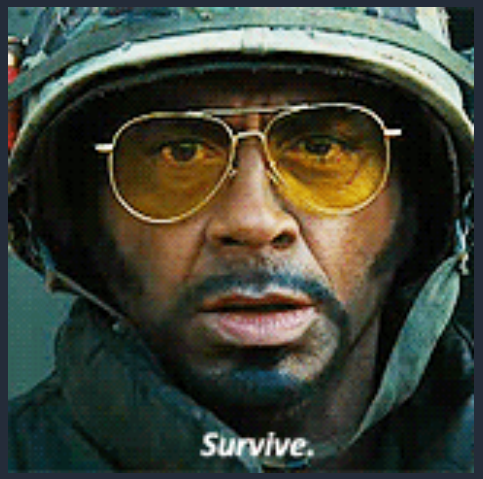
Surviving to survival analysis in R

Mathilde CHEN

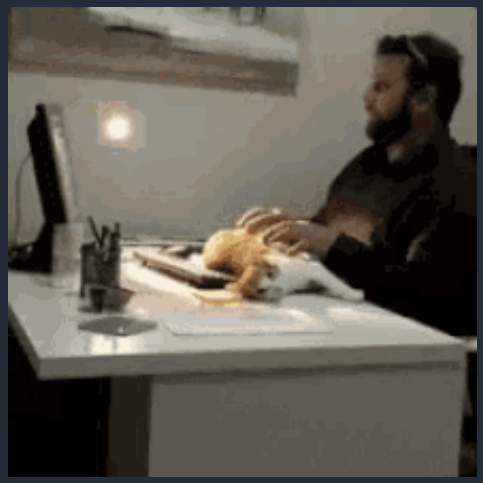
RLadies Paris

2019-06-14

Part I
What are survival analyses?



Part II
Implementation in R

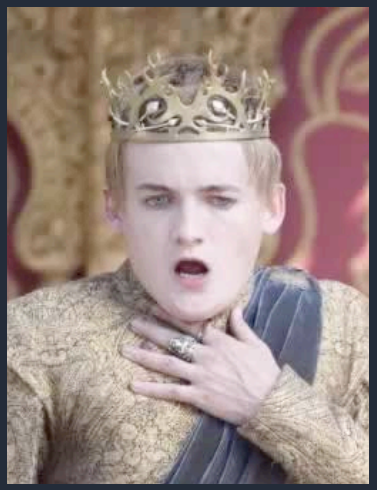


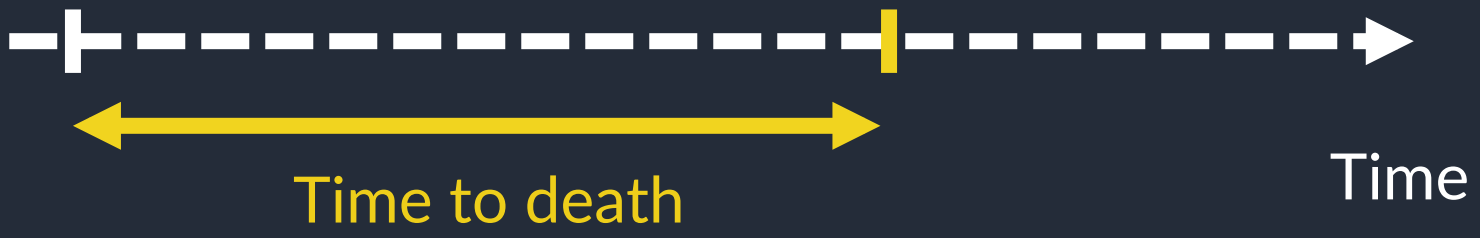
The primary purpose of a survival analysis is to model and analyze time-to-event data

The primary purpose of a survival analysis is to model and analyze time-to-event data

= data that have as endpoint the time when a given event occurs.

Time-to-event data





Other examples: Time until an electronic component fails, length of remission after initial treatment, time of learning a skill etc.

