ECO372 Assignment #1

Patrick Blanchenay

Due 2021/01/31, 11:59PM

1 Assignment presentation

This assignment tests your comprehension of Random Assignment, and how you can use Stata to statistically analyze data based on random assignment. You must answer the questions in Section 7. You will have to submit three files:

- A PDF document giving the answers to the exercises
- The unique Stata do-file that you used to generate the answers to all questions.
- The unique log file produced by Stata when running the do-file

Each of these elements is detailed in Section 4.

2 Warning: plagiarism and academic offenses

This is an individual assignment: do not share you work with anyone, and do not use anyone else's assignment work in yours. I count on your help to make sure the course is interesting and fair. I am saying this because there have been cases of Academic misconduct in the past; I hate reporting them, it takes some of my valuable time, but I have to in order to maintain fairness and integrity in the course. Some students have faced drastic consequences from their ill-considered decisions, that they regret almost immediately. Let's avoid all of this. Do your part: play fair and square.

The files you submit will normally be checked in Turnitin for plagiarism at the time of upload. To limit possible infractions, you must cite any sources that you use. Any elements taken from the papers must be cited between quotes (" "). You should not need any external sources, but if you do use them, you must cite them adequately.

Any suspicious similarities with other submissions or any existing work will be carefully examined. Since the assignment is worth 10% of your final grade, I have to report any suspicion of academic offense to the Undergraduate Chair.

3 Before you start: preparing the files

1. Decide, on your computer, which folder will be your working directory for this project. It is recommended that you use a folder with automatic backup such as Dropbox, OneDrive, Google Drive, iCloud Drive.

2. Download the Zip archive `ECO372_Assignment1_2021Jan.zip`, from Quercus in Assignments > Assignment 1, and save it in the working directory.

3. **Extract** the archive `ECO372_Assignment1_2021Jan.zip` in your working directory. This should extract the files and create the following folder structure:

```
[Working directory]
ECO372_Assignment1_2021Jan.pdf
ECO372_Assignment1_SURNAME_FirstName.do
|- papers
    BurdeLinden_2013_AEJ_EducAfghanistan.pdf
|- datasets
    BurdeLinden_2013.dta
```

4. Rename the do-file template `ECO372_Assignment1_SURNAME_FirstName.do` by replacing "SURNAME" and "FirstName" by your surname and first name, as they appear on ACORN. For instance, mine would be called `ECO372_Assignment1_BLANCHENAY_Patrick.do`. Note that there is a 10 point penalty for failing to name your do-file appropriately.

5. Open the newly-renamed do-file:

   - On line 24, set the working directory to the folder where the do-file is on your computer.
   - On lines 27 and 30, replace BLANCHENAY by your last name as it appears on ACORN, and Patrick by your first name as it appears on ACORN.
   - On line 33, replace 12345678 by your student number.
   - Save the do-file before doing any further changes.

4 Documents to upload

4.1 Results PDF document

Filename: `ECO372_Assignment1_SURNAME_FirstName.pdf`

The Results PDF should be a single document with your answers to all questions. Some questions will require to perform analyses using Stata and provide suitable explanations and interpretations of the results. You can copy and paste results directly from Stata output window (beware of the keeping the font monospaced).

The answers you provide should only use results that are directly produced by your do-file. Conversely, you should not copy-paste ALL Stata output into your Results document. Only put the parts that are used to answer the questions.

Answers will be graded based on the quality of the explanations. It is not enough to paste Stata output. You have to explain how the output answers the specific question.

There are points allocated to the formatting of the PDF, see section 6.1.

Format

- PDF only. No other file type will be accepted (in particular, no MS Word document).
- Letter-sized. Font should be at least 10 points, everything should be easily readable, including the Stata output.
- Top line of the document should contain : [SURNAME] [First name] - ECO372 Assignment 1
- Second line: Student Nb: [Student Nb]
- Answers should be clearly numbered; do not copy the text of the questions.
- Filename should be: `ECO372_Assignment1_SURNAME_FirstName.pdf`. For instance, mine would be called `ECO372_Assignment1_BLANCHENAY_Patrick.pdf`.

