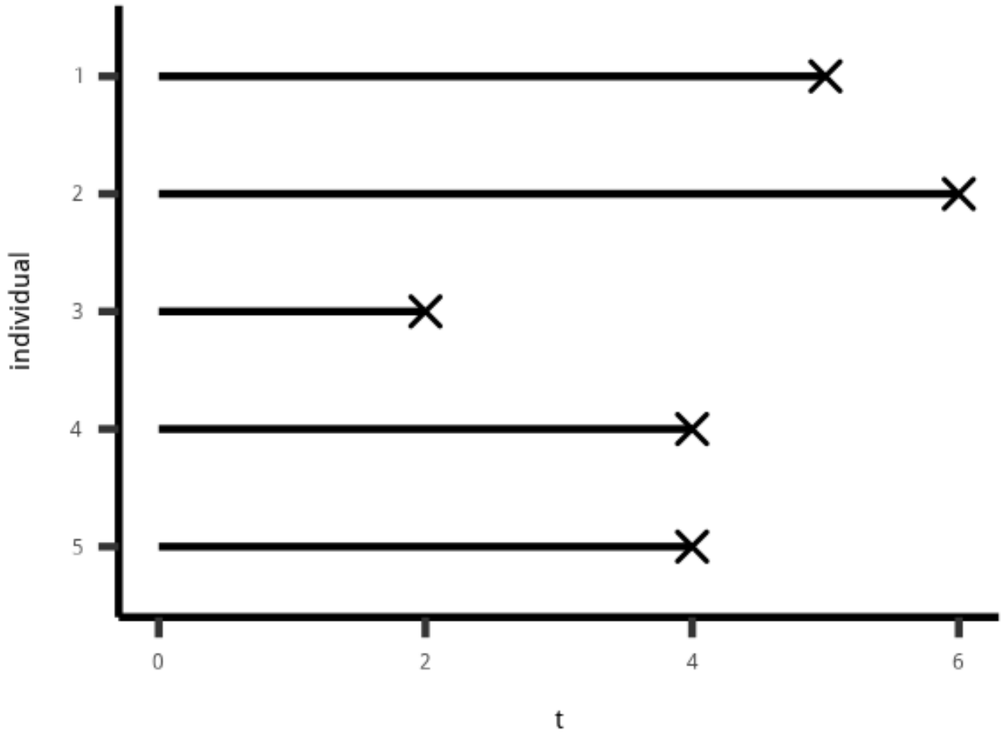
SURVIVAL ANALYSIS IN R

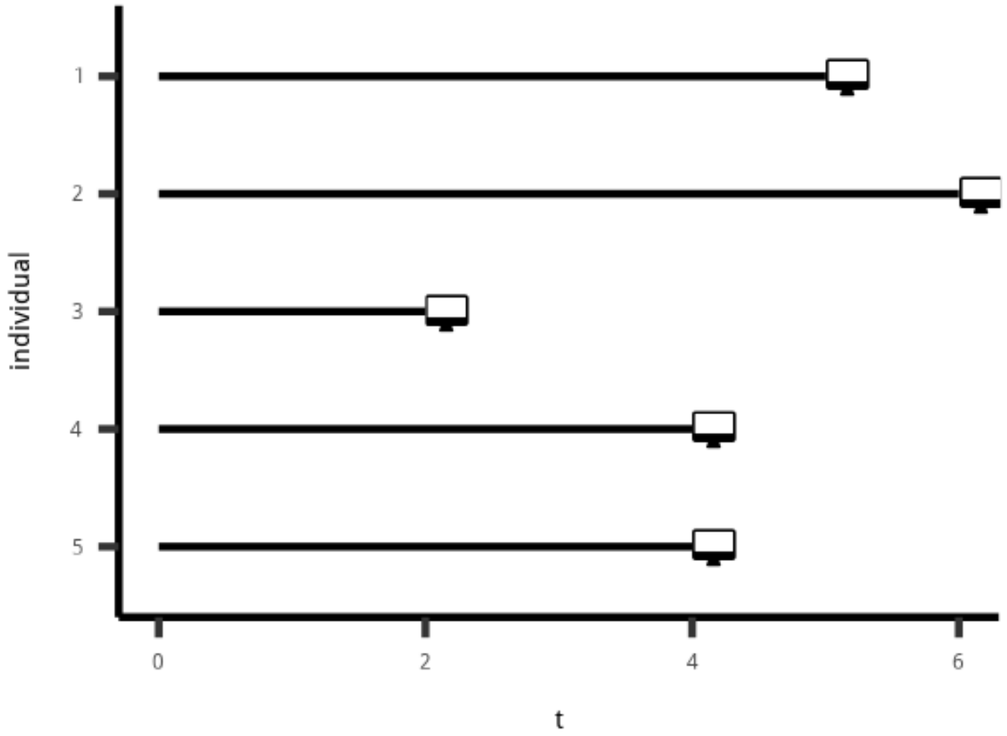
Survival Analysis / Time-to-Event Analysis in R

