

An introduction to survival analysis

Georg Wölflein

School of Computer Science, University of St Andrews

January 15, 2022



University
of
St Andrews

Contents

- 1 Time-to-event data
- 2 Survival and hazard functions

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 $(time, \overset{\text{yes/no}}{\text{event}})$ 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:

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)

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	?	?
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

Let T be a continuous random variable representing time.

The **survival function** $S(t)$ is the probability that an individual will survive for a time greater than t .

Survival function

Let T be a continuous random variable representing time.
The **survival function** $S(t)$ is the probability that an individual will survive for a time greater than t .

Survival function

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

Hazard function

The **hazard function** expresses the *instantaneous rate of occurrence* of the event.

Hazard function

The **hazard function** expresses the *instantaneous rate of occurrence* of the event.

Supposing an individual survived until time t , it expresses the probability of dying within a short additional time dt , per unit time.

Hazard function

The **hazard function** expresses the *instantaneous rate of occurrence* of the event.

Supposing an individual survived until time t , it expresses the probability of dying within a short additional time dt , per unit time.

Hazard function

$$\lambda(t) = \frac{\Pr(\text{ } | \text{ })}{\text{ }}$$

Hazard function

The **hazard function** expresses the *instantaneous rate of occurrence* of the event.

Supposing an individual survived until time t , it expresses the probability of dying within a short additional time dt , per unit time.

Hazard function

$$\lambda(t) = \frac{\Pr(t \leq T \leq t + dt | \quad)}{dt}$$

Hazard function

The **hazard function** expresses the *instantaneous rate of occurrence* of the event.

Supposing an individual survived until time t , it expresses the probability of dying within a short additional time dt , per unit time.

Hazard function

$$\lambda(t) = \frac{\Pr(t \leq T \leq t + dt | T \geq t)}{dt}$$

Hazard function

The **hazard function** expresses the *instantaneous rate of occurrence* of the event.

Supposing an individual survived until time t , it expresses the probability of dying within a short additional time dt , **per unit time**.

Hazard function

$$\lambda(t) = \frac{\Pr(t \leq T \leq t + dt | T \geq t)}{dt}$$

Hazard function

The **hazard function** expresses the *instantaneous rate of occurrence* of the event.

Supposing an individual survived until time t , it expresses the probability of dying within a short additional time dt , per unit time.

Hazard function

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T \leq t + dt | T \geq t)}{dt}$$

Hazard function

The **hazard function** expresses the *instantaneous rate of occurrence* of the event.

Supposing an individual survived until time t , it expresses the probability of dying within a short additional time dt , per unit time.

Hazard function

$$\begin{aligned}\lambda(t) &= \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T \leq t + dt | T \geq t)}{dt} \\ &= \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T \leq t + dt)}{dt \cdot \Pr(T \geq t)}\end{aligned}$$

Hazard function

The **hazard function** expresses the *instantaneous rate of occurrence* of the event.

Supposing an individual survived until time t , it expresses the probability of dying within a short additional time dt , per unit time.

Hazard function

$$\begin{aligned}\lambda(t) &= \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T \leq t + dt | T \geq t)}{dt} \\ &= \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T \leq t + dt)}{dt \cdot S(t)}\end{aligned}$$