Quantitative Methods for Program Evaluation

Ford School of Public Policy

University of Michigan

**Assignment 1**

Due: Friday, January 26th at 11:59pm

This assignment aims to get you comfortable with sample comparison with t tests and bivariate regression mechanics using real world examples (and also a review of Stata/R).

A few reminders with regard to formatting your assignment:

* Type your answers, convert to a single PDF file and post to the course Canvas website under Assignments  Assignment 1.
* *Collaboration*
* *With your study group* discuss the problem set and work on Stata/R code. You’ll learn Stata/R faster if you work alone sometimes.
* *By yourself*, type up the answers to the problem set, in your own words. Do *not* exchange your typed answers with your study group. Do *not* work on a shared text file.
* A few useful Stata commands: use, log, gen, egen, describe, summarize, gsort, tabulate, ttest, list. The most helpful command: help. There are also more Stata and R resources on the Canvas page. See the files folder, and Stata/R Help.

**Part 1 - Labor Market Discrimination**

The data set ***assignment1race.dta*** contains data from [Bertrand and Mullainathan, 2004. “Are Emily and Greg More Employable than Lakisha and Jamal? Evidence on Racial Discrimination in the Labor Market from a Large Randomized Experiment,” September 2004, *American Economic Review.*](https://www.aeaweb.org/articles?id=10.1257/0002828042002561) Variable definitions and a more detailed description can be found in their paper. Open up the dataset and dig around - take some means, do some tabulations, get to know the data. Then get started with the questions below.

1. **Test for observable differences between treatment and control baseline characteristics**

First you will determine whether the treatment group and the control groups differ in their characteristics. For the variables listed below, determine whether the difference in means between resumes assigned a Black vs. White name is statistically significant at the 95% level of significance. You can perform a difference-in-means t-test to do these tests (you will have to determine whether the treatment group and the control groups have equal variances).

* college - has a college degree (hint: Create this variable as an indicator variable taking on the value of 1 if the individual has a college degree and zero otherwise; use the education variable to create this new variable)
* yearsexp - years of work experience
* linc - log per capita income in applicant's zip (note: this variable has missing values for some observations, so will have fewer observations than the other tests).

To present your results, create a table that contains the following items for each variable above:

* mean for White names
* standard error of mean for White names
* mean for Black names
* standard error of mean for Black names
* difference in means between White and Black names
* standard error of the difference in means
* t-statistic
* “Reject” or “Don’t reject” for null hypothesis that difference in means is zero

*Notes: All of the information above can be found using the* ttest *command. You should put standard errors in parenthesis below the respective mean or mean difference (as done in Table 3) in their paper.*

1. **Test for differences in baseline characteristics for subgroups**

Repeat the analysis above (and create a similar table), but only for applicants with military experience (military == 1).

1. **Discussion of balance of baseline characteristics**

In a paragraph, discuss the results of the analysis presented in the two tables above. You should address the following:

* Does it look like the randomization worked overall?
  + Yes overall. Blah blah passed the t-tests so we can assume they are balanced and randomly assigned.
* Does it look like the randomization worked for the military?
* Are there any differences that are statistically significant?
* Are there any differences that are large but not statistically significant? You can address this by comparing the treatment-control differences to the mean or standard deviation of the variable, as we did in lecture
* How do you explain any significant differences between the two racial groups?

1. **Test for differences in outcomes between treatment and control groups**

The main outcome of interest in this study is the call-back rate, indicated by the variable call. Determine whether the difference in mean call-back between resumes assigned a black vs. white name is statistically significant at the 95% level of significance, both for the full sample and the military subsample. To present your results, you should create a table, which contains the following items for the full sample and the military subsample:

* mean call-back for White names
* standard error of mean call-back for White names
* mean call-back for Black names
* standard error of mean call-back for Black names
* difference in call-back means between White and Black names
* standard error of the difference in call-back means
* t-statistic
* “Reject” or “Don’t reject” for null hypothesis that difference in call-back means is zero

*Notes: All of the information above can be found using the* ttest *command. You should put standard errors in parenthesis below the respective mean or mean difference (as done in Table 3) in their paper.*

1. **Discussion**

Discuss your results in 1-2 paragraphs. Be sure to discuss what you find in terms of differences in call-back rates between Black and White sounding names for the two groups (all people, military only). For which of the groups do you think we will be able to make causal claims about this difference? For which of the groups above should we be hesitant to make causal claims about this difference? Identify one limitation of the study from the perspective of identifying racial discrimination by employers.

**Part 2- College Scholarships in Michigan**

The state of Michigan is considering implementing a new college scholarship program for high-achieving students from economically disadvantaged backgrounds. To qualify for the scholarship, students must have a 4.0 grade point average in high school, and have family income below $65,000. A generous donor has agreed to fund the scholarship program, which would provide $20,000 scholarships to graduating high school students to attend a public college or university in the state of Michigan. The state can only afford to award 500 of these scholarships in a given year, but there are 2,000 students who meet the criteria (grade point average, family income, intentions to attend a public institution in Michigan).

Below are three evaluation designs that have been proposed to allocate the scholarships and test whether the scholarship program increases college enrollment rates. For each, briefly (a few sentences) describe one or more concerns the design raises from an identification strategy perspective.