Heidi Seibold
Statistician at LMU Munich

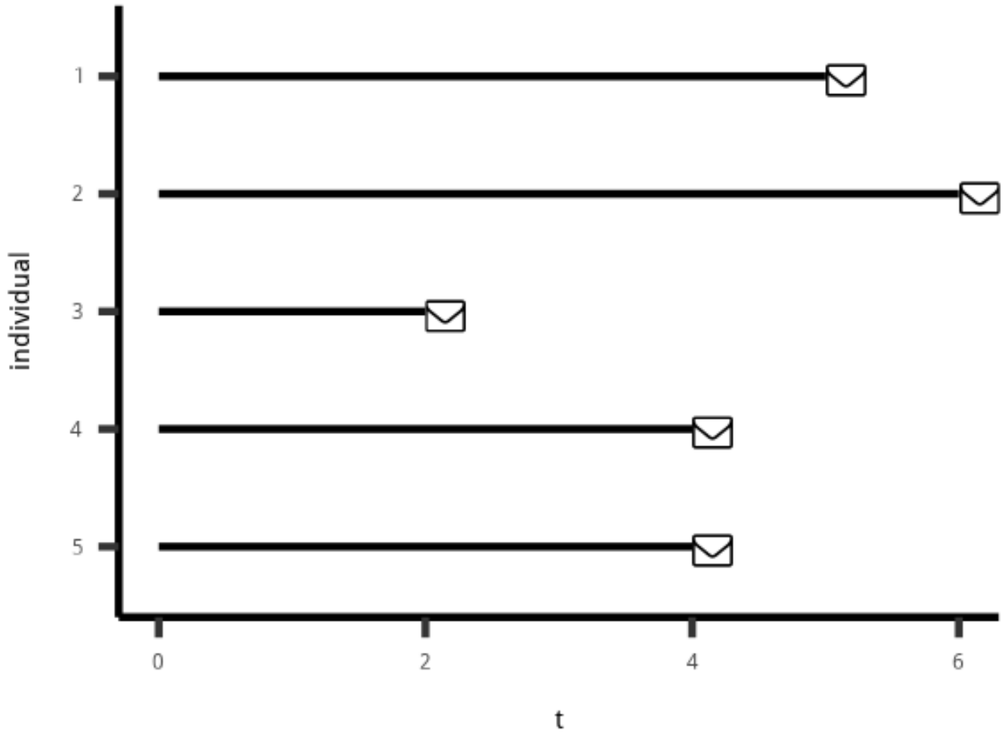
The term survival analysis



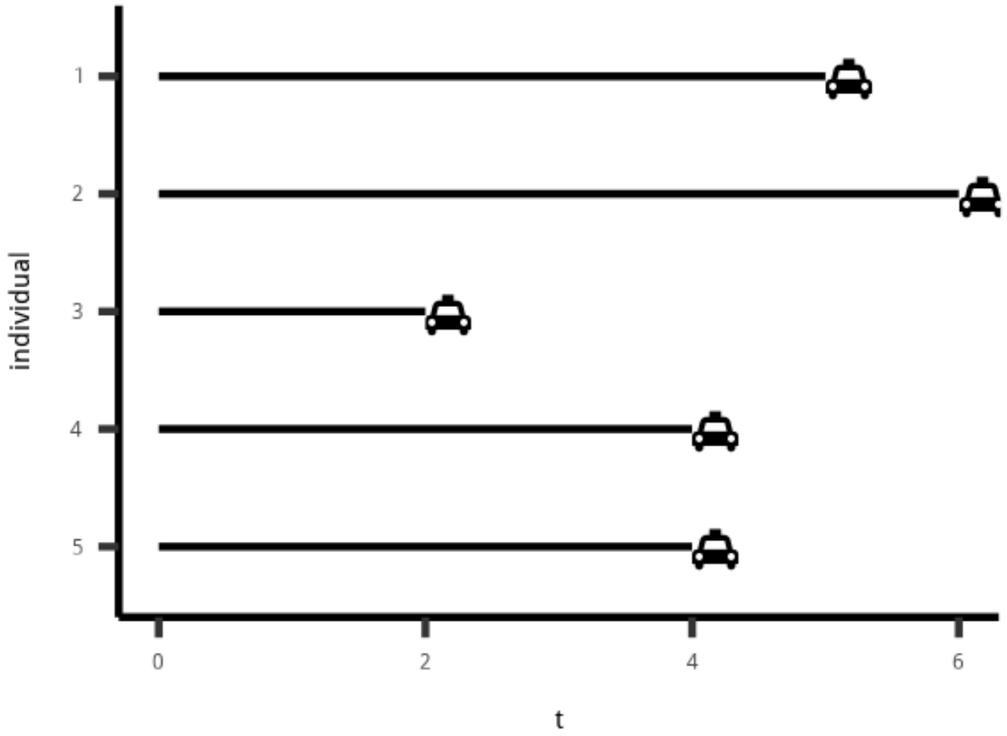
The term survival analysis



The term survival analysis



The term survival analysis



Data sets we will be using

GBSG2: time to death of 686 breast cancer patients

```
data(GBSG2, package = "TH.data")
```

