

# Assignment 2: Potential outcomes

✓ Published

 Edit

⋮

## Instructions

All the problems in this assignment should be solved and handed in **individually**. You should be prepared to answer questions about your solutions yourself. The full set of solutions should be submitted as a single PDF document in Canvas. Feel free to use any software of your choosing (or pen and paper) for preparing illustrations and drawings.

## Problem

The government is investigating the effects of an information campaign aimed at **increasing the number of hours** citizens spend **exercising** per year. The campaign was performed and some results were collected about participants who remained in the study by the end. In particular, analysts collected a dataset about *who was exposed* to the campaign, their *age* (in years) and the *number of hours* they exercised in a year. You will find this data in [data\\_hours\\_2021.csv](#) . You are now tasked with analyzing this data and figuring out whether the campaign was worth the effort.

- A. Give notation to the variables in the dataset
- B. Based on the given data, can you tell whether the observations came from a randomized experiment? If you knew that the only possible cause of treatment assignment was "age", how would you check whether the intervention was randomized? What does the observed data tell you?
- C. Under the assumption that the data comes from a randomized experiment with consistency and without interference:
  - I. What is the average treatment effect? How certain are you about this estimate?
  - II. Was the campaign worth it for middle-aged people? What about people over 60 years of age?
  - III. For whom did the campaign have the largest effect?
- D. You suspect that a fraction of exposed subjects under 40 years of age may have found out that they were part of an experiment and left the study out of embarrassment if their exercise hours were lower than the study average. If true, how has that impacted your results? Show using derivations and/or numerical simulations what the effect of such a pattern may be. Be clear about any assumptions you make.
- E. Discuss other potential difficulties with identifying & estimating the effect of this campaign.

For all parts in C, define the causal parameters you want to estimate and describe your identification & estimation strategy.