Experimental Methods in Development Research
Project/Homework

Mauricio Romero

March 20, 2020

1 General instructions

- You may use Stata, R, Python or a program of your choice.
- For each section, submit the final result in a pdf file (as a memo describing broadly what you did and what you found), the raw data, and all code and scripts used to produce the results. Please put each section in a different folder, and all the folders in one zipped file. Your work should be completely replicable using the code and raw data you submit.
- Email the zipped file to: mtromero+asistente@itam.mx
- You will be graded not only on the answers you give, but also on the process that you take to arrive at your answers (i.e., your code file and comments).
- It is imperative to write your code files and comments clearly and efficiently so that it is easy for you and others to understand and follow your work.
- We often work on ongoing projects, and so we will rerun the code on updated versions of the same data. As much you can, write code that will produce correct answers even as your dataset changes over time.
- Using outside resources is encouraged (and often necessary!) but please mention what resources you have used if you do so.
- Please send the homework before Friday, March 27, 2020, at 10:00 am (CET).
- Good luck!

2 Factorial Designs

Consider next the power implications of leaving the interaction cell empty. Specifically, consider a factorial experiment with the following sample sizes and consider the following regression model

$$Y = \beta_0^* + \beta_1 T_1 + \beta_2 T_2 + \varepsilon. \tag{1}$$

		T_1	
		No	*Yes*
T_2	*No*	N_c	N_T
	Yes	N_T	0

1. Let $\hat{\beta}_1$ denote the OLS estimator. What is the variance of $\hat{\beta}_1$?

2. What is the optimal allocation of the sample to maximize the power of the t-test for $\hat{\beta}_1$ (i.e., what is the sample allocation that minimizes the variance of $\hat{\beta}_1$)?

3. Let $\hat{D} = \hat{\beta}_1 - \hat{\beta}_2$ denote the OLS estimator of the difference between the treatment effects. What is the variance of $\hat{D}$?

4. What is the optimal allocation of the sample to maximize the power of the t-test for $\hat{D}$ (i.e., what is the sample allocation that minimizes the variance of $\hat{D}$)? (Hint: Remember that if $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$, then $X - Y \sim N(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2 - 2cov(X, Y))$

3 Liberia Public-Private education experiment

The Partnership Schools for Liberia (PSL) program is a public-private partnership (PPP) for school *management*. The Government of Liberia contracted multiple non-state providers to run ninety-three existing public primary and pre-primary schools. PSL schools remain public schools and all grades are required to be free of charge and non-selective (i.e., providers are not allowed to charge fees or to discriminate in admissions). Below, we are going to walk through the steps to design a field experiment, and then analyze the data that actually came out from the field experiment.

There are in total 195 schools in the experiment, but 93 will be treated. The government allocated rights to eight providers to manage public schools under the PSL program. The organizations are as follows: Bridge International Academies (23 schools), BRAC (20 schools), Omega Schools (19 schools), Street Child (12 schools), More than Me (6 schools), Rising Academies (5 schools), Youth Movement for Collective Action (4 schools), and Stella Maris (4 schools).

1. First, we need to design the experiment.
   (a) Discuss the merits of randomly assigning schools to different treatment organizations (i.e., randomly assigning a schools to one of eight possible treatments or the control, where the treatments are which organization controls the school)
   (b) Discuss the merits of a design where we first split the sample of 195 schools in eight groups (based on the preferences on the regions where each provider wants to work) and then we randomly assign schools to treatment and control within each group.
   (c) Which design would you prefer and why?

2. We went for the second scenario. That means, that we stratify by "providers preferences". Thus, we have eight disjoint groups of schools and we randomly assign treatment within each group. The main outcome of interest are at the student level (enrollment status and test scores).

3. Before we run any power calculation/analysis, what is the regression you would use to estimate outcomes (please provide indicators of at what levels you would cluster your standard errors, if you would include any fixed effects or any other controls, etc.)

