APSY-UE 25
RESEARCH METHODS I
PROBLEM SET #5

Instructions:

- This problem set aims to assess your understanding of the material covered in **classes 18 to 23**. It draws on the **readings, lectures, and recitations**.
- You must turn in your answers by **Wednesday, April 29, at 5pm EST** electronically through the Assignments tab of the course site.
- You may **complete the problem set with others**, but you must **write up your own answers**.
- If you work with others in the class, you should **acknowledge their contributions** using a footnote stating "In collaboration with [First name] [Last name]". This will help the teaching team understand similarities between assignments by different students.
- You may ask the teaching team **questions about the instructions** via e-mail or during office hours. You may not, however, ask for hints or whether your answers are on the right track.
- Once all problem sets have been graded, an **answer sheet** will be posted to the course site and you may request the teaching team to **discuss the answers during office hours**.
- Submit your answers to the questions as a **.doc or .docx file** with the name "[First name] [Last name] – Problem set 5".
- Submit your Stata commands as a **.do file** with the same name. You do **not need to turn in tables or graphs**. You must turn in **a single .do file**, regardless of how many datasets you are using. Your .do files should **call in the dataset(s)** with which you are working (using the "use" command in Stata).
- If your answers involve computations, please **show your as much work as you can**.
- Remember to **check your spelling and grammar** before submitting your problem set. Problem sets with more than five spelling or grammar mistakes will be deducted five points.
- Respect the **sentence limit** for each question. Answers that go over the sentence limit will be deducted one point per question.

---

1. Download an abridged version of the dataset from an impact evaluation of a scholarship and mentoring program in the Province of Buenos Aires, Argentina. (You can read the working-paper version of the study at: https://www.alejandroganimian.com/journal-articles). The dataset includes the responses of 401 secondary-school students to the Impulsivity Scale for Children (ISC) developed by Tsukayama and colleagues. You can download it at: https://bit.ly/2VuIzf9. Please, answer the questions below in **no more than two sentences**.
   a. Calculate the split-half reliability of this instrument in Stata using two arbitrarily-defined halves. Generate a score with the first four items and another score with the last four items. Estimate the correlation between both halves. Use the Spearman-Brown prophecy formula to estimate the split-half reliability of the instrument. *[10 points]*
   b. Would you say the estimate from question 8) is high, medium, or low and why? *[5 points]*
   c. Calculate the split-half reliability of this instrument in Stata using the two sub-scales in the journal article above. Generate a score with items 2, 3, 5, and 8 and another score with items 1, 4, 6, and 7. Estimate the correlation between both halves. Use the

Spearman-Brown prophecy formula to estimate the split-half reliability of the instrument. *[10 points]*
d. Would you say the estimate from question 10) is high, medium, or low and why? *[5 points]*
e. Which of your estimates from parts a) and c) would you favor and why? *[5 points]*
f. Estimate Cronbach's alpha for the instrument in Stata. When doing so, keep the items coded as they are and display the item-level breakdown of the reliabilities. *[5 points]*
g. If you had to remove one item from the instrument, which would it be and why? Identify the item clearly as well as your criterion for dropping this item, using the output from Cronbach's alpha in Stata. *[10 points]*

2. Download an abridged version of the dataset from an impact evaluation of an extra-teacher for pre-school centers in the state of Tamil Nadu, India. (You can read more about the study at: https://www.alejandroganimian.com/ongoing-projects). The dataset includes the numeracy, literacy, and executive function scores for children ages 3 to 6 in the control and treatment groups of the experiment. You can download it at: https://bit.ly/351b1bE. Please, answer the questions below in **no more than two sentences**.
   a. The variable "per_exec2" indicates the share of questions that each child answered correctly in the assessment of executive function administered at midline (it ranges from 0 to 1). The variable "treat" indicates whether a child was assigned to the control or treatment groups (0 refers to the control group and 1 to the treatment group). If you wanted to compare the performance of control and treatment children on this assessment, would you use a z-test, a t-test, or an analysis of variance? Explain why. *[5 points]*
   b. Use Stata to implement your preferred approach from part a). *[5 points]*
   c. According to your results from part b), do children in the control or treatment groups perform better, and if so, by what amount? *[5 points]*
   d. Use Stata to employ a simple regression to estimate the effect of treatment (as measured by the variable "treat") on children's executive function at midline (as measured by "per_exec2"). *[5 points]*
   e. Are your results from parts b) and d) consistent? How can you tell? *[5 points]*
   f. According to your results from part d), what share of variation in children's executive function scores at midline is explained by their belonging to experimental groups? Remember to show your work. *[5 points]*
   g. Imagine that you also wanted to compare the performance of children in the control and treatment groups on two other outcomes: numeracy and literacy at midline, as measured by the variables "per_math2" and "per_lang2", respectively. If you performed each of these three hypothesis tests at a 5% significance level, what would be the family-wise error rate across all three tests? *[5 points]*
   h. In the scenario from part g), what would be the Bonferroni-adjusted per-comparison error rate for each test? *[5 points]*
   i. Use Stata to estimate the effect of treatment on children's numeracy and literacy scores at midline. *[5 points]*
   j. According to your results from parts d) and i), and using the Bonferroni-adjusted per-comparison error rate from part h), did the treatment have a statistically significant effect on numeracy and literacy? How can you tell? *[5 points]*