

# Homework 3

Introduction to Causal Inference 096260  
Winter 2018/2019

**Submission date: January 10 , 2019**

In this homework assignment you will experience working with actual (partially simulated) data, as preparation for the final project. You are more than encouraged to work in pairs.

## The Data

Together with this document are two data files ( “*data1.csv*” and “*data2.csv*”). Each contains  $n = 4802$  samples, with the following properties:

1.  $X_1 \dots X_{58}$  : Covariates of the model (same in both datasets).
2.  $T$ : The treatment assignment of the model (binary).
3.  $Y$ : The outcome of the model.

## Requirements

Your task is to estimate the Average Treatment effect on the Treated ( $ATT$ )<sup>1</sup>, in both of the datasets. For each dataset you should first calculate the propensity score, and then estimate the ATT using the following approaches:

- (i) Inverse Propensity Score Weighting (IPW)
- (ii) S-learner
- (iii) T-learner
- (iv) Matching

The grade of this assignment will have two parts: 85% of the grade will be on your ability to calculate estimates for each of the models. Then, we ask you to give your best estimate of the true ATT, using any of the methods above or any other method you wish to use. The remaining 15%

---

<sup>1</sup> $ATT \equiv \mathbb{E}[Y_1 - Y_0 | T = 1]$

will be competition between the groups on the most accurate final estimate.

You are free to use any programming language (we recommend on using Python or R) and any model for estimating the propensities, the outcomes, and matching. Please think carefully which model you wish to use, as this is what will make a difference between the more accurate and less accurate estimates.

## Submission Instructions

You are required to submit a single **zip** file containing:

1. the ATT results in a single **csv** file named **ATT\_results.csv** with the format:

$$\begin{array}{l} Type, data1, data2 \\ 1, ATT_1^1, ATT_1^2 \\ 2, ATT_2^1, ATT_2^2 \\ 3, ATT_3^1, ATT_3^2 \\ 4, ATT_4^1, ATT_4^2 \\ 5, ATT_5^1, ATT_5^2 \end{array}$$

where  $ATT_i^j$  is the ATT of approach  $i$  on dataset  $j$ , and  $ATT_5^j$  is your final estimate (for the competition).

2. The propensity scores in a single **csv** file named **models\_propensity.csv** with the format:

$$\begin{array}{l} data1, e_1^1, \dots, e_n^1 \\ data2, e_1^2, \dots, e_n^2 \end{array}$$

where  $e_i^j$  is the propensity score of sample  $i$  in dataset  $j$ .

3. All of your code should be under a directory named **code**.