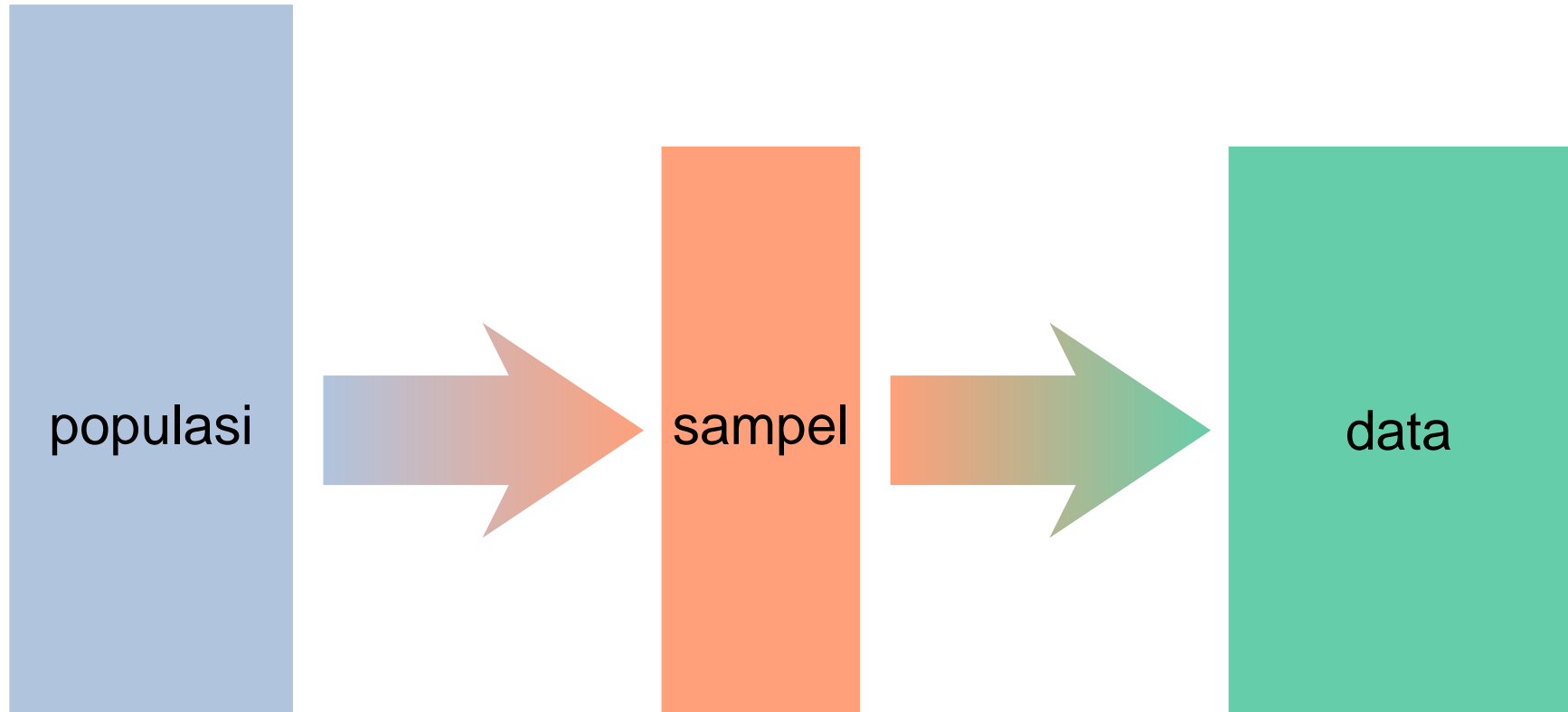
Type Penelitian



Type Penelitian



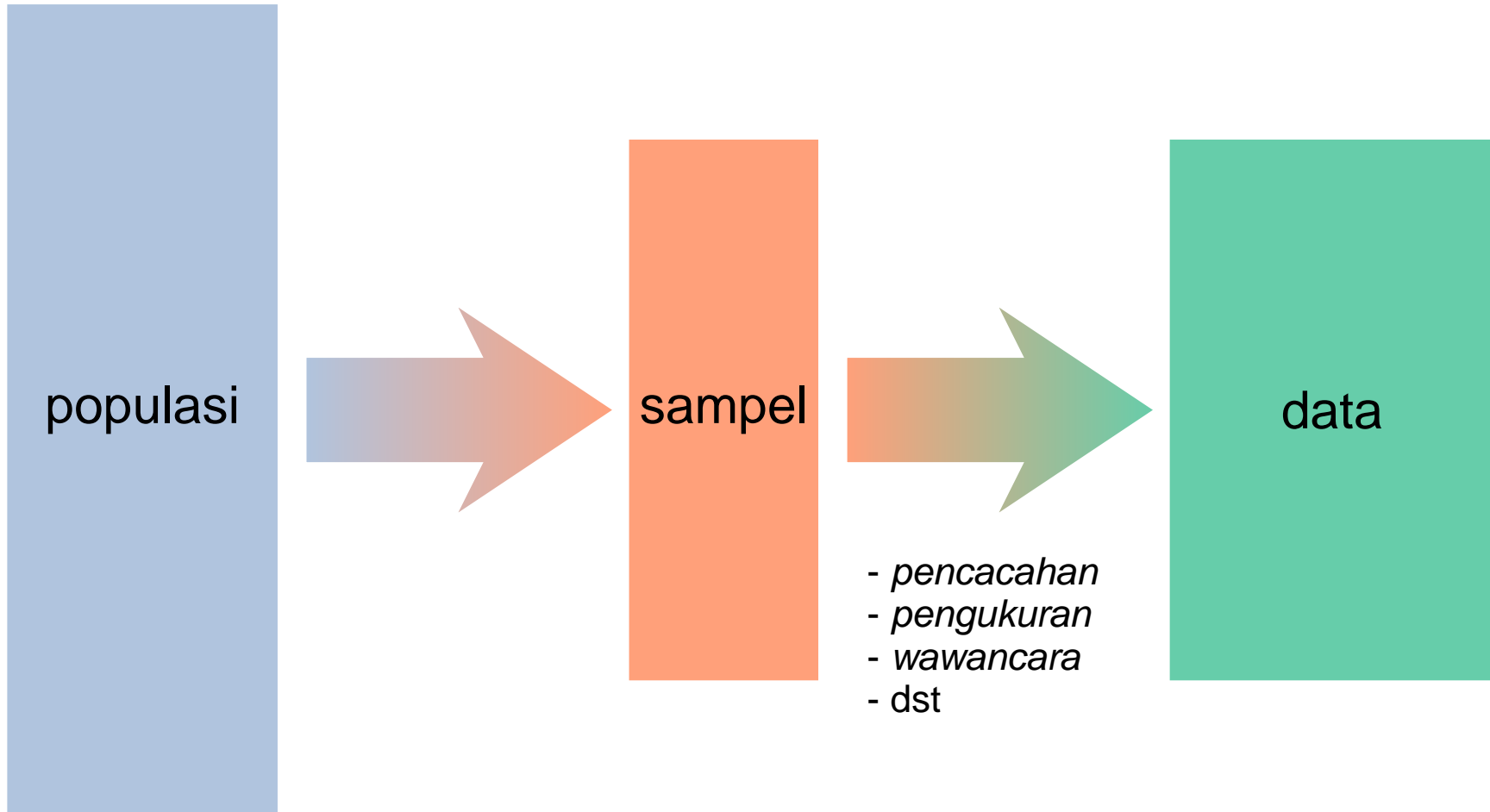
Type Penelitian



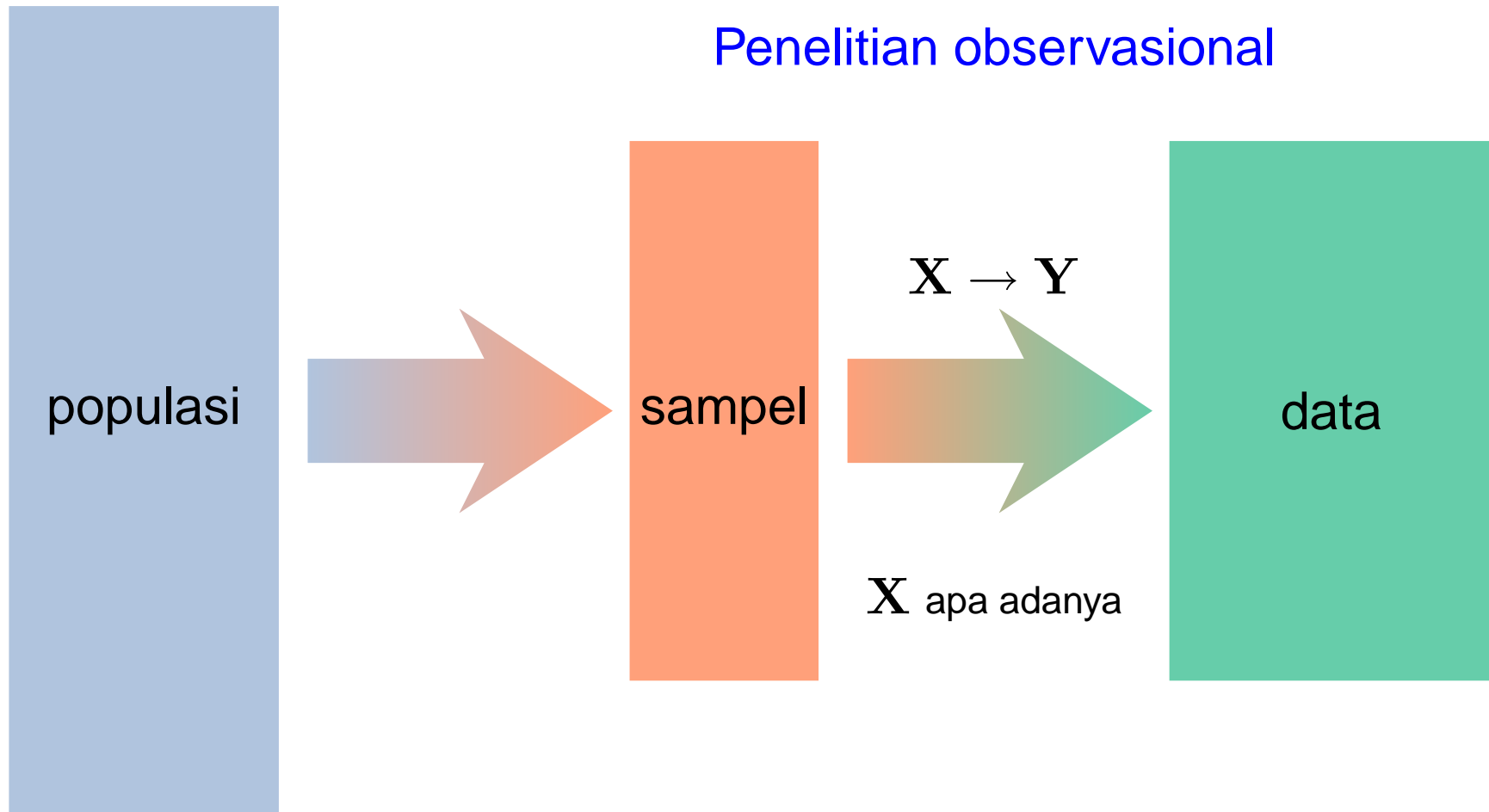
x_1, x_2, \dots

y_1, y_2, \dots

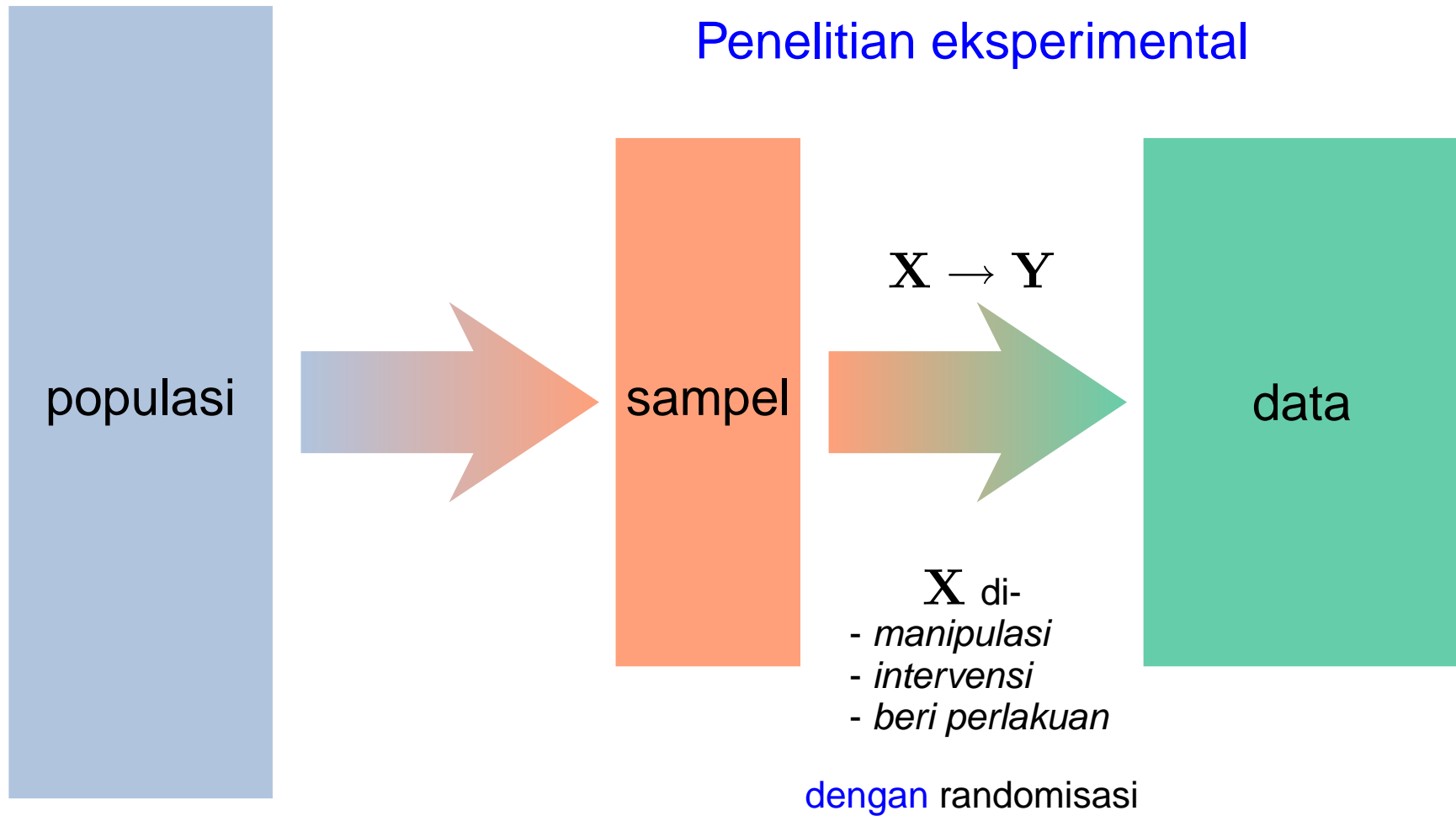
Type Penelitian



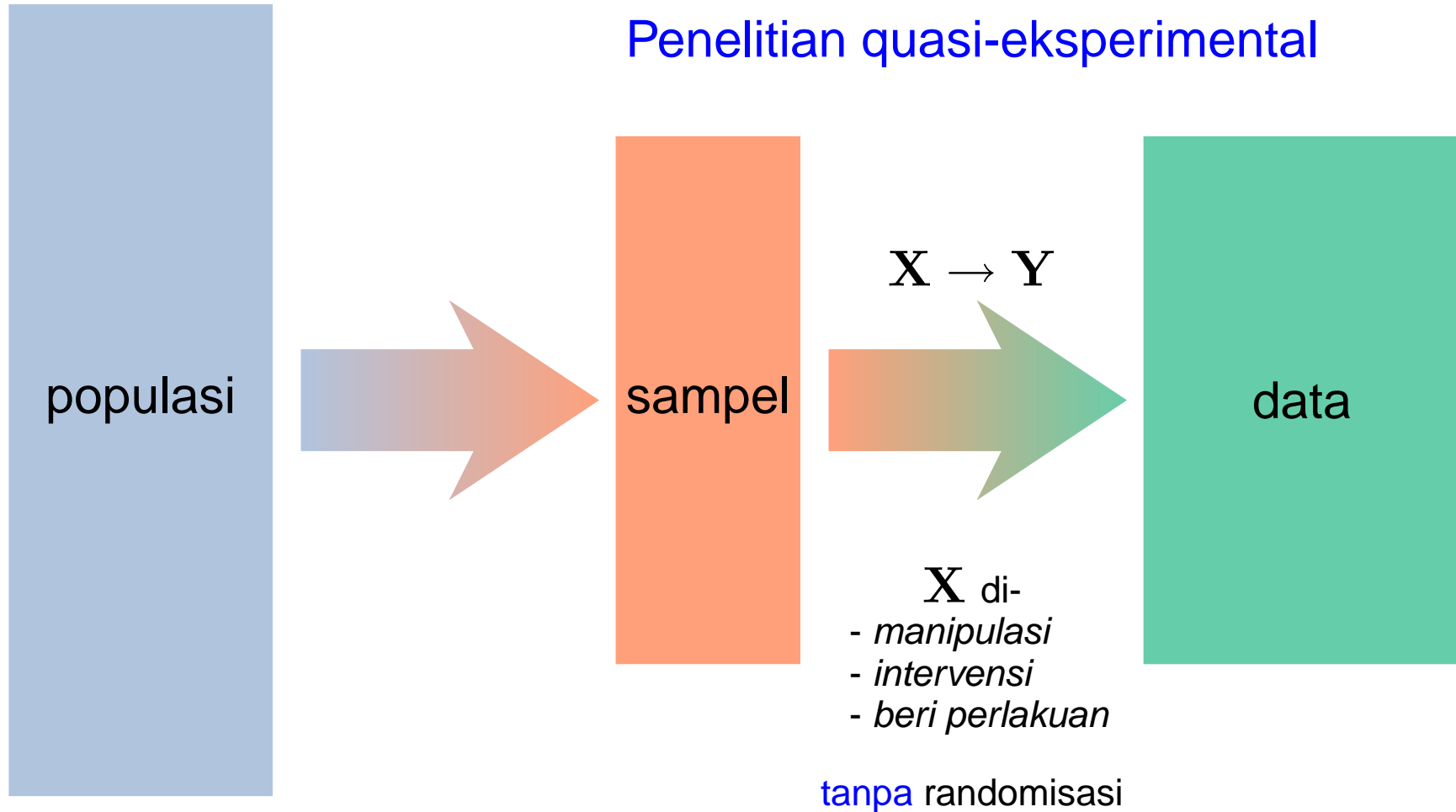
Type Penelitian



Type Penelitian



Type Penelitian



Desain Penelitian

Individu

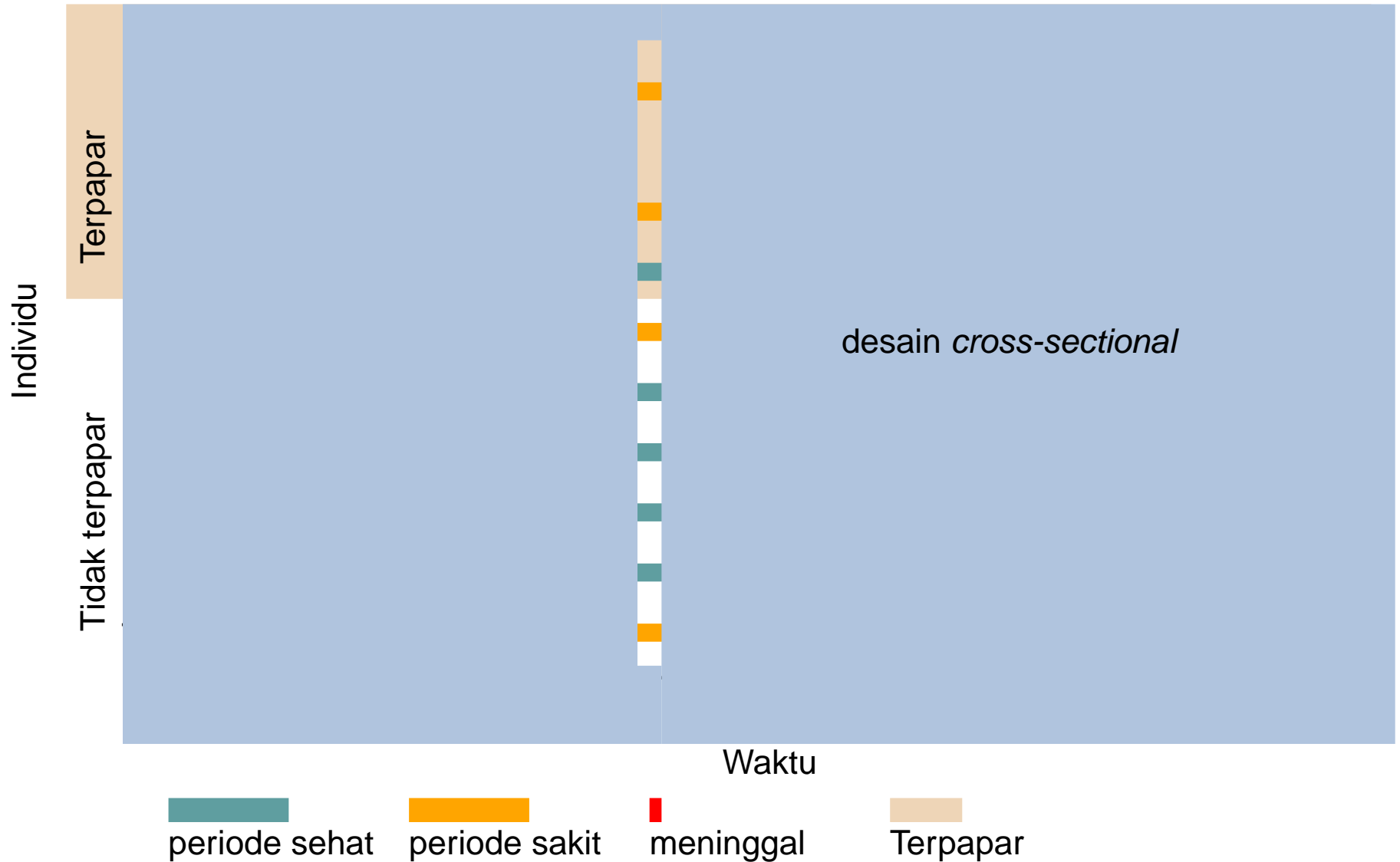


Waktu

Desain Penelitian



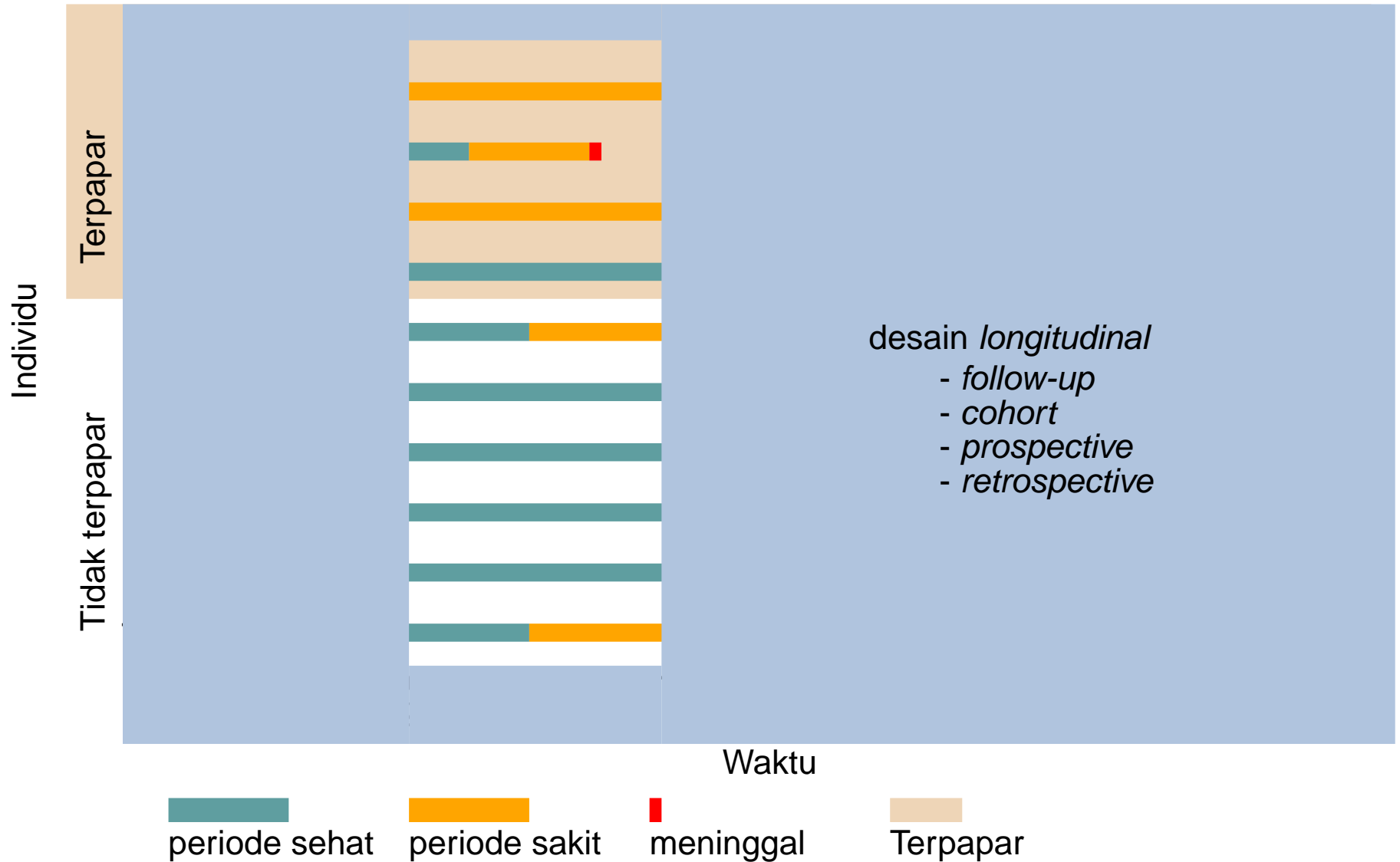
Desain Penelitian



Desain Penelitian

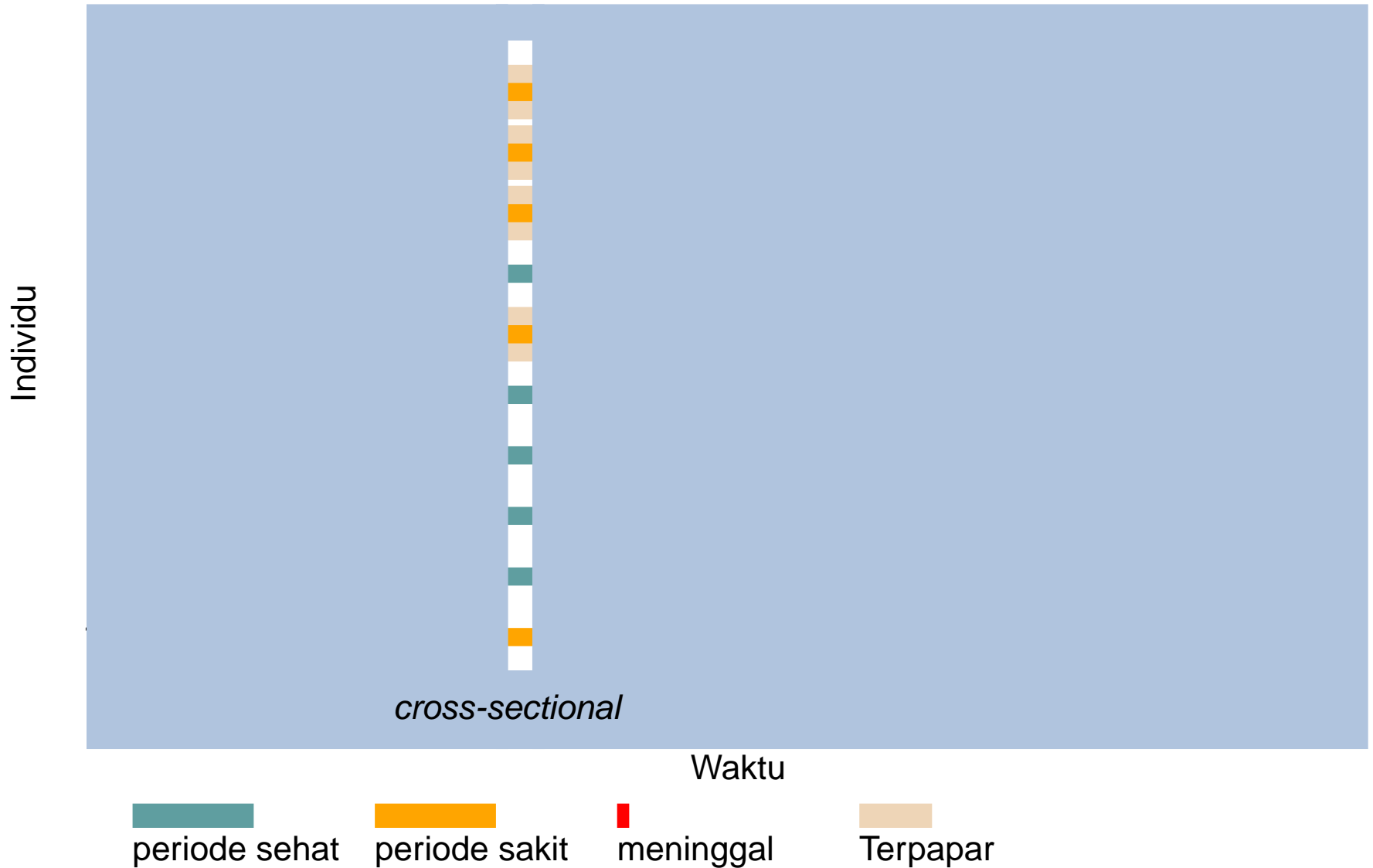


Desain Penelitian



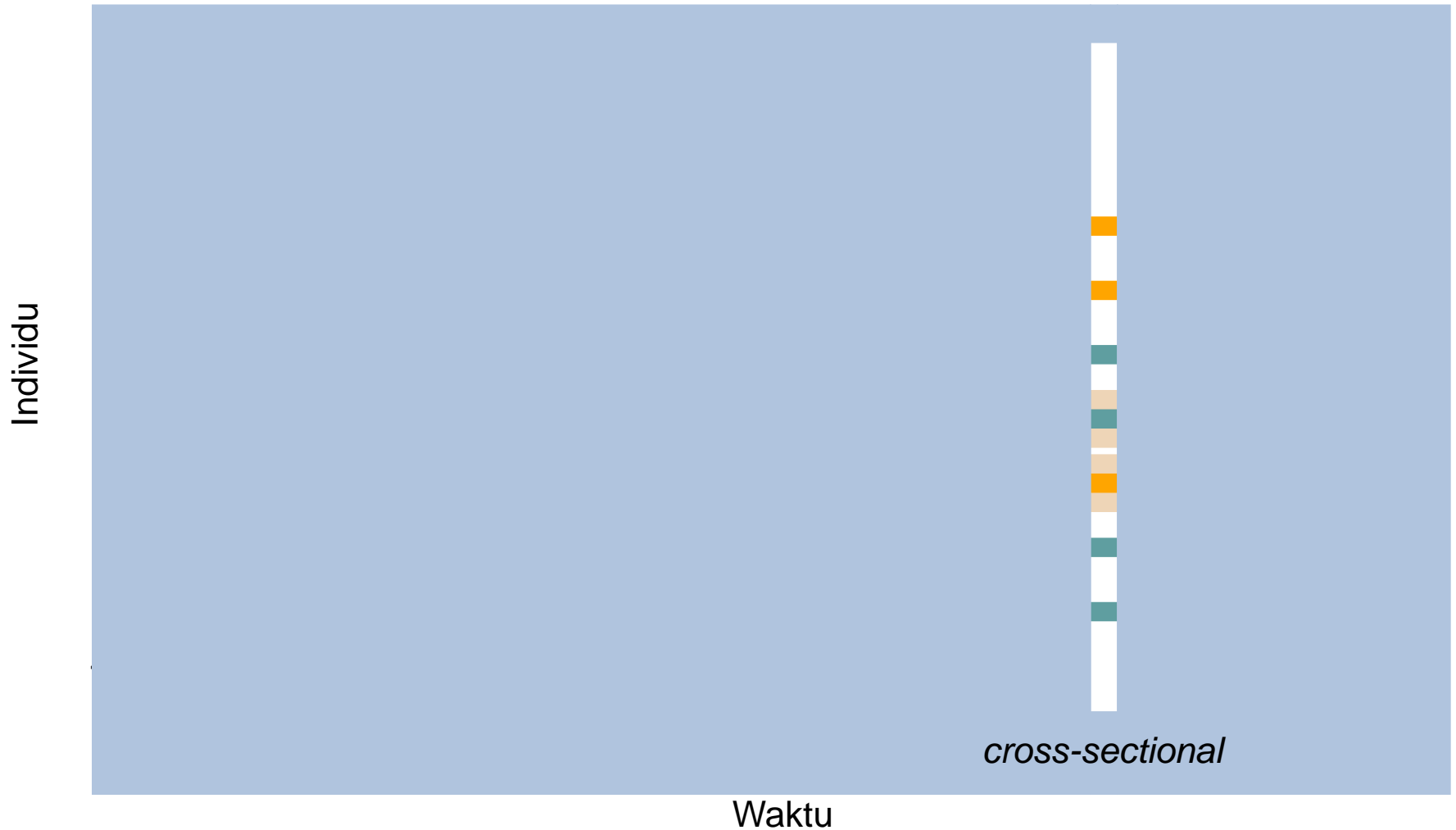
Desain Penelitian

Apakah **terpapar** menyebabkan **sakit** ?



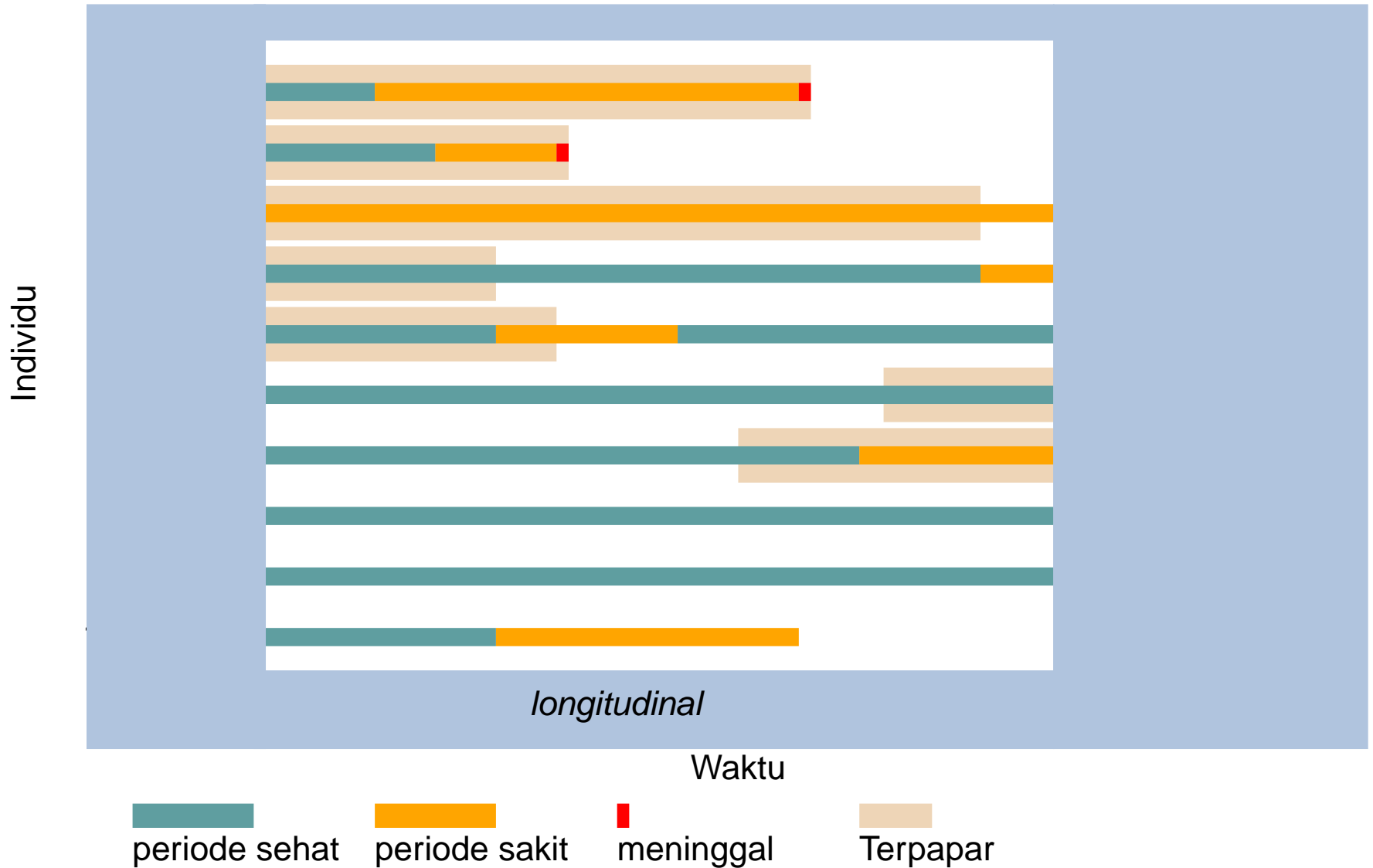
Desain Penelitian

Apakah **terpapar** menyebabkan **sakit** ?



Desain Penelitian

Apakah **terpapar** menyebabkan **sakit** ?



Statistik dan Ukuran dalam Epidemiologi

Prevalensi adalah banyaknya subyek yang mengalami kejadian tertentu atau menderita penyakit tertentu pada suatu waktu tertentu

Insidensi adalah banyaknya subyek yang mengalami kejadian baru atau mendapatkan penyakit baru dalam suatu interval waktu tertentu

Tingkat (*rate*) adalah banyaknya perubahan kuantitatif yang terjadi terkait dengan waktu

Prevalensi

$$P = \frac{d}{N}$$

d: banyaknya subyek yang mengalami kejadian tertentu atau menderita penyakit tertentu pada suatu waktu tertentu

N: banyaknya subyek pada suatu waktu tersebut

Faktor-faktor yang mempengaruhi prevalensi:

Insidensi

Insidensi Kumulatif

$$IK = \frac{d}{N_0}$$

d : banyaknya subyek yang mengalami kejadian tertentu atau menderita penyakit tertentu dalam suatu interval waktu tertentu

N_0 : banyaknya subyek yang belum mengalami kejadian tertentu atau menderita penyakit tertentu pada awal interval waktu tersebut

Insidensi

Insidensi (*Incidence rate*)

$$I = \frac{d}{NT}$$

d : banyaknya subyek yang mengalami kejadian tertentu atau menderita penyakit tertentu dalam suatu interval waktu tertentu

NT : Total waktu subyek yang belum mengalami kejadian tertentu atau menderita penyakit tertentu dalam interval waktu tersebut

(Istilah lain: *person-time incidence rate*, *instantaneous incidence rate*, *force of morbidity*, *incidence-density*, *hazard*)