4.2 Stata Do-file

Filename: `ECO372_Assignment1_SURNAME_FirstName.do`

> **Use the provided do-file template as a starting base.**

You can insert your commands in the space indicated in the provided template. Your code should produce all analyses and output necessary for all exercises and questions, from one single do-file.

Your do-file must be able to run in one go if placed on a computer with the same datasets available. The only thing I should need to change in your do-file, to reproduce exactly your results, is to change the working directory. In particular, this requires to keep the do-file in your working directory, and for the `/datasets/` folder to be in your working directory. If you're not sure, try on a classmate's computer. If you get error when running your do-file (red lines in Stata output), correct the errors, then re-run the do-file again, until the whole do-file can execute in one pass.

Comment your code. You do not need to comment every instruction, but you should comment the big steps, or the big blocks of code. Explain why you are doing such or such instructions, and what you expect Stata to do. Indentation is also useful to make your code more readable.

Part of your grades depends on code formatting & commenting, see section 6.1. You can use the provided do-file as an indication of a well-formatted, well-commented code.

Format

- Only ASCII characters should be used; no accented characters, no characters from extended alphabets or writing systems.
- Filename: `ECO372_Assignment1_[SURNAME]_[FirstName].do`, e.g. `ECO372_Assignment1_BLANCHENAY_Patrick.do`.

4.3 Stata log file

Filename: `EC0372_Assignment1_SURNAME_FirstName.log`

If you followed the steps in Section 3, your log file should be created automatically when you run your do-file. And it will be automatically named `EC0372_Assignment1_SURNAME_FirstName.log`, where SURNAME and Firstname have been appropriately replaced by your ACORN surname and your ACORN first name. For instance, mine would be called `EC0372_Assignment1_BLANCHENAY_Patrick.log`. Again, this should happen automatically if you are using the do-file template provided, and if you have configured it appropriately (see step 5 in Section 3).

Anything in your log file should come from your do-file, not from instructions typed in Stata command window. That is, if I re-run your do-file, I should obtain exactly the same log file (apart from the path to the working directory).

If you get error when running your do-file, correct the errors, then re-run the do-file in its entirety to generate an error-free log file.

Format

- Text file only, not in SMCL.
- Filename should be: `EC0372_Assignment1_SURNAME_FirstName.log`.

5 Submission instructions

By 2021/01/31, 11:59PM, you should have uploaded all three documents, and only those. Failure to include any of the three required files will count as late submission.

Do not include the datasets in your submission. Do not group files in a zip file.

No submission will be accepted on paper, or by email, regardless of any technological problem. Late penalties start applying immediately at deadline; there is no leniency for "my submission is only 1 minute late" requests.

6 Grading

6.1 Expectations

The results PDF should be properly spaced, easy to read, and have a professional appearance. Sentences should be grammatically correct, short, and to the point, using arguments from the course, and results from your data analysis. Stata results that directly help answering the questions should be included in the PDF file (either numbers, or whole tables, as appropriate).

Note that your explanations are as important as your statistical analyses, and therefore proper English is crucial to get your point across. You will lose points if what you write is unclear. Proper English helps clarity.

The do-file should be based on the template provided; it must be executable in one pass by only changing the working directory, without generating errors. The code must be commented to indicate the question you are answering, or the type of manipulation/analysis you are doing. The code must be properly indented if necessary.

6.2 Rubric

The assignment is worth 100 points, graded according to the following rubric.

Item	Points	Item	Points
Question a	5	Question h	5
Question b	8	Question i	6
Question c	6	Question j	6
Question d	5	Question k	6
Question e	8	Question l	4
Question f	8	Question m	5
Question g	5	Question n	7
Code formatting & commenting	8	PDF formatting	6

For exercise questions, you will be graded on the quality of the answers to the questions. Each exercise subquestion a, b, c, . . . is graded as a whole. Emphasis will be put on clear and concise answers that address specifically the question, and show your understanding of the topic and the statistical issues it raises. Appropriate use of the Stata output in the answer will also be taken into account: use what is necessary, leave out the irrelevant. All results file will be checked. Some do-files and log-files will be checked at random.