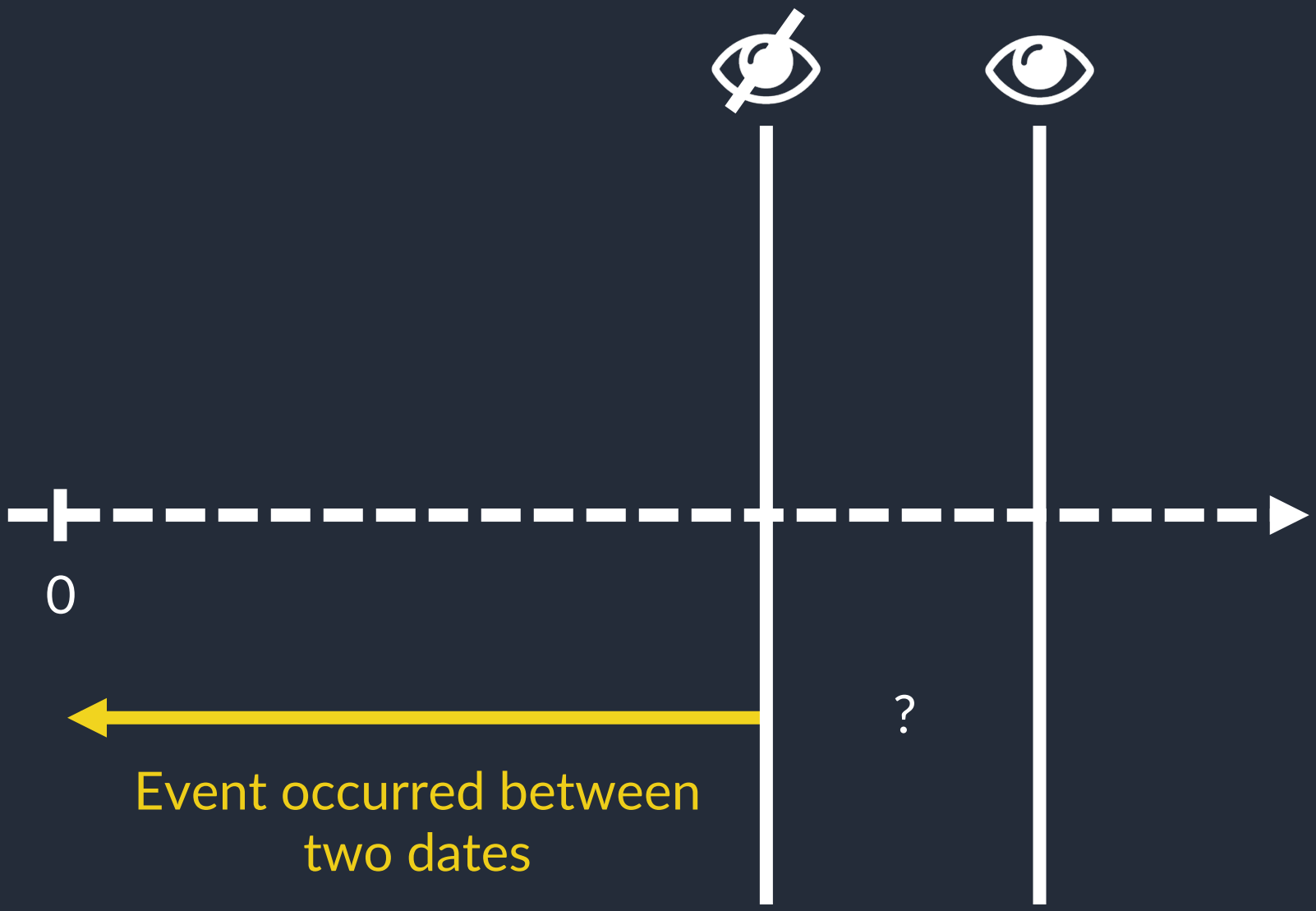
However, the event may not be observed within the follow-up period, producing “censored observations”.

Right-censored data

Left-censored data

The event occurred before
a specific time

Interval-censored data



Censoring occurs in many fields

- Biomedical sciences

  (toxicology, epidemiology, oncology etc.)
- Social sciences
- Engineering
- Ecology, agriculture
- Insurance

What to do?

→ Loss of partial information

→ Introduction of bias

What to do?