UnempDur: time to re-employment of 3343 unemployed patients

```
data(UnempDur, package = "Ecdat")
```

Pro tip: to learn about a dataset in R, use the `help` function

```
help(UnempDur, package = "Ecdat")
```

SURVIVAL ANALYSIS IN R

Let's practice!

SURVIVAL ANALYSIS IN R

Why do we need special methods for time-to-event data?

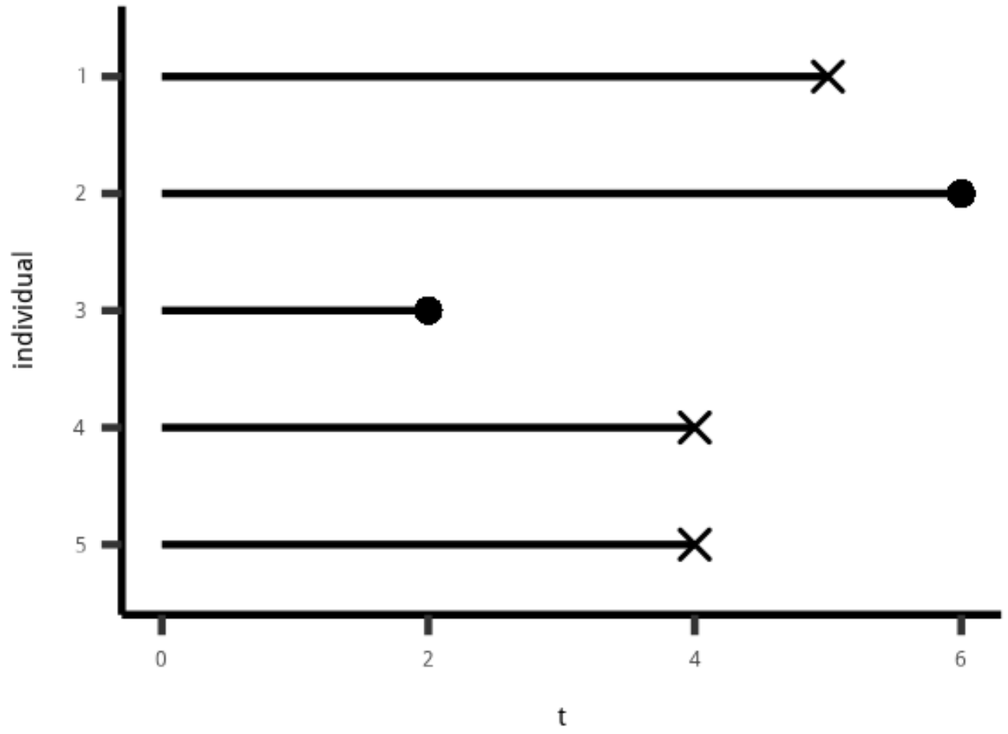
Heidi Seibold
Statistician at LMU Munich

