Bocconi University, Microeconometrics (cd. 20295)
Professor: Thomas Le Barbanchon and T.A.: Erick Baumgartner
Problem Set 1: Experiments and Randomization Inference

Due: Monday March 8^{th} by 11:59pm. Submit by email to
erick.baumgartner@unibocconi.it.

Problem Set 1

We will replicate and re-analyze the discussion initiated by LaLonde (1986) about how regressions with controls (i.e., non-experimental methods) do at replicating experimental results.

In a second stage of this problem set, we will revisit LaLonde (1986) to go through recent econometric advancements on statistical inference of randomized experiments. In particular, we will draw on Athey and Imbens (2017) to implement a randomization-based inference routine over LaLonde (1986)'s experimental data.

Exercise 1

Short Summary of Discussion

LaLonde (1986) presented a highly influential comparison between experimental and non-experimental methods. He studied the results of the National Supported Work (NSW) demonstration project - a subsidized work experience program that operated for 4 years at 15 locations throughout the US.

Between the rounds of 1975-1977, the treatment was randomized in 10 locations. The program served four target groups: **(i)** female long term aid recipients, **(ii)** ex-drug addicts, **(iii)** ex-offenders and **(iv)** young school dropouts. It provided trainees with work in a sheltered training environment and then helped them find a regular job. Eligibility criteria for participants included: being currently unemployed, with no more than 3 months of work in the last 6 months.

For the non-experimental results, instead of using the randomized control group, LaLonde (1986) contrasted outcomes from the NSW randomized treated group with those for control groups drawn from two national surveys (PSID and CPS). Since he found extremely different results, he concluded that observational methods are unreliable.

Data

We will use two datasets (posted on Blackboard):

(i) `jtrain2.dta` is a restricted sample from the experimental data used by LaLonde (1986). The original dataset from LaLonde (1986) consisted of two randomly allocated groups: 297 treated and 425 controls. In the restricted dataset that is available we only have a subset of those, 185 treated males who received treatment during 1976-1977 and 260 controls. Thus, it is natural to expect some imbalances.

(ii) `jtrain3.dta` includes the 185 treated from the NSW data that were part of the experiment, and a group of 2,490 men (under 55 who are not retired) drawn from the PSID who were not part of the experiment, and are used as a non-experimental control group.

Variables

All variables have self-explanatory labels in the dataset (to confirm it, type `describe`): **(1)** the *outcome of interest* `re78` are real earnings during 1978, measured in 1982's US$1000 (take this into account when you interpret the results)[1]; **(2)** the *treatment variable* `train` is an indicator variable equal to 1 when the individual participates in the NSW training group.

Instructions

No need to produce a pdf file with your answers. Save all plots and tables requested (e.g., as `pset_1_question_1_a.xlsx`) and a do-file summarizing all of your work in a zipped folder identifying your group and the problem set (e.g., as `pset_X_group_Y.zip`) to `erick.baumgartner@unibocconi.it`. In the sub-questions where you are asked to write, please add your answer as a comment in your do file.

Hint: Type * to add a comment in your do-file (e.g., `* this is a comment line *`).

Hint: You can choose your preferred way of preparing tables: **(1)** one option is to use the command `outsheet` to construct the tables of summary statistics and the command `outreg2` to construct the regression tables (you can use them to export results of summary statistics and regressions to an excel file); **(2)** another option is to save results using the command `eststo` and then export these directly to a .tex (latex) file using the command `esttab`. Read carefully through the `help` vignettes for each command you choose and try different options, so to have well-formatted tables.

Questions

1. Use the file `jtrain2.dta`.

(a) Construct a table checking for balance across treatment and control for the following covariates: `age educ black hisp nodegree re74 re75`.

Name it `TABLE_1`.

[1] Note that all other earnings-related variables within both datasets are also measured in 1982's US$1000.

Present for each variable: mean for treated, mean for controls, standard deviations for treated, standard deviations for control, difference in means between treatment and control, appropriate standard errors for difference in means.

Comment on how many variables are balanced or not. Is it what you expected?

(b) Regress `re78` on `train`.

Save the estimate and the standard error of the coefficient on `train` as scalars.

Interpret the coefficient.

(c) Construct a table by sequentially adding the output of the following regressions to each column:

(1) `re78` on `train`;

(2) `re78` on `train age educ black hisp`;

(3) `re78` on `train age educ black hisp re74 re75`;

Add rows to the table with the number of controls and treated in each regression.

Name it `TABLE_2`.

Are your results sensitive to the introduction of covariates?

(d) `dfbeta` is a statistic that measures how much the regression coefficient of a certain variable changes in standard deviations if the i-th observation is deleted.

Type `help dfbeta` and discover how to estimate this statistic after a regression.

Generate a variable named `influence_train` storing the `dfbeta`s of `train` of the last regression you did in point **(c)**.

Redo the last regression you did in point **(c)** but removing the observations with the 3, 5, and 10 lowest and largest values in `influence_train`.

Are your results sensitive to influential observations?

