

# CIV 337 Homework 7

due: 10/16/24

Instructions: Please submit solutions on canvas. Only a knitted pdf of an Rmarkdown file will be accepted.

**Problem 1:** Financial literacy – the degree to which an individual is educated in budgeting, life-cycle economics, and basic finance – is typically thought to be a key predictor of future economic success and happiness. To date, books on financial wellness are almost as ubiquitous as those on diet and weight loss.<sup>1</sup>

The **National Financial Capability Study** (NFCS) is a large-scale survey that aims to measure financial capability across the United States. It was first administered in 2009 and every three years thereafter. The data from 2015 and 2018 are combined in the following csv: `finlit15and18.csv`. There are several variables, many of which are described in Table 4 of the footnoted paper above. See also the **2018 codebook** for the survey for detail about each variables values. We will be focusing on the following subset below:

<i>Variable</i>	<i>Description</i>
Y	Perceived economic condition, positive and negative values of Y are good and bad, respectively
literacy	Measure of financial literacy ranging from 0-6, 6 is most literate
A5_2015	Level of education
A3A	Age
J2	When thinking of your financial investments, how willing are you to take risks?
A3	Gender
A8	Household's income
E15	How many times have you been late with your mortgage payment?
E20	Do you owe more on your home than it is worth?
F2_2	Over the past 12 months, have you carried a balance and were charged interest?
F2_3	Over the past 12 months, in some months I paid the minimum payment only
F2_4	Over the past 12 months, I incurred credit card late fee
F2_5	Over the past 12 months, I was charged an over-the-limit fee for exceeding my credit line
F2_6	Over the past 12 months, I used my card for a cash advance

- First, how many observations are in the data? Second, describe the survey respondents across some of the variables. What is the gender and age breakdown? What is the distribution of respondents across household income (remember that household income, although coded numerically, is a categorical variable. Be sure to reference the **codebook** for category descriptions)?
- Compute the average `literacy` difference between females and males. Is this significantly different from zero (use the bootstrap to characterize the sampling distribution of this difference)? Provide a 95% confidence interval around the estimate of this difference. Treating `J2` as a numerical variable, conduct the same analysis for the average difference in `J2` between genders. What conclusions do you draw from these results?

---

<sup>1</sup>This exercise is based on the paper: **Financial Literacy and Perceived Economic Outcomes (2022)**.

- c. Fit a simple linear regression model of `literacy` on `gender` (A3). Report the coefficient and standard error on the `gender` variable. Run a bootstrap of this coefficient to characterize the sampling distribution. How does the standard deviation of the sampling distribution compare to the standard error from the regression output? How does the coefficient and sampling distribution spread compare to the results from the first part of (b)? Remember that in order to regress an outcome `Y` on a factor variable `X` (binary or categorical), you need to specify that the covariate is a factor in R. The R code to do this is: `fit = lm(Y ~ factor(X))`.
- d. Investigate the effect of financial literacy on perceived economic condition. One approach is to start with relatively small regression models (with a couple variables) and move to a large model with all 12 variables listed above. How well do your models describe the variation in perceived economic condition? Does the literacy effect change across models? If so, how and why?

**Problem 2:** This problem develops a nice causal inference approach using linear regression called **regression discontinuity**. Section 4.3.4 in QSS will be helpful here.

In this exercise, we estimate the effects of increased government spending on educational attainment, literacy, and poverty rates.<sup>2</sup>

Some scholars argue that government spending accomplishes very little in environments of high corruption and inequality. Others suggest that in such environments, accountability pressures and the large demand for public goods will drive elites to respond. To address this debate, we exploit the fact that until 1991, the formula for government transfers to individual Brazilian municipalities was determined in part by the municipality's population. This meant that municipalities with populations below the official cutoff did not receive additional revenue, while states above the cutoff did. The data set `transfer.csv` contains the variables:

<i>Variable</i>	<i>Description</i>
<code>pop82</code>	Population in 1982
<code>poverty80</code>	Poverty rate of state in 1980
<code>poverty91</code>	Poverty rate of state in 1991
<code>educ80</code>	Average years education of state in 1980
<code>educ91</code>	Average years education of state in 1991
<code>literate91</code>	Literacy rate of state in 1991
<code>state</code>	State
<code>region</code>	Region
<code>id</code>	Municipal ID
<code>year</code>	Year of measurement

- a. Begin by creating a variable that determines how close each municipality was to the cutoff that determined whether states received a transfer or not. Transfers occurred at three separate population cutoffs: 10,188, 13,584, and 16,980. Using these cutoffs, create a single variable that characterizes the difference from the closest population cutoff. Following the original analysis, standardize this measure by dividing the difference with the corresponding cutoff and multiply

<sup>2</sup>This exercise is based on Litschig, Stephan, and Kevin M Morrison (2013). [The Impact of Intergovernmental Transfers on Education Outcomes and Poverty Reduction](#). *American Economic Journal: Applied Economics* 5(4): 206-40.

it by 100. This will yield a normalized percent score for the difference between the population of each state and the cutoff relative to the cutoff value.

- b. Subset the data to include only those municipalities within 3 points of the funding cutoff on either side. Using regressions, estimate the average causal effect of government transfer on each of the three outcome variables of interest: educational attainment, literacy, and poverty. Give a brief substantive interpretation of the results.
- c. Visualize the analysis done in the previous question by plotting data points, fitted regression lines, and the population threshold. Briefly comment on the plot.
- d. Instead of fitting linear regression models, we compute the difference in means of the outcome variables between the groups of observations above the threshold and below it. How do the estimates differ from what you obtained in the earlier question? Describe the assumptions for both approaches and how they differ. Which estimates are more appropriate? Please discuss.