Why survival analysis

- Times are always **positive**
- Different **measures** are of interest
- **Censoring** almost always an issue

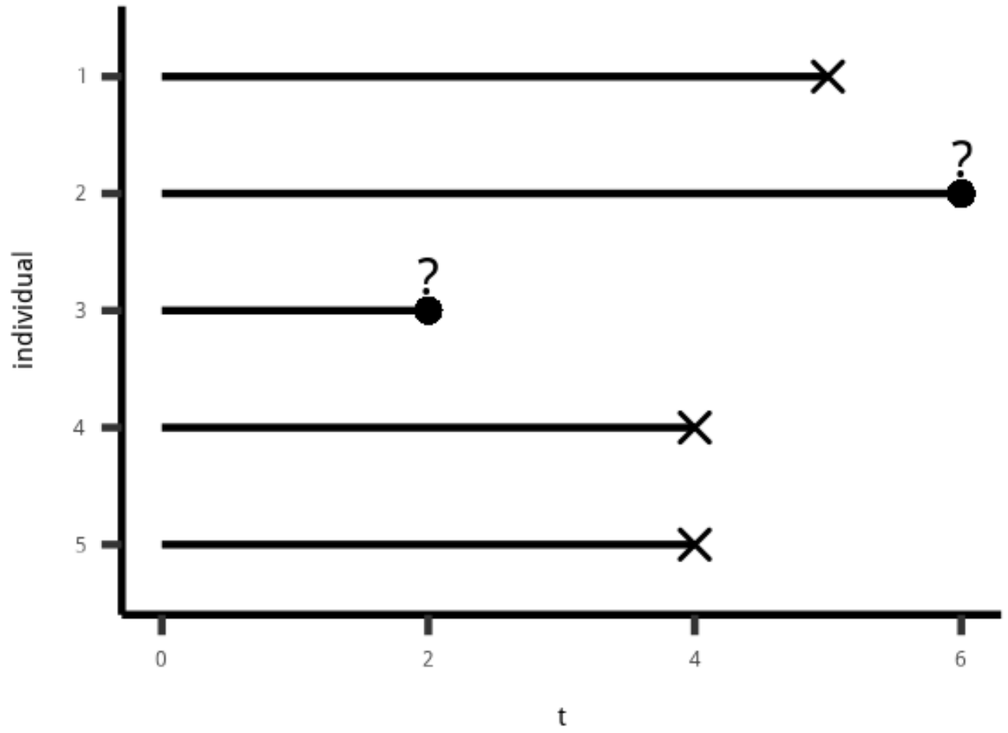
Why survival analysis

- Times are always **positive**
- Different **measures** are of interest
- **Censoring** almost always an issue

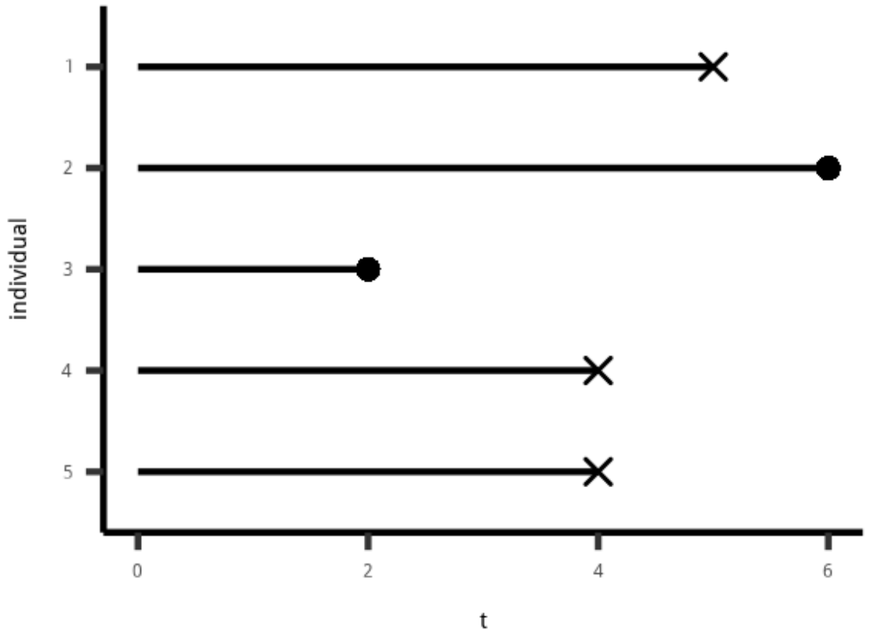


Why survival analysis

- Times are always **positive**
- Different **measures** are of interest
- **Censoring** almost always an issue