4. Do some power calculations using the formula's we studied in class. Please include MDEs for the overall treatment effect, and for the treatment effect of each organization. Assume that $\alpha = 0.05$, and provide MDEs for different scenarios (e.g., $\beta = 0.9$ and $\beta = 0.8$, $\rho = 0.1$ and $\rho = 0.3$, $n = 20$ and $n = 30$).

5. Now repeat the power calculations, using a randomization inference approach.

6. Which is more powerful, OLS, or RI in this setting? How does your answer vary across sample sizes?

7. For logistical reasons, we actually had to do a match-pair design. See the file "school_info.csv" for the actual assignment across match pairs (groupid_nopii identifies the different pairs). Using randomization inference, please update the power calculations for the the overall treatment effect and the eight different organizations.

8. Some schools did not comply with the treatment (turnedpsl gives the actual compliance with the treatment). Essentially, the organizations decided not to take control of some of the schools randomly assigned to them. Please give power estimate for the treatment-on-the-treated using both the formulas we saw in class and randomization inference.

9. Let's take a step back before we estimate treatment effects. Do you think that randomly assigning treatment and measuring learning outcomes is going to yield an unbiased estimate of the "intensive margin" treatment effect on test scores (i.e., the effect on test scores for the students that were already in the school)? Hint: Think of parent's decisions to enroll their kids, as well as the organization's incentives to enroll/unenroll certain kids.

10. Considering your answer above, please give a couple of alternatives of how to design your data collection or the experiment to be able to identify the intensive and the extensive margin effects

11. Somehow, we actually managed to collect data that is unbiased from any extensive margin effects. This is in the file "Students.csv"

    (a) Estimate the ITT and TOT effect on student's learning on both subjects using OLS
    (b) Estimate the ITT and TOT effect on student's learning on both subjects using randomization inference
    (c) Estimate the ITT and TOT effect on student's learning on both subjects using OLS for each organization
    (d) Estimate the ITT and TOT effect on student's learning on both subjects using randomization inference for each organization
    (e) Discuss the difference between OLS and RI in the different cases above
    (f) Discuss the difference between the ITT and TOT in the different cases above.
        i. Do you think the assumptions to be able to estimate the TOT are met?
        ii. If you are the Ministry of Education, which estimator is more relevant to you? Why?
        iii. If you are a donor seeking to fund one of these eight organizations, which estimator is more relevant? Why?
    (g) Provide an analysis of heterogeneity by student's age, gender, and grade.
    (h) Adjust the heterogeneity analysis for multiple hypothesis testing

12. We actually have student non-compliance too (in addition to student-level non-compliance). Some students, left the school before the end of the school year.

    (a) Discuss the theoretical difference between the school and the student level non-compliance and how these result in different ToT estimators
    (b) Do you think we can use the treatment assignment as an instrument to estimate the ToT effect at the student level (the one we estimate before estimate the ToT effect at the school level). Provide as many econometric/statistic arguments as you can

13. In addition to data on student outcomes, we have some data at the school level ("school_data.csv"). We can use this data to study some of the mechanisms behind the treatment. Please estimate the overall treatment effect on the likelihood the school is open

14. We also collected data on classroom observations when the school is open. Estimate the treatment effect on the proportion of the class that the teacher spends actively teaching

15. Is the treatment estimate above biased by attrition? If so, please explain why

16. Provide Lee bounds to the treatment estimates on the classroom observations. Are the treatment effects still positive and significant?

4 Power Calculations

You partner with a government agency to do a field experiment. The idea is to do the experiment at the municipality level, and there are 400 municipalities. The outcomes are at the municipality level (so you don't need to worry about intra-cluster correlation and such). The idea is to collect data in each municipality for 18 months. Thus, in the end you will have a municipality-month panel data, with a total of 7,200 observations (18×400). The idea is to randomly roll-out the treatment, such that some municipalities randomly get the treatment before others. Once a municipality gets treated in a certain month, is treated in every month after it.

