Econ 270 / GSB 603   **Department of Economics**

Fall 2025   **Stanford University**

PROBLEM SET II

DUE: SUNDAY, OCTOBER 5$^{\text{th}}$, 2025, 6PM.

Be concise but clear as to what numbers you are reporting, and answer in full sentences. You should also hand in supporting code, but all answers should be in a PDF or Word document.

Use the experimental data from the Riverside GAIN experiment data posted on the course website under `riverside_2025.txt`. There are eight variables, the latter five of which are pre-treatment variables:

- the binary treatment indicator (1 for the individuals who were assigned to the training program, and 0 for those who were not);
- earnings one year after the program, in thousands of dollars;
- earnings four years after the program, in thousands of dollars;
- a binary indicator for high school graduation;
- a binary indicator for being female;
- age, in years;
- a binary indicator for having at least one child under the age of six; and
- a binary indicator for being single.

There are 5,419 observations in the sample. We are interested in analyzing the effect of the program on the two outcome variables (earnings one and four years after the program).

1. (a) Estimate the average treatment effect and construct 90% confidence intervals using the Neyman variance estimator and the homoskedastic variance.

   (b) Divide the sample into four subsamples, with the four subsamples defined by whether or not individuals have a high school degree and whether or not they have at least one child under the age of six. Estimate the average effect for each of the four subsamples, and calculate the Neyman standard error.

   (c) Combine the four estimates to get an estimate of the overall average effect and its standard error. How do you interpret the difference between this standard error and the ones derived in (a)?

2. Theoretical Calculations

    (a) Suppose we have a completely randomized experiment with N total units, M of which are treated. I am interested in the average effect for the treated. What is the true variance? Find an unbiased estimator for that variance.

    (b) I am interested in the average effect for the controls. What is the true variance? Find an unbiased estimator for that variance.

    (c) *(Bonus question)* Given that I have an unbiased estimator for the variance for the average effect for the treated, and an unbiased estimator for the variance of the average effect for the controls, can I use those to estimate the variance of the average effect for full sample?