Creating Surv objects




```
time <- c(5, 6, 2, 4, 4)
event <- c(1, 0, 0, 1, 1)

library("survival")
Surv(time, event)
```

R packages

For all kinds of analyses:

```
library("survival")
```

For pretty visualisations:

```
library("survminer")
```

For more, see CRAN Task View: Survival Analysis

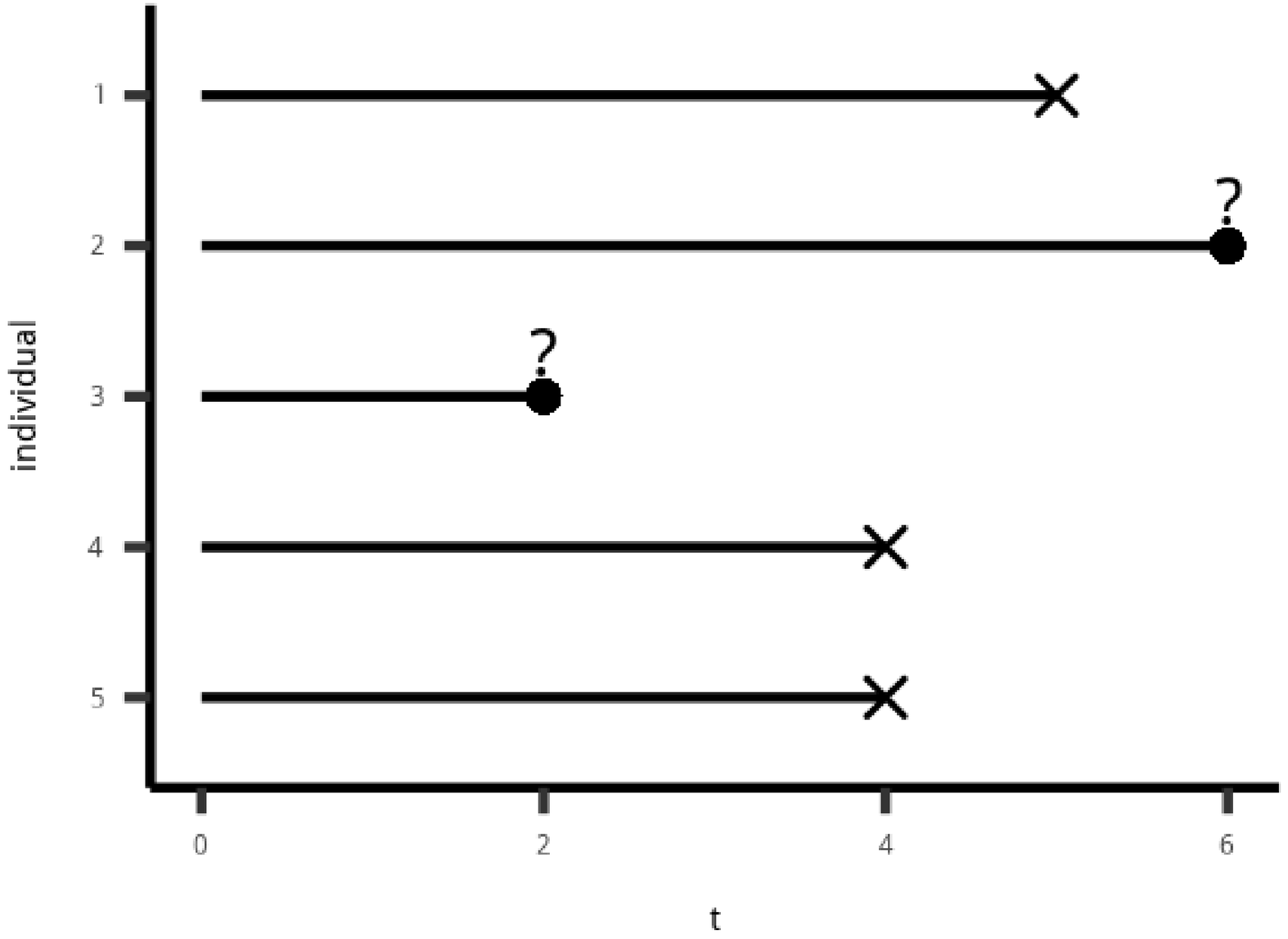
SURVIVAL ANALYSIS IN R

Let's practice!