Hubungan antara Prevalensi dgn. Insidensi

Bila prevalensi kecil dan tidak berubah menurut waktu

$$\text{prevalensi} \approx \text{insidensi} \times \text{durasi}$$

Faktor yang mempengaruhi Prevalence

naik karena

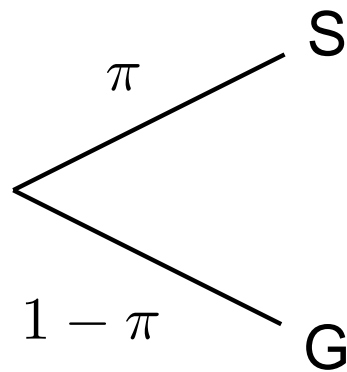
durasi penyakit yg panjang
pasien hidup lama
insidensi meningkat
in-migrasi kasus
out-migrasi penduduk sehat
in-migrasi orang yg rentan
meningkatnya diagnosis

turun karena

durasi penyakit yg pendek
pasien hidup singkat
insidensi menurun
in-migrasi penduduk sehat
out-migrasi kasus
out-migrasi orang yg rentan
meningkatnya kesembuhan

Risk dan Odd

Model probabilitas biner (*Bernoulli*)



π adalah probabilitas sukses (*risk*)

Odd $\pi : (1 - \pi)$ atau

$$\Omega = \frac{\pi}{1 - \pi}$$

Likelihood

Dari $n = 10$ orang diketahui outcome sukses (S) dan gagal (G) SSGSGGGSGG (misalnya sukses adalah terkena penyakit tertentu dan gagal adalah tidak terkena penyakit tertentu).

Seberapa mungkin data ini berasal dari model binomial dengan

(i) $\pi = 0,1$; (ii) $\pi = 0,5$?

Likelihood

Dari $n = 10$ orang diketahui outcome sukses (S) dan gagal (G) SSGSGGGSGG (misalnya sukses adalah terkena penyakit tertentu dan gagal adalah tidak terkena penyakit tertentu).

Seberapa mungkin data ini berasal dari model binomial dengan

(i) $\pi = 0,1$; (ii) $\pi = 0,5$?

$$\begin{aligned}L(\pi \mid \text{data}) &= \pi\pi(1 - \pi)\pi(1 - \pi)(1 - \pi)(1 - \pi)\pi(1 - \pi)(1 - \pi) \\ &= 0,1^4 \times 0,9^6 \\ &= 5,31 \times 10^{-5}\end{aligned}$$

Likelihood

Dari $n = 10$ orang diketahui outcome sukses (S) dan gagal (G) SSGSGGGSGG (misalnya sukses adalah terkena penyakit tertentu dan gagal adalah tidak terkena penyakit tertentu).

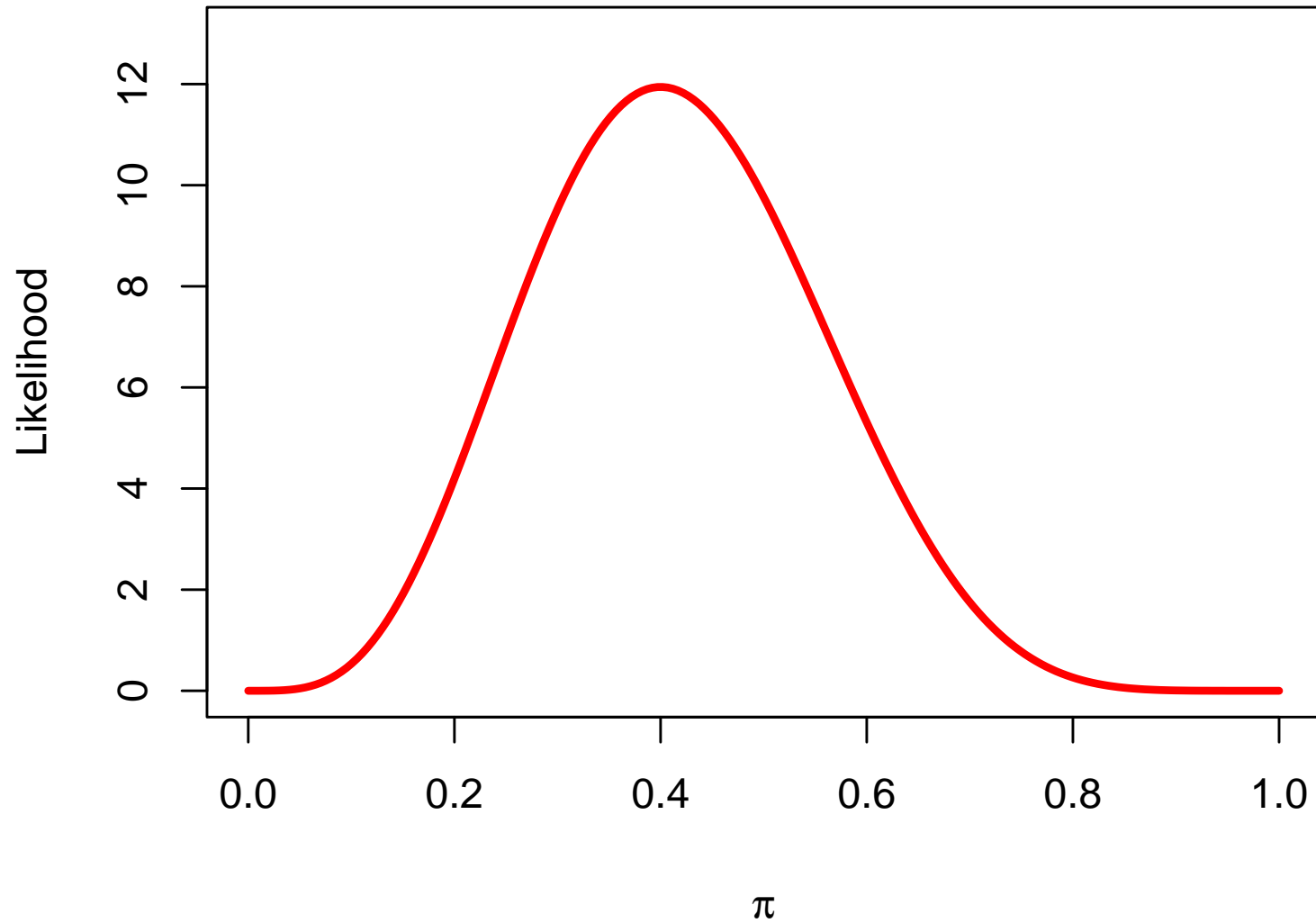
Seberapa mungkin data ini berasal dari model binomial dengan

(i) $\pi = 0,1$; (ii) $\pi = 0,5$?

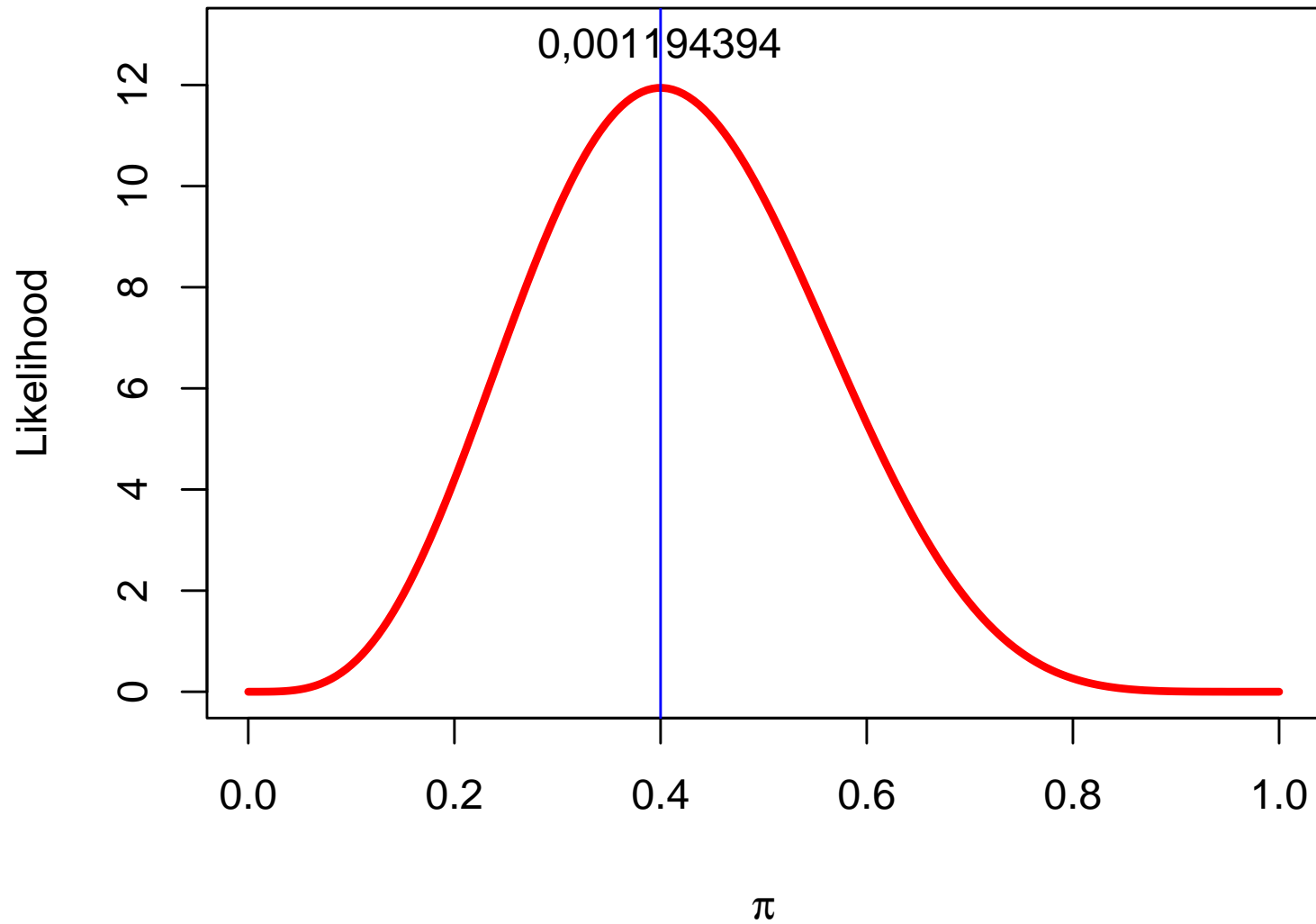
$$\begin{aligned}L(\pi \mid \text{data}) &= \pi\pi(1 - \pi)\pi(1 - \pi)(1 - \pi)(1 - \pi)\pi(1 - \pi)(1 - \pi) \\ &= 0,1^4 \times 0,9^6 \\ &= 5,31 \times 10^{-5}\end{aligned}$$

$$\begin{aligned}L(\pi \mid \text{data}) &= \pi\pi(1 - \pi)\pi(1 - \pi)(1 - \pi)(1 - \pi)\pi(1 - \pi)(1 - \pi) \\ &= 0,5^4 \times 0,5^6 \\ &= 9,77 \times 10^{-4}\end{aligned}$$

