

Econ 184b, Econometrics, Assignment 5

The raw .rmd file for this assignment are available at: https://raw.githubusercontent.com/flkidd/Econ184/main/Assignment/Assignment_5.Rmd

Note: For the math problems, you can either (1) type your solution and compile it with RMarkdown; (2) type your solution use another word-processing software and convert it into a PDF file; (3) write your solution on paper, scan it and make it into a legible PDF file. TAs can decide, at their discretion, that the submitted file is illegible and thus give zero credit to a question or the entire problem set. If you type your solutions (options 1 and 2 above), you will get **a bonus equal to 5% of your performance** on the problem set. No credit will be given if you only report the final answers without showing intermediate steps whenever appropriate.

For the programming problems, you need to submit both the compiled pdf/html file and the RMarkdown (.rmd) file on LATTE. You will get **an additional bonus equal to 5% of your performance** if your RMarkdown file is completely reproducible with minimal alteration (install packages, etc.). That is, our team (or anyone else) should be able to recompile your Rmarkdown file and reach *the same* result. You need to explicitly write out your answers - just showing programming outputs will receive zero credit.

You will be receiving a total of 10% bonus if you submit one single pdf/html compiled directly from Rmarkdown, along with the .rmd file.

Question 1:

A researcher believes that traffic fatalities increase when roads are cold and icy and thinks that therefore states that are colder will have more fatalities than other states. Comment on the following methods designed to estimate the effect of snow on fatalities:

- The researcher collects data on the average temperature for each state and adds this regressor ($AverageTemp_i$) to the regression estimated two-way fixed effect: $FatalityRate_{it} = BeerTax_{it} + DrinkingAge_{it} + v_i + \lambda_t + u_{it}$
- The researcher collects data on average temperature in each state for each year in the sample ($AverageTemp_{it}$) and adds this regressor to the regression in part (a).
- The researcher collects data on temperature anomalies (difference between the annual temperature and the long-run average in that state) for each year in the sample ($TempAnomaly_{it}$) and adds this regressor to the regression in part (a).

Question 2 (Adapted version of Acemoglu, Johnson, Robinson, and Yared 2008 AER):

Do citizens demand more democracy and political freedom as their incomes grow? That is, is democracy a normal good? The data file “Income_Democracy.xlsx” contains a panel data set from 195 countries for the years 1960, 1965, . . . , 2000. A detailed description of the dataset is given here.

The dataset contains an index of political freedom/democracy for each country in each year, together with data on the country’s income and various demographic controls. (The income and demographic controls are lagged five years relative to the democracy index to allow time for democracy to adjust to changes in these variables.)

Part a: The index of political freedom/democracy is labeled Dem_ind.

