

Introduction to Statistics

lab 4

Johan A. Elkind
jos.elkind@ucd.ie

3 October 2016

Downloading the data

For this lab we will use the replication data from Ross (2004), which can be accessed on the course teaching data page¹ and is listed as “taxation & representation” – download only the 1997 data. See the instructions for Lab 1 on how to open this file, which is also in Stata format. Remember to always save the file first, then open the statistical package of your choice, then open the file – double-clicking the file name does not work. Make sure you also open the description file, so that you have a list of variables in the data set, with brief descriptions.

For R users, I will assume you opened the data using “ross” as the data set name, i.e.

```
ross <- read.dta(...)
```

Democracy and development

We will look at the relationship between economic development (as measured by GDP per capita) and democracy. As usual we take the logarithm of GDP per capita, which in this case is already a variable in the data set.

1. Calculate the mean and standard deviation of **loggdppc** and **democracy**.
2. Calculate Pearson’s correlation between the two variables.
3. Regress **democracy** on **loggdppc**.²

SPSS: REGRESSION /DEPENDENT democracy /METHOD = ENTER loggdppc.

R: `lm(democracy ~ loggdppc, data = ross)`, however, you get better information when you also use `summary()`: `summary(lm(democracy ~ loggdppc, data = ross))`.

¹<http://www.joselkink.net/teaching/teaching-data/>

²Note that in the formulation “regress Y on X”, Y is the dependent variable and X the independent variable(s).

Furthermore, you would probably want to save the regression output for subsequent steps, so the complete command is: `summary mdl <- lm(democracy ~ loggdppc, data = ross)` (whereby `mdl` is an arbitrary object name).

Stata: `reg democracy loggdppc`

4. Produce a scatter plot of the two variables.

5. Add the regression line to the scatter plot.

SPSS: In SPSS, you cannot use the old-fashioned syntax to include the regression line. The replacement of the `GRAPH` command, `GGRAPH`, can do this, but the code is quite complicated.³ Instead, you can create the scatter plot, then double-click on the plot, select “chart” from the top menu, then “options”, then under “Fit line” there is a checkbox with “Total”. Then close the Chart Editor and you should see the fitted regression line.

R: After the plot command, you can use: `abline mdl`.

Stata: `graph twoway (lfit democracy loggdppc) (scatter democracy loggdppc)`.⁴

6. What do you conclude about the relationship between democracy and development? Make sure you can identify and interpret the intercept, the slope coefficient, and R^2 .

No taxation without representation

The main purpose of Ross (2004) is to evaluate whether there is, indeed, a relationship between democracy and taxation. He tests whether higher taxation does indeed lead to democracy, but perhaps the argument should be the other way around, that to generate high tax revenues, one needs democracy.

1. Calculate the mean and standard deviation of **taxrevenue** and **democracy**.
2. Calculate Pearson’s correlation between the two variables.
3. Regress **taxrevenue** on **democracy**.
4. Produce a scatter plot of the two variables.
5. Add the regression line to the scatter plot.
6. What do you conclude about the relationship between democracy and tax revenue? Make sure you can identify and interpret the intercept, the slope coefficient, and R^2 .
7. Repeat, using only the observations from Sub-Saharan Africa. What do you conclude? Make sure you can identify and interpret the intercept, the slope coefficient, and R^2 .

³See: <http://www.ats.ucla.edu/stat/spss/faq/scatter.htm>.

⁴In Stata it is particularly easy to also add confidence intervals: `graph twoway (lfitci democracy loggdppc) (scatter democracy loggdppc)`. See <http://www.stata.com/manuals13/g-2graphtwowaylfitci.pdf>.

Democracy and country size

One might argue that in larger countries, the challenges of government are sufficiently larger that democracy is more difficult to sustain.

1. Calculate the mean and standard deviation of **population** and **democracy**.
2. Calculate Pearson's correlation between the two variables.
3. Regress **democracy** on **population**.
4. Produce a scatter plot of the two variables.
5. Add the regression line to the scatter plot.
6. What do you conclude about the relationship between democracy and population size? Make sure you can identify and interpret the intercept, the slope coefficient, and R^2 .
7. Repeat, using the logarithmic transform of population size.

References

Ross, Michael. 2004. "Does taxation lead to representation?" *British Journal of Political Science* 34:229–249.