

Political Analysis: Lab 4

Hilary Term 2015

NOTE: a PDF version of this document is available at <http://andy.egge.rs/teaching/lab4.pdf>.

Regression

In this lab we will examine the relationship between environmental performance and consensus democracy.

1. Load the Lijphart dataset.

```
L <- read.csv("http://andy.egge.rs/data/L.csv")
```

2. Inspect the environmental performance measure that Lijphart uses.

- Look at the values of the variable. How many are missing? For which countries? *Hint:* You can simply look at the variable in the upper-left pane of RStudio, but it might be easier to look at the two key columns in isolation using a command like

```
L[, c('country', 'enviro_performance_index_2010')]
```

The `is.na()` command may also be helpful, in conjunction with `table()` and/or subsetting of the country variable, e.g.

```
L$country[is.na(L$enviro_performance_index_2010)]
```

- What country has the highest value of the environmental index?
 - What is the average value of the environmental index? *Hint:* Don't forget to specify `na.rm = TRUE` to tell R to ignore missing values.
 - Plot a histogram of the environmental index.
3. Create a scatterplot (using the `plot()` command) in which environmental performance (`enviro_performance_index_2010`) is on the vertical axis and Lijphart's executive/parties measure of consensus democracy (`exec_parties_1945_2010`) is on the horizontal axis.) Does the correlation between the two variables look positive or negative?
 4. Calculate the correlation between the two variables. *Hint:* The syntax here is:

```
cor(first_variable, second_variable, use = 'complete')
```

5. Use the `lm()` command to regress the environmental performance index (dependent variable, i.e. Y) on Lijphart's executive/parties measure of consensus democracy (independent variable, i.e. X). *Hint:* The syntax here is:

```
lm(dependent_variable ~ independent_variable)
```

- What is the estimate of the intercept in this regression? What does this tell you?
- What is the estimate of the coefficient on the independent variable (consensus democracy)? What does this tell you?
- What is the standard error of the coefficient on the independent variable? What does this tell you? *Hint:* You can get regression coefficients simply by entering

```
lm(dependent_variable ~ independent_variable)
```

but the easiest way to see detailed regression results is by wrapping that command in the `summary()` command, as follows:

```
summary(lm(dependent_variable ~ independent_variable))
```

- What is the p -value associated with the coefficient on the independent variable? What does this tell you? (As above, you can most easily see this p -value by using the `summary()` command.)
6. Create a variable that takes the value `TRUE` for countries in Europe and `FALSE` otherwise. *Hint:* There are many ways to do this, but the easiest is probably this:

```
L$europe = L$country %in% c('AUT', 'BEL', 'DEN', 'FIN', 'FRA', 'GER',  
                           'ICE', 'IRE', 'ITA', 'LUX', 'MAL', 'NET',  
                           'NOR', 'POR', 'SPA', 'SWE', 'SWI', 'UK')
```

You can copy and paste into your code, assuming your Lijphart dataframe is called `L`.

7. What is the mean value of the environmental policy index for countries in Europe? How about for countries not in Europe? *Hint:* Try this:

```
mean(L$enviro_performance_index_2010[!L$europe], na.rm = T)
```

8. Rerun the regression, this time controlling for whether a country is in Europe.
- What is the estimate of the intercept? What does this tell you?

- What is the estimate of the coefficient on consensus democracy? What does this tell you?
 - Is the relationship between environmental policy and consensus democracy statistically significant when we control for whether a country is in Europe?
 - If we want to understand the effect of consensus democracy on environmental policy, is it appropriate to control for whether a country is in Europe?
9. Rerun the original regression, but this time restrict the analysis to countries in Europe. One way to do this:

```
lm(enviro_performance_index_2010 ~ exec_parties_1945_2010,
    data = L[L$europe==TRUE,])
```

Another way to do this:

```
lm(L$enviro_performance_index_2010[L$europe==TRUE] ~
    L$exec_parties_1945_2010[L$europe==TRUE])
```

- What is the estimate of the intercept? What does this tell you?
- What is the estimate of the coefficient on consensus democracy? What does this tell you?
- Is the relationship between consensus democracy and environment policy statistically significant when we focus only on Europe? (*Hint*: look at the `summary()` of the regression output.)