Assignment #1

ECON 5783 — University of Arkansas

Prof. Kyle Butts

These assignments should be completed in groups of 1 or 2 but submitted individually. My preference is for you to use Rmarkdown files to have your code, results, and your answers to the questions intermixed. Since I am not requiring you to code in R for these assignments, you can use latex or microsoft word to write up answers alternatively. Not that in these cases, I would like you to upload your code seperately.

Theoretical Questions

1. For the following examples, I will describe an observational study where we observed 'treatment' D_i and outcomes y_i. I want you to think about the untreated potential outcomes (what the outcome would be for units in the absence of treatment) for the treated and the control groups. Do we think the average for one group is higher than the other and why? If you do think so, which way do you think the bias would move your difference-in-means estimate of the ATT (positively or negatively)?

    i. D_i an indicator variable that equals 1 if a worker went to college; y_i a variable measuring the worker's health

    ii. D_i is an indicator variable that equals 1 if the Amazon user was experimentally shown a banner of type 'A'; y_i is whether the user purchased the product

    iii. D_i is a variable if a school district contains an arts program; y_i is the school district's average ELA score

    iv. D_i is a variable that equals 1 if a census tract has a Dollar General in it; y_i measures the census tracts' rate of obesity.

    v. D_i is a variable that equals 1 if a driver has tinted windows; y_i is a measure of "response time" that measures visual performance.

2. In your own words, describe why A/B testing with a large amount of arms and few trials per arm can lead to problems with inference.

- If you were to recommend to a tech firm how to fix this, what might you suggest?

Coding Exercise

This exercise involves a randomized control trial done by the US Government.[1] The National JPTA Study was a large scale experiment done by the U.S. Department of Labor to evaluate the effectiveness of employment and training programs under the Job Training Partnership Act of 1982. The RCT randomly assigned economically disadvantaged adults and youths to a treatment and control arm where the treated group received JTPA training and services and the control group did not.

A cleaned version of the dataset appears in this folder `data/national_jtpa_study.csv`. A codebook for the dataset is at the end of the document.

1. First, let's see if we think randomization was done correctly. Do so by assessing if `age`, `priorearn`, `educ`, `female`, and `nonwhite` are balanced between the treated and control groups (Hint: use a regression estimator). Discuss whether you think the treated and the control group 'look the same'.

2. Second, calculate the difference-in-means estimator for the effect of JTPA training and services on the probability a worker finds a job "by hand".

3. Finally, calculate the difference-in-means estimator via regression and report robust (`HC1`) standard errors. Is the estimate statistically significant?

Variable	Description
`foundjob`	Indicator for finding a job
`treat`	Treatment indicator
`age`	Age of worker
`priorearn`	Earnings prior to the program
`educ`	Years of education
`female`	Indicator for being a female worker
`nonwhite`	Indicator for being a non-White worker
`married`	Indicator for being a married worker

---

1. See https://www.upjohn.org/data-tools/employment-research-data-center/national-jtpa-study for details.