Survival Analysis and Censored Data

Lecture 12

Xiaofei Shi

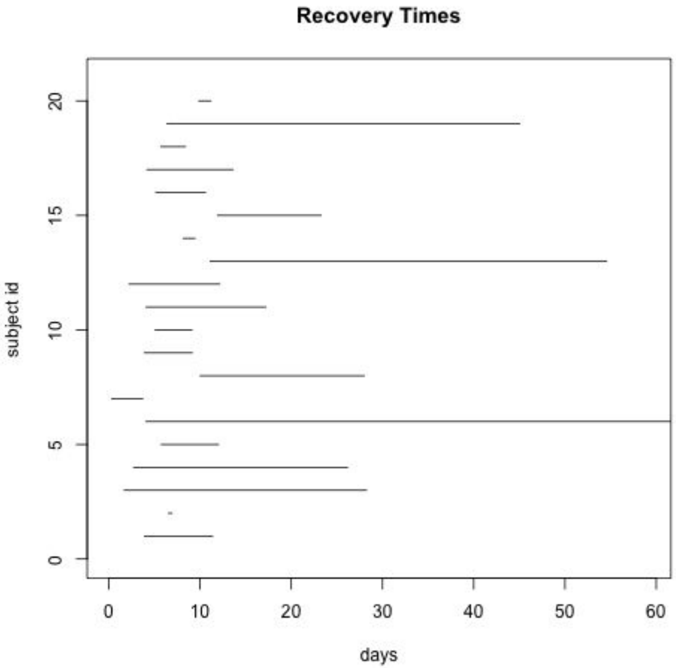
Learning Objectives

- Understand the properties of survival data
- Understand the challenges with censored data
- Understand the basic models in survival analysis

Survival data

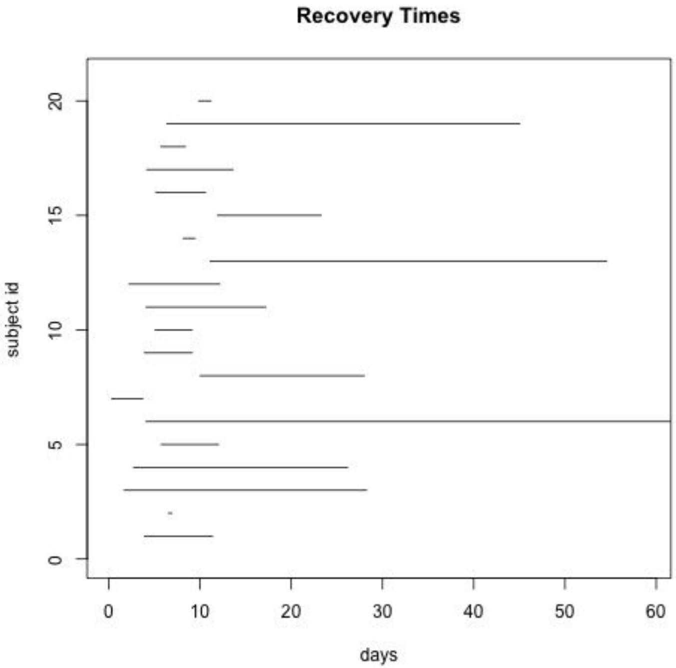
- Time to part failure
- Time to recovery
- Time to death from a disease

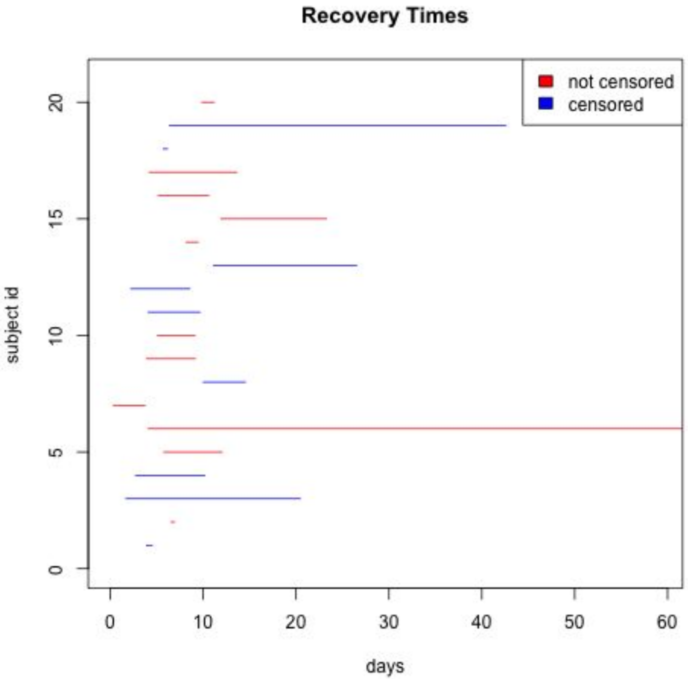
Simulated Survival Data



- Recruiting window
- Duration of recovery
- Perfect information on start/end times

Simulation - right censoring in survival data





What are the impacts of censoring?

- If you’re just estimating the recovery time?
- If you’re running an experiment?
- Is it better to just drop the censored data points?

Let's simulate some data!

- Simulate survival data
  - Make sure the survival time is positive
  - Make the survival times random
  - How would you do this?

Source of bias from censored data

- Censored data only informs you that $T_i > t$ instead of T_i
- Censored data should not be treated as “recovered”

Asking the right question in survival analysis

- Given you are sick, are you more likely to ask?
  - $P(T_i > t)$
  - $P(t < T_i <= t + \varepsilon \mid T_i > t)$
  - $P(T_i <= t)$
  - $P(t < T_i <= t + \varepsilon \mid T_i <= t + \varepsilon)$

Kaplan-Meier Estimator

$$P(T_i > t) = P(T_i > t | T_i > \alpha t) P(T_i > \alpha t) \text{ where } \alpha < 1$$

Kaplan-Meier Estimator

$$P(T_i > t) = P(T_i > t | T_i > \alpha t) P(T_i > \alpha t) \text{ where } \alpha < 1$$

$$\hat{P}(T_i > t | T_i > \alpha t) = \frac{\sum_i 1[T_i > t]}{\sum_i 1[T_i > \alpha t]}$$

Dealing with censoring - what goes into the sum?

$$\hat{P}(T_i > t | T_i > \alpha t) = \frac{\sum_i 1[T_i > t]}{\sum_i 1[T_i > \alpha t]}$$

if T_i is censored at C_i

- $\alpha t < t < C_i$
- $C_i < \alpha t < t$
- $\alpha t < C_i < t$

Trust but verify with simulations!

- Simulate survival data
- Demonstrate that the Kaplan-Meier estimator “works”

Hazard function

$$h(t) = \lim_{\epsilon -> 0} \frac{P(t < T < t + \epsilon | T > t)}{\epsilon} = \frac{f(t)}{S(t)}$$

where $f(t) = \frac{\partial}{\partial t} F(t) = -\frac{\partial}{\partial t} S(t)$

Cox Proportional Hazard Model

$$h(t|X) = \lambda(t)\exp(X\beta)$$

Cox Proportional Hazard Model

$$h(t|X) = \lambda(t)\exp(X\beta)$$

$$\frac{h(t|X=1)}{h(t|X=0)} = \frac{\lambda(t)\exp(\beta)}{\lambda(t)\exp(0)} = \exp(\beta)$$

What would you validate for the model?

- What assumptions do we have?