Home-Assignment

This is an individual home assignments. You are allowed to support each other, but **identical or copied submissions will fail**. The home assignment (code and answers to exercise below) has to be submitted until Sunday June 13, 2021, by email to anthony.strittmatter@unibas.ch.

Data Description:
Low voter turnout can be a problem for the legitimation of political representation. There are attempts to increase voter turnout with so-called get-out-the-vote (GOTV) calls. A GOTV call includes information about the election date and the importance of voting. In 2002, an experiment was conducted in Michigan and Iowa to examine the effect of GOTV calls on voter turnout.[1] Electorates were randomly divided into two groups. One group with 60'000 electorates received a GOTV call, but not all of them were able to answer the call. The other group of almost 2 million electorates was the control group, which did not receive a call.

You find in the Github folder the data set `mobilization.RData`, which contains the following variables:

- *vote02*: voted in 2002, 0 = no, 1 = yes
- *treat_real*: 1 = received a GOTV call, 0 = did not receive a GOTV call
- *contact*: 1 = answered a GOTV call, 0 = not reached by telephone (or not called)
- *persons*: number of persons in the household
- *newreg*: newly registered voter
- *age*: age (in years)
- *female*: 1 = women, 0 = men
- *vote00*: voted in 2000, 0 = no, 1 = yes
- *vote98*: voted in 1998, 0 = no, 1 = yes
- *state*: 0 = Michigan, 1 = Iowa
- *competiv*: 1 = competitive district, 2 = uncompetetive district
- *comp_mi*: = competiv*(1-state)
- *comp_ia*: = competiv*state

[1] Arceneaux, K., Gerber, A. S. and Green, D. P. 2006. *Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment* Political Analysis, Vol. 14, No. 1, pp. 37-62

Exercise:

Answer all questions below using the `mobilization` data.

1. How many people voted in the 2002 election?

2. What is the share of electorates who answered the GOTV call?

3. Why do we use different samples to train and test prediction models?

4. Predict the probability to vote in the 2002 election based on the observed covariate using Classification Tree (you find information about how this works in the help file of the package *rpart*). Which model accuracy measure do you use? How accurate is your predicted model in the test data? Which tuning parameters do you have and how do you select them?

5. Explain the concept of selection bias at the example of answering the GOTV call (e.g. think of the causal effect of answering the GOTV call on the probability to vote in the 2002 election).

6. Estimate the causal effect of answering the GOTV call on the probability to vote in the 2002 election using either the Post-Double-Selection procedure or Double/Debiased-Machine-Learning. Explain briefly how the chosen estimation approach works. How large is the estimated effect?