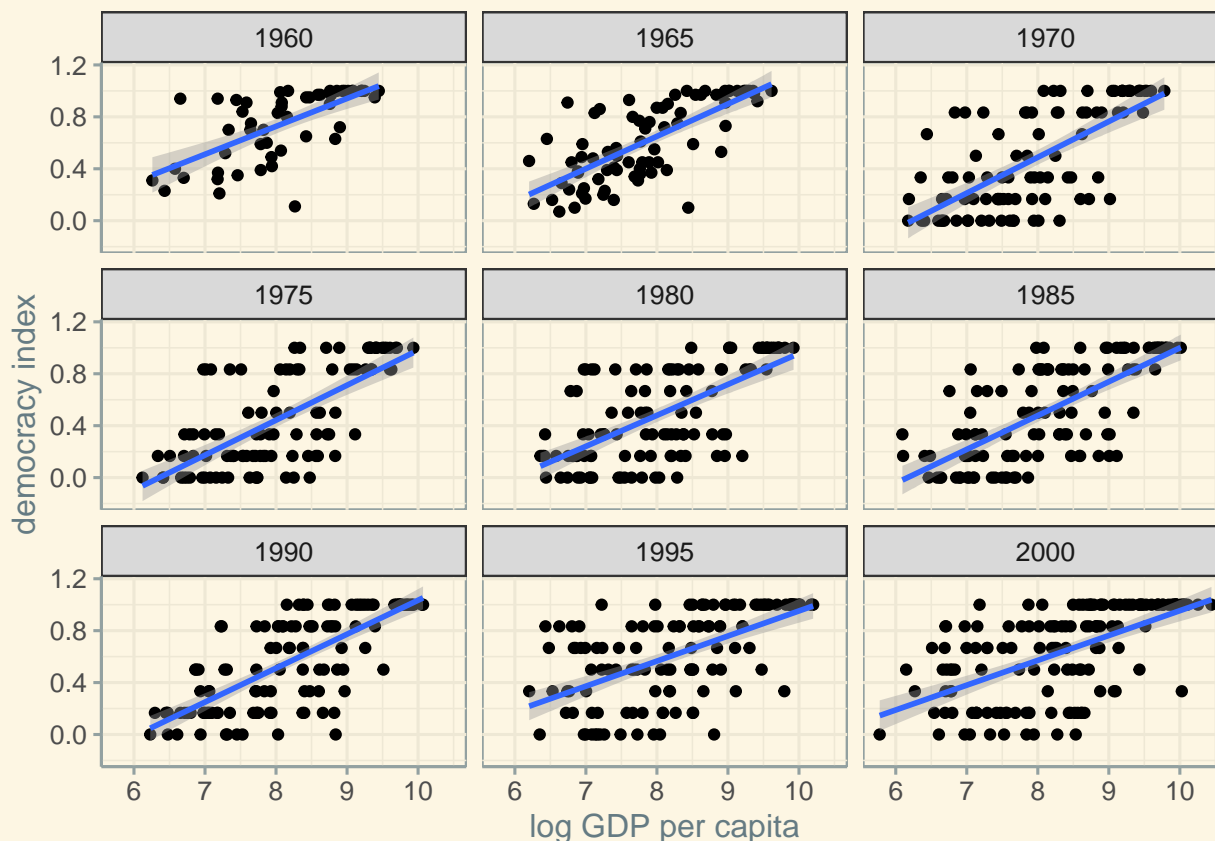
1. What are the minimum and maximum values of Dem_ind in the data set? What are the mean and standard deviation of Dem_ind?
2. What is the value of Dem_ind for the United States in 2000? Averaged over all years in the data set?
3. What is the value of Dem_ind for Libya in 2000? Averaged over all years in the data set?

Part b:

1. Plot the bivariate relationship between log GDP per capita (on the x axes) and the democracy score (on the y axes) by pooling all time periods together. Your plot should resemble this:



2. Plot a panel of bivariate relationships between log GDP per capita (on the x axes) and the democracy score (on the y axes), separating the nine time periods. Your plot should resemble this:



Part c: The logarithm of per capita income is labeled `Log_GDPPC`. Regress `Dem_ind` on `Log_GDPPC`. Use standard errors that are clustered by country.

1. How large is the estimated coefficient on `Log_GDPPC`? Is the coefficient statistically significant?
2. If per capita income in a country increases by 20%, by how much is `Dem_ind` predicted to increase? What is a 95% confidence interval for the prediction? Is the predicted increase in `Dem_ind` large or small? (Explain what you mean by large or small.)

Part d:

1. Suggest a variable that varies across countries but plausibly varies little—or not at all—over time and that could cause omitted variable bias in the regression in (c).
2. Suggest a variable that varies over time but plausibly varies little—or not at all—across countries and that could cause omitted variable bias in the regression in (c).

Part e:

1. Estimate the regression in (c), allowing for country fixed effects.
2. There are additional demographic controls in the data set. Should these variables be included in the regression? If so, how do the results change when they are included?

Part f:

1. Which of the models you just ran consistently estimate the effect of income on democracy? Why?
2. Based on your analysis, what conclusions do you draw about the effects of income on democracy?

Your code could start like this:

```
# place the data file in the same directory as your code
library(readxl)
data = read_excel("income_democracy.xlsx")

## to merge different datasets, use the left_join function
# e.g. data %<>% left_join(another_data,by=c("country","year"))
```