1. The state would offer scholarships to the 500 poorest students who meet the achievement criteria set out by the scholarship program. Researchers would then compare the college-going rates of the 500 students who were offered the scholarship to the 1,500 students who were not.
2. The state would open an application system requiring students to verify their income, submit their high school transcripts, and write a 2-page personal statement on what a college degree means to them. Students must also pay a $25 fee to process their applications. The first 500 students who complete applications for the scholarship are offered the scholarship. Researchers would then compare the college-going rates of those first 500 students, to the 1,500 students in the state who met the criteria, but who were not one of the first 500 students to submit an application.
3. The state would assign each of the 2,000 students a random number between 1 and 2,000 and draw 500 numbers out of a hat. Students whose number was drawn from the hat will receive the scholarship. Researchers would then compare the college-going rates of the students who were offered the scholarship to the students who were not.

**Part 3 - Charter Schools**

In this assignment, you will examine the relationship between charter school status and student achievement. The data (***mich\_charters\_2014.dta***) is a sample of public elementary and middle schools in Michigan during the 2013-14 school year. The main causal question we want to answer is: “What is the effect of charter school status on student achievement?”

1. **What is an *ideal* experiment that would answer the causal question above?**

- take the entire current PS student population of michigan

- randomly assign a portion of the students to charter school

- the only difference could be attributed to the treatment itself

1. **Using the summary statistics below, test whether the average proficiency rate in math is different for charters vs. traditional public schools.** You can assume the samples are independent. State your null hypothesis, your alternative hypothesis, your test statistic, and your conclusion.

Table 1. Summary statistics

|  |  |  |  |
| --- | --- | --- | --- |
|  | School average proficiency rate in math | | Number of Schools |
|  | Mean | (S.D.) |  |
| Charter schools | 0.282 | (0.177) | 177 |
| Traditional public schools | 0.430 | (0.183) | 1,877 |

1. **Explore data and test for observable differences in other characteristics between districts with different grade spans.** 
   1. How many schools are in the data set?
   2. What fraction of observations are charter schools?
   3. Create a crosstab of school type and urbanicity: “tab urbanicity charter, missing”
   4. Create a table that compares the student demographics and financial expenditures in charters and traditional public schools similar to what you created in Part I above. Use the variables: enroll, per\_fl, per\_as, per\_hi, per\_bl, per\_wh, pp\_curr\_opp\_exp.
   5. Create a table that compares the teacher characteristics in charters and traditional public schools. Use the variables: pct\_teach\_yr1, median\_experience, pct\_competitive, pct\_new\_to\_school.
2. **Discussion.**

Drawing on the results from 3c-3e above, discuss whether you believe that converting a school from a traditional to a charter school would increase student achievement? 1-2 short paragraphs.

- cannot conclude charter increase achievement

- differences between treatment and control

- difference in urbanicity

- 3d: significant between % demographics

- charters more likely to have higher free lkunch

- each differences could contribute to difference in achievement

- lower operating costs

- significant diff between new to school and year 1 teachers

-

**Part 4. Random Assignment (NOT REQUIRED)**

The key strength of a randomized trial is that the randomization creates two groups that are similar to each other (up to a known degree of sampling error) on both observable and unobservable characteristics. This exercise helps you get familiar with the randomization process.

In this exercise, you will randomly assign schools to a treatment and control group and test whether the two groups are indeed similar to each other. You will do this randomization repeatedly, to measure how frequently the two groups are statistically indistinguishable and how frequently they differ.

***Step 1: Randomly assign observations to treatment and control***

Randomly assign the schools in ***mich\_charters\_2014.dta*** into charters and traditional schools by using the following commands. In order to mirror the proportion of charters in the actual data, we will assign 177 schools to our “random charter” group. (Set the seed to some number before you create the random number so that results are replicable.):

. set seed 1492

. gen index=uniform()

. sort index

. gen random\_charter=(\_n<=177)

You have created a new variable!

***Step 2: Test for differences between treatment and control***

You will now estimate whether the treatment group and the control group differ in their characteristics. For the variables listed below, record whether the difference in means between the treatment and control groups is statistically significant at the 95% level of significance. You can use a difference-in-means t-test to do these tests.

1. Current operating expenditures per student
2. Percent free lunch
3. Median teacher experience

***Step 3: Re-randomize and test for differences between newly-defined treatment and control groups***

Repeat Step (1) and Step (2) nine times, for a total of *ten iterations*. It is critical that you re-randomize every time.

Here is example Stata and R code to get you going. It loops 10 times through the process of randomly assigning treatment status. (Note: if you want to do step 3 within the same do file as step 1 and 2 you do not need to generate the index or treatment variable again. This will result in an error.)

set more off

cd "C:\My Documents\Assignment 1"

capture log close

log using ps1.log,replace

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Ps1.do

Stata do-file to Randomly Generate Treatment and Control Groups

and Test for Differences between Them

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

gen index=.

gen random\_charters=.

local i=1

while `i' <= 10 {

replace index= uniform()

sort index

replace random\_charter=(\_n<=177)

display `i'

tab random\_charter, m

ttest expn\_stu, by (treatment) unequal

***(insert your additional statistical tests here)***

local i =`i' + 1

}

log close

R

DATA <- DATA[order(DATA$observat),]

set.seed(1492)

for (i in 1:10){

DATA$randnumber <- runif(n = dataleng)

DATA <- DATA[order(DATA$randnumber),]

DATA$random\_charter <- c(rep(1, 177), rep(0, dataleng - 177))

print(paste("Iteration", i, sep = " "))

print("Treatment Frequency")

print(table(DATA$random\_charter))

print("t-test, expn\_stu")

print(t.test(expn\_stu ~ random\_charter, data = DATA))

***[insert your additional statistical tests here]***

}

**Step 4: discussion**

Discuss your results in 3-4 sentences. Address the following: Did randomization create comparable treatment and control groups? How do you explain any statistically significant differences between the two groups? Were any differences large from a policy perspective? Hint: you can address this last question by comparing the treatment-control differences to the mean or standard deviation of the variable.