Survival Analysis[1]

STA312 Spring 2019

[1] See last slide for copyright information.

Survival Time Data

- Data could be from cancer patients who have undergone some treatment.
- The outcome (dependent variable, response variable) will be length of time until death.
- Does not have to be a bad event, and it does not have to be medical.
- Response variable could be time required for an unemployed person to find work.
- Length of time a prisoner actually spends in jail.
- Time required for a lightbulb to burn out.
- Generally, "time-to-event" data.
- But much of the terminology and notation assumes it's survival time – time until death.

What makes survival analysis special?

- Time T is a continuous random variable with $P(T > 0) = 1$.
- That is, it's positive-valued.
- More important is *censoring*.

Censoring

Some of the times cannot be observed exactly.

- In a cancer study, it's impractical to wait for all the patients to die.
- Patient 172 was diagnosed 6 years ago, and she's still alive.
- The study's over.
- What do you record for Patient 172? A six?
- That's no good. We know she lived more than 6 years.
- We just don't know how much more.
- Giving her a 6 would bias the results downward, toward shorter average survival times.
- How about discarding her data altogether, because it's incomplete?
- That's even worse.

What do censored data look like?

- Let T^* represent time to failure.
- Let U represent censoring time.
- We observe $T = \min(T^*, U)$.
- And an indicator for whether failure occurred: $\delta = 1$ if the uncensored, and zero if censored.
- All this is $i = 1, \ldots, n$, so we observe n pairs $(T_1, \delta_1), (T_2, \delta_2), \ldots, (T_n, \delta_n)$.
- A simple data file might look like this:

Patient	Time	Uncensored (delta)
1	5	1
2	6	0
3	8	1
4	3	1
5	22	1

Types of censoring

- **Right censoring**: Event happened after some specified time.
- Example: Ex-con dies of drug overdose before being re-arrested.
- **Left censoring**: Event happened before some specified time.
- Example: Person is known to have died before some date because that's when the obituary appeared.
- **Interval censoring**: Event happened between time A and time B.
- Example: Cancer came back between two monthly check-ups.

More types of censoring

- **Type I censoring**: Censoring times are pre-specified. In a smoking cessation study, participants are followed until they relapse or 180 days, whichever comes first.
- **Type II censoring**: Subjects are followed until a specified fraction fail. Mostly used in industrial settings.
- **Random censoring**: Drop-out happens according to some probability distribution, usually unknown.

Random censoring

- In random censoring, it is critical that the cause of drop-out be unrelated to the process being studied.
- For example, if a paient withdraws from the study (before death) because she is too sick to participate, the result can be biased estimation and inference.
- A good design is common in medical studies. Patients are continuously enrolled in the study, and anyone still alive when the study ends is censored.

- In this course, we will focus on *right censored* data with *random censoring*.

Roadmap

- Single sample, parametric.
- Single sample, non-parametric.
- With explanatory variables, parametric.
- With explanatory variables, non-parametric.

Tools

- Maximum likelihood.
- A little large sample theory.
- R.

Copyright Information