Empirical Exercise

This document describes the technical exercise for the research assistant position. Your responses will allow us to see the amount of prior exposure you have had to data work and the fit with the position. Be as careful and efficient in the treatment of the data as you can when answering the questions below, and be sure to document the steps of your analysis in detail so that we know what you were trying to do. Please send your log files or any program files along with the actual output in a compressed zip folder, and label each file with your name (last name, first name). Ideally, this will include a single pdf document which summarizes all of the key results. Your write-up needn't use Latex, but your results should be clean and presentable. **You will have a 5-hour window to complete the assignment, however it shouldn't take you more than 2.5 hours to finish it.**

Our objective is to assess the current status of your qualitative and quantitative skills related to empirical economics research, but a "perfect" score on this task is not a prerequisite to be considered for the job. With that in mind, please do this work yourself, without input from other people.

1. Download the Summary Extract Public Data from the Survey of Consumer Finances for the last 3 cycles (2010/13/16) and combine these into one dataset. You may download the raw data, many links to documentation, and sample code to reproduce various published statistics here: https://www.federalreserve.gov/econres/scfindex.htm. Please note that the SCF uses a weighted sampling methodology, so please be sure to use the correct sample weights (column wgt) in **all** calculations below. Also, there are numerous broken links to an article about the summary extracts which may be found here: https://www.federalreserve.gov/publications/files/scf17.pdf

2. Plot the distribution of total household income (column income) by year. Please make sure that they are representative of the population averages. Please make two versions of these plots with both the level of income and a logarithmic scale on the axis, respectively.

3. Regenerate the plots of the distribution of total household income separately for men (hhsex = 1) versus women (hhsex = 2) as heads of household. Please use a log scale and overlay the graphs for men and women onto a single plot.

4. Imagine you were asked to analyze how total household income varies with the age of the household. Please describe the analysis you would perform and implement it with the SCF data that you have. Your answer should involve interpretation of output from a linear regression model.

5. To your model from the previous part, please also add controls for (i) the number of kids (column kids) a household has, and (ii) indicator variables for each survey year. Can you provide an economic justification for the coefficient on the number of children? How much should we infer about the effect of having more children on income from this estimate? Explain your reasoning.

6. Re-estimate the same regression from part 5 using the natural logarithm of household income as the dependent variable.
    a. Discuss why it may be preferable to run the regression in logs instead of levels?
    b. In words, what is the interpretation of the coefficient on number of kids in this case?
    c. Report the number of observations. Why is it different from the previous regression? Can you recommend an alternative transformation of income that does not reduce the number of observations, and explain the pros and cons of your transformation versus the natural logarithm?

7. Now imagine you were asked to control for birth cohort effects when analyzing the relation between income and age. Which of the control variables from the previous regression (question X) do you need to drop to implement this analysis and why? Please describe the analysis you would perform and implement it with the SCF data that you have.

8. Finally, how would you test if the income-age profiles of households that have women as heads of household are different from those with men as the household head?