→ Loss of partial information

→ Introduction of bias

→ Require special care

Survival function

Let T >= 0: a a random variable, representing the survival time

Survival function

Let T >= 0: a a random variable, representing the survival time

→ We want to know the probability that an individual survives beyond time t

Survival function

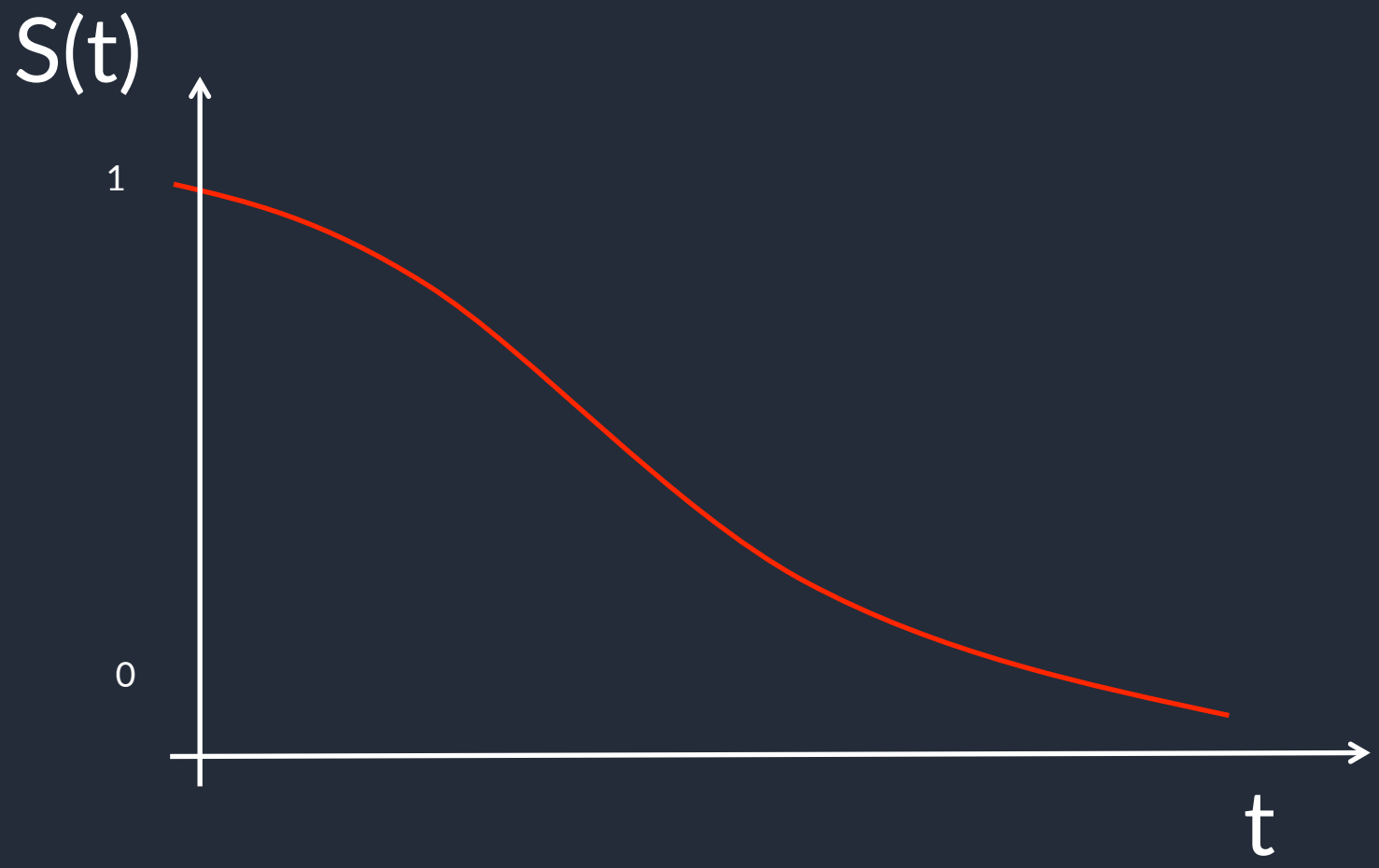
Let T >= 0: a a random variable, representing the survival time