SURVIVAL ANALYSIS IN R

The survival function

Heidi Seibold
Statistician at LMU Munich


individual
1
2
3
4
5
?
?
0
2
4
6
t

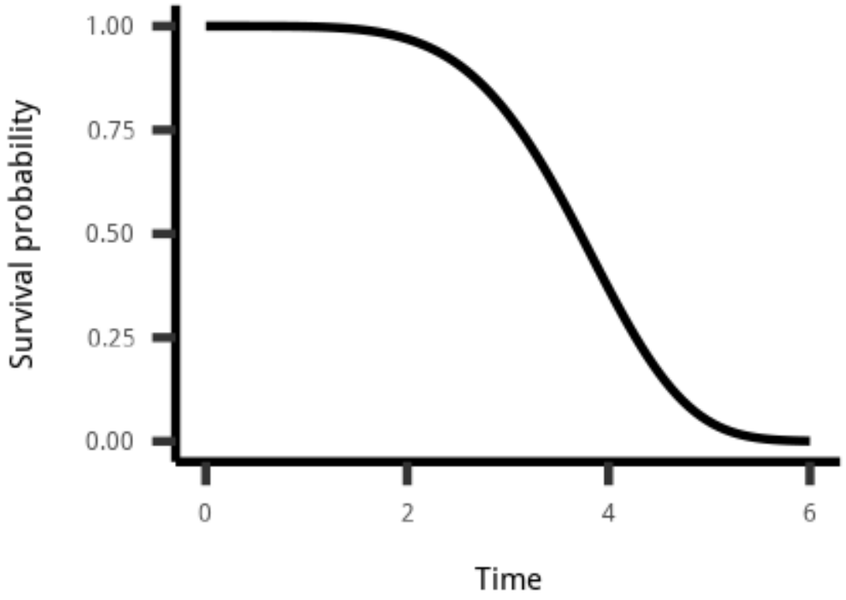
Survival analysis questions

- What is the probability that a breast cancer patient survives longer than 5 years?
- What is the typical waiting time for a cab?
- Out of 100 unemployed people, how many do we expect to have a job again after 2 months?

Survival function

THEORY

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



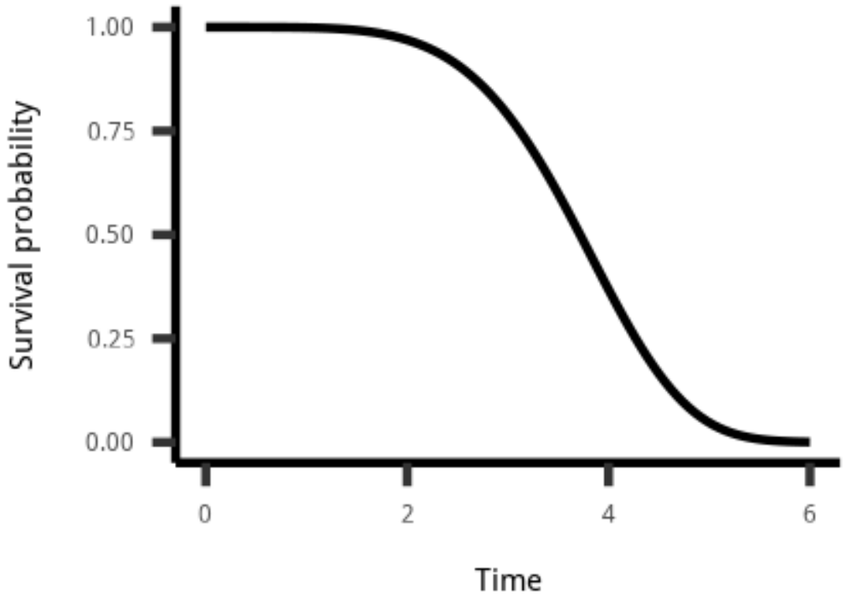
INTERPRETATION

Probability that duration is longer than t.