6.3 Penalties

Note the penalties below, as they can quickly lower your grade:

Problem	Penalty
Late submission (starting immediately at deadline)	10pts per 24hrs
File names do not follow the prescribed pattern	5pts per file
Do-file generates errors after modifying working directory	10pts
Do-file does not run in one go after modifying working directory	10pts
Log file does not correspond to do-file	10pts
Results are used that are not reproducible with the do-file	10pts

7 Questions

Exercise : Burden & Linden (2013)

In the exercise, you are to re-analyze some of the findings of the 2013 paper by Dana Burde and Leigh L. Linden about schooling in Afghanistan.[1] The study took place in rural Afghanistan, and tries to find the effect of placing a school in a village, on academic participation and performance among all children.

Note: In some of the questions, you will not find exactly the same numerical values as in the paper, which uses multi-variate regressions to perform the analyses, while we are using t-tests to do comparisons between treatment and control villages.

  a. Explain briefly the research question, the treatment and outcomes that the researchers are investigating. (Max. 4 sentences.)

  b. Use the `BurdeLinden_2013.dta` dataset to answer each of these questions in one sentence:

      (i) What is the level of observation in the dataset? (I.e. what does one observation represent in the dataset?)
     (ii) How many observations are there?
    (iii) What variable denotes whether participants were in a village where a school was placed or not?
     (iv) How many participants were were in a village where a school was placed, and how many were not?
      (v) What variables record, respectively, being enrolled in a formal school in Fall 2007? Test score in Fall 2007? Test score in Spring 2008?
     (vi) What variable records whether the participant is a boy or girl?

  c. Imagine we were to find a difference, in the Fall 2007 test scores, between children of villages where a school was placed and the others. Could we interpret such difference as the effect of placing that school? Explain.

  d. Construct a bar chart comparing the test scores in Fall 2007 of children in the control, and children in the treatment group.

  e. Is there a statistically significant difference in Fall 2007 test scores between treatment and control? Use a t-test to (significance level : $\alpha = 0.05$) to answer. Interpret your finding.

  f. In Table 4, in column (3), locate the numbers 0.691, and 0.424. Provide a full interpretation of each of these numbers and a quick comment on their relative magnitude.

  g. Explain why your estimate from Question e. falls between these two numbers.

  h. Plot the distribution (density) of the Fall 2007 test scores *for girls only*. Do you notice anything? Can you think of a reason why that is?

  i. Using a t-test, re-obtain the number 0.691 as close as possible. You will not be able to get the exact same number but you should be able to get within 0.06 of that. Compare your standard error to that of the table (you do not need to explain why they are different).

---

[1] Burde, Dana, and Leigh L. Linden. 2013. "Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools." *American Economic Journal: Applied Economics*, 5 (3): 27-40. The paper is included with the assignment.

j. Suppose that you found that parents of children in the treatment group are *much more* educated than those of children in the control group. Would this affect your previous finding of Questions c. and e.? If so, how? If not, why not? Explain.

k. Find what variable in the dataset records the level of education of the head of household (measured in Fall 2007). Use an appropriate statistical analysis (significance level : $\alpha = 0.05$) to find out whether there is a difference between the treatment and control group. Explain your finding in line of your previous answers.

l. The `runiform()` function draws a random number between 0 and 1. Generate a new variable `randvar` equal to a random number drawn between 0 and 1. Then generate another variable equal to 1 whenever `randvar` is greater or equal to 0.5 , and 0 when `randvar` is smaller than 0.5.[2] Suppose the dummy variable `randvar` you just created defines which of two groups an individual belongs to (you can call them group-0 and group1). How many individuals are there in each of these two groups?

m. Is there a statistically significant difference (at significance level $\alpha = 0.05$) in Fall 2007 test scores between the two groups group-0 and group-1 you defined in Question l.? Is this what you expected? Explain why or why not.

n. There are 180 students registered in ECO372. Between 0 and 180, how many students do you expect to find a statistically significant difference (significance level : $\alpha = 0.05$) in Question m.? Explain your reasoning.

---

[2]   In essence, you have just created a random variable which, for each observation, takes a value of 1 with probability 50%, and a value of 0 with probability 50%.