# Homework Solutions on Matching

Padma Tanikella

2024-04-30

1. *Assume that you conduct a different study and collect the data* `lalonde2`*. In this new study, the treatment is assigned at random. How can you estimate the causal effect of training program on the real earnings of individuals in 1978?*

Ans: Since the treatment assignment is random, it is safe to assume that the covariates are distributed "uniformly" across the treatment and control groups. Therefore, it is safe to assume that there is no confounding variable involved *due to random assignment*. Let the participants that were assigned treatment have response variable values equal to $T_1, T_2, \cdots, T_{307}$. Let $C_1, C_2, \cdots, C_{307}$ denote the control group's real income in 1978 (i.e. the response variable). Then one possible causal estimate ,$\tau$, of the treatment is:

$$\tau = \frac{\sum_{i=1}^{n} T_i}{n} - \frac{\sum_{i=1}^{n} C_i}{n} \text{ where } n = 307$$

2. *While working with the* `lalonde` *data set, can we use the technique of the first question to estimate the causal effect of the training program on the real earnings of individuals in 1978? Why or why not? Explain in a couple of sentences.*

Ans: No, we cannot use the technique from the first question. This is because we are working with an observational study i.e. the treatment is not randomly assigned. Therefore, the problem of confounders hasn't been taken care of.

3. *Let us now perform* ***exact matching*** *based on the covariates* `race`*,* `married` *and* `nodegree`*. The table below summarizes the number of subjects included in each of the ten groups created. Based on the table, can we claim that each subject was assigned exactly one group? In other words, is there a subject assigned to two or more groups?*

Ans: Yes, each subject was assigned to exactly one group since the sum of subjects in the twelve groups adds up to 614.

4. *Propose a different way of matching that one can use while working with the* `lalonde` *data set.*

Ans: There are different ways one can perform matching. One example is to perform propensity score (PS) based matching. Set an upperbound $\delta$ so that two subjects A and B with PS values $e_A$ and $e_B$ respectively are "close" or "similar" when $D_{AB} = |e_A - e_B| \leq \delta$.