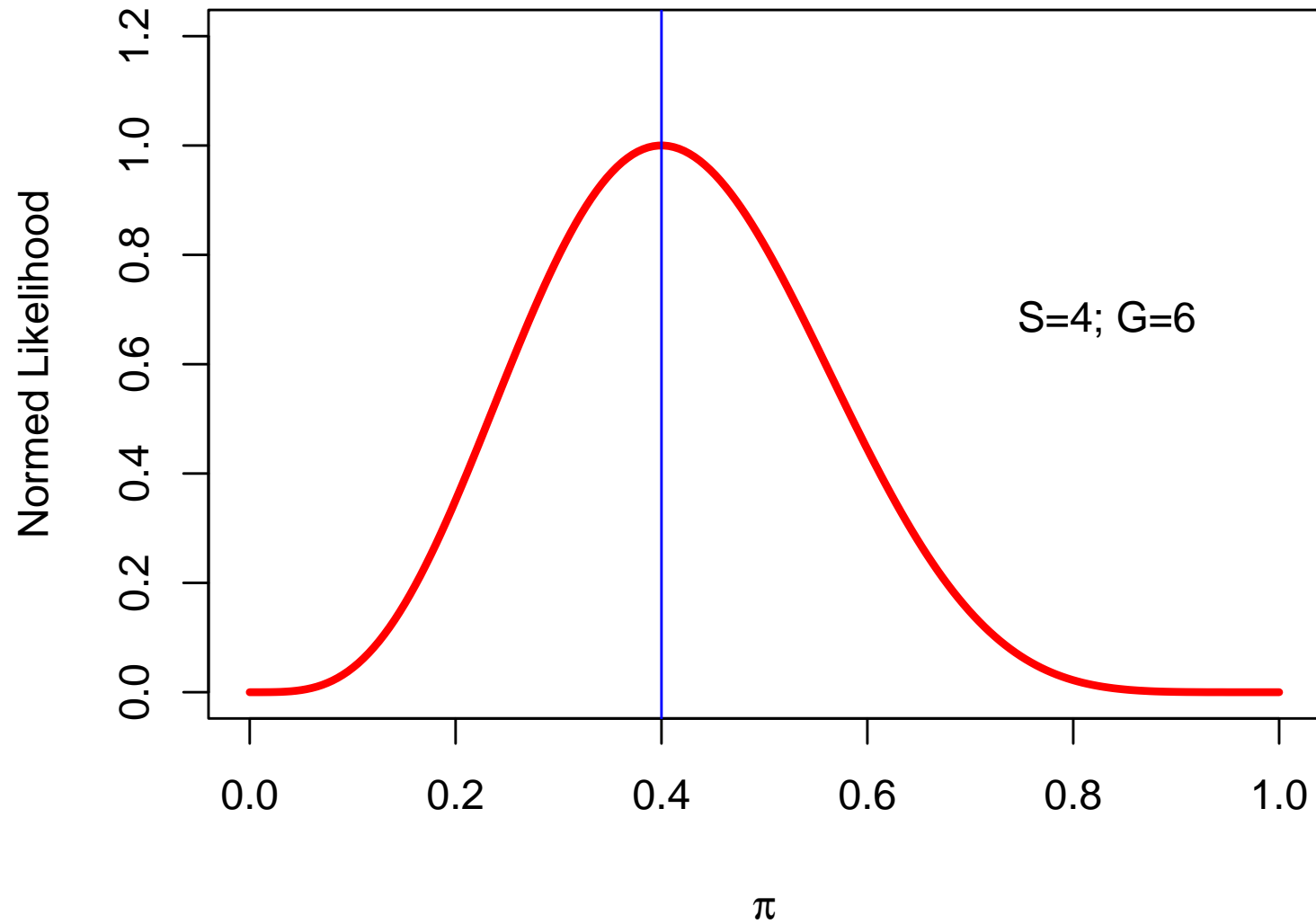
Likelihood



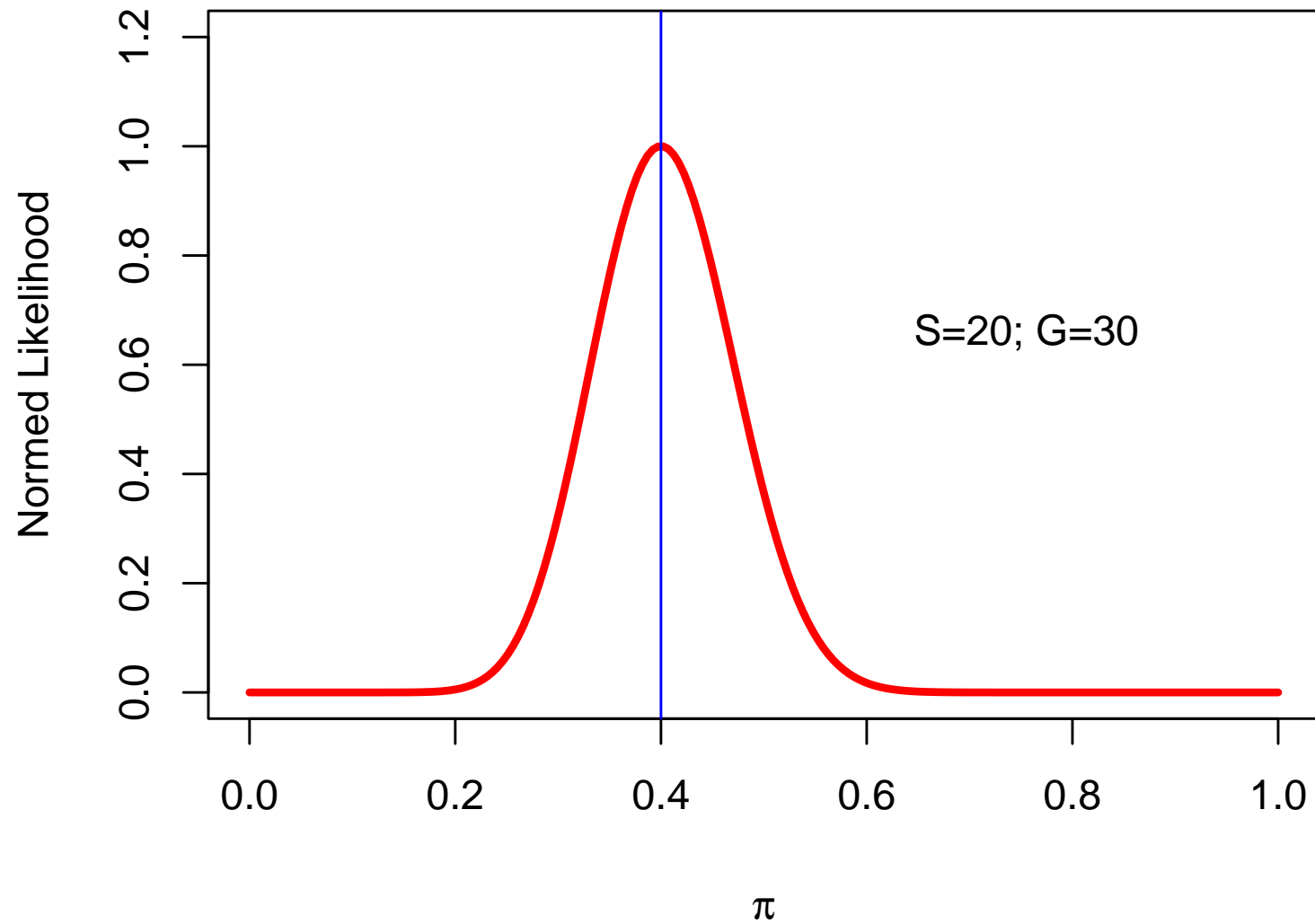
Likelihood



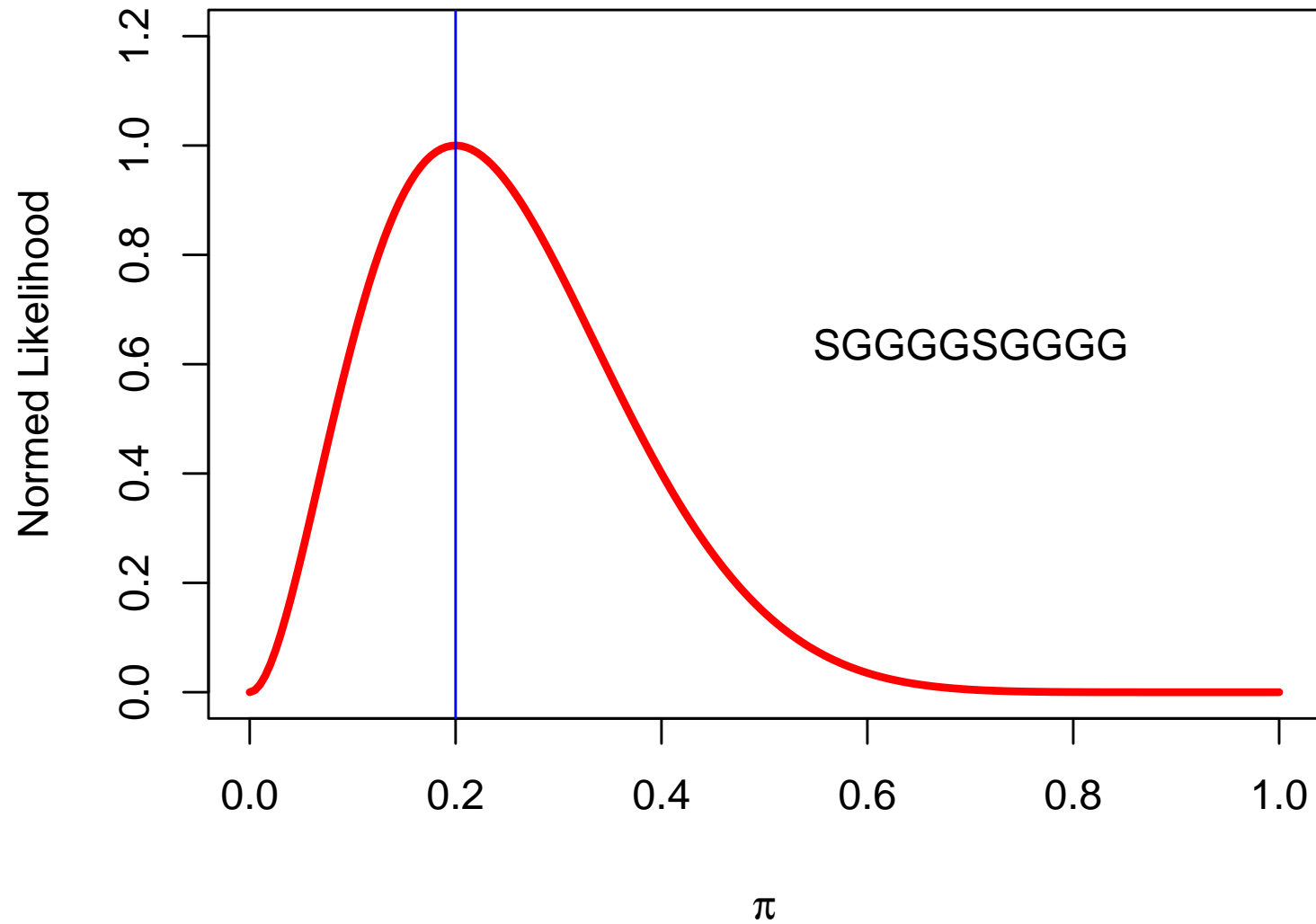
Likelihood



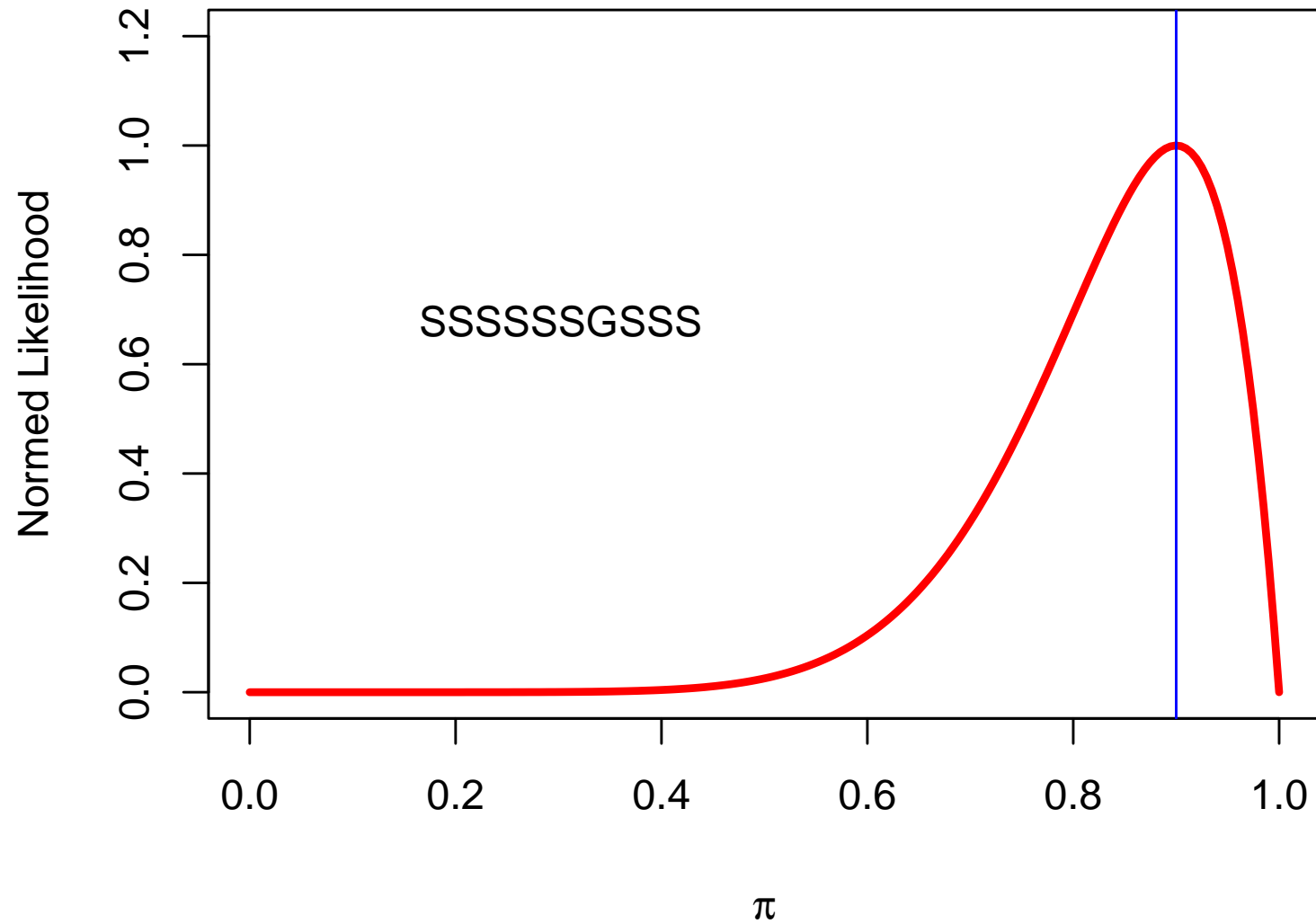
Likelihood



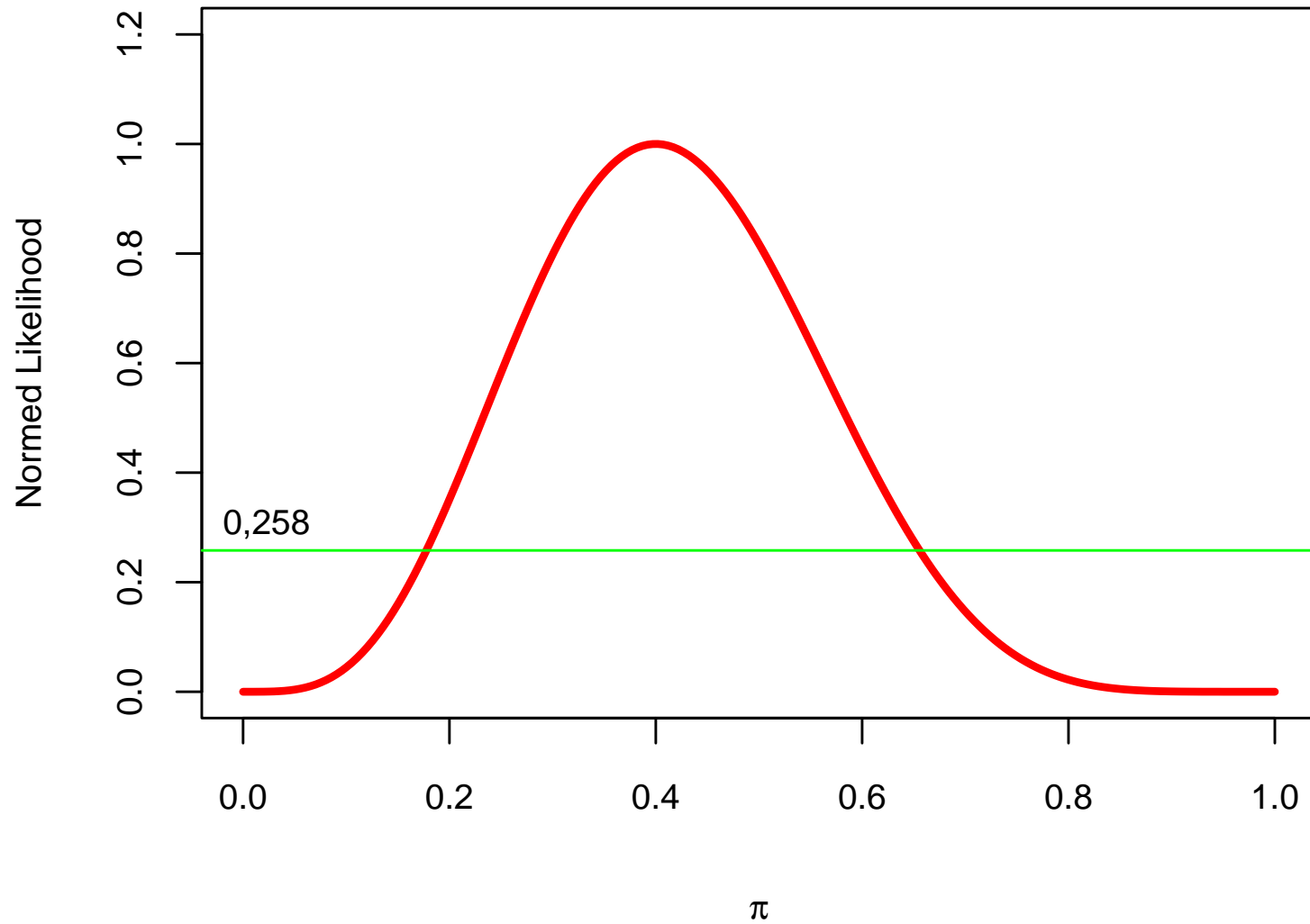
Likelihood



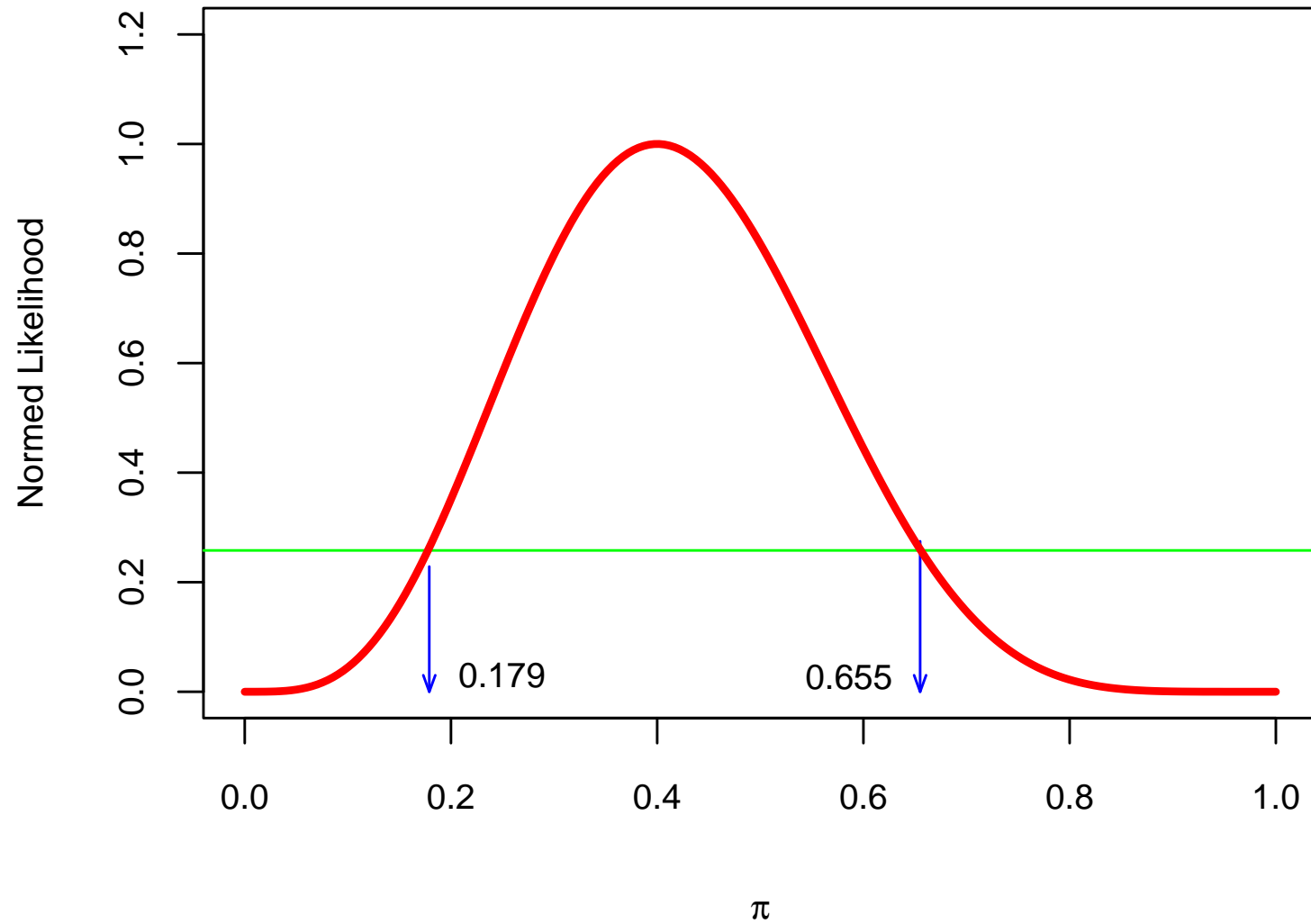
Likelihood



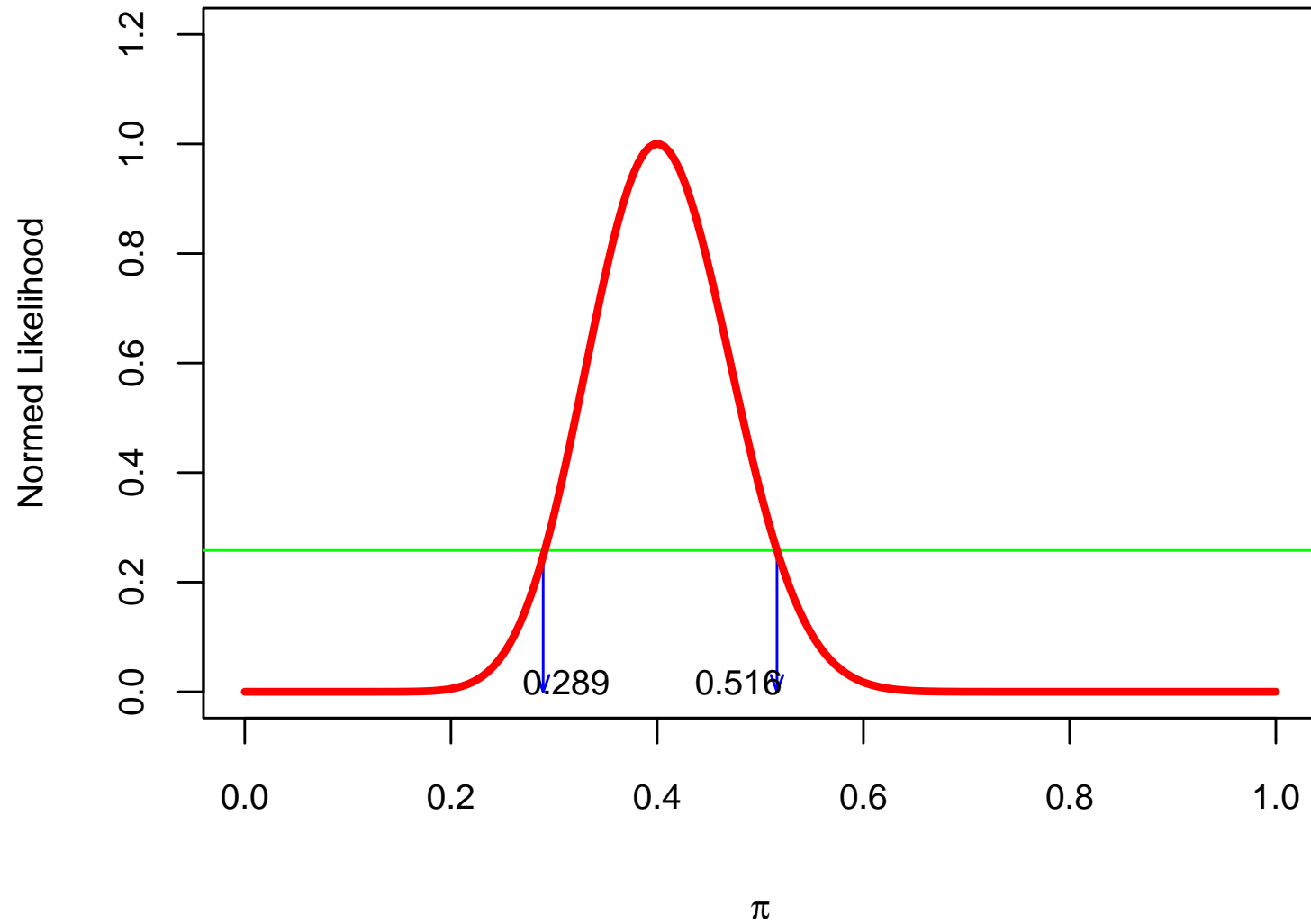
Likelihood Ratio



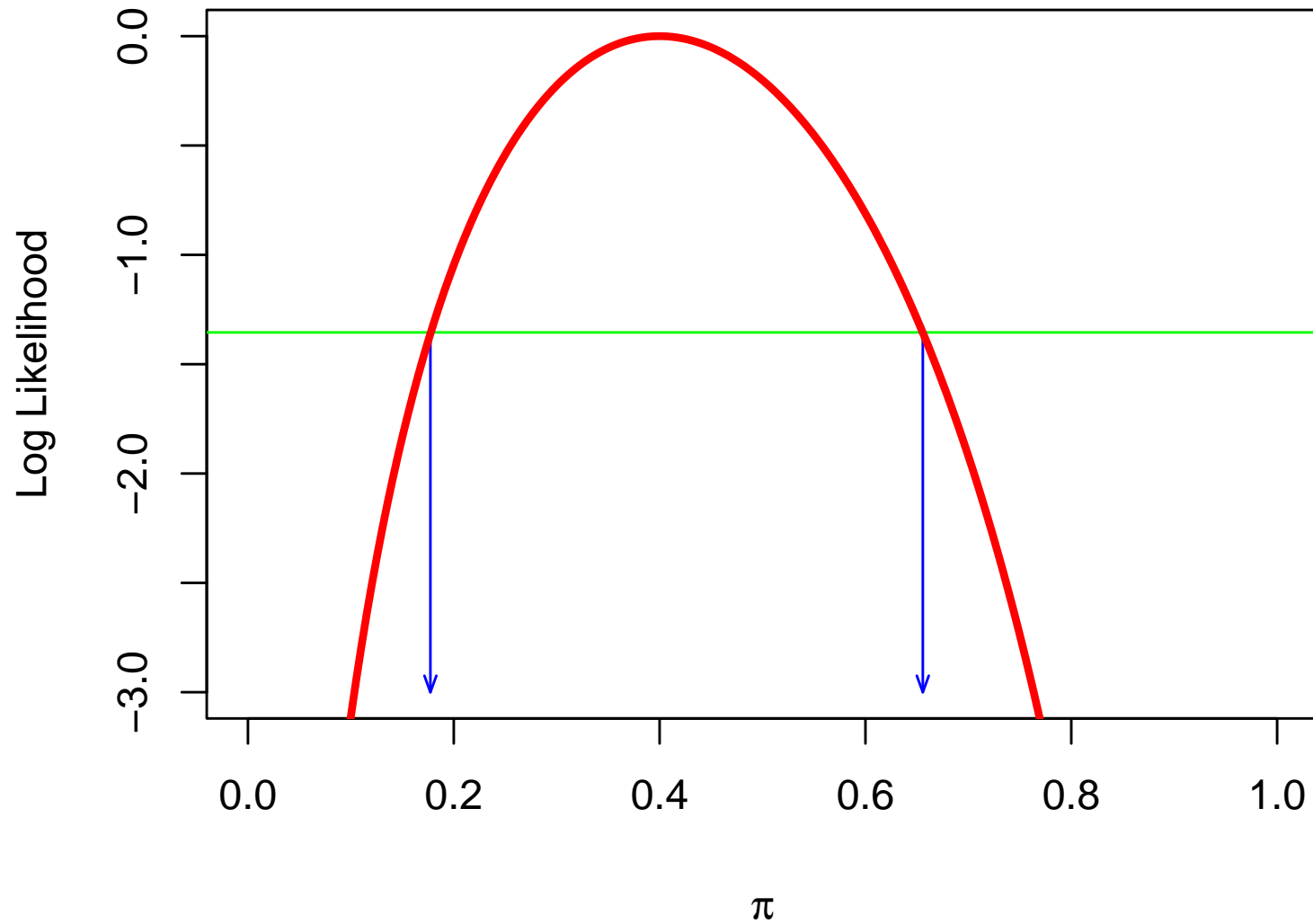
Likelihood Ratio



Likelihood Ratio



Log Likelihood



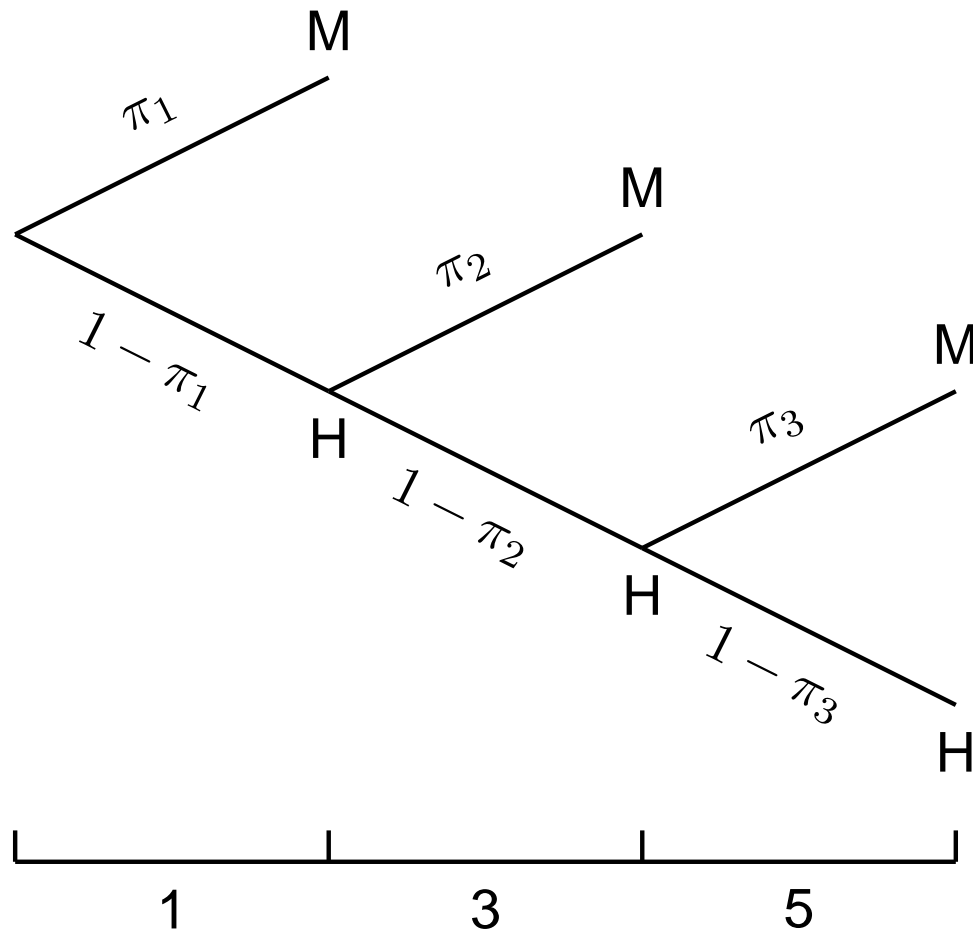
Konsep Inferensi

Frequentist interval (konfidensi) untuk suatu parameter (*confidence interval*) diinterpretasikan sebagai *coverage probability* nilai parameter terletak dalam interval, bila sampel (studi) diulang banyak kali (interpretasi frekuensi relatif)

Bayesian interval (konfidensi) untuk suatu parameter diinterpretasikan berdasarkan anggapan bahwa parameter adalah variabel random yang nilainya terletak dalam suatu *credible interval*

Likelihood interval (konfidensi) untuk suatu parameter disusun berdasarkan *likelihood ratio* yang menunjukkan seberapa besar kemungkinan suatu model didukung oleh data yang diperoleh

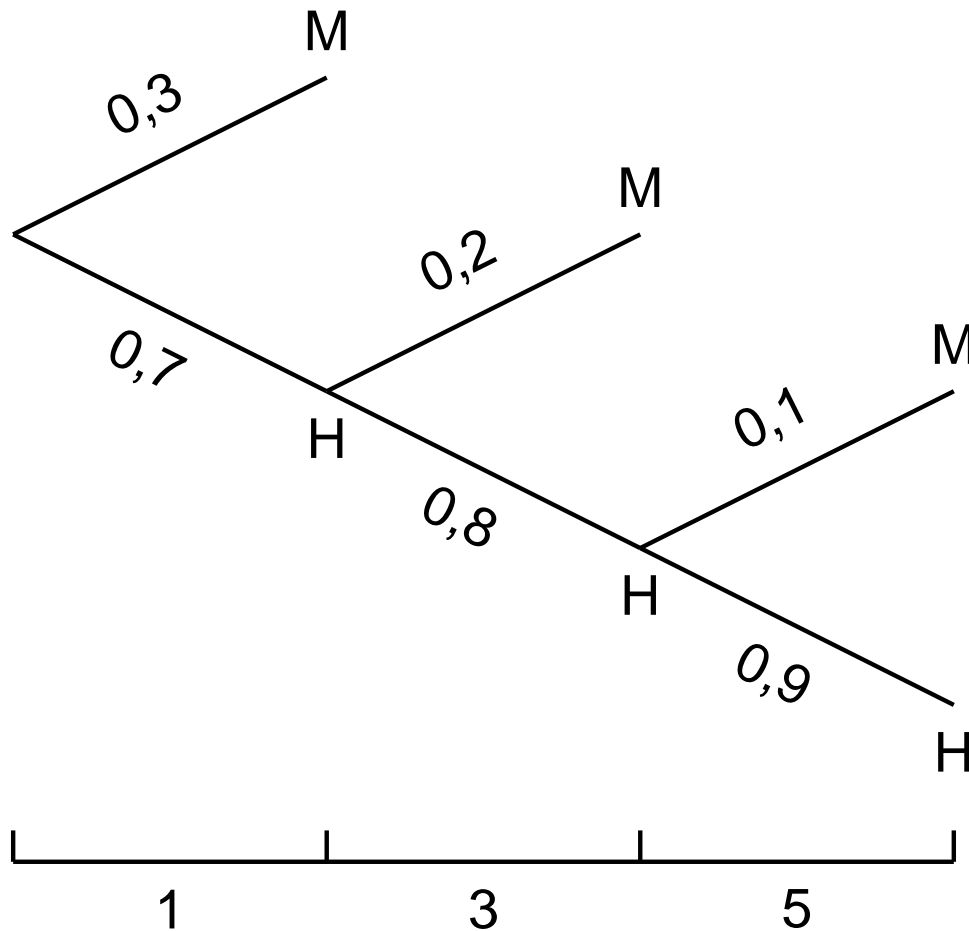
Likelihood untuk Insidensi



M: mati

H: hidup

Likelihood untuk Insidensi



M: mati

H: hidup

Probabilitas untuk 4 outcome :

0,3;

$0,7 \times 0,2$;

$0,7 \times 0,8 \times 0,1$;

$0,7 \times 0,8 \times 0,9$

***Likelihood* untuk Insidensi**

Untuk interval yang semakin sempit, probabilitas kondisional (untuk M) menjadi semakin kecil pula, dan konvergen ke *hazard rate* (*force of mortality*)

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

Likelihood untuk Insidensi

Poisson Likelihood

Likelihood untuk λ dapat diturunkan dari likelihood binomial dengan menganggap probabilitas sukses adalah λh dengan h kecil,

$$L(\lambda) = \lambda^D \exp(-\lambda Y)$$

dengan D adalah banyaknya kejadian, Y adalah total waktu observasi.