→ We want to know the probability that an individual survives beyond time t

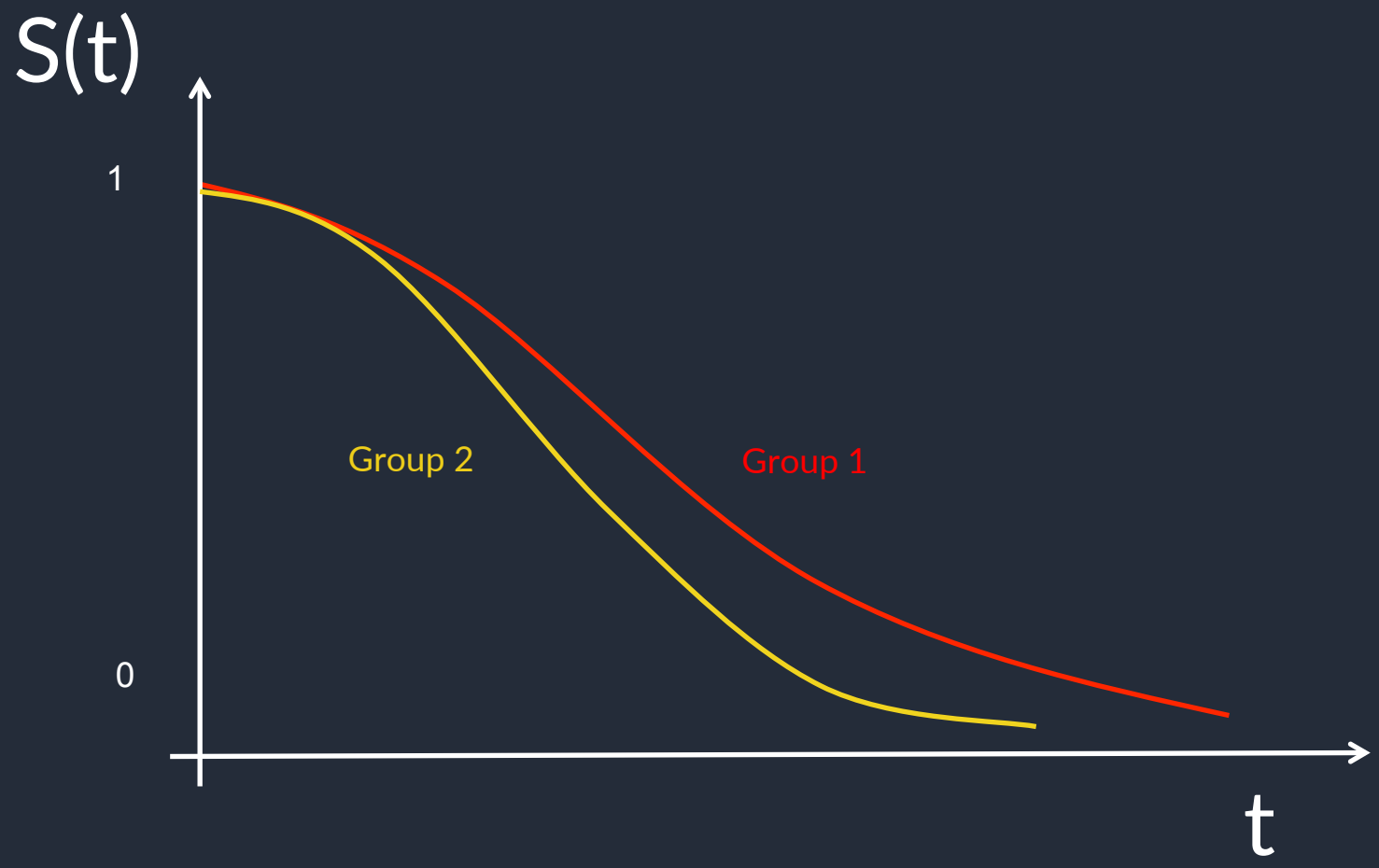
$$S(t) = P(T > t)$$

With P (T > t) : the probability that T exceeds t ($0 < t < \infty$)

