Survival Analysis

Alexis Destefano (she/her),
working with
Ethan Ancell and Sherry Ren

1. Introduction

- Survival analysis–time to event data
- Our focus: the math of it all
- Sherry Ren will present an application

2. Basics

- Time to event data
  - beginning of study -> event of interest occurs
  - example: first illegal pain pill usage -> first usage of heroin
- Useful in biostatistics

3. The variables

- τ_i-> true time at which *i*th observation dies
- C_i -> time at which *i*th observation is notes as censored
- $\tilde{\tau}_i$->
    - last time at which *i*th observation is noted to be alive, whether it “died” or was censored
    - minimum of C_i, point at which data is noted to be censored, and τ_i

4. Survival Curve

- Denoted $S(t)$
- Complement of CDF: $S(t) = 1 - F(t)$
- Represents probability of survival at every point on x-axis
- Is central object of study in survival analysis

5. Hazard Function

- Represents instantaneous risk of experiencing event at time *t* among those still at risk--rate of change of survival function
- Low when survival curve isn't dropping much
- Spikes when survival function has steep decline
- Not our focus, but still useful for understanding time to event data

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

6. In practice

- We have access to: $\tilde{\tau}_i$ and C_i
- $\tilde{\tau}_i$: last time at which *i*th observation was noted to be alive
- C_i: time when *i*th observation is censored
- With these, we can *estimate* survival curve
- Estimate options: Naive and Kaplan Meier estimators

$$\tilde{\tau}_j = \min(\tau_j, c_j)$$

7. Naive estimate

- Number of objects in study that haven't "died"/been censored divided by number of objects that haven't been censored at time *t*
- Ignores observations with censoring times preceding *t*
- Inspires Kaplan Meier estimator on next slide

$$\hat{S}_{\text{naive}}(t) = \frac{1}{m(t)} \sum_{k:c_k \geq t} X_k = \frac{|\{1 \leq k \leq n \, : \, \tilde{\tau}_k \geq t\}|}{|\{1 \leq k \leq n \, : \, c_k \geq t\}|} = \frac{|\{1 \leq k \leq n \, : \, \tilde{\tau}_k \geq t\}|}{m(t)}$$

8. Kaplan Meier

- Better for more censoring
- Recursive relationship–$S(t_{prev})$

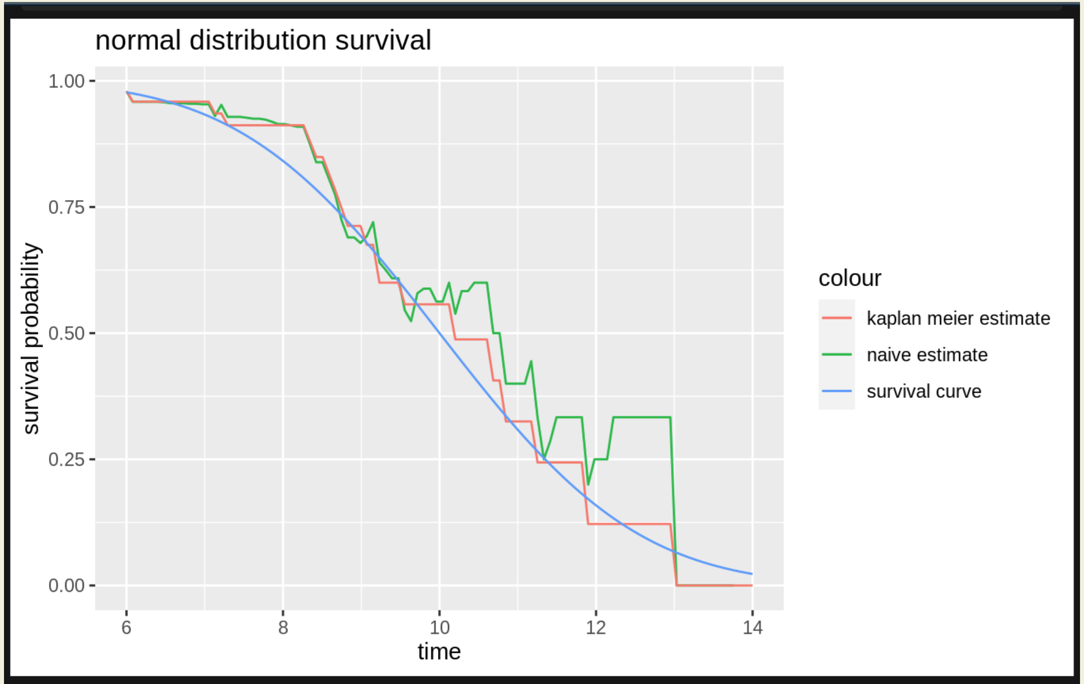
$$
\begin{aligned}
S(t) &= P(T > t) \\
&= P(T > t | T \geq t) P(T \geq t) \\
&= [1 - P(T \leq t | T \geq t)] \, P(T \geq t) \\
&= [1 - P(T = t | T \geq t)] \, P(T \geq t) \\
&= [1 - P(T = t | T \geq t)] \, P(T > t_{\text{prev}}) \\
&= [1 - P(T = t | T \geq t)] \, S(t_{\text{prev}})
\end{aligned}
$$

9. Kaplan Meier, defined

Probability an object will be alive at time *t*
One minus the proportion of events at time *t* among those at risk, multiplied by probability of survival at previous time

10. Our measurements

Kaplan Meier estimate wins!



11. Questions?

12. Thank you:)