Log-likelihood untuk λ

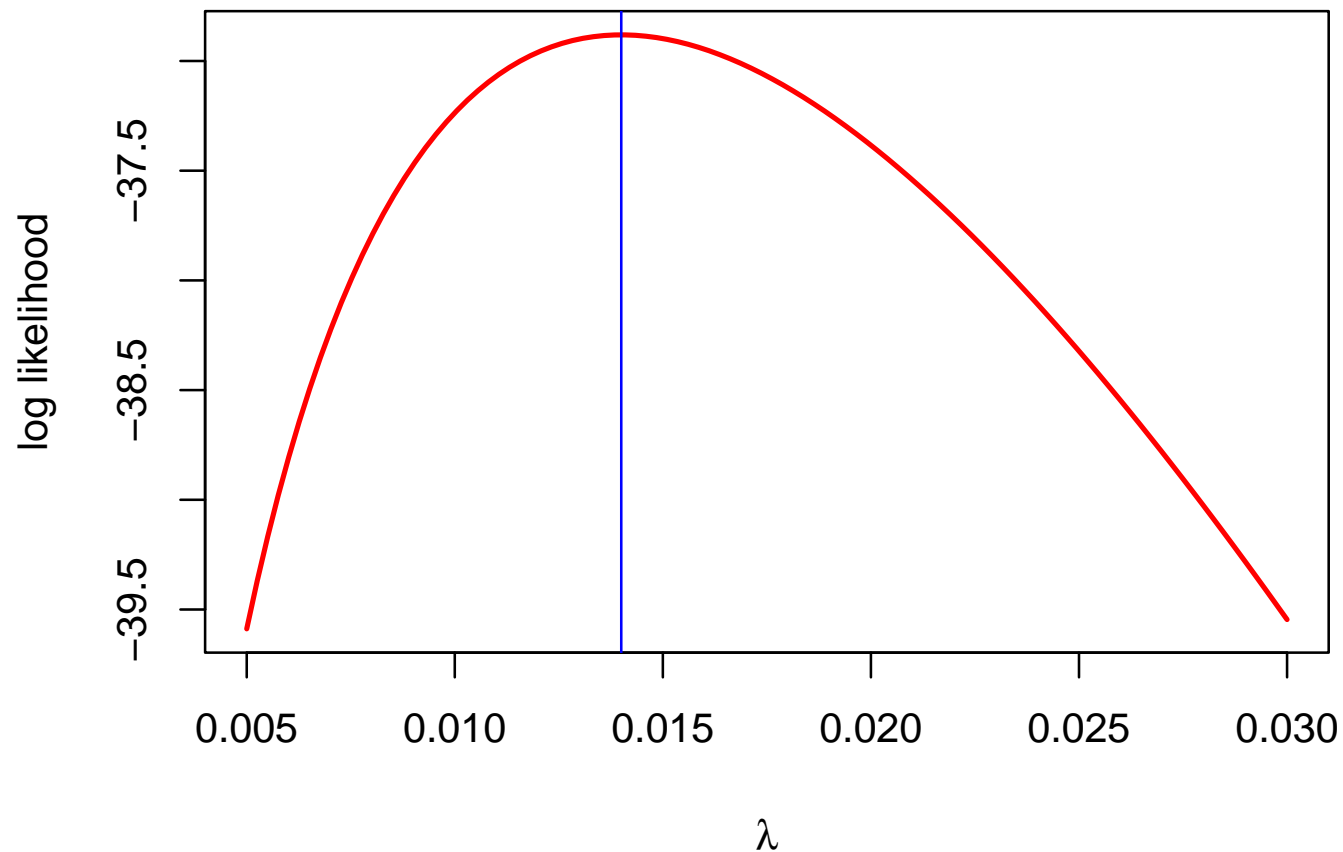
$$\ell(\lambda) = D \log(\lambda) - \lambda Y$$

Penduga untuk λ , $\hat{\lambda} = D/Y$

Likelihood untuk Insidensi

Misalkan ada 7 observasi dengan total waktu observasi 500 orang-tahun (*person-years*). Log-likelihood untuk λ

$$\ell(\lambda) = 7 \log(\lambda) - 500\lambda$$



Ukuran untuk Pengaruh Faktor

- Tidak hanya melihat *outcome* saja namun juga faktor (v. independen, v. penjelas, paparan/*exposure*)
- Bergantung pada desain penelitian

Ukuran untuk Pengaruh Faktor

- Selisih resiko (*risk difference*)
- Rasio resiko (*risk ratio*)
- *Odds ratio*

Ukuran untuk Pengaruh Faktor

Desain cohort

Data terobservasi

		D		
		1	2	
E				
1	n_{11}	n_{12}	N_1	
2	n_{21}	n_{22}	N_2	



Waktu

Model probabilitas

		D		
		1	2	
E				
1	π_1	$1 - \pi_1$	1	
2	π_2	$1 - \pi_2$	1	



Waktu

$$RD = \pi_1 - \pi_2$$

Ukuran untuk Pengaruh Faktor

Desain cohort

Data terobservasi

		<i>D</i>		
		1	2	
<i>E</i>				
1		n_{11}	n_{12}	N_1
2		n_{21}	n_{22}	N_2



Waktu

Model probabilitas

		<i>D</i>		
		1	2	
<i>E</i>				
1		π_1	$1 - \pi_1$	1
2		π_2	$1 - \pi_2$	1



Waktu

$$RR = \frac{\pi_1}{\pi_2}$$

Ukuran untuk Pengaruh Faktor

Desain cohort

Data terobservasi

		<i>D</i>		
		1	2	
<i>E</i>				
1	n_{11}	n_{12}	N_1	
2	n_{21}	n_{22}	N_2	



Waktu

Model probabilitas

		<i>D</i>		
		1	2	
<i>E</i>				
1	π_1	$1 - \pi_1$	1	
2	π_2	$1 - \pi_2$	1	



Waktu

$$OR = \frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)}$$

Inferensi untuk OR, RR dan RD

Estimasi titik untuk OR

$$\widehat{OR} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

alternatifnya, untuk menghindari masalah bila ada $n_{ij} = 0$

$$\widehat{OR} = \frac{(n_{11} + 0,5)(n_{22} + 0,5)}{(n_{12} + 0,5)(n_{21} + 0,5)}$$

galat standar (*standard error*) untuk $\log \widehat{OR}$

$$\hat{\sigma}(\log \widehat{OR}) = \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}} \right)^{1/2}$$

Interval konfidensi $(1 - \alpha)100\%$ untuk $\log \widehat{OR}$

$$\log \widehat{OR} \pm Z_{\alpha/2} \hat{\sigma}(\log \widehat{OR})$$

Inferensi untuk OR, RR dan RD

Estimasi titik untuk RD

$$\widehat{RD} = \hat{\pi}_1 - \hat{\pi}_2$$

galat standar (*standard error*) untuk $\hat{\pi}_1 - \hat{\pi}_2$

$$\sigma(\hat{\pi}_1 - \hat{\pi}_2) = \left[\frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2} \right]^{1/2}$$

Interval konfidensi $(1 - \alpha)100\%$ untuk $\pi_1 - \pi_2$

$$(\hat{\pi}_1 - \hat{\pi}_2) \pm Z_{\alpha/2} \hat{\sigma}(\hat{\pi}_1 - \hat{\pi}_2),$$

$\hat{\sigma}(\hat{\pi}_1 - \hat{\pi}_2)$ sama seperti $\sigma(\hat{\pi}_1 - \hat{\pi}_2)$ dengan π_i diganti $\hat{\pi}_i$

Inferensi untuk OR, RR dan RD

Estimasi titik untuk RR

$$\widehat{RR} = \frac{\hat{\pi}_1}{\hat{\pi}_2}$$

galat standar (*standard error*) untuk $\log \widehat{RR}$

$$\sigma(\log \widehat{RR}) = \left(\frac{1 - \pi_1}{\pi_1 n_1} + \frac{1 - \pi_2}{\pi_2 n_2} \right)^{1/2}$$

Interval konfidensi $(1 - \alpha)100\%$ untuk $\log RR$

$$\log \widehat{RR} \pm Z_{\alpha/2} \sigma(\log \widehat{RR})$$

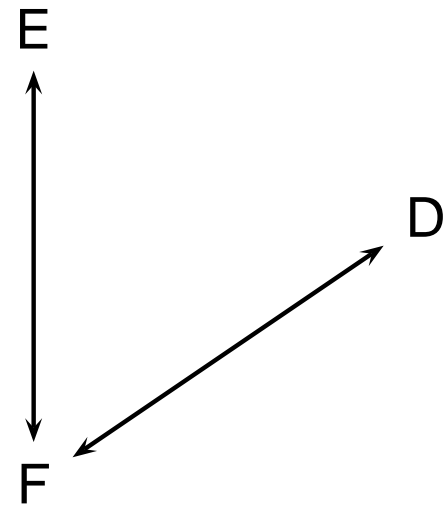
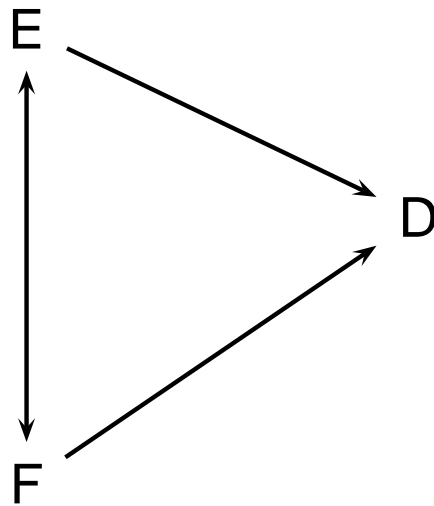
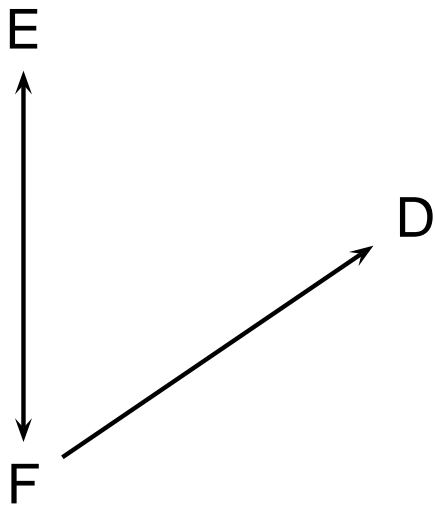
Perancuan (*confounding*)

Variable perancu adalah variabel yang memenuhi dua kondisi:

- merupakan faktor resiko
- mempunyai hubungan dengan variabel paparan tapi bukan merupakan konsekuensi dari variabel paparan

Perancuan (*confounding*)

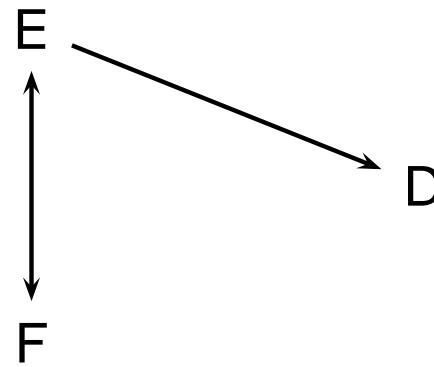
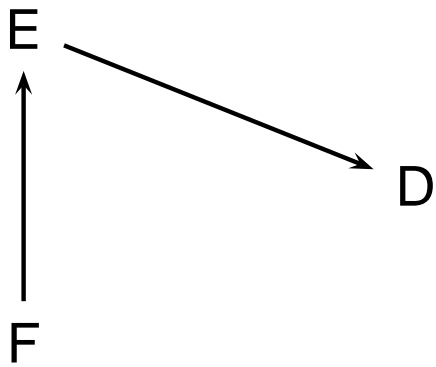
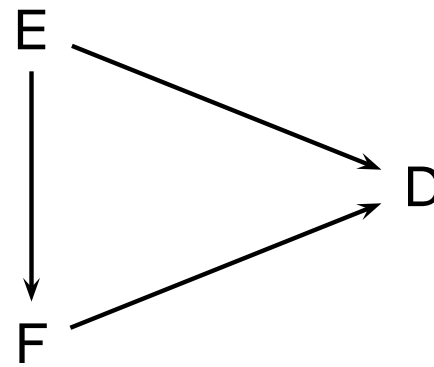
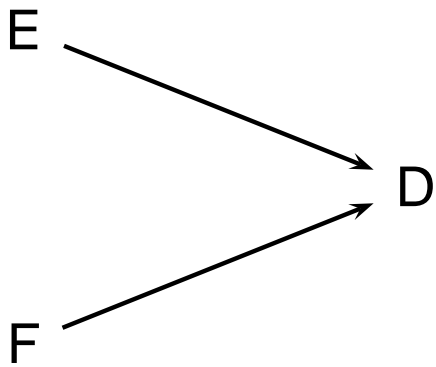
F adalah perancu antara D dengan E



D : variabel respon
E : variabel paparan
F : variabel perancu

Perancuan (*confounding*)

F bukan perancu antara D dengan E



Perancuan (*confounding*)

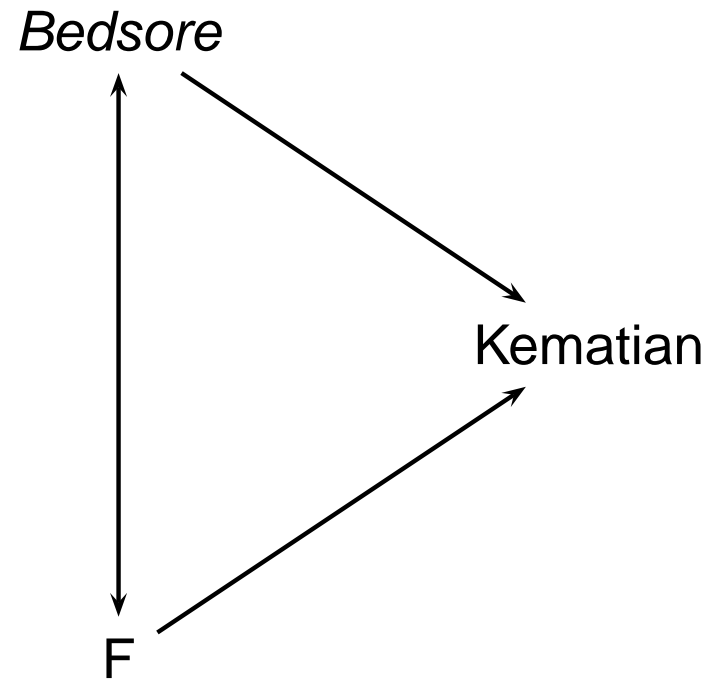
Bedsore study. Manula yang mengalami kecelakaan, seperti terjatuh, seringkali menjadi tidak dapat bangun dan bergerak dalam waktu lama. Hal ini dapat mengakibatkan *bedsores*, yaitu luka pada kulit yang dapat berlanjut ke otot dan tulang dan dapat berakibat fatal.

	Meninggal	hidup	Total
<i>Bedsore</i>	79	745	824
tidak <i>Bedsore</i>	286	8.290	8.576
Total	365	9.035	9.400

$$RR = \frac{79/824}{286/8576} = 2,9$$

Perancuan (*confounding*)

Adakah kemungkinan variabel perancu?



Perancuan (*confounding*)

Tingkat keparahan tinggi akibat penyakit lain

	Meninggal	hidup	Total
<i>Bedsore</i>	55	51	106
tidak <i>Bedsore</i>	5	5	10
Total	60	56	116

$$RR = \frac{55/106}{5/10} = 1,04$$

Tingkat keparahan rendah akibat penyakit lain

	Meninggal	hidup	Total
<i>Bedsore</i>	24	694	718
tidak <i>Bedsore</i>	281	8.285	8.566
Total	305	8.979	9.284

$$RR = \frac{24/718}{281/8566} = 1,02$$

Metode Analisis

Respon	Variabel penjelas	Metode
Kontinu	Biner Nominal, 2 kategori atau lebih Ordinal Kontinu Nominal dan kontinu Kategorik dan kontinu	t-test, z-test ANOVA ANOVA Regresi Ganda Analisis Kovariansi Regresi Ganda
Biner	Kategorik Kontinu Kategorik dan kontinu	Tabel kontingensi Regresi Logistik Regresi Logistik, probit atau model dose-response Regresi Logistik
Nominal, 2 kategori atau lebih	Nominal Kategorik dan kontinu	Tabel kontingensi Regresi Logistik Nominal
Ordinal	Kategorik dan kontinu	Regresi Logistik Ordinal
Cacah	Kategorik Kategorik dan kontinu	Model Log-linear, Regresi Poisson Regresi Poisson

Metode Analisis

Respon	Variabel penjelas	Metode
Durasi (survival)	Biner Kategorik dan kontinu	Log-rank test Survival analysis
Respon berkorelasi	Kategorik dan kontinu	Generalized Estimating equation Multilevels model Analisis Data Longitudinal Analisis Data Panel

Metode Analisis

Masih banyak variasi metode yang belum disebutkan;

masih banyak masalah yang memerlukan pengembangan metode baru atau modifikasi metode!

Metode Analisis

Langkah-langkah umum Analisis statistik:

1. Deskripsi data
2. Spesifikasi model
3. Estimasi parameter model
4. Uji Kecocokan Model
5. Inferensi

Sensitivity, Specificity & Predictive Value

Dalam diagnosa atau *screening*,

$T+$: diagnosa atau *screening* menunjukkan tes positif

$T-$: diagnosa atau *screening* menunjukkan tes negatif

$D+$: kenyataannya positif ada penyakit

$D-$: kenyataannya tidak ada penyakit (negatif)

$$\textit{Sensitivity} = P(T+ | D+)$$

$$\textit{Specificity} = P(T- | D-)$$

$$\textit{Predictive Value}+ = P(D+ | T+)$$

$$\textit{Predictive Value}- = P(D- | T-)$$

Sensitivity, Specificity & Predictive Value

Menggunakan Teorema Bayes:

$$\begin{aligned} \text{Predictive Value}_+ &= P(D+ | T+) \\ &= \frac{P(D+ \cap T+)}{P(T+)} \\ &= \frac{P(D+)P(T+ | D+)}{P(D+)P(T+ | D+) + P(D-)P(T+ | D-)} \\ &= \frac{\text{Prevalence} \times \text{Sensitivity}}{\text{prev.} \times \text{sens.} + (1 - \text{prev.}) \times (1 - \text{spec.})} \end{aligned}$$

Sensitivity, Specificity & Predictive Value

Menggunakan Teorema Bayes:

$$\begin{aligned} \text{Predictive Value-} &= P(D- | T-) \\ &= \frac{P(D- \cap T-)}{P(T-)} \\ &= \frac{P(D-)P(T- | D-)}{P(D-)P(T- | D-) + P(D+)P(T- | D+)} \\ &= \frac{(1 - \text{Prevalence}) \times \text{Specificity}}{(1 - \text{prev.}) \times \text{spec.} + \text{prev.} \times (1 - \text{sens.})} \end{aligned}$$

Sensitivity, Specificity & Predictive Value

Contoh:

Suatu tes sitologi (*cytological test*) dilakukan untuk *screening* kanker rahim pada wanita. Diperoleh data 24.103 wanita yang terdiri atas 379 wanita yang diketahui sudah menderita kanker rahim (dengan tes yang dianggap sebagai *gold standar*):

	$T-$	$T+$	Total
$D-$	23.362	362	23.724
$D+$	225	154	379

Hitung *sensitivity* dan *specificity* tes tersebut!

Sensitivity, Specificity & Predictive Value

	$T-$	$T+$	Total
$D-$	23.362	362	23.724
$D+$	225	154	379

$$\text{sensitivity} = \frac{154}{379} = 0,406 = 40,6\%$$

$$\text{specificity} = \frac{23.362}{23.724} = 0,985 = 98,5\%$$

Interpretasi:

- Jika tes digunakan untuk wanita yang tidak menderita kanker rahim, tes hampir pasti akan negatif (*specificity* = 98,5% cukup besar)
- Jika tes digunakan untuk wanita yang menderita kanker rahim, peluang tidak terdeteksi besar (*sensitivity* = 40,6 % rendah; *false negatif* 59,4%)

Sensitivity, Specificity & Predictive Value

Predictive Value + :

$$\begin{aligned} PV+ &= \frac{\textit{Prevalence} \times \textit{Sensitivity}}{\textit{prev.} \times \textit{sens.} + (1 - \textit{prev.}) \times (1 - \textit{spec.})} \\ &= \frac{\textit{Prevalence} \times 0,406}{\textit{prev.} \times 0,406 + (1 - \textit{prev.}) \times (1 - 0,985)} \end{aligned}$$

Predictive Value - :

$$\begin{aligned} PV- &= \frac{(1 - \textit{Prevalence}) \times \textit{Specificity}}{(1 - \textit{prev.}) \times \textit{spec.} + \textit{prev.} \times (1 - \textit{sens.})} \\ &= \frac{(1 - \textit{Prevalence}) \times 0,985}{(1 - \textit{prev.}) \times 0,985 + \textit{prev.} \times (1 - 0,406)} \end{aligned}$$

Sensitivity, Specificity & Predictive Value

PV+ dan *PV-* untuk berbagai nilai prevalensi dengan *spec*=98,5% dan *sens*=40,6%

prevalensi	<i>PV+</i>	<i>PV-</i>
0,0010	0,0264	0,999
0,0157	0,3015	0,990
0,0500	0,5876	0,969
0,1000	0,7505	0,937
0,5000	0,9644	0,624

Kurva ROC

- ROC = *receiver operating characteristic*
- berguna untuk respon diagnosis (*screening test*) yang kontinu atau mempunyai lebih dari dua nilai (jenis respon)
- menghubungkan *sensitivity* dengan $1 - \textit{specificity}$
- Area di bawah kurva ROC dapat digunakan untuk menilai keakuratan suatu diagnosis (tes)

Kurva ROC

Diketahui probabilitas skor CT *image* (*computed tomographic image*) untuk pasien syaraf oleh seorang radiolog adalah sebagai berikut:

Status	Skor dari radiolog *					
	(1)	(2)	(3)	(4)	(5)	
Penyakit (D)						
Normal ($D-$)	0,303	0,055	0,055	0,101	0,018	0,532
Abnormal ($D+$)	0,028	0,018	0,018	0,101	0,303	0,468

* (1) hampir pasti normal; (2) mungkin normal; (3) tidak dapat ditentukan
(4) mungkin abnormal; (5) hampir pasti abnormal

- Nilai probabilitas pada tiap sel adalah irisan peristiwa $D+$, $D-$ dengan $T_{(1)}, T_{(2)}, T_{(3)}, T_{(4)}, T_{(5)}$, misalnya $P(D+ \cap T_{(1)}) = 0,028$, $P(D- \cap T_{(5)}) = 0,018$
- $P(D+) = 0,468$ adalah prevalensi

Kurva ROC

Diketahui probabilitas skor CT *image* (*computed tomographic image*) untuk pasien syaraf oleh seorang radiolog adalah sebagai berikut:

Status	Skor dari radiolog *					
	(1)	(2)	(3)	(4)	(5)	
Penyakit (D)						
Normal ($D-$)	0,303	0,055	0,055	0,101	0,018	0,532
Abnormal ($D+$)	0,028	0,018	0,018	0,101	0,303	0,468

* (1) hampir pasti normal; (2) mungkin normal; (3) tidak dapat ditentukan
(4) mungkin abnormal; (5) hampir pasti abnormal

Kriteria tes positif: skor ≥ 2

$$sens = (0,018 + 0,018 + 0,101 + 0,303) / 0,468 = 0,94$$

$$spec = 0,303 / 0,532 = 0,57$$

Kurva ROC

Diketahui probabilitas skor CT *image* (*computed tomographic image*) untuk pasien syaraf oleh seorang radiolog adalah sebagai berikut:

Status	Skor dari radiolog *					
	(1)	(2)	(3)	(4)	(5)	
Penyakit (D)						
Normal ($D-$)	0,303	0,055	0,055	0,101	0,018	0,532
Abnormal ($D+$)	0,028	0,018	0,018	0,101	0,303	0,468

* (1) hampir pasti normal; (2) mungkin normal; (3) tidak dapat ditentukan
(4) mungkin abnormal; (5) hampir pasti abnormal

Kriteria tes positif: skor ≥ 3

$$sens = (0,018 + 0,101 + 0,303) / 0,468 = 0,90$$

$$spec = (0,303 + 0,303) / 0,532 = 0,67$$

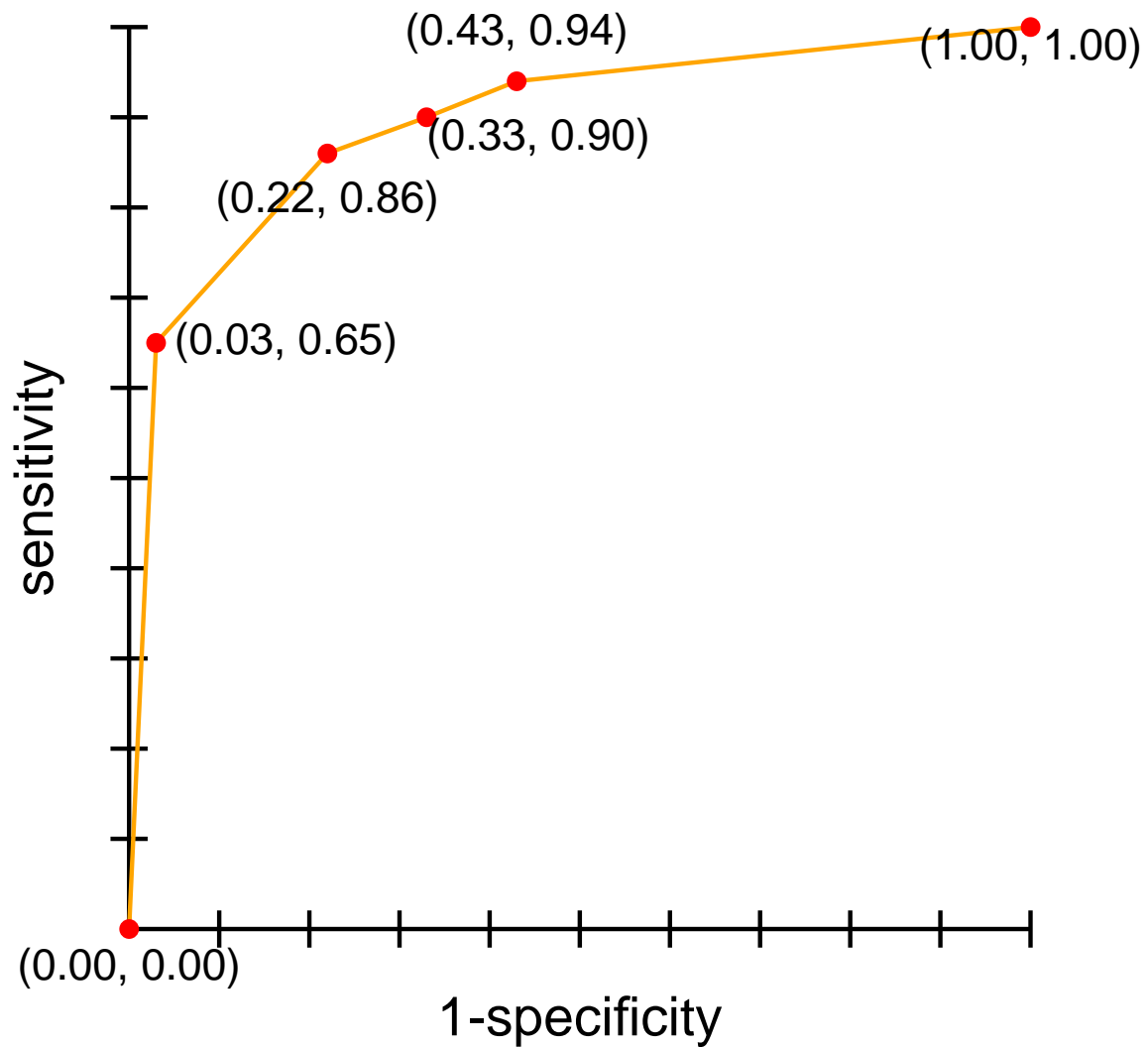
Kurva ROC

Sensitivity dan *Specificity* berdasarkan beberapa kriteria tes positif.

