

Problem Set 1

1. For each of the following research problems, specify an appropriate unit of analysis, the dependent variable, the independent variable. For each, use the approach discussed in the first class to write down the start of a table summarizing how the data would be organized if you collected it (you can just make up the values of the variables).
 - (a) You want to know whether men and women differ in their willingness to violate civil liberties in the fight against terrorism.
 - (b) You want to know whether the number of “great powers” in an international system is related to the frequency of great power (interstate) war.
 - (c) You want to know whether the percentage of Democrats in the military has been decreasing over time. (Note: No independent variable here.)
 - (d) Choose a *qualitative* work of political science that have read in another political or social science course. Identify what you think is the author’s main dependent variable (or just a major dependent variable), the unit of analysis, and one or more independent variables the author uses in his/her explanation. Begin a list like those given in class, identifying a few cases and values on the dependent and independent variables.
2. Use **dyteach.zip** (which unzips into a .dta file) to answer the questions below. (Note: **dyteach.dta** is very large, about 14K. You will need to **set mem 20000** before **use dyteach** in order to allocate enough memory. Possibly some of the machines in the computer cluster won’t accomodate this file, I don’t know.)
 - (a) For this historical period, are democracies more or less likely to be contiguous with each other? (show the relevant cross tab, and do not worry yet about formal tests of statistical significance).
 - (b) Are democracies more or less likely to be allies than other kinds of dyads?
 - (c) Are contiguous states more or less likely to be allies?
 - (d) What might the high proportion of contiguous allies in this period reflect? Should we infer that contiguity causes alliances?
3. Again using **dyteach.dta**,
 - (a) How much less likely were allied states to have (cow) wars with each other than are non-allied states in this period? Does anything surprise you about this finding? Should we infer that alliances are very poor reflections of whether states have common interests?
 - (b) Are contiguous allies more or less likely to have wars than nonallied states?
 - (c) Are non-contiguous allies more or less likely to have wars than nonallied states?
 - (d) Are contiguous states more or less likely to have wars?

- (e) Return to a and reconsider what might be going on here. Why does the relationship in b and c appear to be much stronger than that suggested by a?

4. *A McGlobalization Indicator.* There is a great deal of interest these days in “globalization.” This problem proposes two indicators or measures for globalization and asks you to put together a Stata data set using these indicators. You will do a little bit of analysis of this data on this problem set, and more on others.

The indicators are (1) the number of ©McDonalds restaurants operating at present (or at end of 2000) in a country, and (2) the year the first McDonalds opened in a country. Go first to the course website and download the data set **mcDs.dta**. This contains a list of countries, a country code for each, an estimate of per capita income in 1998, and an estimate of total population (in 1000s) in 1998 for each country. Next go to the McDonalds website (mcdonalds.com) and go through the country web sites one by one writing down the total number of restaurants and the year of first opening. Finally, **use mcDs** in Stata, and then type **edit country** to open the Stata spreadsheet. You can now input the data you collected. Be sure to click preserve and to save the data when you are done. Leave countries with no McDonalds (or no info on the website) as missing data (*be sure to treat the US as missing data*, for now). To name the variables something other than **var6**, etc., doubleclick on the variable name at the top of the column and a naming window should open.

(Note: If you are daunted by this data collection and entry effort, pool your resources with the people you are talking with about the problem set. It shouldn’t be bad though – I got the data from the websites and into Stata in about 50 minutes.)

- (a) Generate a new variable that equal the aggregate GDP of each country (i.e., per capita income times total population (in 1000s)). Then generate a new variable that is the base 2 log of aggregate GDP.
- (b) For countries with at least one McDonalds in these data, produce a scatterplot of the number of restaurants against the (base 2) log of aggregate GDP. Use Stata’s labelling and symbol features to make the graph look nice.
- (c) Use Stata’s **regress** command to compute the regression of the number of restaurants on the log of aggregate $\overline{\text{GDP}}$ for the thirty or so countries with data. Draw (on paper) the regression line indicated by the results. (If you are feeling ambitious, figure out how to get Stata to draw the line on the scatterplot. What you need to do is generate a new variable from the equation for the line, type **graph [restaurants] [new variable] [ln2GDP]** and then add an option to connect the dots for the new variable.)
- (d) Interpret the substantive meaning of the regression coefficient estimated for the independent variable base 2 log of aggregate GDP. That is, answer the question “A doubling of GDP is associated, on average, with how great an increase in the number of McDonalds?”
- (e) One indication of how “globalized” a country is might be how many more McDonalds the country has than one would expect based on country GDP. “Eyeballing” the scatterplot, which countries appear more “globalized” than you would expect based on this measure? (Hint: You might want to redo the scatterplot with the condition **if country = “Japan”**, which is an “outlier.”)
- (f) (extra credit). Compute the regression of (base 2) log of the number of restaurants on the base 2 log of aggregate GDP, and repeat the interpretation asked for in (d) above. Hint: you should now be answering the question “A doubling of GDP is associated, on average, with an increase *by what factor* in the number of McDonalds?”