Introduction to survival analysis

Jorge N. Tendeiro

Department of Psychometrics and Statistics
Faculty of Behavioral and Social Sciences
University of Groningen

j.n.tendeiro@rug.nl
www.jorgetendeiro.com
jorgetendeiro/Seminar-2020-Survival-Analysis

Plan for today

Gentle introduction to survival analysis.

Source:
Harrell, F. E., Jr. (2015). *Regression Modeling strategies*, 2nd edition. Springer

Chapters:
17, 18, and 20.

Survival analysis (SA)

Data:
For which the *time until the event* is of interest.

- This goes beyond *logistic regression*, which focuses on the *occurrence* of the event.

Outcome variable:

- T = Time until the event.
- Often referred to as *failure time*, *survival time*, or *event time*.

Advantages of SA over typical regression models

- SA allows modeling units that did not fail up to data collection (*censored on the right* data).
- Regression could be considered to model the expected survival time. *But*:
  - ✓ Survival time is often not normally distributed.
  - ✓ $P(\text{survival} > t)$ is often more interesting than $\mathbb{E}(\text{survival time})$.

Censoring

We focus on

Three main functions

Recall that the outcome variable is $T =$ time until event.

- Survival function:

$$S(t) = P(T > t) = 1 - F(t),$$

  where $F = P(T \leq t)$ is distribution function of T.

- Cumulative hazard function:

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

- Hazard function:

$$\lambda(t) = \Lambda'(t)$$

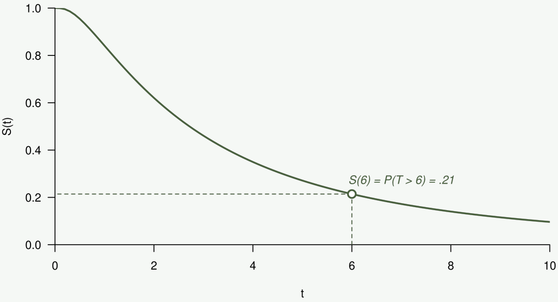
Survival function

$$\boxed{S(t) = P(T > t) = 1 - F(t)}$$

Example:
If event = death, then $S(t)$ = prob. that death occurs after time t.

Properties:

- $S(0) = 1$.
- Non-increasing function of t.



Cumulative hazard function

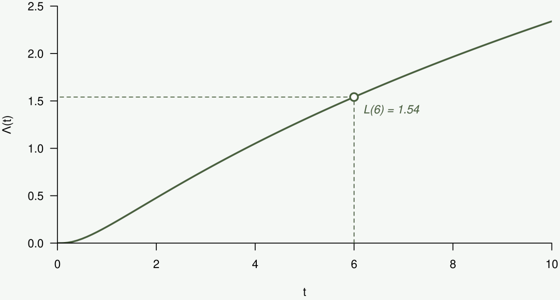
$$\boxed{\Lambda(t) = -\log(S(t))}$$

Idea:

Accumulated risk up until time t.

Properties:

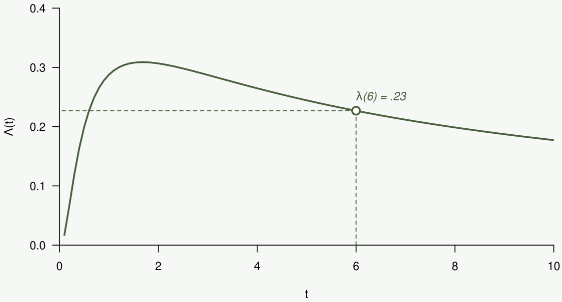
- $\Lambda(0) = 0$.
- Non-decreasing function of t.



Hazard function

$$\boxed{\lambda(t) = \Lambda'(t)}$$

Idea:
Instantaneous event rate.



Relation between the three function

All functions are related:
Any two functions can be derived from the third function.

- The three functions are equivalent ways of describing the same random variable ($T =$ time until event).

Given...	$S(t)$	$\Lambda(t)$	$\lambda(t)$
$S(t) =$	$\cdot$	$\exp(-\Lambda(t))$	$\exp\left(-\int_0^t \lambda(v)dv\right)$
$\Lambda(t) =$	$-log(S(t))$	$\cdot$	$\int_0^t \lambda(v)dv$
$\lambda(t) =$	$-\frac{S'(t)}{S(t)}$	$\Lambda'(t)$	$\cdot$