kriteria tes positif	<i>sensitivity</i>	<i>specificity</i>	1- <i>specificity</i>
1 \leq skor	1,00	0,00	1,00
2 \leq skor	0,94	0,57	0,43
3 \leq skor	0,90	0,67	0,33
4 \leq skor	0,86	0,78	0,22
5 \leq skor	0,65	0,97	0,03
5 < skor	0,00	1,00	0,00

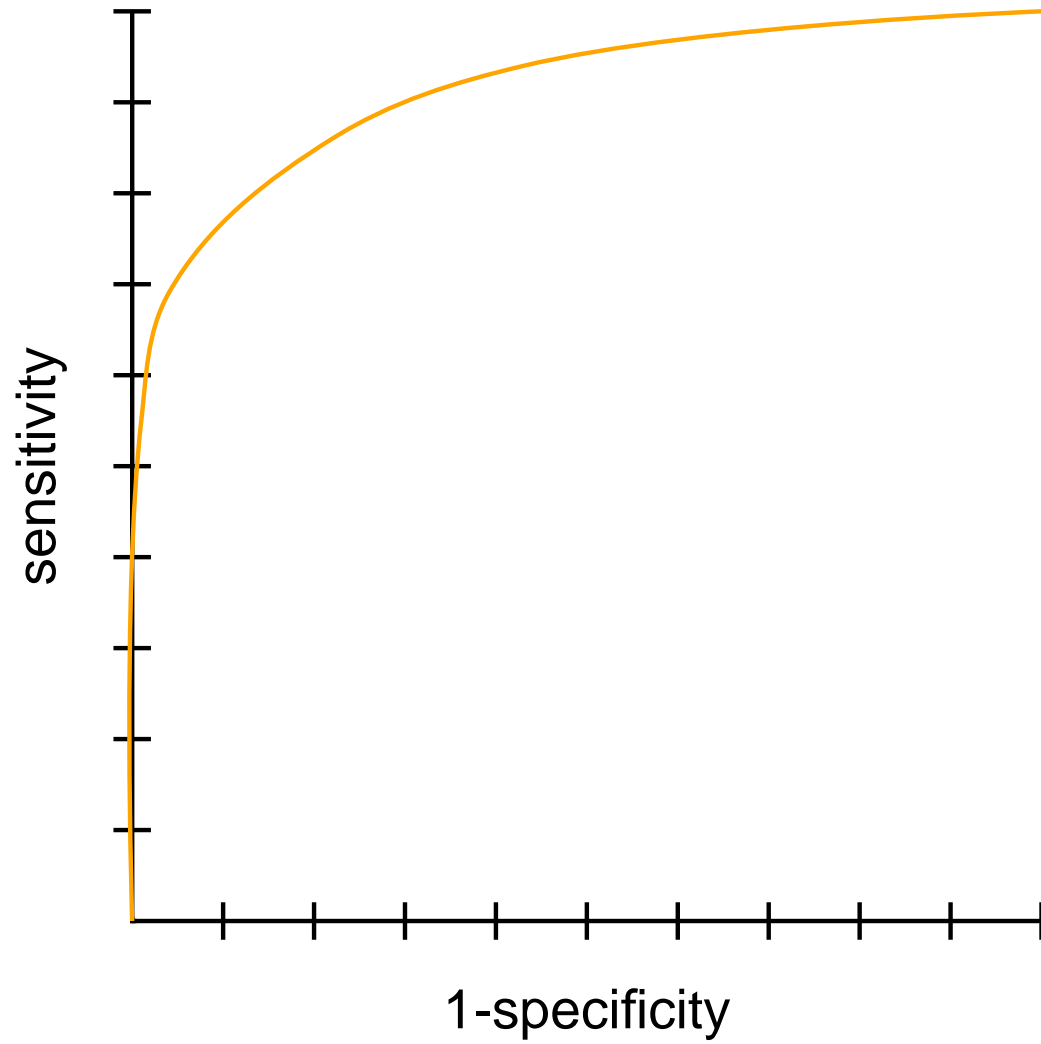
Kurva ROC

Kurva ROC



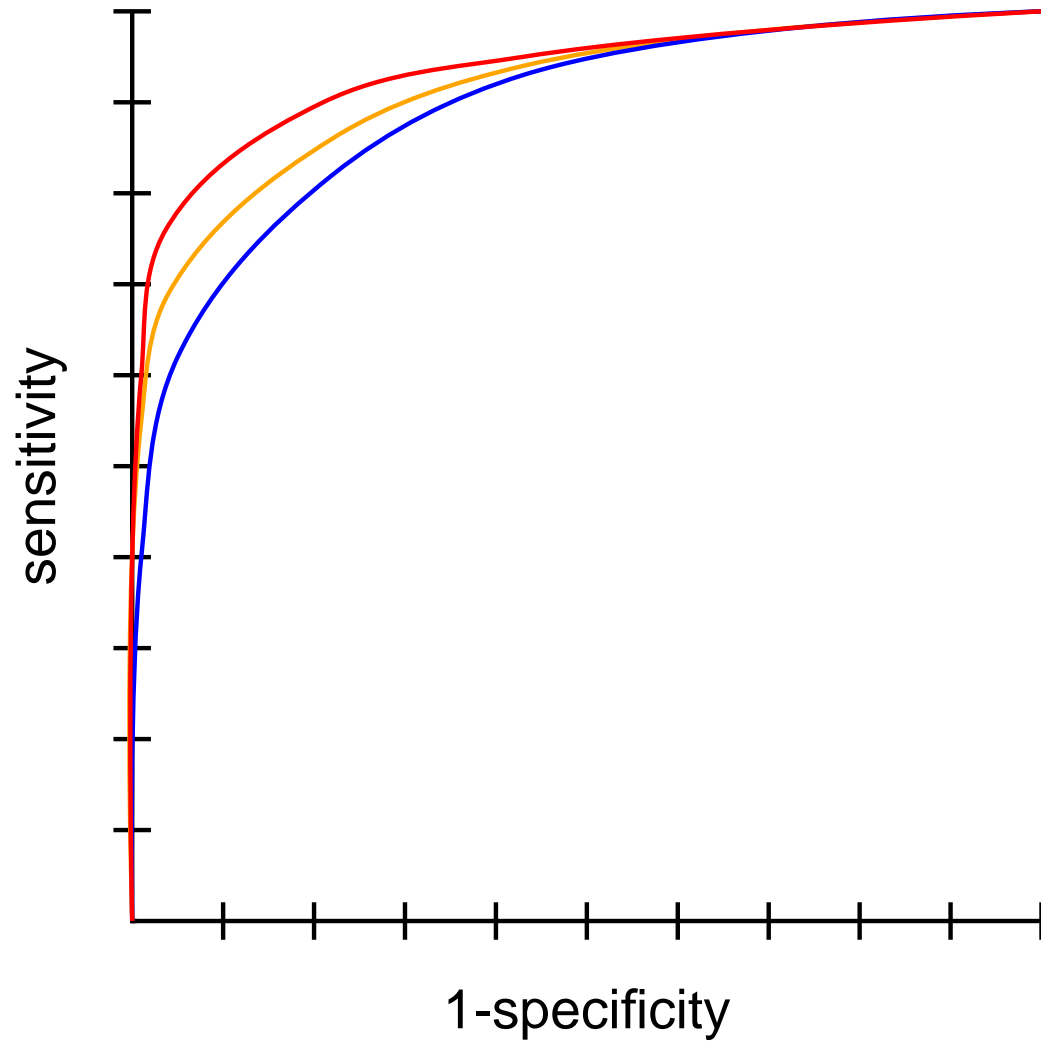
Kurva ROC

Kurva ROC



Kurva ROC

Kurva ROC



Regresi Logistik

Data studi tentang hubungan Penyakit jantung koroner dengan Tekanan pekerjaan:

Tertekan krn. Pekerjaan	Penyakit jantung koroner		Total
	Ya	Tidak	
Ya	97	307	404
Tidak	200	1409	1609

Risk ratio:

$$\begin{aligned}\widehat{RR} &= \frac{97/404}{200/1609} \\ &= 1,932\end{aligned}$$

Regresi Logistik

Menggunakan R:

```
> twoby2(dt$E,dt$D)
```

```
2 by 2 table analysis:
```

```
-----  
Outcome      : Y
```

```
Comparing    : Y vs. T
```

	Y	T	P(Y)	95% conf. interval
Y	97	307	0.2401	0.2009 0.2842
T	200	1409	0.1243	0.1091 0.1413

	95% conf. interval
Relative Risk:	1.9316 1.5554 2.3987
Sample Odds Ratio:	2.2259 1.6956 2.9221
Conditional MLE Odds Ratio:	2.2249 1.6754 2.9429
Probability difference:	0.1158 0.0674 0.1693

```
Exact P-value: 0
```

```
Asymptotic P-value: 0
```

Regresi Logistik

Y_i adalah variabel random Bernoulli untuk individu i

$$P(Y_i = y_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i}, \quad y_i = 0, 1$$

Setiap individu i mempunyai karakteristik berupa kovariat x_i yang mempengaruhi π_i dalam bentuk

$$\pi_i = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_i))}$$

Secara umum π_i adalah fungsi logistik

$$f(Z) = \frac{1}{1 + e^{-Z}}, \text{ atau } f(Z) = \frac{e^Z}{1 + e^Z}$$

dengan $Z = \beta_0 + \beta_1 x_1 + \beta_1 x_1 + \dots + \beta_p$ adalah fungsi linear dari p variabel penjelas.

Regresi Logistik

Model

$$\log \frac{\pi_i}{1 - \pi_i} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{pi}$$

atau

$$\text{logit}(\pi) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{pi}$$

Estimasi $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ dapat diperoleh dengan MLE untuk fungsi likelihood

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n P(Y_i = y_i) \\ &= \frac{[\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{pi})]^{y_i}}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{pi})} \end{aligned}$$

Regresi Logistik

Interpretasi parameter
Untuk model

$$\begin{aligned}\text{logit}(\pi_i) &= \beta_0 + \beta_1 x_i \\ \frac{\pi_i}{1 - \pi_i} &= \exp[\beta_0 + \beta_1 x_i] \\ \text{odds}_{x_i} &= \exp[\beta_0 + \beta_1 x_i]\end{aligned}$$

dengan

$$x_i = \begin{cases} 0 & i \text{ tdk terpapar} \\ 1 & i \text{ terpapar} \end{cases}$$

Sehingga

$$\text{OR} = \frac{\text{odds}_1}{\text{odds}_0} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

Regresi Logistik

Estimasi parameter dan OR

Program statistika seperti R, SPSS, Epi-Info, STATA menyediakan fasilitas untuk estimasi $\hat{\beta}$.

Estimasi titik dan interval konfidensi $(1 - \alpha)100\%$ untuk OR:

$$\widehat{OR} = \exp(\hat{\beta})$$

$$\exp(\hat{\beta} \pm Z_{\alpha/2} \text{SE}(\hat{\beta}))$$

Regresi Logistik

Contoh dengan R:

```
> m<-glm(D~E,family=binomial(link=logit), data=dt)
> round(ci.logistik(m),digits=3)
      coef.p  s.err      L      U  ecoef.p      eL      eU
(Intercept) -1.952 0.076 -2.100 -1.804   0.142 0.122 0.165
E            0.800 0.139  0.528  1.072   2.226 1.696 2.922
```

Diperoleh interval konfidensi untuk OR 2,226 (1,696 – 2,922)

RR dan RD ? Gunakan estimasi probabilitas $\pi(x)$ dari model regresi.

Probabilitas mendapatkan penyakit jantung untuk individu yang terpapar $P(y_i = 1 \mid x_i = 1)$ adalah

```
> predict(m,newdata=data.frame(E=1),type="response")
[1] 0.240099
```

Regresi Logistik

Penggunaan Regresi Logistik berdasarkan desain tertentu

- Dapat digunakan untuk menghitung RR, RD, OR dalam desain cohort (probabilitas tiap-tiap individu, $\hat{\pi}(x_i)$, dapat diestimasi)
- Hanya dapat digunakan untuk menghitung OR dalam desain case-control

Regresi Poisson

Distribusi Poisson

$$P(X = x) = \frac{\theta^x e^{-\theta}}{x!}, \quad x = 0, 1, 2, \dots$$

mempunyai mean dan variansi θ

Regresi Poisson

Contoh:

Untuk menyelidiki infeksi pada suatu populasi organisme tertentu, sering tidak mungkin untuk meneliti tiap-tiap individu. Organisme tersebut dibagi dalam kelompok-kelompok dan kelompok tersebut dianggap sebagai unit.

N = banyaknya organisme

n = banyaknya kelompok

m = banyaknya organisme tiap kelompok, $N = nm$ (dengan menganggap m sama untuk tiap kelompok)

Misalnya X adalah banyaknya organisme yang tidak terinfeksi, variabel random X kemungkinan besar dapat dimodelkan dengan Poisson,

Regresi Poisson

Data:

y_i banyaknya observasi cacah pada unit i ; s_i ukuran tiap unit i ; dan karakteristik tiap unit (kovariat) \mathbf{x}_i , $i = 1, 2, \dots, n$

Model:

$$\begin{aligned} E(Y_i | X_i) = \mu_i &= s_i \lambda(x_i) \\ &= s_i \exp(\beta_0 + \beta_1 x_i), \quad \text{atau} \\ \log \mu_i &= \log s_i + \beta_0 + \beta_1 x_i \end{aligned}$$

dengan $\lambda(x_i)$ dinamakan resiko unit i .

s_i dapat berupa:

- banyaknya anggota populasi
- interval waktu
- luasan
- *exposure time*

Regresi Poisson

Dengan asumsi $Y_i \sim \text{Poisson}$, diperoleh fungsi likelihood:

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n P(Y_i = y_i) \\ &= \prod_{i=1}^n \frac{[s_i \lambda(x_i)]^{y_i} \exp[-s_i \lambda(x_i)]}{y_i!} \end{aligned}$$

Dapat digunakan beberapa program statistika seperti R, STATA, SAS untuk estimasi β

Regresi Poisson

Interpretasi parameter:
Untuk model

$$\log \mu_i = \log s_i + \beta_0 + \beta_1 x_i$$

dengan

$$x_i = \begin{cases} 0 & i \text{ tdk terpapar} \\ 1 & i \text{ terpapar} \end{cases}$$

Sehingga

$$\begin{aligned} \text{RR} &= \frac{E(Y_i | X_i = 1)}{E(Y_i | X_i = 0)} \\ &= \frac{s_i \exp(\beta_0 + \beta_1)}{s_i \exp(\beta_0)} \\ &= e^{\beta_1} \end{aligned}$$

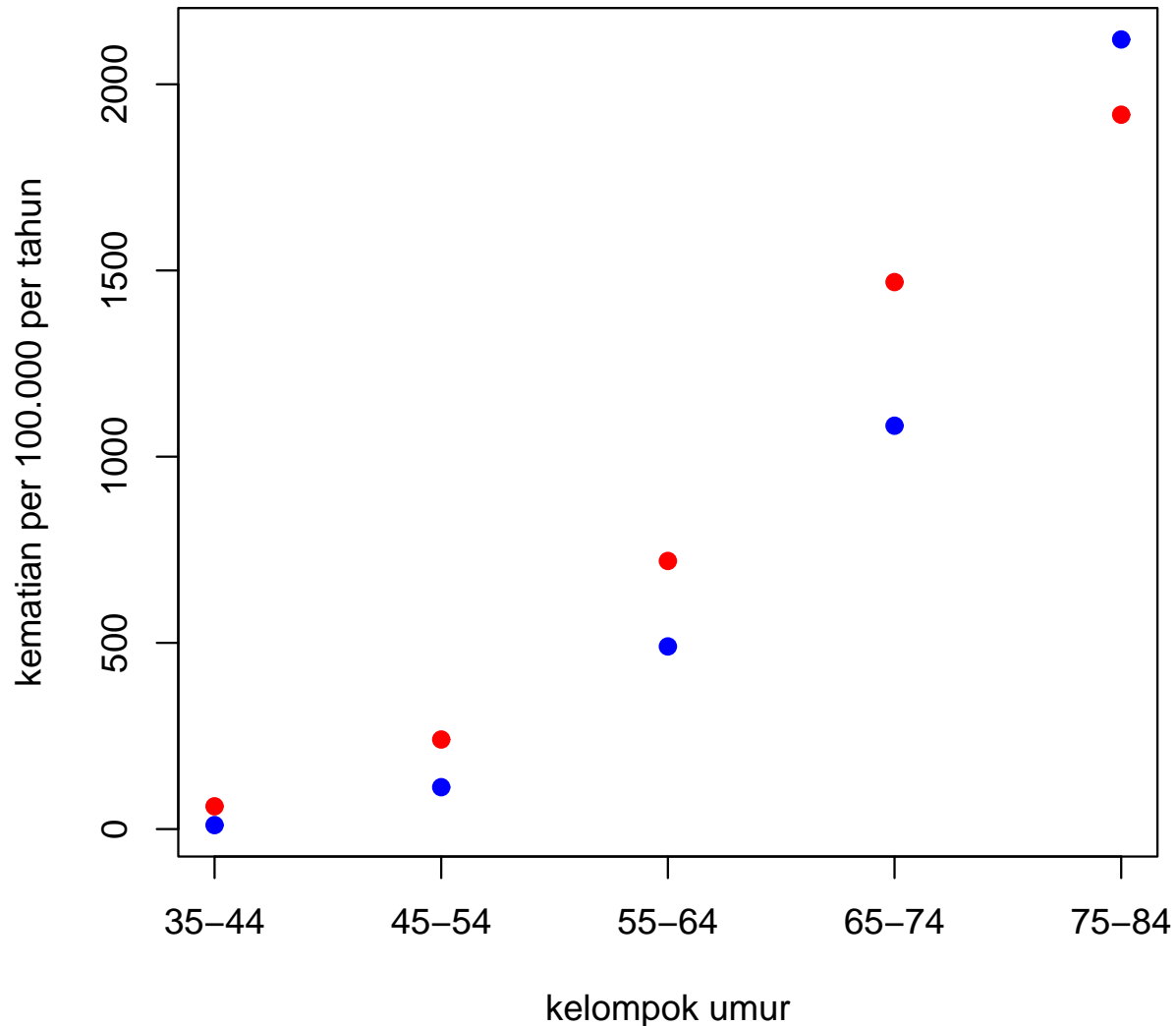
Regresi Poisson

Data diperoleh dari studi awal tentang akibat buruk merokok bagi kesehatan pada tahun 1951. Kematian akibat penyakit jantung koroner dikategorikan menurut umur dan status merokok.

Kel. Umur	perokok		bukan perokok	
	kematian	<i>person-years</i>	kematian	<i>person-years</i>
35 – 44	32	52407	2	18790
45 – 54	104	43248	12	10673
55 – 64	206	28612	28	5710
65 – 74	186	12663	28	2585
75 – 84	102	5317	31	1462

Regresi Poisson

Tingkat kematian akibat penyakit jantung koroner per 100.000 *person-years* untuk perokok ● dan bukan perokok ●



Regresi Poisson

Model 1:

$$\log \mu_i = \log(s_i) + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} \times x_{2i} + \beta_4 x_{1i}^2, \quad i = 1, \dots, 10$$

dengan

- μ_i : mean dari kematian
- s_i : *person-years*
- x_{1i} : perokok atau bukan;
- x_{2i} : usia 1, 2, 3, 4, 5 ;
- $x_{1i} \times x_{2i}$: interaksi (hasil kali) antara x_{1i} dengan x_{2i} ;
- x_{1i}^2 : kuadrat umur

Regresi Poisson

Model 1:

Parameter	Estimasi	SE	RR	Int-konf. 95% RR
β_0	-10,79	0,450		
β_1	1,44	0,372	4,22	2,04 – 8,76
β_2	2,37	0,207	10,77	7,16 – 16,18
β_3	-0,19	0,027	0,82	0,78 – 0,87
β_4	-0,30	0,097	0,74	0,61 – 0,89

Regresi Poisson

Model 2:

$$\begin{aligned}\log \mu_i &= \log(s_i) + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \\ &\quad \beta_6 x_{1i} x_{2i} + \beta_7 x_{1i} x_{3i} + \beta_8 x_{1i} x_{4i} + \beta_9 x_{1i} x_{5i} \\ &\quad i = 1, 2, \dots, 10\end{aligned}$$

dengan

- μ_i : mean dari kematian
- s_i : *person-years*
- x_{1i} : perokok atau bukan;
- $x_{ki}, k = 2, 3, \dots, 5$: kelompok umur 35 – 44, 45 – 54, ..., 75 – 84
- $x_{1i} x_{ki}, h = 2, 3, \dots, 5$: interaksi (hasil kali) antara x_{1i} dengan kelompok umur x_{ki}

Regresi Poisson

Model 2:

Model 1:

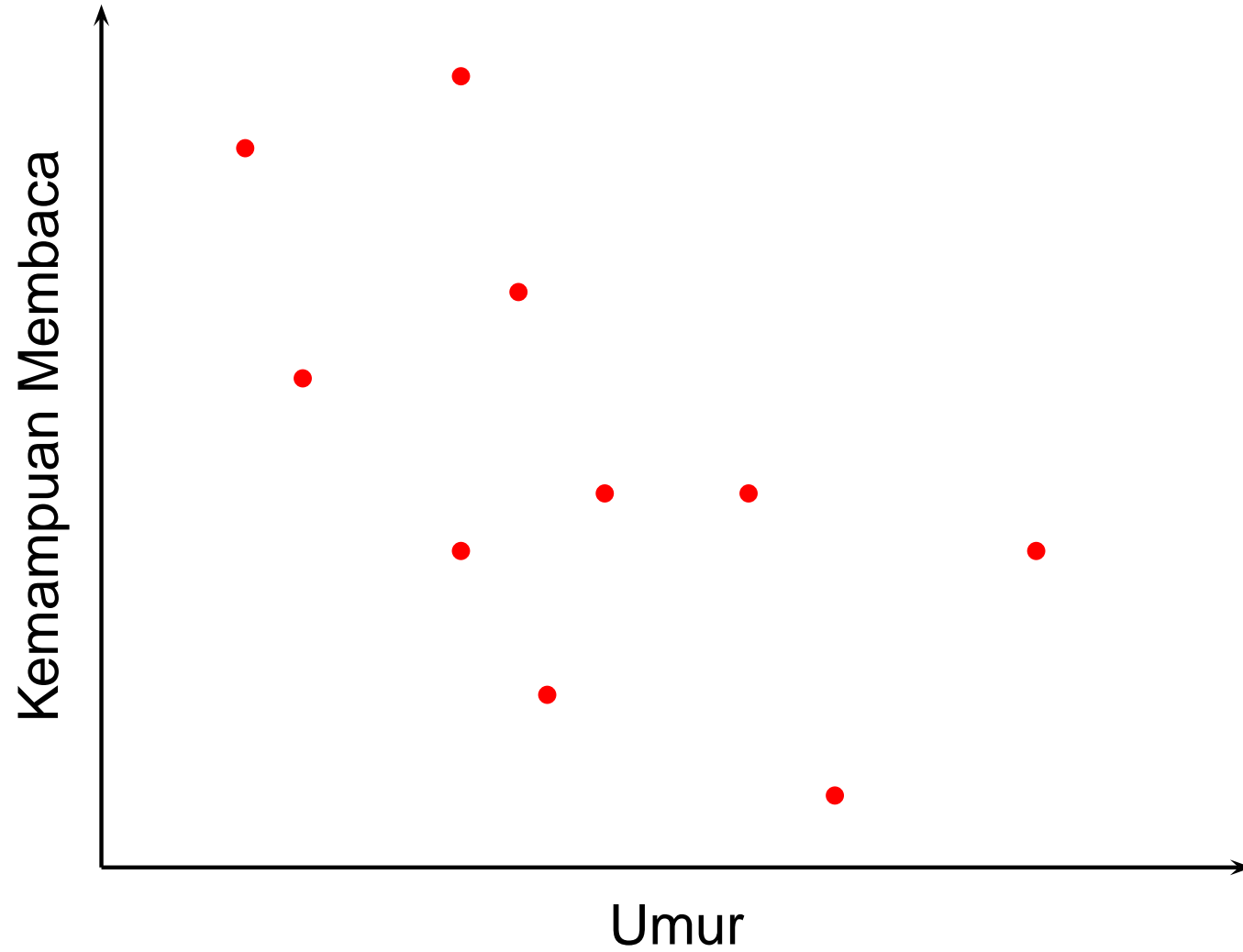
Parameter	Estimasi	SE	RR	Int-konf 95% RR
β_0	-9,15	0,71	0,00	0,00 – 0,00
β_1	1,75	0,73	5,74	1,37 – 23,94
β_2	2,36	0,76	10,56	2,36 – 47,20
β_3	3,83	0,73	46,07	10,97 – 193,39
β_4	4,62	0,73	101,76	24,24 – 427,18
β_5	5,29	0,73	199,21	47,68 – 832,36
β_6	-0,99	0,79	0,37	0,08 – 1,75
β_7	-1,36	0,76	0,26	0,06 – 1,13
β_8	-1,44	0,76	0,24	0,05 – 1,04
β_9	-1,85	0,76	0,16	0,04 – 0,70

Data Longitudinal

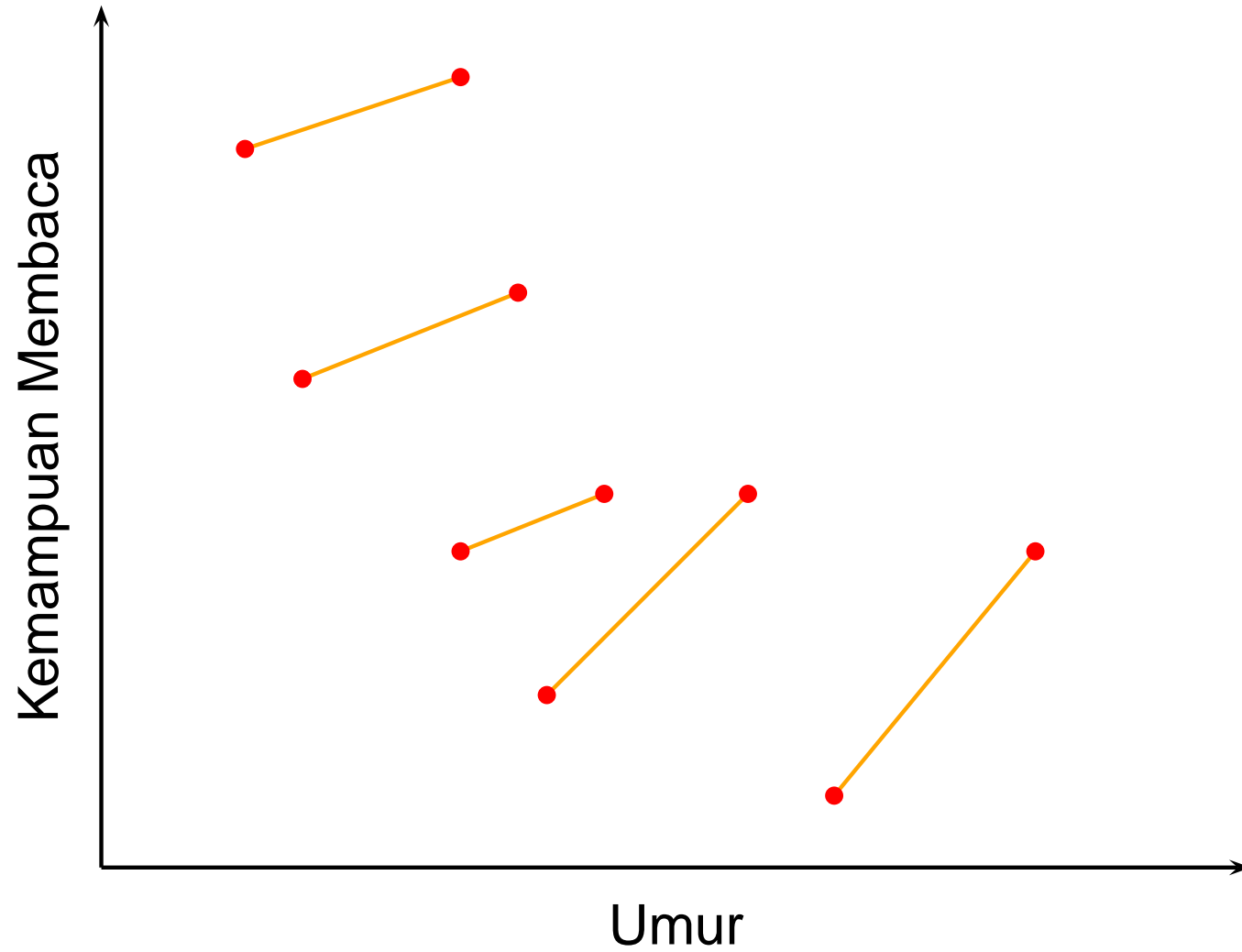
Karakteristik data longitudinal

- Individu (subyek, unit sampel) diamati dalam suatu periode waktu tertentu lebih dari satu kali
- Pengukuran berulang pada suatu individu (subyek, unit sampel)

