A-164/API-211 Program Evaluation (Spring 2025)

Rucinski

**Problem Set 2:**

**Random Assignment and Power Calculations**

**Problem Set Policies:**

* This assignment is due by **5pm on Friday, February 21**, uploaded to Canvas a Word doc or PDF. A 24-hour extension is granted automatically, no questions asked. To request an extension longer than 24 hours, email [melanie\_rucinski@gse.harvard.edu](mailto:melanie_rucinski@gse.harvard.edu). We encourage you to submit by the deadline to give yourself a break over the weekend.
* Please type your solutions. You can use the equation editor/insert equation tool in Word to help you typeset any equations or math. **Include your Stata or R code at the end of your submission**. You can copy and paste the code directly into Word.
* You are welcome to discuss the problem set in small groups, and to submit questions to the teaching team and other students via email or Slack. Please write the names of any students you worked closely with at the top of your submission. **You must submit individually, and** **all work you submit should be your own**. To avoid misunderstandings, we ask that you **do not share any *written* answers to these questions** in shared documents, emails, chats, etc.
* Please familiarize yourself with [HGSE policy on the use of generative AI](https://registrar.gse.harvard.edu/AI-policy). You may use ChatGPT, Claude, etc. to refine or translate your ideas and to help with writing code. You may **not** use AI to *generate* ideas for this or any other assignment. If you do use AI to assist you in completing this assignment, please credit the tool you used as you would a human collaborator, and explain how you used it.
* If you have questions about appropriate use of AI or find yourself turning to AI to write answers for this assignment, or if you are considering intentionally copying an answer written by another student, please be in touch with a member of the teaching team. We would far prefer to help you in office hours or grant you an extension than to have a conversation about academic misconduct.
* If you encounter any issues or questions regarding the coding for this problem set that you are unable to answer with the help of the internet or your peers, please post to the #coding-helpdesk channel on Slack. **Please do not email or DM members of the teaching team with coding questions.**
* Once your assignment is graded, please pay careful attention to the written feedback, and seek to address this feedback in future assignments.

**TN STAR Data**

We will use the provided dataset *krueger\_class\_size.dta* to answer the following questions. This file contains the actual data from the Tennessee STAR experiment, in which students and teachers were randomly assigned to various class sizes in 80 different schools in Tennessee.

The questions make use of the following key variables:

pread1g: end-of-year reading percentile score in first grade

pmath1g: end-of-year math percentile score in first grade

cltype1: class type in first grade (remember that this identifies three distinct groups)

schid1n: the ID for the school students were in in first grade

white: indicator for whether or not student identifies as white

black: indicator for whether or not student identifies as black

trace1: race of the student’s first-grade teacher

**Part 1: Power and Effect Size in TN STAR**

1. Suppose you are designing the Tennessee STAR experiment. Based on the literature on impacts of class size, you hypothesize that there will be a difference in means of 0.20 standard deviations between small and large classes. How large a sample size would you need in order to detect a 0.20 difference in means with (1-β)=80% power if you allow for α=5% chance of a Type I error?

*(Hint: Stata users, use the power or sampsi commands. R users, use the pwr command.)*

1. Using the same assumptions as above, how large a sample size would you need to detect a difference half as large (0.10 SDs) at the same power and using the same significance level (alpha)? Is the required sample size equal to, more, or less than twice as large? Explain why.

*(Hint: Stata users, use the power or sampsi commands. R users, use the pwr command.)*

1. Tennessee STAR was randomized at the class level, but suppose instead it had been randomized at the school level. In this version of the experiment, 30 schools were randomly assigned to administer only small classes, while 30 schools were randomly assigned to administer only large classes. The average school enrolls 250 students in grades K-5. Examining data from prior school years, you find that 18% of the total variation in student achievement, as measured by the state assessment, can be attributed to between-school variance. (This is another way of saying that the intra-class correlation is 0.18.)

Taking this information into account, what is the minimum detectable effect size (i.e., the smallest effect size will you be able to detect) with 80% power? What would have been the minimum detectable effect size if you had been able to randomize individual students with the same sample size?

**Part 2: Interpreting Treatment Effects in TN STAR**

1. Use a regression to test whether there was any difference in math or reading percentile scores at the end of first grade for those assigned toregular sized classrooms versus those assigned to small or regular/aide classrooms (you should run a single regression for each subject with “regular” as the omitted category). Report your estimated differences in scores. You do not need to interpret the differences.

Note: Following Krueger, you should use school fixed effects in your regression. To do so in Stata, use the areg command with the option absorb(schid1n). In R, add factor(schid1n) to your regression as a control.

1. In question 4, you measured the effect of *assignment* to a small or regular/aide class. Is this treatment effect an intent-to-treat, a local average treatment effect, or a treatment on the treated? Explain.
2. In Table V, the reduced form estimates (corresponding to your estimates in question (4)) for first grade are consistently smaller than the OLS estimates. What does this imply about selection into classrooms of different sizes after randomization?

**Part 3: Student-Teacher Race Match Effects in TN STAR**

The following question extends on what we have explicitly covered in class. It will be graded for effortful completion rather than correctness.

1. The Tennessee STAR experiment has been leveraged to estimate the impact of same race teachers on students’ academic achievement. [[1]](#footnote-1) The following questions will guide you to do similar analysis and understand how the estimates should be interpreted.
2. Create a new binary variable which is equal to 1 if the student and the teacher are of the same race for first grade students. (Hint: Combine use of generate command with if statements.)
3. Estimate the difference in first grade math percentile scores associated with:
   1. Having an African American teacher for African American students
   2. Having a White teacher for White students

Include school fixed effects in your regression. (You can do (i) and (ii) in two separate regressions, or in a single regression with an interaction term.)

1. Interpret your estimates in part (b). Do these estimates represent causal effects of student-teacher race match? Why or why not?

1. Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of economics and statistics*, *86*(1), 195-210. [↑](#footnote-ref-1)