1. First, the agency tells you that they want to do a randomized staggered rolled out. Specifically, they want to randomize the order in which municipalities are treated, and every 2 months they want to treat 50 additional municipalities. That is, at $T = 0$ no one is treated. At $T \geq 2$ 50 municipalities are randomly treated, at $T \geq 40$ another 50 municipalities are randomly treated (and thus at $T = 4$ and $T = 5$ 100 municipalities are treated), and so on. Note that by the time $T = 16$ and $T = 17$ everyone will be treated. Please conduct a simulation to estimate the power of this experiment. Assume the treatment is constant across time and across municipalities. (The model you want to estimate is essentially a difference-in-difference model with municipality fixed effects and month fixed effects).

2. The agency then changes their mind and wants to randomize half of the municipalities into treatment at $T = 0$ and don't treat the other half (at least until your study is concluded, so they will never be treated in the 18 months you collect data). Please conduct a simulation to estimate the power of this experiment. Assume the treatment is constant across time and across municipalities.

3. The agency then changes their mind again and wants to randomize half of the municipalities into treatment at $T = 8$ and don't treat the other half (at least until your study is concluded, so they will never be treated in the 18 months you collect data). Please conduct a simulation to estimate the power of this experiment. Assume the treatment is constant across time and across municipalities.

4. The agency then changes their mind again (this happened to me in real life) and wants to randomize half of the municipalities into treatment at $T = 6$ and treat the other half at $T = 12$. Note no municipality will be treated for $T < 6$ and all municipalities will be treated for $12 < T \leq 17$. Please conduct a simulation to estimate the power of this experiment. Assume the treatment is constant across time and across municipalities.

5. Which design has more power? Give an intuition for your result.

5 Spillovers

I want you to go over some of the analysis in "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya" by Haushofer and Shapiro published in 2016 in the Quarterly Journal of Economics(https://www.princeton.edu/haushofer/publications/Haushofer_Shapiro_UCT_QJE_2016.pdf). The data to replicate the study can be found here https://www.princeton.edu/haushofer/publications/Haushofer%20Shapiro%20QJE%202016%20Data%20and%20Code.zip

A good idea might be to read this blog by Berk Özler before: https://blogs.worldbank.org/impactevaluations/givedirectly-three-year-impacts-explained

Use the file "UCT_FINAL_CLEAN.dta" for the analysis below.

As a summary, this experiment featured a partial population design, with some households randomly assigned to treatment in villages randomly assigned to treatment, some households randomly assigned to control in villages assigned to treatment, and a pure control group with households in villages assigned to control.

The treatment was an unconditional cash transfer.

1. First, let's replicate Column 2 of Table 2. For this the outcome variables are: asset_total_ppp cons_nondurable_ppp ent_total_rev_ppp fs_hhfoodindexnew med_hh_healthindex ed_index psy_index_z ih_overall_index_z. You also need to remember they are dropping the pure control villages, and comparing the treated households within a village, to the untreated households within the same village (i.e., they include village fixed effects). The standard errors are clustered at the household level. You can check the code "UCT_Endline_Regs_Main.do" to see exactly what they do.

2. Great you replicated their table. Now this table assumes there are exactly zero spillovers. Re-run the regression, but using the original randomization design. That is, compare both treated and untreated households in treated villages, to pure control households. You must cluster at the village level (the first-stage unit of randomization). Can you conclude that there are no spillovers?

3. The authors are implicitly undertaking model selection: they test whether spillover are significant, if not they choose one specification — the one in the paper — and if they were significant they would choose a different one (namely, the one you estimated in the previous point). For the case of non-durable expenditure (cons_nondurable_ppp1) please conduct some simulations assuming the true treatment effect is the one you estimated in the previous point, and show the bias and type-I error of this type of model selection in this particular case (use a simulation starting that matches the structure of the data, that is, the same household-village structure, and the same mean and standard deviation for the outcome).

4. Did the authors followed the rule "As ye randomized, ye shall analyze"?

5. What do you conclude from this exercise? Reflect a bit on the blog above, as well as the following two blogs: https://blog.givewell.org/2018/05/04/new-research-on-cash-transfers/ and https://www.cgdev.org/blog/cash-transfers-cure-poverty-side-effects-vary-symptoms-may-return-when-treatment-stops