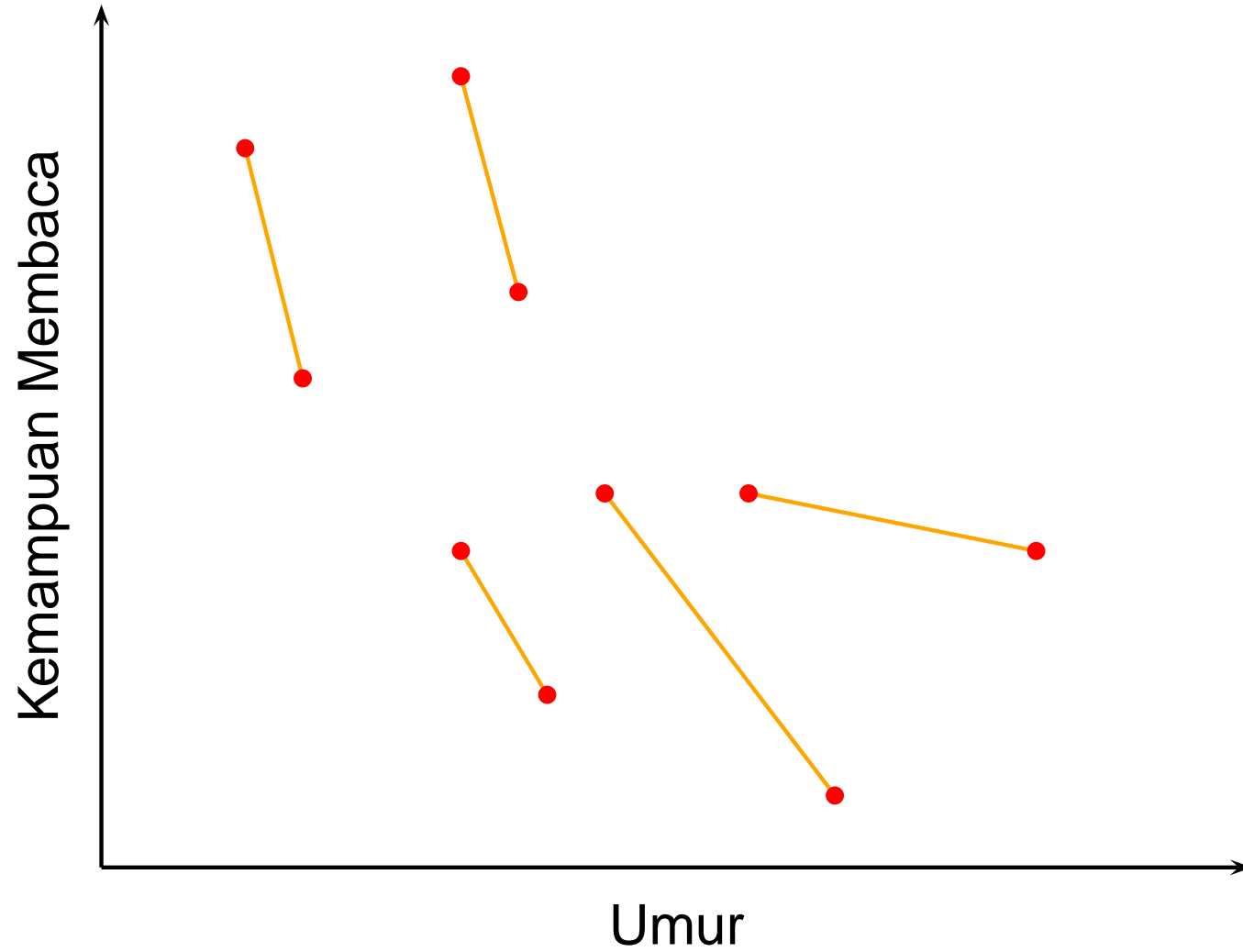
Data Longitudinal



Data Longitudinal



Data Longitudinal



Data Longitudinal

Jenis data yang berkaitan dengan data longitudinal:

- Data Panel
- Data Survival, Antar Kejadian (*Event History*)
- Data Runtun Waktu

Keuntungan

- Dapat digunakan untuk mengetahui pola perubahan
- Setiap individu dapat menjadi kontrol bagi dirinya sendiri
- Dapat membedakan efek dari umur dengan efek dari cohort maupun efek dari periode
- Memungkinkan untuk meneliti kausalitas

Eksplorasi Data

Metode Statistik:

- Eksploratori (deskriptif)
- konfirmasi (inferensial)

Eksplorasi Data

Metode Statistik:

- Eksploratori (deskriptif)
- konfirmasi (inferensial)

Eksplorasi untuk data longitudinal:

- tampilkan sebanyak mungkin data mentah daripada hanya ringkasannya
- tonjolkan pola atau ringkasannya
- identifikasilah baik pola *cross-sectional* maupun longitudinal
- identifikasilah individu atau observasi yang tidak biasa (*outliers*)

Bentuk Umum Data Longitudinal

subyek	observasi	waktu	respon	kovariat		
1	1	t_{11}	y_{11}	x_{111}	\dots	x_{11p}
1	2	t_{12}	y_{12}	x_{121}	\dots	x_{12p}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
1	n_1	t_{1n_1}	y_{1n_1}	x_{1n_11}	\dots	x_{1n_1p}
2	1	t_{21}	y_{21}	x_{211}	\dots	x_{21p}
2	2	t_{22}	y_{22}	x_{221}	\dots	x_{22p}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
2	n_1	t_{2n_1}	y_{2n_1}	x_{2n_11}	\dots	x_{2n_1p}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
m	1	t_{m1}	y_{m1}	x_{m11}	\dots	x_{m1p}
m	2	t_{m2}	y_{m2}	x_{m21}	\dots	x_{m2p}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
m	n_m	t_{mn_1}	y_{mn_1}	x_{mn_11}	\dots	x_{mn_1p}

Bentuk Umum Data Longitudinal

Contoh :

Suatu studi dilakukan untuk merehabilitasi pasien stroke. Ada 3 perlakuan dalam studi ini, yaitu:

- A Terapi yang baru
- B Program rehabilitasi yang sekarang digunakan dalam rumah sakit yang sama
- C Program perawatan biasa yang dilakukan dalam rumah sakit yang lain

Setiap kelompok perlakuan terdiri dari 8 pasien yang diamati selama 8 minggu. Respon yang diperoleh adalah *Bartel index*, yaitu skor yang menunjukkan kemampuan fungsional pasien, nilai yang tinggi menunjukkan kemampuan yang baik (maksimum 100).

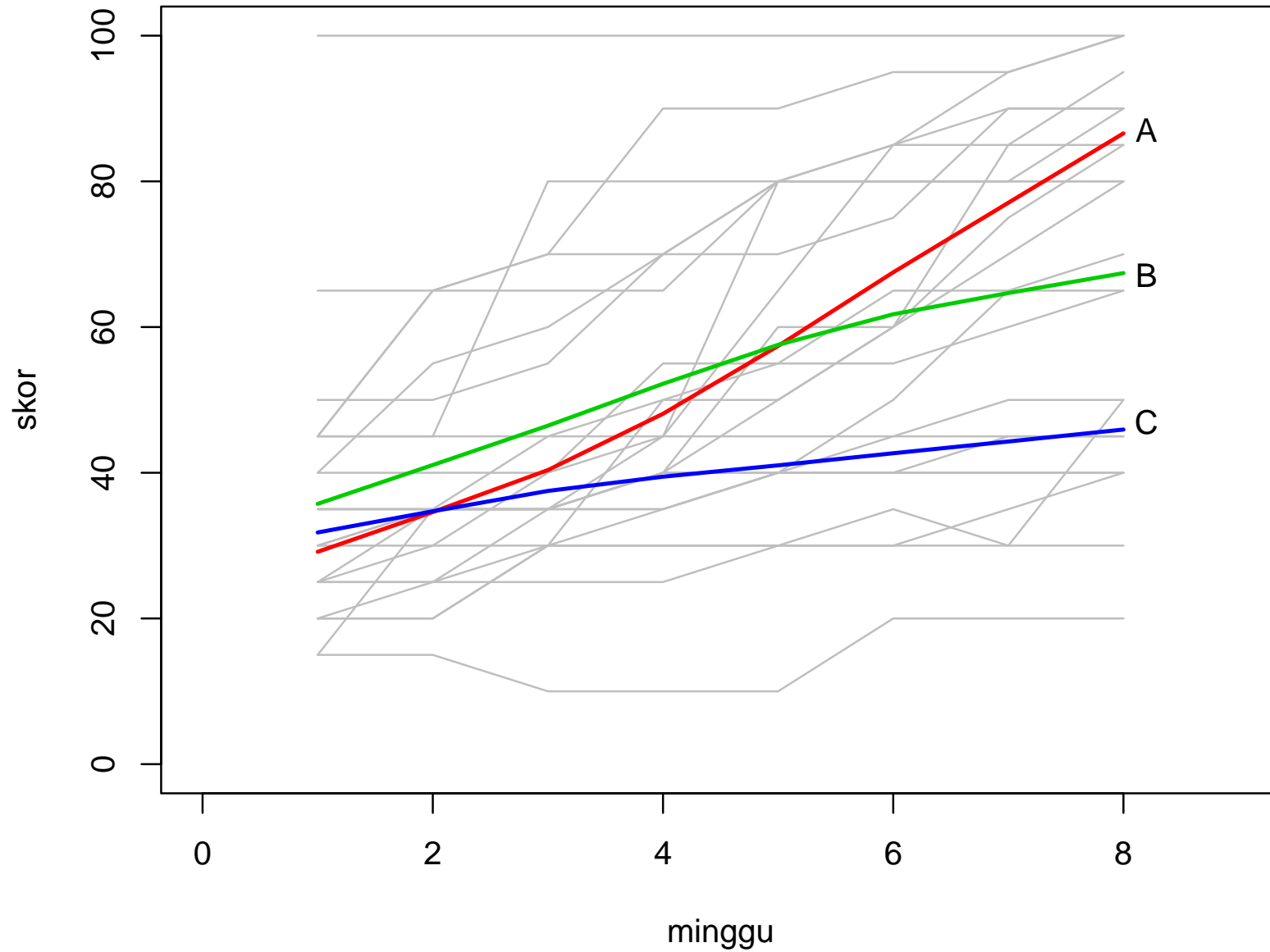
Bentuk Umum Data Longitudinal

subyek	kovariat (perlakuan)	respon (pada minggu ke-)							
		1	2	3	4	5	6	7	8
1	<i>A</i>	45	45	45	45	80	80	80	90
2	<i>A</i>	20	25	25	25	30	35	30	50
3	<i>A</i>	50	50	55	70	70	75	90	90
⋮									
9	<i>B</i>	40	55	60	70	80	85	90	90
⋮				⋮					

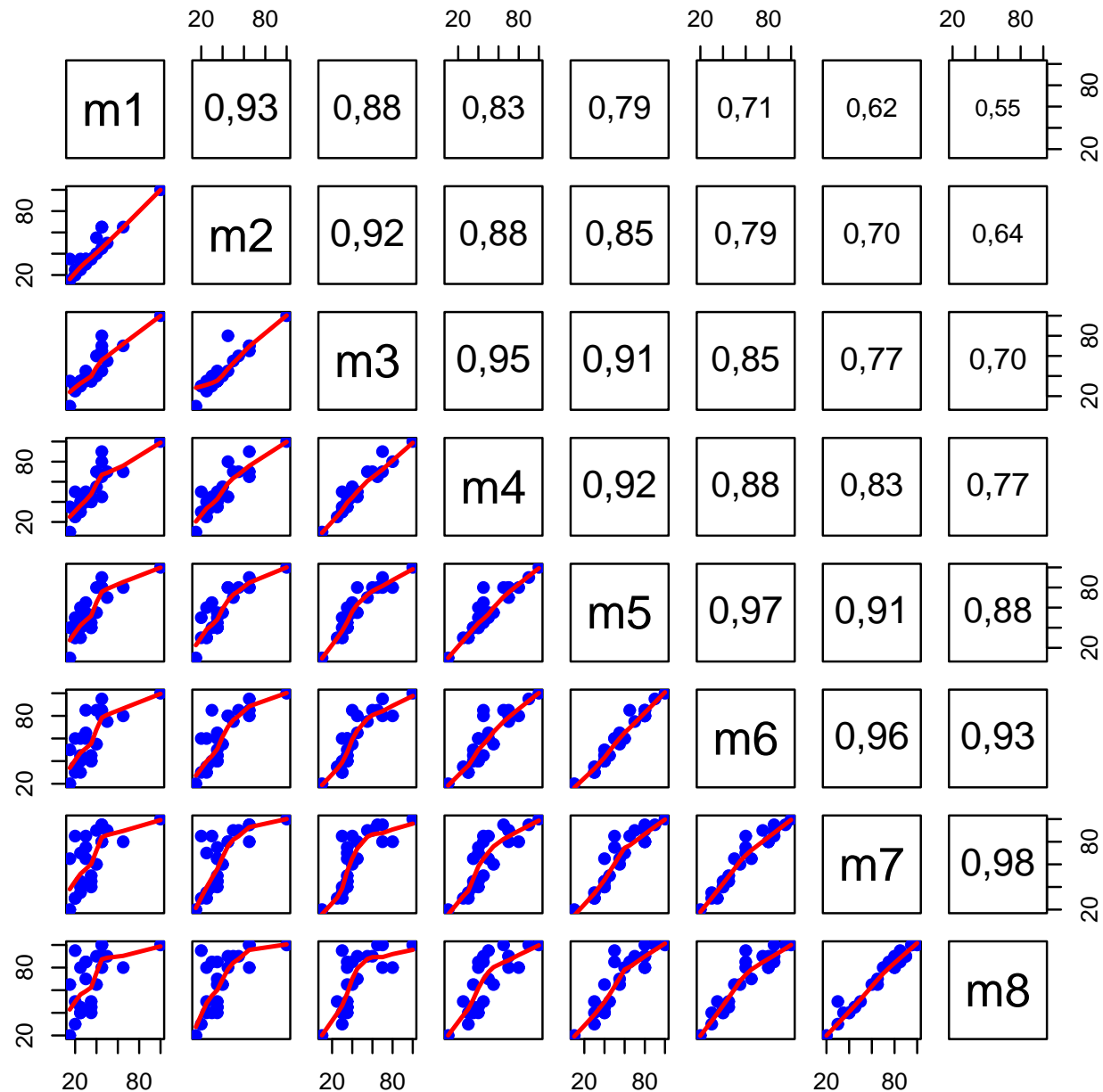
Bentuk Umum Data Longitudinal

subyek	observasi (minggu ke-)	respon	kovariat (perlakuan)
1	1	45	<i>A</i>
1	2	45	<i>A</i>
1	3	45	<i>A</i>
1	4	45	<i>A</i>
1	5	80	<i>A</i>
1	6	80	<i>A</i>
1	7	80	<i>A</i>
1	8	90	<i>A</i>
⋮		⋯	
9	1	40	<i>B</i>
9	2	55	<i>B</i>
9	3	60	<i>B</i>
9	4	70	<i>B</i>
9	5	80	<i>B</i>
9	6	85	<i>B</i>
9	7	90	<i>B</i>
9	8	90	<i>B</i>
⋮		⋯	

Bentuk Umum Data Longitudinal



Bentuk Umum Data Longitudinal



Pemodelan Data Longitudinal

Prinsip Pemodelan:

- Permasalahan ilmiah diformulasikan sebagai model regresi
 - var respons \leftarrow var penjelas
 - variable penjelas awal (*baseline*) yang tetap sepanjang waktu
 - variabel yang berubah sepanjang waktu (*time-varying expl. variables*)
- Korelasi (asosiasi) karena pengukuran berulang pada individu yang sama, atau observasi berulang

Notasi

- Individu: $i = 1, \dots, m$
- Observasi pada individu i : $jh = 1, \dots, n_i$
- Total observasi: $N = \sum_{i=1}^m n_i$
- Waktu observasi aktual: t_{ij}

- Variabel respon:

variabel random

respon observasi

Y_{ij}

y_{ij}

$\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})$

$\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})$

$\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_m)$

$\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_m)$

- Variabel penjelas:

$\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})^T$, vektor berukuran $p \times 1$

$\mathbf{X}_i = (x_{i1}, \dots, x_{in_i})$, matriks berukuran $n_i \times p$

Notasi

- Mean \mathbf{Y}_i untuk individu i : $E(\mathbf{Y}_i) = \boldsymbol{\mu}_i$
- Variansi \mathbf{Y}_i ; Matriks Kovariansi $n_i \times n_i$ untuk individu i :

$$\text{Var}(\mathbf{Y}_i) = \begin{bmatrix} v_{i11} & \dots & v_{i1n_i} \\ \dots & v_{ijk} & \dots \\ v_{in_i1} & \dots & v_{in_in_i} \end{bmatrix}$$

dengan $v_{ijk} = \text{Cov}(Y_{ij}, Y_{ik})$

Pendekatan Pemodelan

- Model linear umum
- Model marginal (*marginal, population average*)
- Model efek random (*random effects; subject specific*)
- Model transisi (*transition*)

Contoh Pemodelan

Naive analysis (pooled analysis)

Model 1:

$$E(Y_i | \mathbf{X}_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i}, \quad i = 1, \dots, n$$

dengan

$$X_{1i} = \begin{cases} 1 & \text{i mendapat perlakuan } B \\ 0 & \text{i mendapat perlakuan selain } B \end{cases}$$

$$X_{2i} = \begin{cases} 1 & \text{i mendapat perlakuan } C \\ 0 & \text{i mendapat perlakuan selain } C \end{cases}$$

$X_{3i} = 1, 2, \dots, 8$ minggu pengamatan

$n = 24 \times 8 = 192$, 24 pasien (terbagi dalam 3 grup perlakuan) diamati

selama 8 minggu

Contoh Pemodelan

Naive analysis (pooled analysis)

Model 2:

$$E(Y_i | \mathbf{X}_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 (X_{1i} \times X_{3i}) + \beta_5 (X_{2i} \times X_{3i}), \quad i = 1, \dots, n$$

dengan

$(X_{1i} \times X_{3i})$ dan $(X_{2i} \times X_{3i})$ adalah interaksi antara perlakuan dengan waktu.

Contoh Pemodelan

Parameter	Estimasi	SE
Model 1:		
β_0	36,84	3,971
β_1	-5,63	3,715
β_2	-12,11	3,715
β_3	4,76	0,662
Model 2:		
β_0	29,82	5,774
β_1	3,35	8,166
β_2	-0,02	8,166
β_3	6,32	1,143
β_4	-1,99	1,617
β_5	-2,69	1,617

Model Linear Normal

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} = \boldsymbol{\mu}, \quad \mathbf{y} \sim N(\boldsymbol{\mu}, \mathbf{V})$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix},$$

dengan p adalah banyaknya parameter model, n adalah banyaknya observasi.

Model Linear Normal

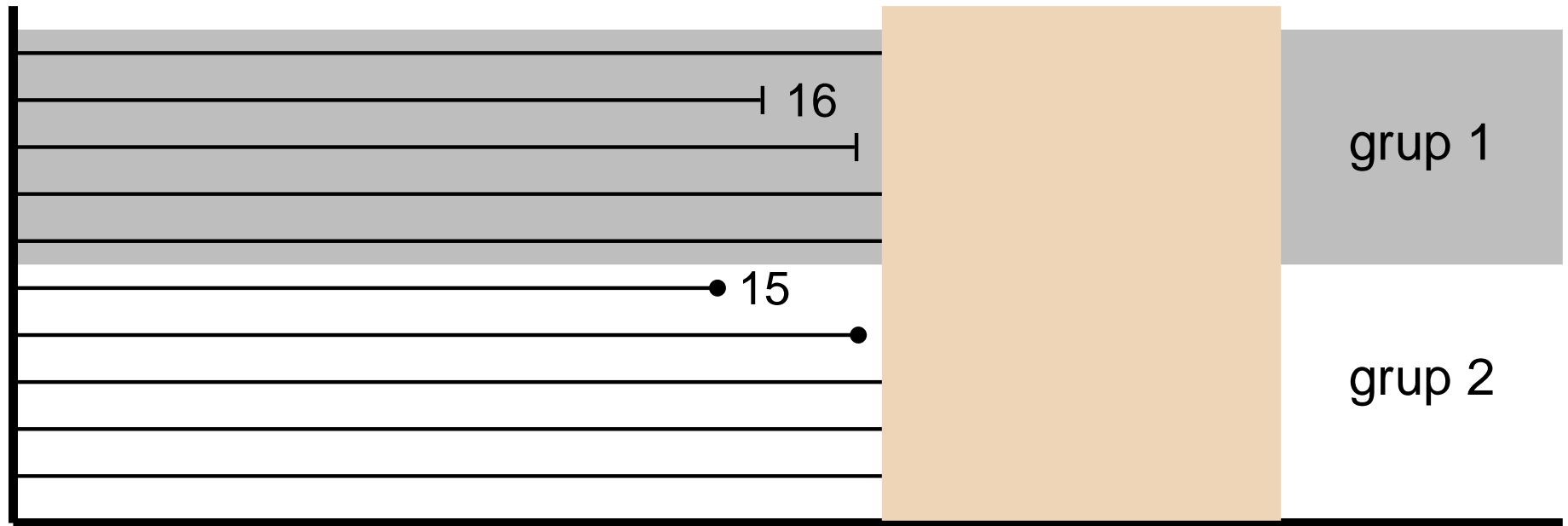
Estimasi parameter model

- *Generalized Least Squares*
- *Maximum Likelihood*
- *Generalized Estimating Equation*

Model Linear Normal

Parameter	Estimasi (SE)		
	<i>Naive</i>	<i>Random effects</i>	GEE-indep
Model 1:			
β_0	36.84 (3.971)	36.84 (7.307)	36.84 (8.003)
β_1	-5.63 (3.715)	-5.62 (10.177)	-5.63 (9.519)
β_2	-12.11 (3.715)	-12.10 (10.177)	-12.11 (9.519)
β_3	4.76 (0.662)	4.76 (0.282)	4.76 (0.628)
Model 2:			
β_0	29.82 (5.774)	29.82 (7.497)	29.82 (10.175)
β_1	3.35 (8.166)	3.34 (10.602)	3.35 (11.633)
β_2	-0.02 (8.166)	-0.02 (10.602)	-0.02 (10.895)
β_3	6.32 (1.143)	6.32 (0.467)	6.32 (1.131)
β_4	-1.99 (1.617)	-1.99 (0.660)	-1.99 (1.477)
β_5	-2.69 (1.617)	-2.68 (0.660)	-2.69 (1.470)

Data Durasi/Survival



Data terobservasi

E	D		
	1	2	
1	0	5	5
2	2	3	5

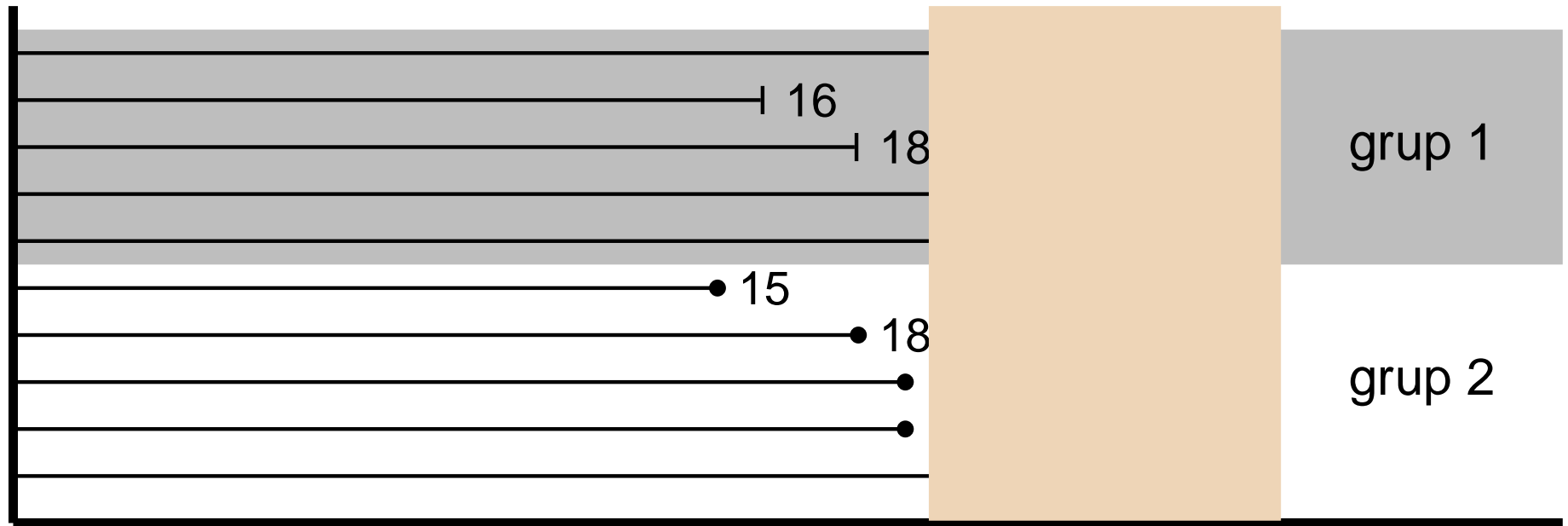
$E = 1$ adalah grup 1

$E = 2$ adalah grup 2

$D = 1$ mendapatkan *event* (misalnya sakit)

$D = 2$ tidak mendapatkan *event*

Data Durasi/Survival



Data terobservasi

E	D		
	1	2	
1	0	5	5
2	4	1	5

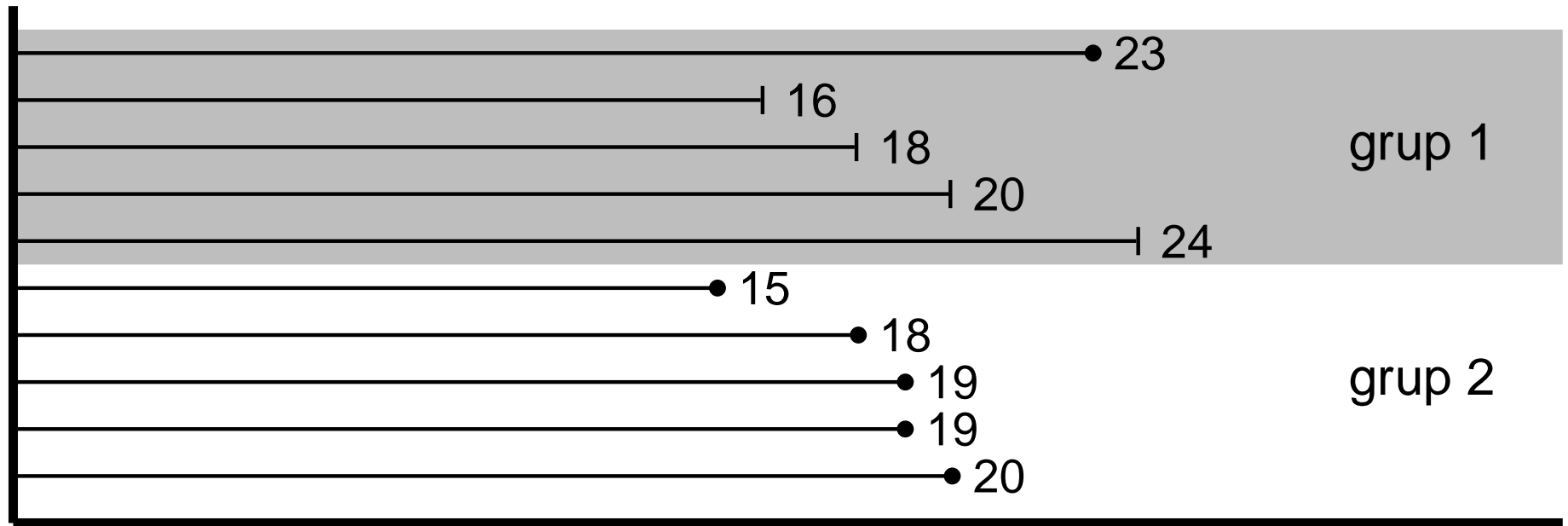
$E = 1$ adalah grup 1

$E = 2$ adalah grup 2

$D = 1$ mendapatkan *event* (misalnya sakit)

