Lab 10: Debt, the Last Seventy Years

23 October 2015

Agenda: Practicing split-apply-combine.

Gross domestic product (GDP) is a measure of the total market value of all goods and services produced in a given country in a given year. The percentage growth rate of GDP in year t is

$$100 \times \left(\frac{GDP_{t+1} - GDP_t}{GDP_t} \right) - 100$$

An important claim in economics is that the rate of GDP growth is closely related to the level of government debt, specifically with the ratio of the government's debt to the GDP. The file `debt.csv` on the class website contains measurements of GDP growth and of the debt-to-GDP ratio for twenty countries around the world, from the 1940s to 2010. Note that not every country has data for the same years, and some years in the middle of the period are missing data for some countries but not others.

1. Load the data into a dtaframe called `debt`. You should have 1171 rows and 4 columns.

2. Calculate the average GDP growth rate for each country (averaging over years). You may use `aggregate`, `split`, and/or `tapply` (and of course `mean`). Don't use something like `mean(debt$growth[debt$Country=="Australia"])`, except to check your work. (The average growth rates for Australia and the Netherlands should be 3.72 and 3.03.) Present a table of the twenty growth rates.

3. Calculate the average GDP growth rate for each year (averaging over countries). Again, don't use something like `mean(debt$growth[debt$Year==1972])` except to check your work. (The average growth rates for 1972 and 1989 should be 5.63 and 3.19, respectively.) Make a plot of the growth rates versus the year. Make sure the axes are labeled appropriately.

4. The function `cor(x,y)` calculates the correlation coefficient between two vectors `x` and `y`.

   a. Calculate the correlation coefficient between GDP growth and the debt ratio over the whole data set (all countries, all years). Your answer should be -0.1995.
   b. Calculate the correlation coefficient separately for each country, and plot a histogram of these coefficients. The mean of these correlations should be -0.1778. Do not use a loop. *Hint*: consider writing a function and then making it an argument to `sapply()`.
   c. Calculate the correlation coefficient separately for each year, and plot a histogram of these coefficients. The mean of these correlations should be -0.1906.
   d. Are there any countries or years where the correlation goes against the general trend?

5. Make a scatter-plot of GDP growth (vertical) against the debt ratio (horizontal). Describe the over-all shape of the point cloud in words. Does its shape match what you expect from problem 4?