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

Fall 2025   **Stanford University**

<div align="center">

PROBLEM SET IV

DUE: TUESDAY, OCTOBER 21th, 2025, 6PM.

</div>

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.

We use the census data from 1990, with about 300,000 observations. We will use the data on education, log wages and age.

1. Construct the age variable as 1990 minus year of birth. Construct the wage variable as $exp(lwage)$. Report the mean and standard deviation from both.

2. To investigate the properties of the homoskedastic and heteroskedastic standard errors regress the wage (not the log wage, the actual wage) on years of education.

3. Calculate the residuals, and their squares. Calculate the average squared residual for all 21 values of years of education, and the standard errors of these averages. Does it look like there is heteroskedasticity?

4. Calculate the homoskedastic and heteroskedastic standard errors for the regression in the first part. Are they different? Which one is bigger? What did you expect?

5. Now draw random samples of size 20 from this population, and calculate the least squares estimates and the standard errors. For what fraction of the random samples does the 95% confidence interval include the value for the population (from the first part of the exercise? is that closer to 0.95 for the homoskedastic or for the heteroskedastic standard errors?

6. Do the same for samples of size 200, and samples of size 2000.