POLS/CSSS 503:
Advanced Quantitative Political Methodology

Problem Set 1

Professor: Chris Adolph, Political Science and CSSS

Spring Quarter 2014

Due in class, 15 April 2014

General instructions for homeworks: Homework can be handwritten or typed. For any exercises done with R or other statistical packages, you should attach all code you have written and all (interesting) output. Materials should be stapled together in order by problem. The most readable and elegant format for homework answers incorporates student comments, code, output, and graphics into a seamless narrative, as one would see in a textbook.

This homework involves many short problems, and aims to build familiarity with basic data acquistion and analysis using R; future homeworks will involve fewer but harder problems which build on the skills you develop here. Before tackling this assignment, you should read the assigned chapters from Zuur and the matrix algebra handout available on the course website. Start soon – this is manageable if you work a bit each day, but it is hard to catch up if you fall behind!

Problem 1: R practice.

For this problem, you will need the `democracy.csv` file in your working directory. This file contains data from Przeworksi, Alvarez, Cheibub & Limongi. *Democracy and Development: Political Insitutions and Well-being in the World, 1950–1990*. The data have been slightly recoded, to make higher values indicate higher levels of political liberty and democracy. Missing values are coded as ".".

Variable	Description
`COUNTRY`	numerical code for each country
`CTYNAME`	name of each country
`REGION`	name of region containing country
`YEAR`	year of observation
`GDPW`	GDP per capita in real international prices
`EDT`	average years of education
`ELF60`	ethnolinguistic fractionalization
`MOSLEM`	percentage of Muslims in country
`CATH`	percentage of Catholics in country
`OIL`	whether oil accounts for 50+% of exports
`STRA`	count of recent regime transitions
`NEWC`	whether county was created after 1945
`BRITCOL`	whether country was a British colony
`POLLIB`	degree of political liberty (1–7 scale, rising in political liberty)
`CIVLIB`	degree of civil liberties (1–7 scale, rising in civil liberties)
`REG`	presence of democracy (0=non-democracy, 1=democracy)

Table 1. Codebook for Problem 1. Data are in `democracy.csv`, and are taken from data from Przeworksi, *et al*, *Democracy and Development: Political Insitutions and Well-being in the World, 1950–1990*.

The answer to each of the following questions can be found using a single line of `R` code. The code you use does not need to be this concise, but if you are writing more than a line or two of code for each unstarred problem, look for a simpler way. And though it may not be obvious at first how, the starred problems should be answered with no more than two to six lines of code, rather than laborious looking up of data.

It is essential you show your work and all code.

a. Load the Democracy dataset into memory as a dataframe.

b. Attach it so that each variable in the dataset is accessible by name.

c. *Check whether `CTYNAME` has been read as a character variable (that is, as unique names for each case), or as factor variable (that is, a categorical variable that takes on a named value out of a menu of options). If it is a factor, convert it to character. (This is necessary for parts m., o., and q. to go smoothly.)

d. Report summary statistics (means and standard deviations, at least) for all variables.

e. Report a correlation matrix of all the variables in the dataset. Watch out for missing values.

f. Create a histograms for political liberties.

g. Create a histogram for GDP per capita.

h. Create a scatterplot of political liberties against GDP per captia.

i. Create a boxplot of GDP per capita for oil producing and non-oil producing nations.

j. On average, how many times smaller or greater is GDP per capita in countries with at least 40 percent Catholics, compared to those with fewer than 40 percent Catholics?

k. On average, how many times smaller or greater is GDP per capita in countries with more than 60 percent ethnolinguistic fractionalization, compared to those with less than 60 percent ethnolinguistic fractionalization?

l. What was the median average years of education in 1985 for all countries?

m. *Which country was (or countries were) closest to the median years of education in 1985 among all countries?

n. What was the median average years of education in 1985 for democracies?

o. *Which democracy was (or democracies were) closest to the median years of education in 1985 among all democracies?

p. What were the 25th and 75th percentiles of ethnolinguistic fractionalization for new countries?

q. Which country-years were nearest to the 75th percentile of ethnolinguistic fractionalization for new countries?

r. Specify a linear regression of your choosing and run it using `lm()`. Print the summary table.

s. What are some possible sources of bias in the regression you ran? Note each briefly, but list as many as you can.

Problem 2: Matrix Algebra

Let

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 5 & 0 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 2 & -3 \\ 7 & 1 & -1 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} 9 & 10 & 1 \\ 1 & 3 & 7 \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \mathbf{f} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$$

Compute each of the following matrices if it is defined. If it is not defined, state why it is not defined.

a. $\mathbf{A} + \mathbf{B}$

b. $\mathbf{AB}$

c. $\mathbf{B} + \mathbf{D}$

d. $\mathbf{AB}'$

e. $\mathbf{D}'\mathbf{f} + \mathbf{f}$

f. $\mathbf{De}$

g. $\mathbf{CC}$

h. $\mathbf{CA}$

i. $\mathbf{AC}$

j. $(\mathbf{C} + \mathbf{A})\mathbf{e}$

k. $\mathbf{e}'\mathbf{e}$

l. $\mathbf{ee}'$

m. $\mathbf{f}'\mathbf{f}$

n. $\mathbf{ff}'$

o. $\mathbf{ACA}$

p. $\mathbf{e}'\mathbf{Ce}$