Project Proposal
Impact of Job Training on Real Earnings
CS396 Causal Inference

May 5, 2024

Jacob John, Shravan Srinivasan, Sai Ganesh Nellore, Michael Hartmann

1 Problem Statement

> *"To what extent can supervised job training improve a male participants' yearly income?"*

2 Causal Questions or Hypotheses

2.1 Causal Question

We define our variables of interest as follows: let $A \in \{0, 1\}$ represent a binary treatment variable, which is 1 when exercise is supervised and 0 otherwise. Let ψ denote a set of outcome variables from the Psychological General Well-Being Index (PGWBI) [?]. $\psi = Y_{ax\_pre}, Y_{ax\_post}$

2.2 Causal Estimate

We will examine $E[Y_i^{a_t}]$, which represents the expected rate of the outcome Y_i as a function of the treatment level A_i, assuming all other factors are held constant. We are trying to measure the casual effect of A on Y, where our counterfactual would be the *low* exercise. Our goal here is to approximate this function by effectively mapping out how changes in daily physical activity influence the expected outcome of MADRS scores differences.

2.3 Comments

(a) The quantification of exercise is challenging due to its subjective nature, and establishing a threshold for what constitutes sufficient exercise is uncertain. Additionally, defining the counterfactual scenario for a "lack of exercise" is problematic. Furthermore, we could also look at sleep and sedentary data via the actigraph (provided we can define them).

(b) One way to do this via matching where we consider two individuals that are similar and we look at their exercise level. We then compare the counterfactuals among these groups.

(c) We hypothesize that our outcome variable, whether it's one measure of depression severity or potentially another, might not be influenced by physical activity. As mentioned previously, depression is a multi-faceted and complex issue and basic exercise might not reduce its onset.

(d) Furthermore, Mayo Clinic's article also highlights that regular physical activity helps alleviate depression but isn't a replacement for therapy and other forms of professional help.

(e) Alternatively, we could explore the causal impact of depression on exercise, hypothesizing that a decrease in depression severity could lead to an increase in physical activity. In this case, we could also say that depression acts as a mediator for some treatment variable like therapy. As therapy improves depression, better physical fitness follows. However, it should be noted if causality works in both directions, this would be more of a correlation analysis.

3 Dataset

We will use the Depresjon Dataset [?], which specifically targets the exploration of depression symptoms and their correlation with physical activity patterns. This dataset is valuable for understanding how motor activity patterns differ between those with depression (both unipolar and bipolar) and healthy[1] controls. Due to missing data, we will ignore the healthy patients.

3.1 Overview

(a)

3.2 Limitations

(a)

3.3 DataFrame Preview

After loading the `scores.csv` into pandas, we see:

```
NUMBER       DAYS  GENDER  AGE     AFFTYPE  MELANCH  INPATIENT  EDU
condition_1  11    2       35-39   2        2        2          6-10
condition_2  18    2       40-44   1        2        2          6-10
condition_3  13    1       45-49   2        2        2          6-10
condition_4  13    2       25-29   2        2        2          11-15
condition_5  13    2       50-54   2        2        2          11-15
...
             MARRIAGE  WORK  MADRS1  MADRS2
condition_1  1         2     19      19
condition_2  2         2     24      11
condition_3  2         2     24      25
condition_4  1         1     20      16
condition_5  2         2     26      26
...
```

3.4 Variable Overview

3.4.1 Primary

i.

3.4.2 Confounders

i.

4 Expectations and Concerns

(a)

---

[1]The dataset doesn't explicitly mention what healthy means, but we'll assume it to be those patients with no signs of depression.

References