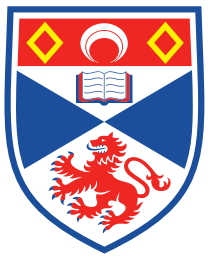
An introduction to survival analysis

Georg Wölflein

School of Computer Science, University of St Andrews

January 15, 2022



Contents

What is time-to-event (TTE) data?

We can measure **time** in:

- years
- months
- seconds

What is time-to-event (TTE) data?

We can measure **time** in:

- years
- months
- seconds

The **event** could be:

- death from disease
- product failure
- losing a customer

What is time-to-event (TTE) data?

We can measure **time** in:

- years
- months
- seconds

The **event** could be:

- death from disease
- product failure
- losing a customer

} must be a binary variable

What is time-to-event (TTE) data?

We can measure **time** in:

- years
- months
- seconds

The **event** could be:

- death from disease
- product failure
- losing a customer

} must be a binary variable

TTE data consists of $(time, \overbrace{event}^{\text{yes/no}})$ tuples.

Time-to-event (TTE) data

TTE analysis is also known as:

- survival analysis
- failure time analysis
- reliability theory (engineering)
- duration modelling (economics)
- event history analysis (sociology)

Time-to-event (TTE) data

TTE analysis is also known as:

- survival analysis
- failure time analysis
- reliability theory (engineering)
- duration modelling (economics)
- event history analysis (sociology)

Use cases for TTE analysis:

- TODO

Example: Covid-19 treatment trial

A randomised controlled trial ($n = 4$) was conducted to assess the efficacy of drug ABC in treating Covid-19. This is what happened to the patients:

patient	received ABC?	outcome
1	yes	died from Covid-19 on day 14
2	no	dropped out of the study after day 3
3	yes	died by a lightning stroke on day 5
4	no	survived the study (90 days)

Example: Covid-19 treatment trial

patient	received ABC?	outcome
1	yes	died from Covid-19 on day 14
2	no	dropped out of the study after day 3
3	yes	died by a lightning stroke on day 5
4	no	survived the study (90 days)

The **time** is the number of days since testing positive for Covid-19.
The **event** is whether the patient died due to Covid-19.

Example: Covid-19 treatment trial

patient	received ABC?	outcome
1	yes	died from Covid-19 on day 14
2	no	dropped out of the study after day 3
3	yes	died by a lightning stroke on day 5
4	no	survived the study (90 days)

The **time** is the number of days since testing positive for Covid-19.
The **event** is whether the patient died due to Covid-19.

Time-to-event data

patient	time	event
1	14	yes
2	?	?
3	?	?
4	?	no

Example: Covid-19 treatment trial

patient	received ABC?	outcome
1	yes	died from Covid-19 on day 14
2	no	dropped out of the study after day 3
3	yes	died by a lightning stroke on day 5
4	no	survived the study (90 days)

The **time** is the number of days since testing positive for Covid-19.
The **event** is whether the patient died due to Covid-19.

Time-to-event data

patient	time	event
1	14	yes
2	$[0, 3]$	no
3	$[0, 5)$	no
4	$[0, 30]$	no

Censoring

We just saw an example of **right-censored** data.

Survival function

$$S(t) = \Pr(T > t)$$

Supposing an individual survived until time t, the **hazard function** expresses the probability of surviving an additional time dt.

Hazard function

$$\lambda(t) = \lim_{dt \to 0} \frac{\Pr(t \leq T \leq dt | T \geq t)}{dt} = \lim_{dt \to 0} \frac{\Pr(t \leq T \leq dt)}{dt \cdot S(t)}$$