Survival function

THEORY

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



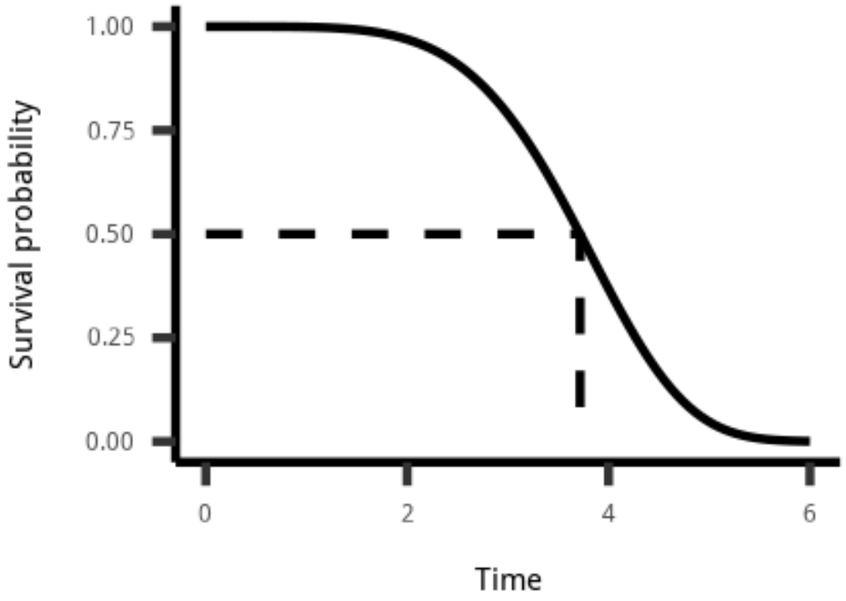
INTERPRETATION

Probability that duration is longer than t.

Examples:

- Probability to survive beyond time point t.
- Probability that the cab takes more than t minutes to arrive.

Survival function



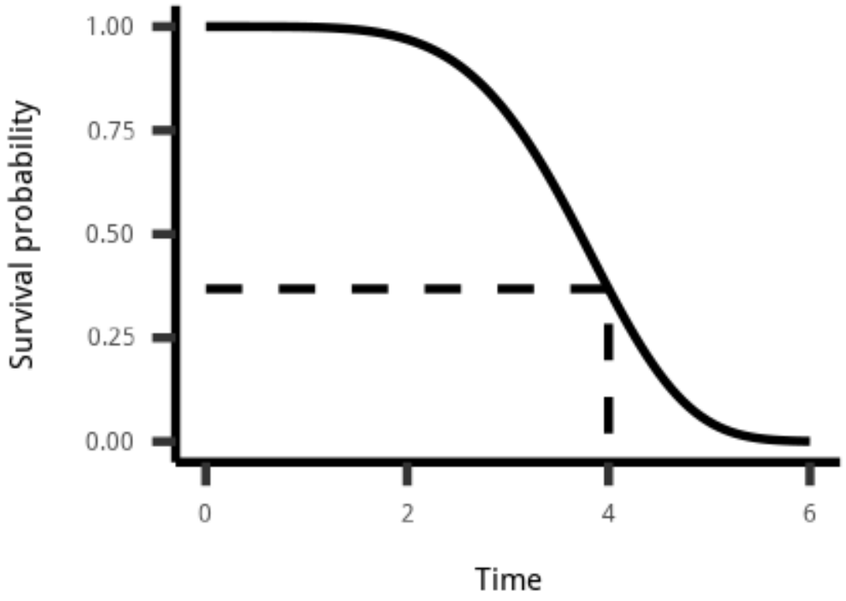
INTERPRETATION

The median duration is t.

Examples:

- The median survival time is 3.7 years.
- Median time until the cab arrives is 3.7 minutes.

Survival function



INTERPRETATION

$100 \cdot \hat{S}(t)$ percent of durations are longer than t.

Examples:

- 37 percent of all patients survive longer than 4 years. 63 percent die within the first 4 years.
- Out of 100 cabs, 37 take more than 4 minutes to arrive.

SURVIVAL ANALYSIS IN R

Let's practice!