Tutorial 2, Econometric Theory I

<u>Guidelines.</u> **(READ CAREFULLY).**

Deadline. You will need to hand in your solution (**text and code**) to Exercises 1 and 2 before 9am on the day of the next tutorial session (**Wednesday Oct. 9**).

Guidance for coding questions. You are welcome to use your favorite language (Stata, R, Python, Julia etc.) to answer the questions of the exercises below. Your code should be commented **in details**, with each step of the code being explained in plain english.
Example (fake, it does not correspond to anything specific):
1. Loop repeating a 1000 times a sampling from super-pop, treatment allocation and computation of diff-in-means estimator + estimator of its variance.
2. Computing the diff-in-means estimator.
3. Computing estimator of variance of the diff-in-means estimator.
4. Storing the result in a dataset containing the value of the diff-in-means estimator and its estimated variance for each Monte-Carlo simulation iteration.
5. Sorting the resulting dataset by X etc.
The written answer to questions involving coding must include all such comments, and the reader should be able to determine whether or not there are any missing steps in your code from this text (so **be precise**).

Guidance for those of you who don't code. Unfortunately, most of the questions require coding, and commenting on the results. Therefore, you won't be able to complete this problem set without coding. **I strongly encourage you to team up with someone who can code.**

Working in groups. You can work in groups of up to 4 people if that includes students who don't code, 2 people if that includes only students who can code. **To those who code: if possible, please be generous and welcome a student who don't code in your group**.

To conclude, this problem set is rather long. Do as much as you can — as it is designed to deepen your understanding of the concepts we studied through practice. Yet do not worry if you do not manage to answer all questions ; you'll get a discussion of the full problem set in the tutorial session (and the grade on problem sets is only there as a bonus for your final grade, it won't penalize you in any case).

Exercise 1: measuring the effect of tracking on students' test scores

This exercise uses data from "Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. The American Economic Review, 101(5), 1739-1774." In this paper, the authors randomly assigned 108 primary schools with 2 1st grade classes in Kenya to a tracking group and to a non-tracking group. In the tracking group, the 2 1st grade classes in the school were formed based on students' ability: students faring below the school median in a test in the beginning of the year go to class A, while students faring above the school median go to class B. In the non-tracking group, classes A and B are randomly formed, and therefore both classes have students below / above the median.

The data set tracking.dta contains data on 5170 students in these 108 primary schools. schoolid is the unique identifier of their school. tracking is an indicator equal to 1 for students whose school was randomly assigned to the tracking group. bottomhalf is an indicator equal to 1 for students who fared below the median of their school in the beginning of first grade test. Finally, scoreendfirstgrade is students' standardized score in the end of first grade test.

Throughout the exercise, you need to bear in mind that the randomization took place at the school level, not at the student level. Therefore, your unit of observation in all of what follows should be schools, not students. Hint: the Stata command "collapse" — or the equivalent in any other software, e.g. "group_by" in R (dplyr package) — might prove useful.

1. Estimate the effect of tracking on students' end of first grade test scores.

2. Use a 10% level randomization inference test to assess whether the finding of question 1 is robust or whether it rests on an unappropriate asymptotic approximation. Conclude based on the result of this randomization test.

3. One might fear that tracking benefits strong students while harming weaker ones. Assess whether this is a legitimate concern using the dataset at hand.

Exercise 2: some properties of the ATE estimator in randomized experiments

1. Prove the following theorem:

Theorem 0.0.1 *Using draws from a uniform distribution to generate draws from other continuous distributions*

Let F denote a strictly increasing cdf. If U follows the uniform $[0,1]$ distribution, then the cdf of $F^{-1}(U)$ is F.

Hint: the cdf of $F^{-1}(U)$ if the function $x \mapsto G(x) = P(F^{-1}(U) \leq x)$. You need to show that $G(x) = F(x)$.

The following questions require using a statistical software.

2. Set the number of observations to 1000, create a variable $y(0) = U$, where U follows a $N(0,1)$ distribution, and create a variable $y(1) = 0.5y(0) + 0.5V + 0.2$, where V is a $N(0,1)$ random variable independent of U. $y(0)$ and $y(1)$ represent the potential outcomes of 1000 units that participate in a randomized experiment. Hint: it follows from the previous question that if ε follows a uniform distribution, $\Phi^{-1}(\varepsilon)$ follows a $N(0,1)$ distribution. The command for Φ^{-1} is invnormal() in Stata, and qnorm() in R.