Survival function



Survival function



Probability density function

= frequency of events per unit of time

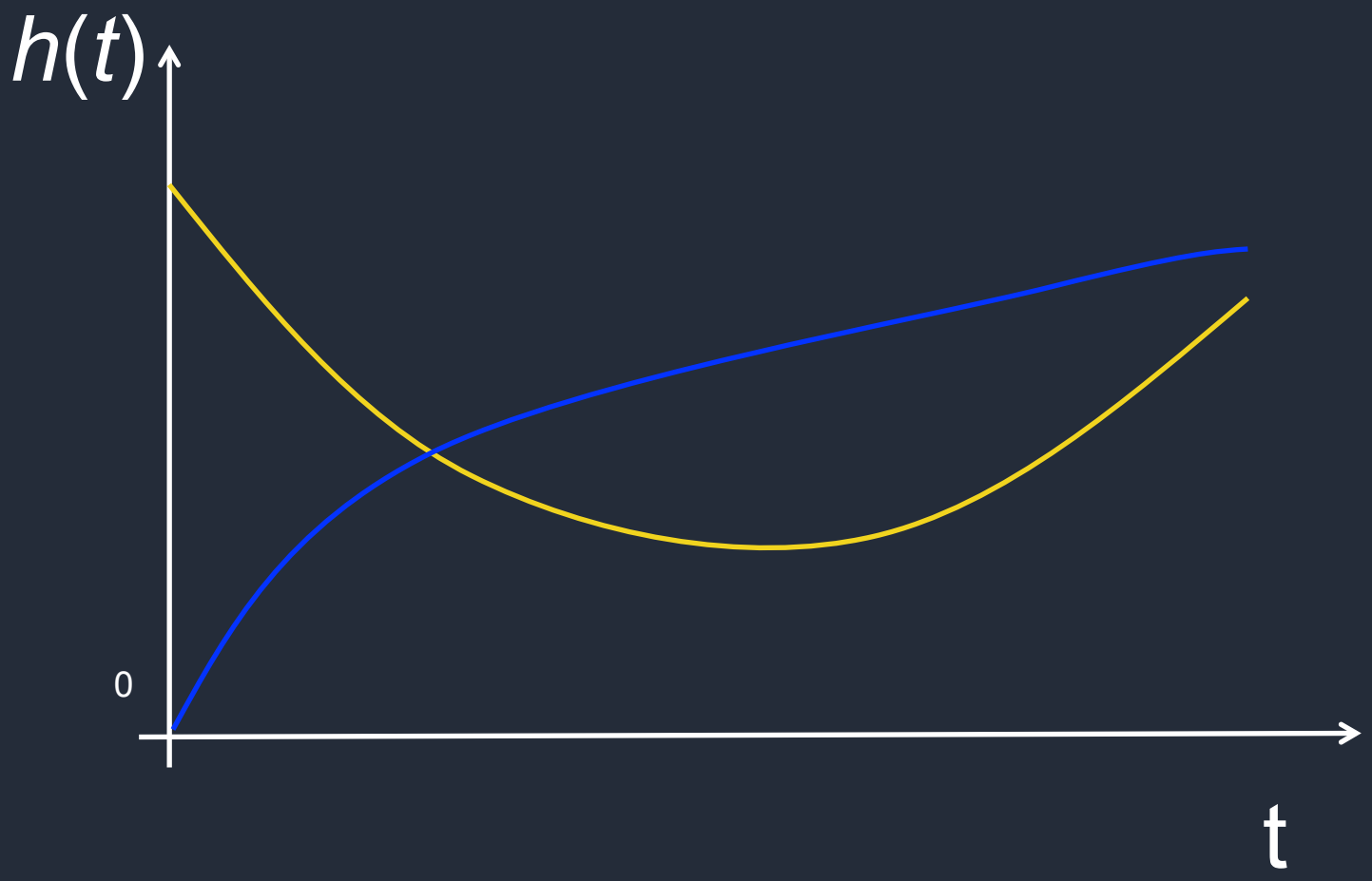
$$f(t) = - dS(t) / dt$$

Hazard function

= instant mortality rate knowing that the individual has survived to time t

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

Hazard function



Let’s practice!

- analysis of patterns of event times,

- comparison of distributions of survival times in different groups of individuals

- examining whether and by how much some factors affect the risk of an event of interest.

Literature

Tableman, 2012. Survival Analysis Using S/R* (very pedagogic)

→ https://tbrieder.org/epidata/course_reading/e_tableman.pdf

Online tutorials for survival analysis and survival curves plotting

→ http://www.sthda.com/english/wiki/survival-analysis-basics
→ http://www.sthda.com/english/wiki/survminer-r-package-survival-data-analysis-and-visualization

Several packages were developed in R

→ https://rviews.rstudio.com/2017/09/25/survival-analysis-with-r/

Thank you for your attention!

Mathilde CHEN
mathilde.chen@inra.fr