$D = 2$ tidak mendapatkan *event*

Data Durasi/Survival



Data terobservasi

E	D		
	1	2	
1	1	4	5
2	5	0	5

$E = 1$ adalah grup 1

$E = 2$ adalah grup 2

$D = 1$ mendapatkan *event* (misalnya sakit)

$D = 2$ tidak mendapatkan *event*

Fungsi Survival

Variabel random T adalah waktu antar kejadian atau durasi,
 $T > 0$

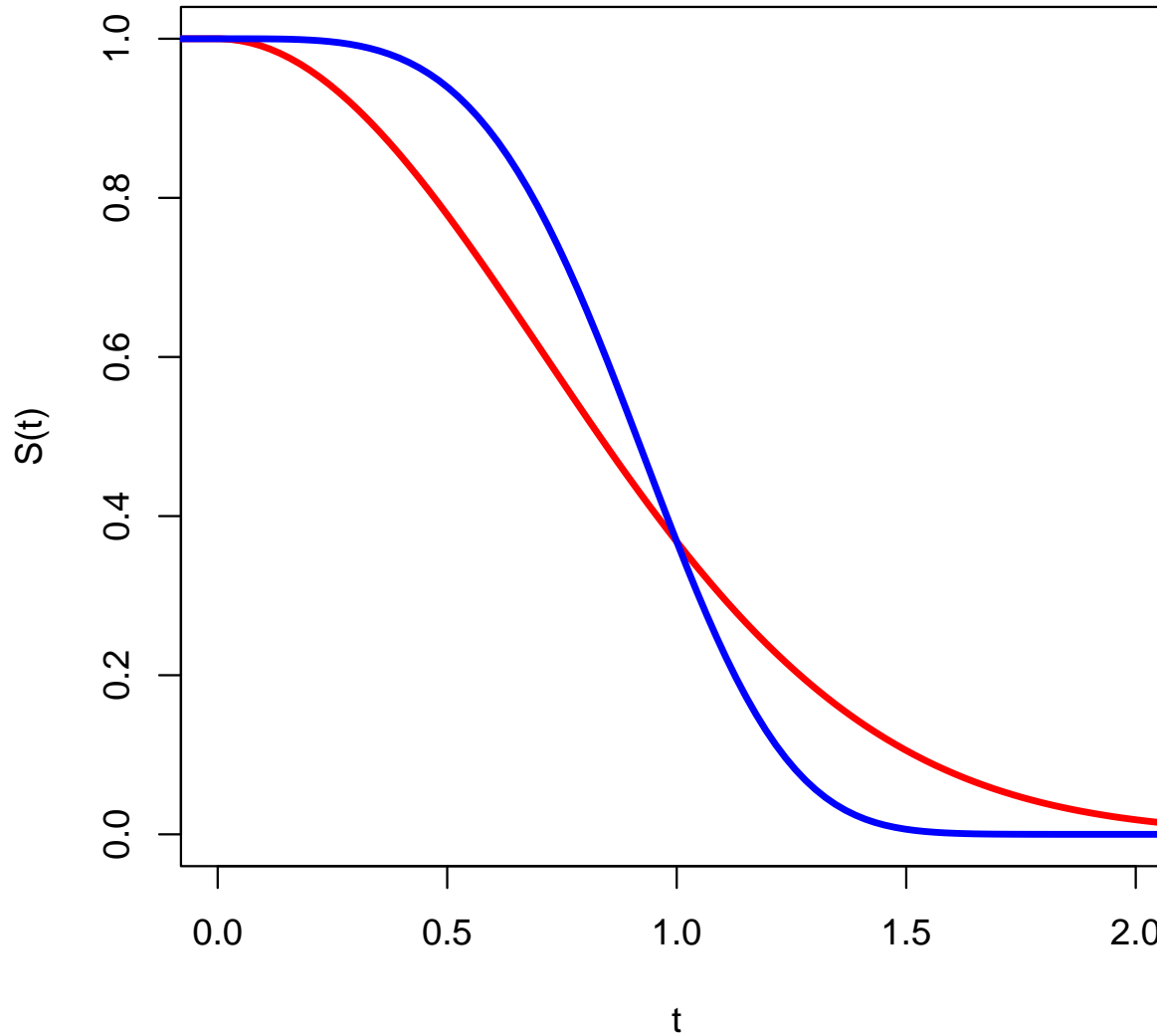
Probabilitas satu individu hidup (tinggal dalam suatu *status*)
lebih lama daripada t

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

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

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

Fungsi Survival



Fungsi 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

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

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

Fungsi Hazard

Tingkat (*rate*) terjadinya suatu *event*

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

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

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

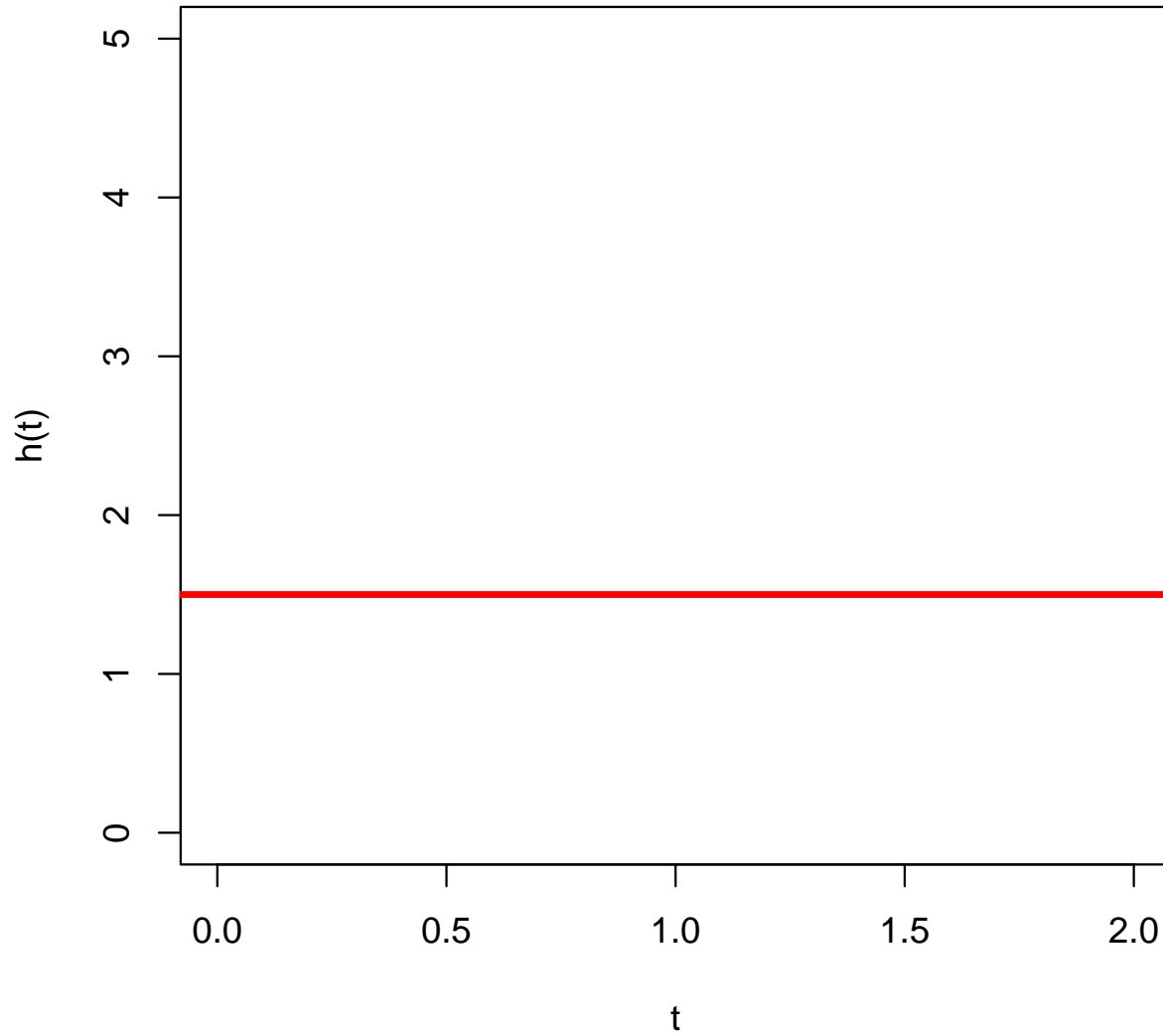
Fungsi *Hazard* Kumulatif

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

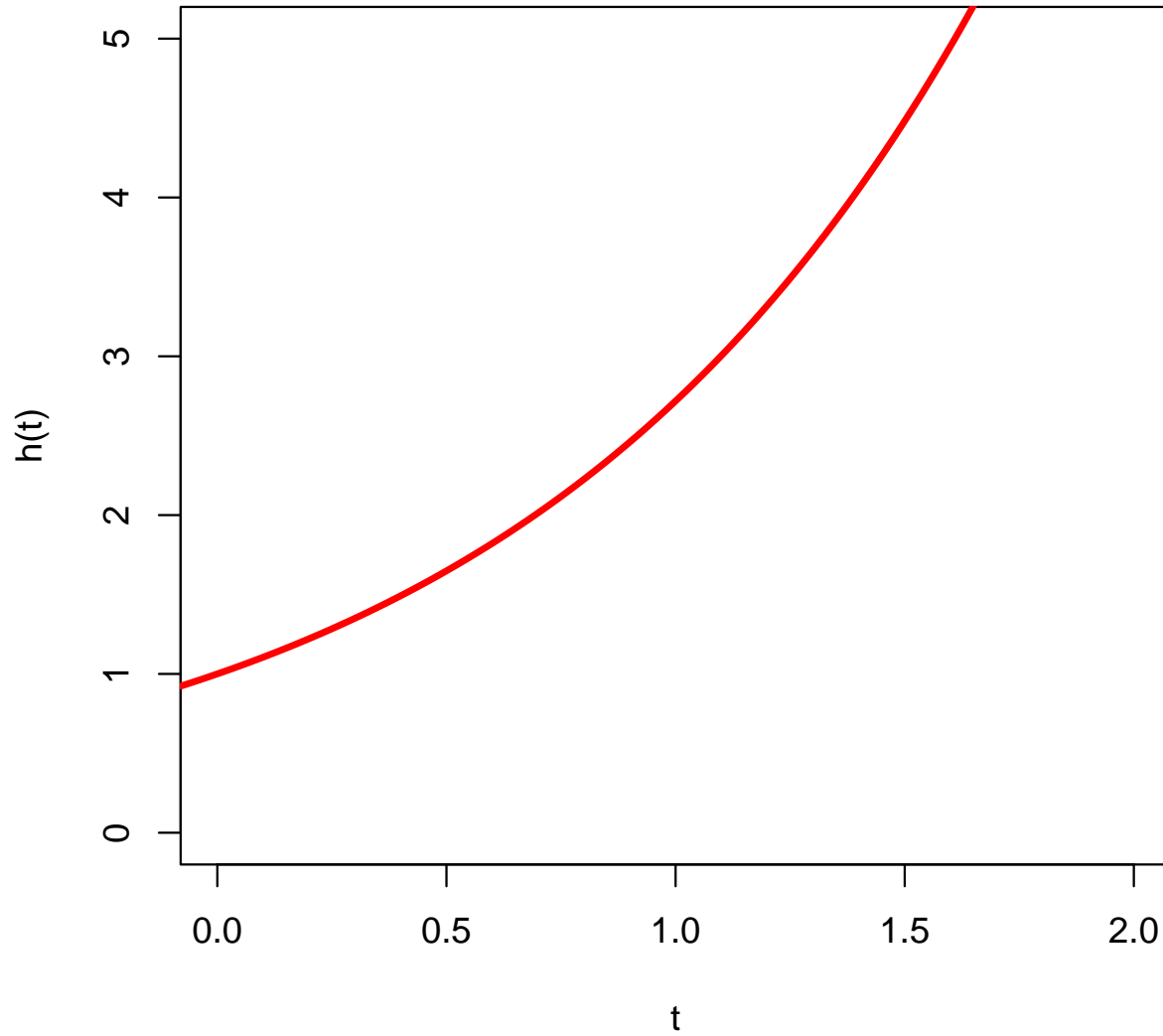
Hubungan $H(t)$ dengan $S(t)$

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

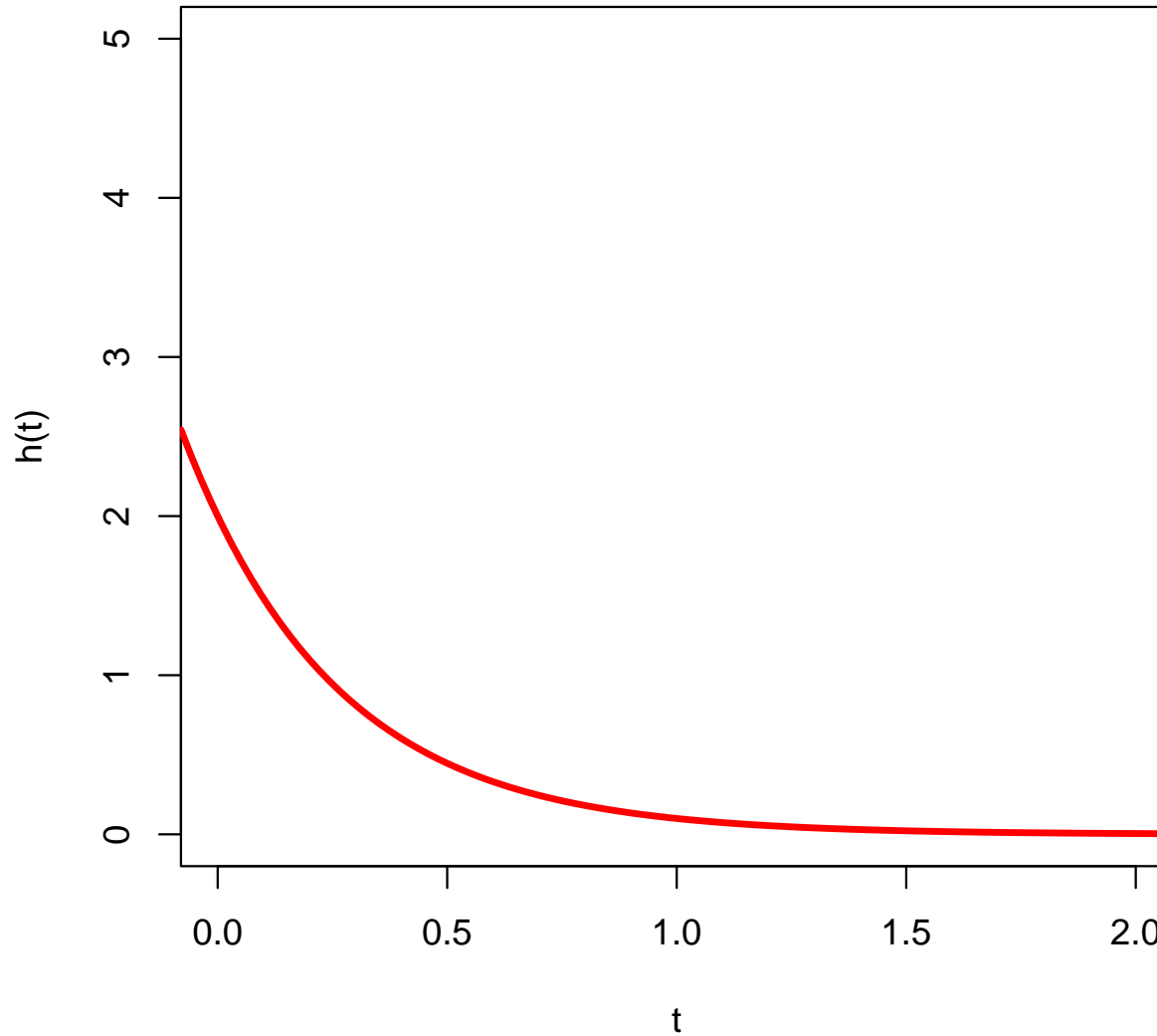
Fungsi Hazard



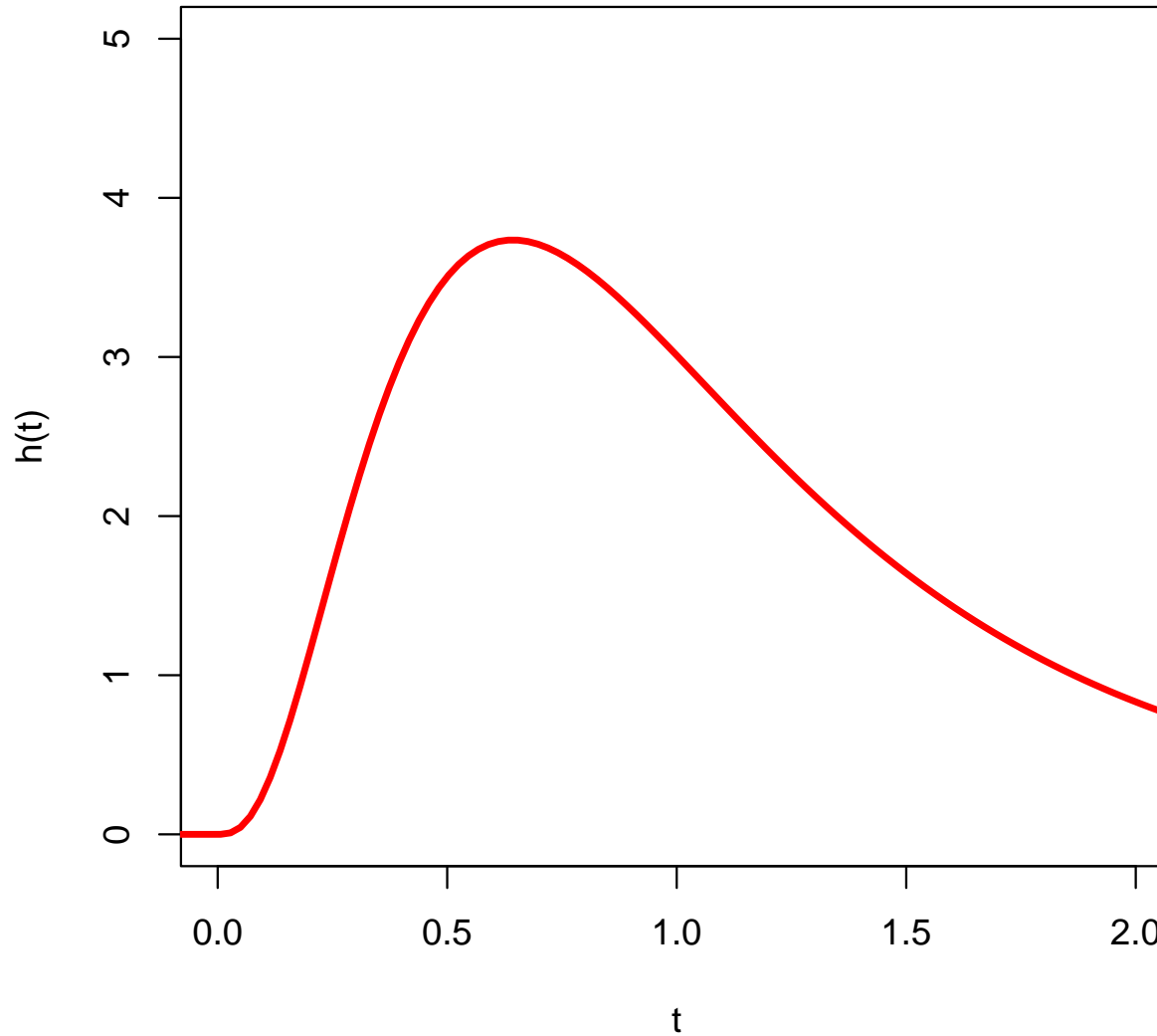
Fungsi Hazard



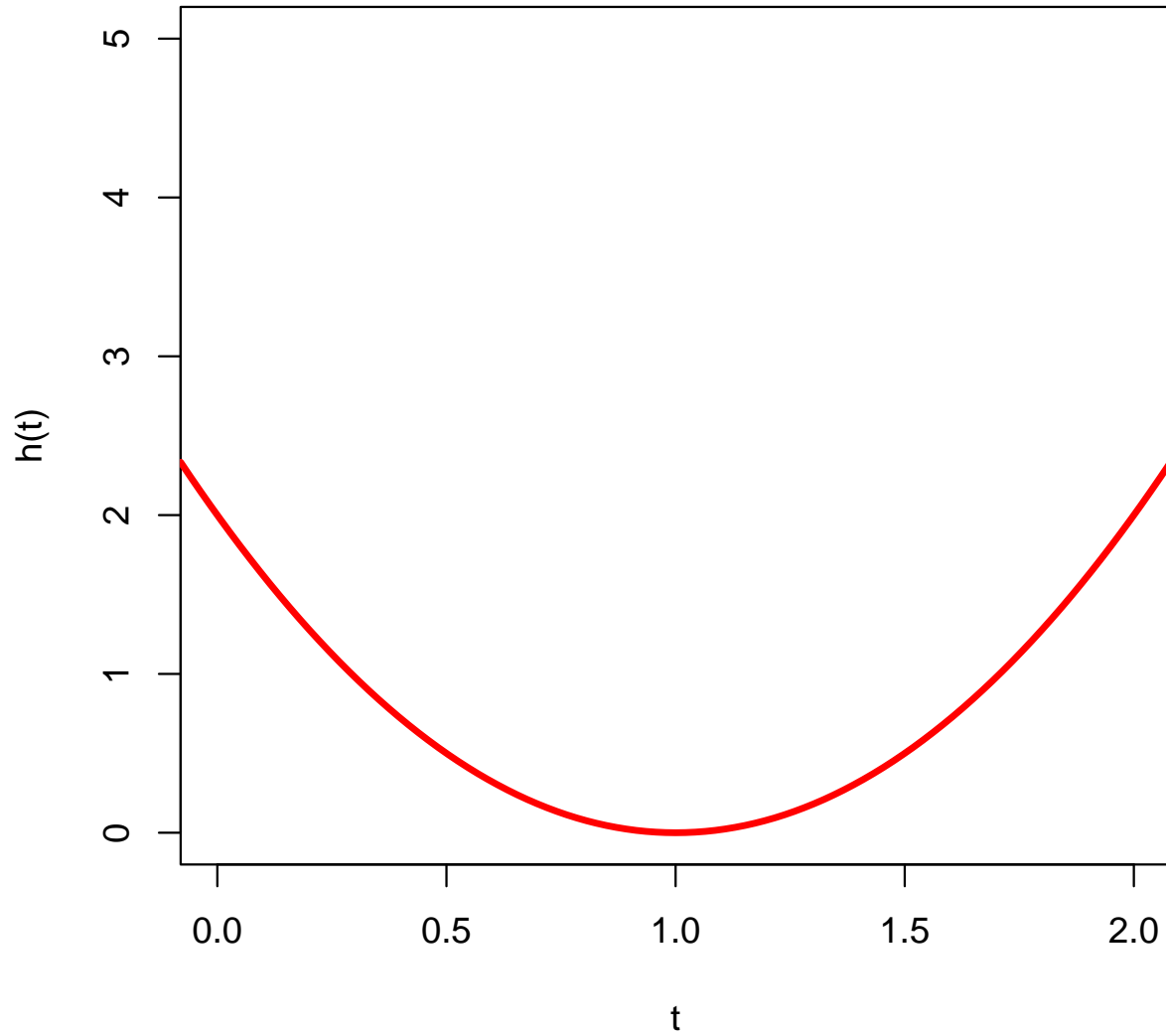
Fungsi Hazard



Fungsi Hazard



Fungsi Hazard



Kaplan-Meier

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

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

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)

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

Alternatif:

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

Kaplan-Meier

Contoh:

Diketahui data survival sebagai berikut: 23 16+ 18+ 20+ 24+ 15
18 19 19 20 (Tanda + menunjukkan tersensor). Hitung estimasi $S(t)$ menggunakan Kaplan-Meier:

t_i	Y_i	d_i	$\hat{S}(t)$	$\sqrt{\text{var}(\hat{S}(t))}$	Int-konf 95% $S(t)$
15	10	1	0,900	0,0949	0,7320 – 1,000
18	8	1	0,787	0,1340	0,5641 – 1,000
19	6	2	0,525	0,1759	0,2722 – 1,000
20	4	1	0,394	0,1742	0,1655 – 0,937
23	2	1	0,197	0,1642	0,0384 – 1,000

Nelson-Aalen

Estimator untuk fungsi hazard kumulatif:

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

dengan variansi

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

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)

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 pengujian:

