

ECON 499: Capstone – Empirical Analysis
Assignment 1
Points: 20

Objective: Estimate the average difference in earnings and employment status between graduates with an economics degree and graduates with another social science degree.

Data: You should download the 2019 Integrated Public Use Microdata Set (IPUMS) sample.¹ You will need to select the following variables for your analysis: STATEFIP, PUMA, EMPSTAT, SEX, AGE, INCEARN, DEGFIELD, **and** one additional demographic variable of your choice.

Estimates: The centerpiece of your assignment will be a table with (exactly) five regression results. The five regressions (columns) are:

1. An OLS regression of log income on a dummy variable indicating if the worker's degree is in economics. Your sample should only include workers living in California with a bachelor's degree in social science. They should be currently employed.
2. Same as #1 above but include age, a dummy variable for female, and your demographic variable of choice as additional regressors, i.e. three additional right-hand side variables.
3. Same as #2 above but restrict the sample to workers living in Alameda or Contra Costa counties.²
4. A probit or logit regression (up to you). The left-hand side variable is a dummy indicating whether the person is employed or unemployed. The right-hand side variables are the same as in #2 above. Include only persons living in California with a social science degree who are in the labor force.
5. Same as #4 above but restrict the sample to persons living in Alameda or Contra Costa counties.

Style: Your assignment must be typed and follow the guidelines discussed in class. In particular, your assignment must include the following (clearly marked) sections:

1. Abstract: State the purpose of the empirical exercise (1 sentence), the data used (1 sentence), and the main results (1-2 sentences).
2. Data: Discuss the data you are using, e.g. what it is, where it is from, and the key variables

¹Steven Ruggles, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler and Matthew Sobek. IPUMS USA: Version 11.0 [dataset]. Minneapolis, MN: IPUMS, 2021.

²The following site will help here: <https://www.census.gov/geographies/reference-maps/2010/geo/2010-pumas/california.html>. If the link doesn't work, search for "census gov california pumas" and the first hit be the right one.

used in the analysis (1 paragraph). Include a table of summary statistics for the variables used in your regressions. Your discussion of the key variables can include references to the summary statistics if appropriate.

3. Empirical strategy: Present the regression equations that you will be estimating and define the variables contained in those equations. If two equations differ only in the number of right-hand side controls (e.g. #1 – #3), present and discuss only one equation.
4. Results: Present and discuss your table of results. Your discussion should include two paragraphs. In the first, discuss the main takeaways from the first three columns. In the second, discuss the main takeaways from the last two columns. In each paragraph, be sure to interpret exactly what the main coefficients mean and if they are statistically significant.
5. References: Please include a reference to the IPUMS data you are using. Use footnote 1 of this assignment as a guide.

Additional guidelines and tips:

- You must include a PDF of your Jupyter Notebook (or equivalent) as an additional attachment.
- It is usually optimal for a paragraph's first sentence to summarize what the rest of the paragraph is about.
- Your write-up must be your own work.
- IPUMS will add some identifier variables (e.g. YEAR, SAMPLE etc.) to your “cart” before you “checkout.” You can delete these variables to download a smaller CSV file if you are running low on space.
- It is a good idea to add comments in the cells of your Jupyter Notebook so anyone (including me) can follow what each cell is supposed to accomplish.