3. Given the way we created $y(0)$ and $y(1)$, is the effect of the treatment homogeneous or heterogeneous across units?

4. Compute ATE_{1000}, the average effect of the treatment in this fixed population of 1000 units. Store this number in a scalar.

5. Compute the correlation between $y(1)$ and $y(0)$ and the variance of $y(1) - y(0)$ in this finite population of 1000 units.

6. Write a 800 iterations loop, where in each iteration you:

   1. Create a variable containing a random number.

   2. Sort the 1000 observations according to that random number.

   3. Create a dummy variable D equal to 1 for the first 500 observations in the sorted dataset.

   4. Create a variable $Y = (1 - D)y(0) + Dy(1)$.

   5. Regress Y on D, using the robust option. Check that the coefficient of D is equal to $\widehat{ATE}_{1000}$. Check that the variance of that coefficient is equal to $\widehat{V}\left(\widehat{ATE}_{1000}\right)^+$.

   6. Compute $IC(0.05)_+ = \left[\widehat{ATE}_{1000} - 1.96\sqrt{\widehat{V}\left(\widehat{ATE}_{1000}\right)^+}, \widehat{ATE}_{1000} + 1.96\sqrt{\widehat{V}\left(\widehat{ATE}_{1000}\right)^+}\right]$, and the indicator $1\{ATE_{1000} \in IC(0.05)_+\}$.

   7. Store $\widehat{ATE}_{1000}$, $\widehat{V}\left(\widehat{ATE}_{1000}\right)^+$, and $1\{ATE_{1000} \in IC(0.05)_+\}$ in a matrix/dataframe.

Compute the mean of $\widehat{ATE}_{1000}$ over those 800 replications, and compare it to ATE_{1000}. Compute the variance of $\widehat{ATE}_{1000}$ over those 800 replications, and compare it to the mean of $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$ over those replications. Compute the mean of $1\{ATE_{1000} \in IC(0.05)_{+}\}$. Explain your results in light of the lecture slides/notes.

7. Write a 800 iterations loop, where in each iteration you:

   1. Create a variable $y(0) = U$, where U follows a $N(0,1)$ distribution, and create a variable $y(1) = 0.5y(0) + 0.5V + 0.2$, where V is a $N(0,1)$ random variable independent of U, with 1000 observations each.

   2. Create a variable containing a random number.

   3. Sort the 1000 observations according to that random number.

   4. Create a dummy variable D equal to 1 for the first 500 observations in the sorted dataset.

   5. Create a variable $Y = (1-D)y(0) + Dy(1)$.

   6. Compute and store the values of $\widehat{ATE}_{1000}$ and $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$ using the formulas in the slides/lecture notes.

   7. Compute $IC(0.05)_{+} = \left[\widehat{ATE}_{1000} - 1.96\sqrt{\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}}, \widehat{ATE}_{1000} + 1.96\sqrt{\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}}\right]$, and the indicator $1\{0.2 \in IC(0.05)_{+}\}$.

   8. Store $\widehat{ATE}_{1000}$, $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$, and $1\{0.2 \in IC(0.05)_{+}\}$ in a matrix/dataframe.

Compute the mean of $\widehat{ATE}_{1000}$ over those 800 replications, and compare it to 0.2. Compute the variance of $\widehat{ATE}_{1000}$ over those 800 replications, and compare it to the mean of $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$ over those replications. Compute the mean of $1\{0.2 \in IC(0.05)_{+}\}$ (i.e. $1\{ATE_{1000} \in IC(0.05)_{+}\}$). Use the lecture notes to explain why the results change wrt to those you obtained in question 6.

8. Write a 800 iterations loop, where in each iteration you:

   1. Create a variable $y(0) = U$, where U follows a $N(0,1)$ distribution, and create a variable $y(1) = y(0) + 0.1771$, with 1000 observations each.

   2. Create a variable containing a random number.

   3. Sort the 1000 observations according to that random number.

   4. Create a dummy variable D equal to 1 for the first 500 observations in the sorted dataset.

   5. Create a variable $Y = (1-D)y(0) + Dy(1)$.

6. Compute and store the values of $\widehat{ATE}_{1000}$ and $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$ using the formulas in the slides/lecture notes.
   Remark (no need to implement that if you don't want to try): we did not show that already, but you can also compute these values $\widehat{ATE}_{1000}$ as the coefficient on D in a linear regression of Y on a constand and D. To compute the $\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}$, you should then ask your statistical software to compute heteroscedasticity robust standard errors — this is done using the command "robust" in Stata for instance.