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

dengan $\chi^2 \sim \text{Chi-square}(df=1)$

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					

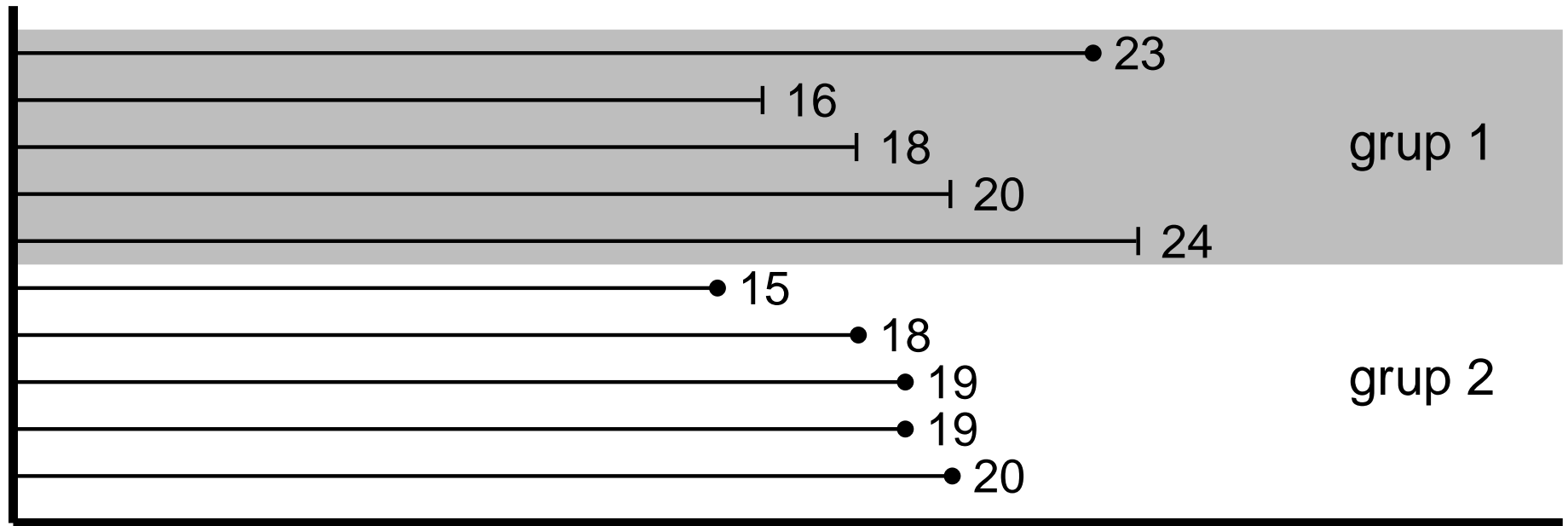
t : *event-time*

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					

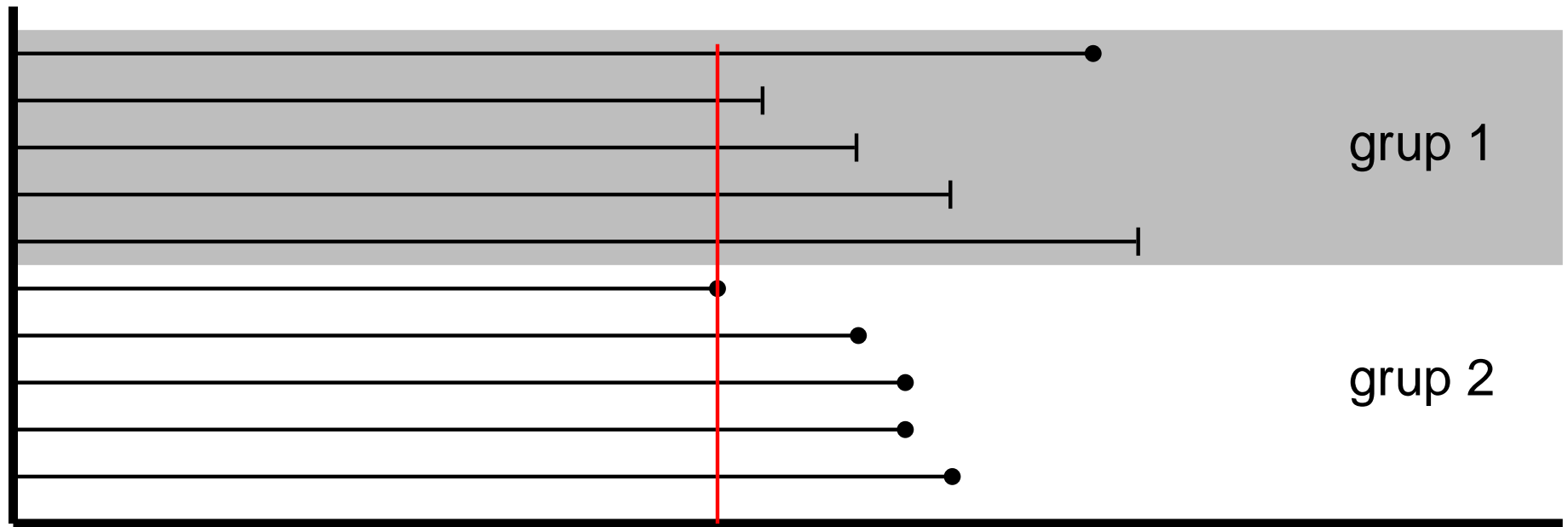
t : *event-time*

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					

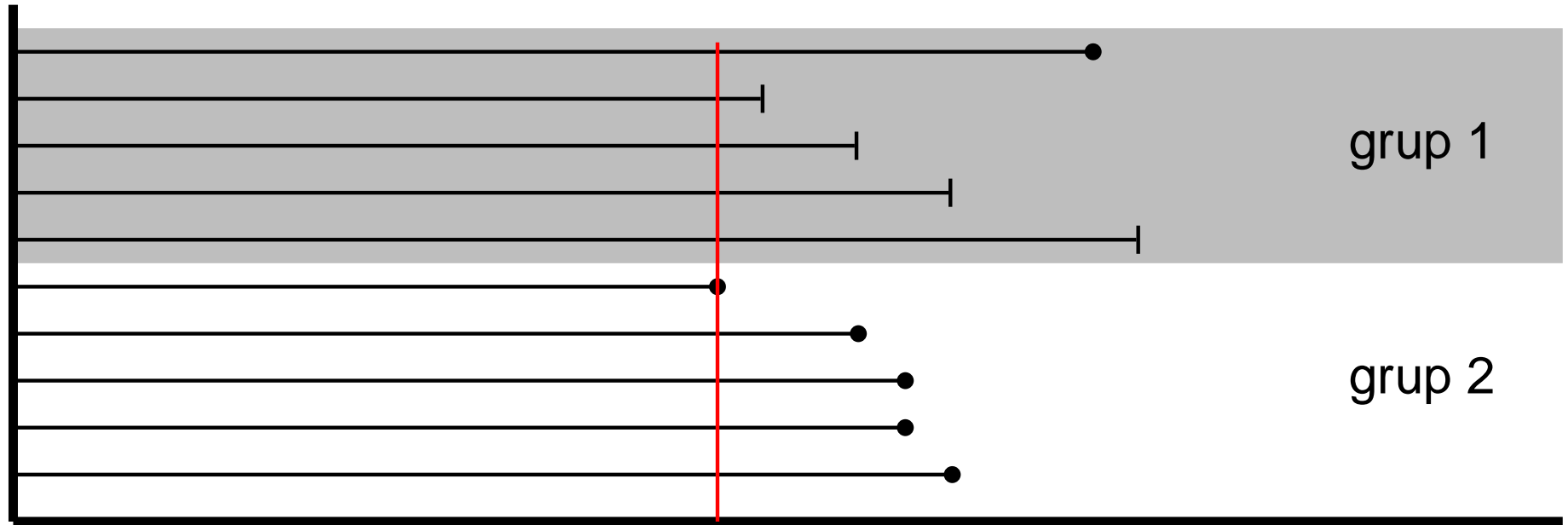
t : *event-time*

d_t : *banyaknya event*

n_1, n_2 : *number at risk*

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

Logrank Test

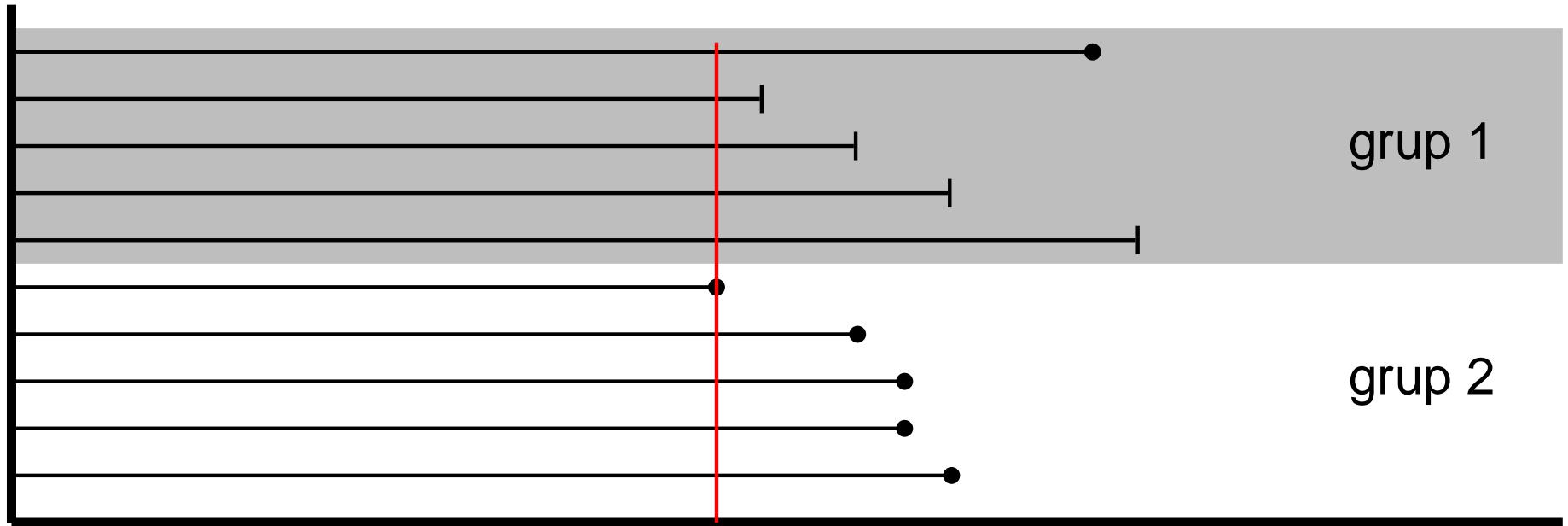


t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5		
18					
19					
20					
23					

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

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

Logrank Test

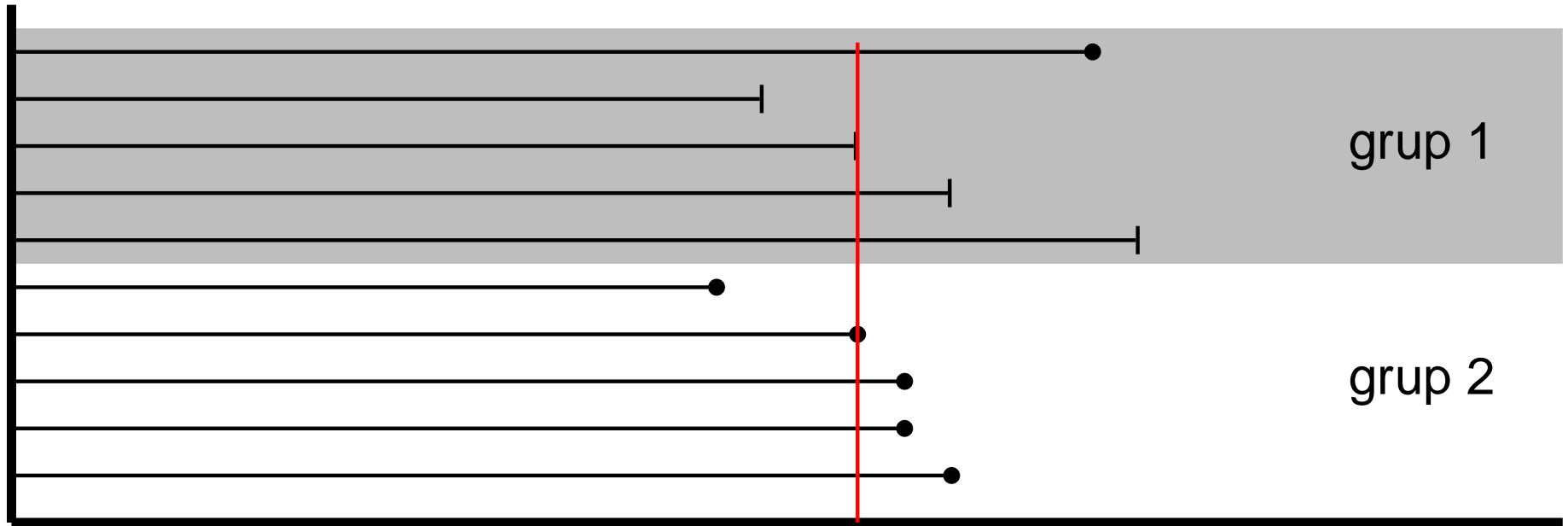


t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5	0,5	0,5
18					
19					
20					
23					

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

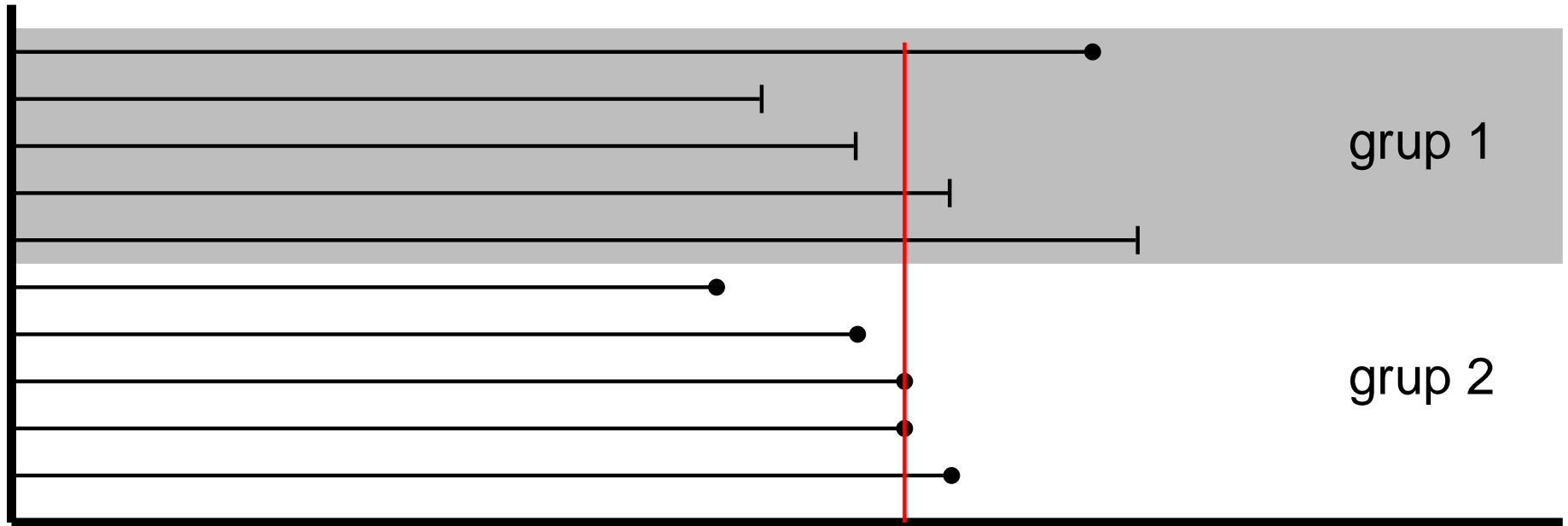
$$e_{2t} = \frac{n_{2t}}{n_{1t} + n_{2t}} \times d_t$$

Logrank Test



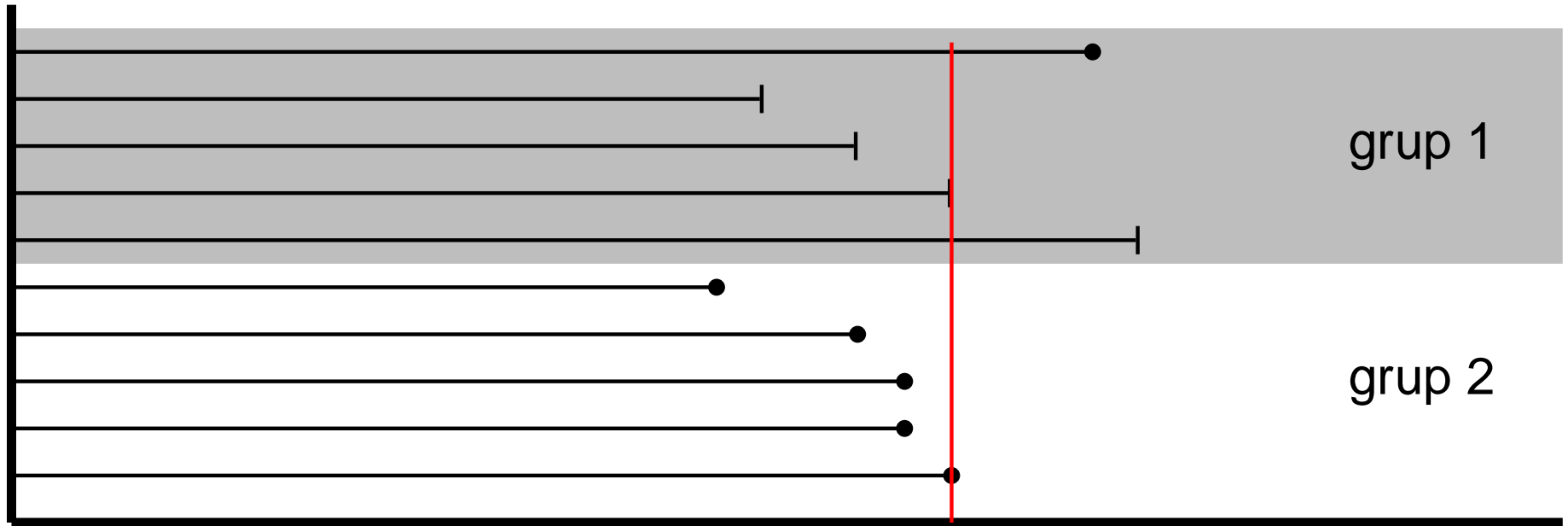
t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5	0,5	0,5
18	1	4	4	0,5	0,5
19					
20					
23					

Logrank Test



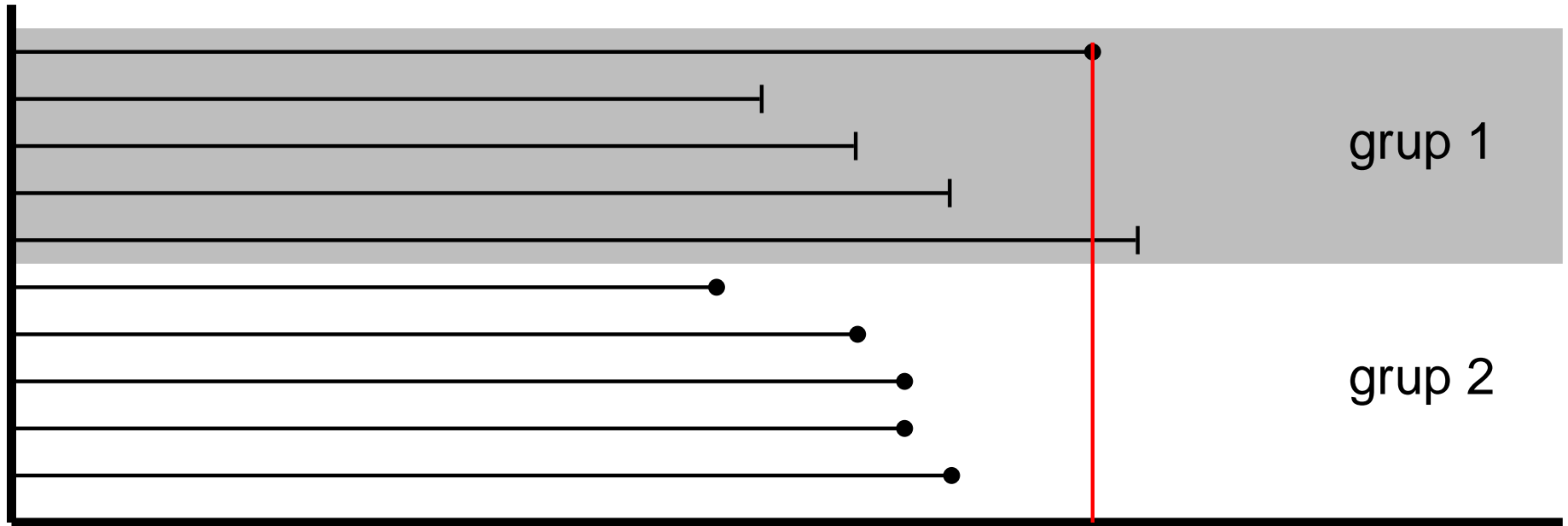
t	d_t	n_{1t}	n_{2t}	e_{1t}	e_{2t}
15	1	5	5	0,5	0,5
18	1	4	4	0,5	0,5
19	2	3	3	1,0	1,0
20					
23					

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+

grup 2: 15, 18, 19, 19, 20

$$O_1 = 1$$

$$O_2 = 5$$

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

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

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

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

$$p\text{-value} = 0,0204 < 0,05$$

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

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

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

- Representasi $\log T$

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

dengan $\boldsymbol{\alpha}^T = (\alpha_1 \dots \alpha_p)$ 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

$$H(t | \mathbf{x}) = H_0(\exp(\mathbf{x}\boldsymbol{\beta})t), \quad \text{untuk semua } t$$

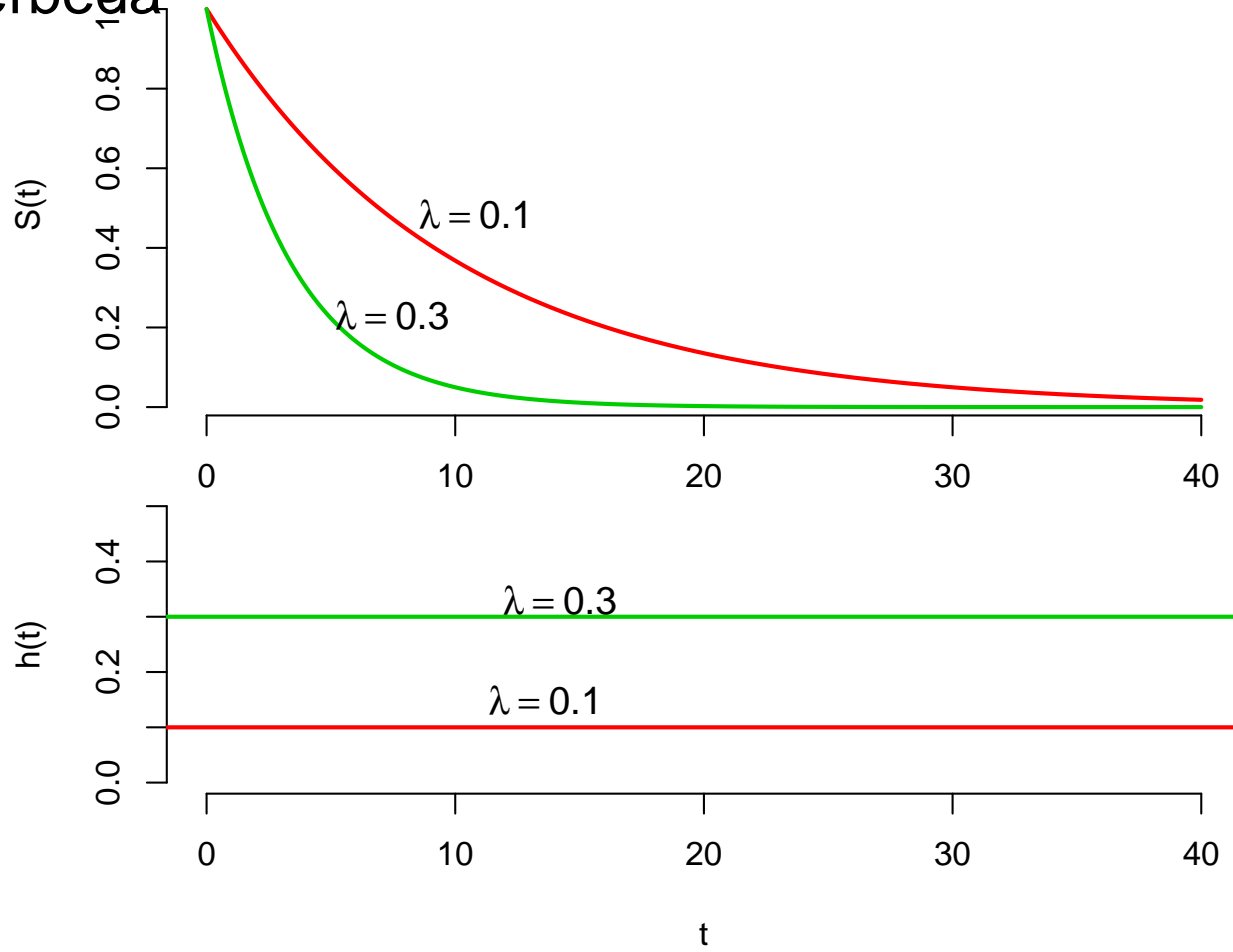
atau

$$S(t | \mathbf{x}) = S_0(\exp(\mathbf{x}\boldsymbol{\beta})t), \quad \text{untuk semua } t$$

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

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):

$$h(t | \mathbf{x}) = h_0(t)\psi(\mathbf{x}, \boldsymbol{\beta})$$

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

Cox's Regression Model

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

$$h(t | \mathbf{x}) = h_0(t)\psi(\mathbf{x}, \boldsymbol{\beta})$$

fungsi hazard
bergantung pada \mathbf{x} = baseline hazard
tdk bergantung pd \mathbf{x} \times fungsi kovariat

Cox's Regression Model

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

$$h(t | \mathbf{x}) = h_0(t)\psi(\mathbf{x}, \boldsymbol{\beta})$$

fungsi hazard
bergantung pada \mathbf{x} = baseline hazard
tdk bergantung pd \mathbf{x} \times fungsi kovariat

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

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

Cox's Regression Model

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

$$h(t | \mathbf{x}) = h_0(t)\psi(\mathbf{x}, \boldsymbol{\beta})$$

fungsi hazard
bergantung pada \mathbf{x} = baseline hazard
tdk bergantung pd \mathbf{x} \times fungsi kovariat

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

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

Cox's Regression Model

Model:

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

Misalkan:

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

Cox's Regression Model

Model:

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

Hazard ratio:

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

Cox's Regression Model

Model:

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

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:

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

Hazard ratio:

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

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

Cox's Regression Model

Model:

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

Hazard ratio:

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

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

Cox's Regression Model

Model:

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

Hazard ratio:

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

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

Cox's Regression Model

Model:

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

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:

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

Dapat dituliskan dalam $H(t | \mathbf{x})$ atau $S(t | \mathbf{x})$

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

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

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

Alat Bantu

- komputer dan *software*
- online reference, documentation
- personal documentation

Alat Bantu

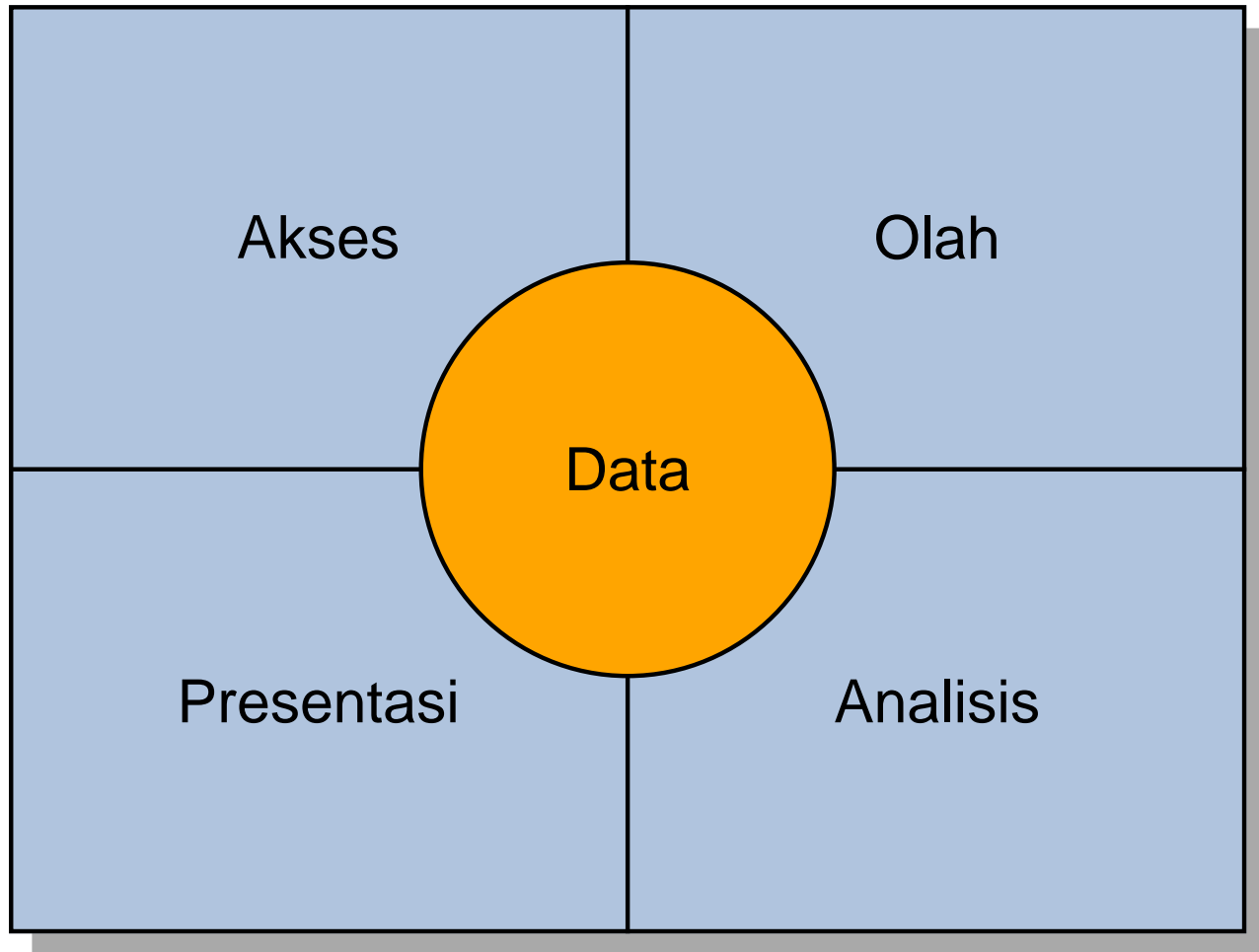
The good news is that statistical analysis is becoming easier and cheaper. The bad news is that statistical analysis is becoming easier and cheaper. (Hofacker, 1983)

Alat Bantu

Keuntungan menggunakan komputer (dibandingkan manual)

- Akurasi dan kecepatan
- Fasilitas dan metode yang digunakan lebih banyak
- Grafik
- Fleksibel
- Manipulasi variabel mudah
- Volume data besar
- Transfer data mudah

Alat Bantu



Alat Bantu

Tahapan analisis menggunakan komputer

1. *Data collection*
2. *Data entry*
3. *Data checking*
4. *Data screening*
5. *Data analysis*
6. *Checking results*
7. *Interpretation*

Alat bantu

Akses: Memasukkan data (entry data), mengambil data (dari format data yang lain)

Olah: Mengurutkan, menyeleksi, mentransformasi, mengambil subset data, menambah data

Presentasi: Membuat deskripsi data, tabel, grafik, ringkasan-ringkasan statistik

Analisis: Melakukan analisis data berdasarkan teori, metode-metode statistika tertentu atau metode-metode kuantitatif yang lain

Alat bantu

Cara eksekusi Paket Statistik :

batch mode berupa program/urutan (sekuen) perintah

non-interactive mode dalam mode ini, biasanya tidak diketahui proses eksekusi, hanya hasil dari eksekusi saja (biasanya dijalankan langsung dari DOS prompt)

interactive line mode perintah dijalankan melalui prompt dari paket statistik per baris

display manager (menu) mode perintah dijalankan interaktif melalui menu-menu atau window

Alat Bantu

Program (paket statistik) utama yang akan digunakan dalam kuliah ini:

- **R**

`http://www.cran.ugm.ac.id`

- **Epi-Info**

`http://www.cdc.gov/EpiInfo/epiinfo.htm`

Sumber informasi internet:

1. `http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed`
2. `http://www.pitt.edu/super1/`

Konsultan Biostatistika

Permasalahan yang biasa dikonsultasikan

- Desain penelitian dan ukuran sampel
- Prosedur analisis statistik yang tepat
- Penggunaan program komputer
- Interpretasi hasil analisis statistik