2. Use the `jtrain3.dta`.

(a) Do a table with the same structure of `TABLE_1` of item **(a)** in question **1** for the following covariates: `age educ black hisp re74 re75` (note that `nodegree` is not present in the current dataset.)

Add the corresponding columns to `TABLE 1`.

(b) Generate a variable named `treated` that randomly allocates half of observations to a (fake) treatment group and the other half to a (fake) control group.

Fix a seed of 5 digits using the command `set seed`.

Hint: Try to do it yourself, but use the following code if you cannot.

```
* generate empty variable to fill with 0's and 1's *
gen treated=.

* fix a seed to make realization of your random variables unique *
set seed 88888

* generate random numbers and order them *
gen random=uniform()
sort random

* create a rank from 1 to N based on the random_order *
egen random_order=rank(random)

* assign the first N/2 obs to treatment and the rest to control *
qui sum random
gen N =r(N)
replace treated=0 if random_order<=(N/2)
replace treated=1 if random_order>(N/2) & random_order<=N
```

(c) Type `ssc install randtreat`. Then, read `randtreat` help file.

Redo point **(b)** using the command `randtreat`.

Name `treated_2` your new (fake) treatment variable.

Check whether the correlation between `treated_2` and `treated` is statistically significant or not. (Hint: use `pwcorr X Y, sig`)

(d) Do a table with the same structure of `TABLE_1` of item **(a)** in question **1.**, but using `treated` instead of train.

Use the same list of covariates of item **(a)** of this question.

Add the corresponding columns to `TABLE_1`.

What you find corresponds to your expectations?

(e) Sequentially add the output of the following regressions to `TABLE_2`:

(1) `re78` on `treated`;

(2) `re78` on `treated age educ black hisp`;

(3) `re78` on `treated age educ black hisp re74 re75`.

Add lines in the table with the number of controls and treated in each regression.

Comment on what you find. Is it what you expected?

(f) Sequentially add the output of the following regressions to `TABLE_2`:

(1) `re78` on `train`;

(2) `re78` on `train age educ black hisp`;

(3) `re78` on `train age educ black hisp re74 re75`.

Add lines in the table with the number of controls and treated in each regression.

Compare the results with the first three columns of `TABLE_2`.

Comment on what you find. Is it what you expected? Are your results sensitive to the introduction of covariates?

3. Use the `jtrain2.dta`.

Read Athey and Imbens (2017) (focus on those sections where the authors discuss how to perform inference in completely randomized experiments; in particular, section 4).

(a) Under which conditions, allowing for heterogeneous treatment effects, is Neyman's inference unbiased? see paper, theory

(b) Describe Fisher's inference and replicate section 4.1 of Athey and Imbens (2017) in Stata. Do you arrive at their same p-value? If not, why? **Hint:** Note that you can draw motivation from third-parties for your own answer; for this case, we suggest that you read Heß (2017).

(c) Read again the randomization plan in LaLonde (1986). On which grounds Athey and Imbens (2017)'s illustration of Fisherian inference on LaLonde (1986)'s paper could be criticized?

(d) The article Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results (Young, 2019) presents the results of an exercise to test the null hypothesis of no treatment effects in a series of experimental papers recently published in AEA journals, showing that many of the coefficients reported in those papers are no longer significant in a randomization test. A critique of this paper has been published by professors Uri Johnson, Leif Nelson and Joe Simmons in their blog, Data Colada. Read their post here and answer the questions below.

(1) Briefly explain the difference between the procedure used as the default in Stata for the calculation of standard errors (HC1) and the one proposed by the Data Colada post (HC3).

(2) Using the dataset `jtrain2.dta`, rerun the analysis you have performed in exercise `1`, now calculating the standard errors based on HC3 (this is done in Stata using the option `vce()` in your regression command).

(3) Perform a third version of your analysis, now based on bootstrapping (use the `bootstrap` command in Stata). Briefly describe how the standard errors are calculated in this approach.

(4) Do any of your conclusions regarding the effect of the training program change based on the analysis performed in this exercise? Based on the discussion provided

in the Data Colada post, can you think of a reason for why your results using HC3 should or shouldn't change for this exercise?[2]

c. Computed from an approximiation of the distribution of the population, that is non-parametric (no assumptions on the population). Whereas HC1/3 needs large sample approximations, small sample estimators, and Fisher does not approximate the population, it counts the distribution as the population.

[2]If you'd like to read further regarding the discussion on robust standard errors, a recent Stata Blog post also covers some simulations and comparisons related to the Data Colada discussion.

References

Athey, S. and Imbens, G. W. (2017). The Econometrics of Randomized Experiments. In *Handbook of Economic Field Experiments*, volume 1, pages 73–140. Elsevier.

Heß, S. (2017). Randomization Inference with Stata: A Guide and Software. *The Stata Journal*, 17(3):630–651.

LaLonde, R. J. (1986). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *American Economic Review*, 76(4):604–620.

Young, A. (2019). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The quarterly journal of economics*, 134(2):557–598.