Confidence intervals with Bootstrapping

a) Introduction

Prof. Dr. Jan Kirenz

We use **randomization** to **test claims**

Randomization tests are best suited for modeling experiments where

- the treatment (explanatory variable) has been **randomly assigned** to the observational units
- and there is an attempt to answer a simple **yes/no** research question

Examples:

1. Does this vaccine make it less likely that a person will get malaria?
2. Does drinking caffeine affect how quickly a person can tap their finger?
3. Can we predict whether candidate A will win the upcoming election?

Now we want to estimate **population parameters**

- Instead of testing a **claim** (yes/no),
- the goal now is to estimate the unknown value of a **population parameter**.

Examples:

1. How much less likely am I to get malaria if I get the vaccine?
2. How much faster (or slower) can a person tap their finger, on average, if they drink caffeine first?
3. What proportion of the vote will go to candidate A?

Bootstrapping is a simulation method

- The focus is on a **single proportion**
- Ideally, sample data was generated through **random sampling** from a population.
- our goal with bootstrapping is to **understand variability** of a **statistic**.

Randomization vs bootstrap

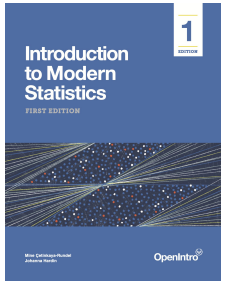
Randomization tests:

- modeled how the **statistic** would **change** if the treatment had been allocated differently

The **bootstrap**

- will model how a **statistic varies** from one **sample** to another **sample** taken from the population.
- This will provide information about **how different** the **statistic** is from the **parameter of interest**.

Resources



The content of this lecture is mainly based on the excellent book (can be accessed for free)

- “Introduction to Modern Statistics” by Mine Çetinkaya-Rundel and Johanna Hardin (2021)

  https://openintro-ims.netlify.app/index.html