7. Compute the indicator $1\left\{\left|\frac{\widehat{ATE}_{1000}}{\sqrt{\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}}}\right| > 1.96\right\}$.

8. Store that indicator in a matrix/dataframe.

Compute the mean of $1\left\{\left|\frac{\widehat{ATE}_{1000}}{\sqrt{\widehat{V}\left(\widehat{ATE}_{1000}\right)^{+}}}\right| > 1.96\right\}$ over those 800 replications. Use the lecture notes to explain your result. Hint: notice that in this simulation design, $V(Y(0)) = V(Y(1)) = 1$. Then, where does 0.1771 (in step 1 of question 8.) come from?

9. Assume you want to use a randomized experiment to measure the effect of a treatment. Your experiment will have 1000 subjects. 500 will be treated, while 500 will remain untreated. Moreover, given the nature of the treatment, you think it makes sense to assume that $V(Y(0)) = V(Y(1))$. No one has ever measured the effect of the specific treatment you are interested in, but a literature review of the effects of treatments with a similar cost shows that they typically produce effects in the range of 10% of a standard deviation of the outcome. Should you embark in this experiment? Hint: revise the lecture notes on MDD (end of first chapter) to refresh your memore on how to proceed when the MDD is expressed as standard deviations of the outcome.

10. Write a 800 iterations loop, where in each iteration you:

1. Create variables $y(0) = U$, where U follows a $N(0, 1)$ distribution, and $y(1) = y(0)$, with 10 observations each.

2. Create a variable containing a random number.

3. Sort the 10 observations according to that random number.

4. Create a dummy variable D equal to 1 for the first 5 observations in the sorted data set.

5. Create a variable $Y = (1 - D)y(0) + Dy(1)$.

6. Regress Y on D. Store $\widehat{ATE}_{10}$, the coefficient of D in that regression.
   Remark: you can also compute $\widehat{ATE}_{10}$ using the formula in the notes if you prefer.

7. Then, write a 100 iterations loop, where in each iteration you:

(a) Create a variable containing a random number.

(b) Sort the 10 observations according to that random number.

(c) Create a dummy variable $\tilde{D}$ equal to 1 for the first 5 observations in the sorted dataset.

(d) Regress Y on $\tilde{D}$. Store the coefficient of $\tilde{D}$ in that regression.
Remark: you can also compute this coef using the formula for $\widehat{ATE}$ in the notes if you prefer.

8. Compute the 5th percentile and the 95th percentile of the coefficients of $\tilde{D}$ in the previous loop. Let $q_{0.05}^{\tilde{D}}$ and $q_{0.95}^{\tilde{D}}$ denote those percentiles.

9. Compute the indicator $1\{\widehat{ATE}_{10} \notin [q_{0.05}^{\tilde{D}}, q_{0.95}^{\tilde{D}}]\}$.

10. Store $1\{\widehat{ATE}_{10} \notin [q_{0.05}^{\tilde{D}}, q_{0.95}^{\tilde{D}}]\}$ in a matrix/dataframe.

Compute the mean of $1\{\widehat{ATE}_{10} \notin [q_{0.05}^{\tilde{D}}, q_{0.95}^{\tilde{D}}]\}$ over those 800 replications. Use the lecture notes to explain your result.

Exercise 3: Partial identification of the ATE without randomization through Manski bounds

This exercise aims at highlighting the fact that even in the absence of randomization, data on the joint distribution of the outcome and treatment status (Y, D) may still be informative about the average treatment effect.

We will not assume anything about the allocation of the treatment:

$$D \not\perp\!\!\!\perp (Y(0), Y(1))$$

However, we will make the following assumption.

Assumption 1 *Bounded outcome.*

$$Y(0) \in [l, h], Y(1) \in [l, h]$$

This kind of conditions would naturally arise if the outcome variable is bounded (e.g., binary), hence the name of this assumption.

1. Show that

$$\begin{aligned} \text{ATE} \equiv E(Y(1) - Y(0)) = & \; E\left(Y|D=1\right)P(D=1) + E\left(Y(1)|D=0\right)P(D=0) \\ & - E\left(Y(0)|D=1\right)P(D=1) - E\left(Y|D=0\right)P(D=0) \end{aligned}$$

For each term on the right hand side, indicate which ones are observable (in the sense that they are *estimands*) and which ones aren't.

2. Propose bounds for the ATE as a function of the estimands involved in question 1.

3. Show that the bounds constructed in question 2. always include 0.