

# Introduction to Statistics

## homework 2

Johan A. Elkink  
jos.elkink@ucd.ie

*Due 30 October 2018, 5 pm*

### Submission instructions

You will submit **two files**: one **PDF** file<sup>a</sup> including all plots, tables and interpretations and one command file (SPSS Syntax file (*.sps*), or Stata do-file, or R-file) with all commands used to answer the exercise and no superfluous commands. R Markdown files in PDF format that include both code and interpretation will also be accepted.<sup>b</sup>

Please ensure that the **file names** contain your student number and the module code (POL40950), in the following format: Homework2\_POL40950\_18123456.pdf and Homework2\_POL40950\_18123456.do (please note absence of spaces and be precise in your naming!).

Your name should not appear in either the file name or the file contents.

Please send both files to [jos.elkink@ucd.ie](mailto:jos.elkink@ucd.ie)

<sup>a</sup>Word files will be sent back—note that newer versions of Word can easily save to PDF format.

<sup>b</sup>If you only have the HTML version, you can open it in a web browser and then print to PDF.

(5%) of the grade for the evaluation of the clarity / presentation of your command file, including the use of comments, clear variable names, and whitespace.

### Data

The data set we will use for this homework is based on the European Social Survey, using a random subset of Irish respondents in the 2010–11 wave of the survey. You can find the data on the teaching data page<sup>1</sup>, including a description of each variable. While we use a subset

<sup>1</sup><http://www.joselkink.net/data.php>

of the data, with fewer variables and all missing data removed, you will be able to find the original data on the BES website.<sup>2</sup>

## Questions

1. (5%) Open the data and make sure the command for doing so is in the command file. The data can be directly opened using the following URL:  
<http://www.joselkink.net/wp-content/uploads/2013/01/ess.dta>.  
Note that in SPSS, the variable labels will be automatically visible, but you might have to adjust measurement levels if they are inappropriately set, for those variables that you use in this homework. Make sure to include this in the SPSS Syntax file when you do so.

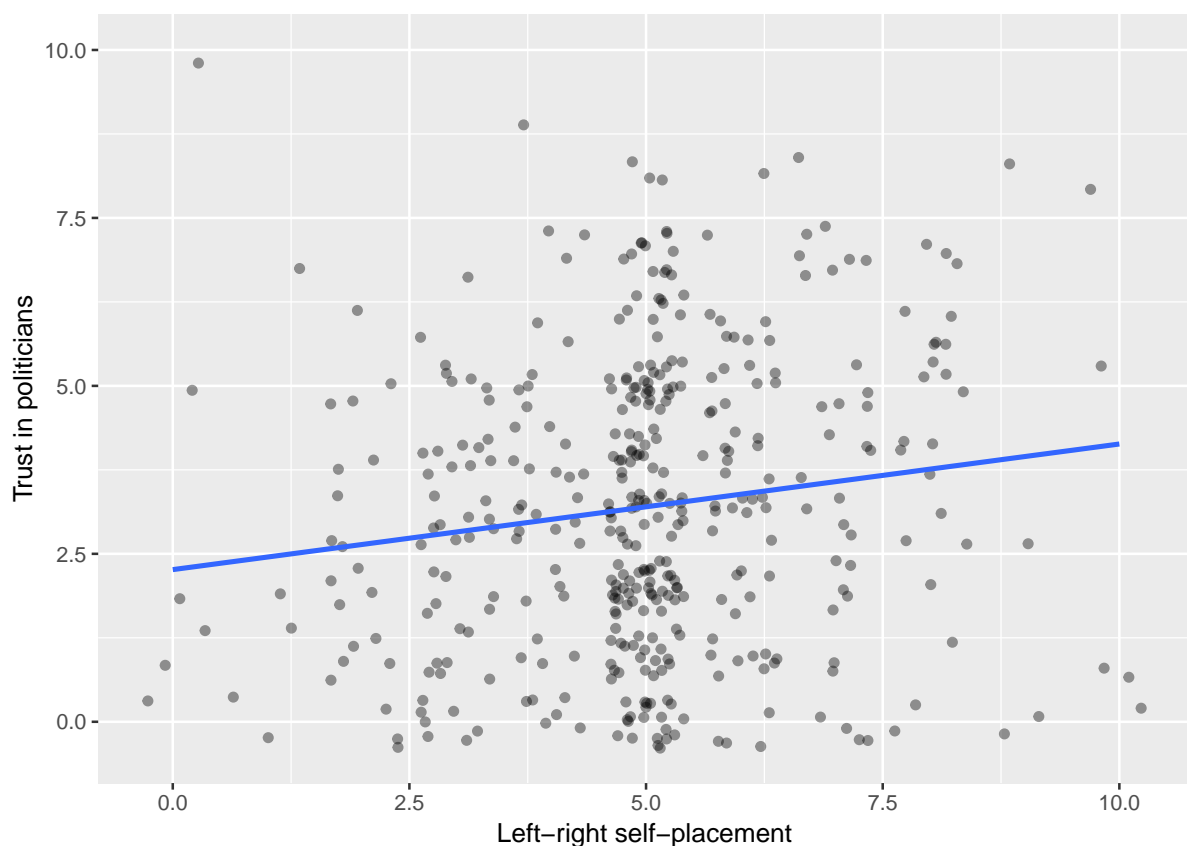


Figure 1: Trust in politicians by left-right self-placement, including linear regression line.

2. Figure 1 shows the relationship between left-right self-placement of respondents (**lr**) and the level of trust in politicians (**trust\_pol**).
  - (5%) Calculate variances and standard deviations for both variables.
  - (5%) Calculate the covariance and correlation coefficient between those two variables.

<sup>2</sup><https://www.europeansocialsurvey.org/data/>

- (5%) Perform the regression analysis that is depicted by the line in Figure 1.
- (5%) Produce a publishable regression table.<sup>3</sup>
- (10%) Fully interpret the regression coefficients and the  $R^2$ . What do you conclude about the relationship between the two variables? (approx. 200 words)

	<i>Dependent variable:</i>
	Government satisfaction
Pro-immigration attitude	0.175*** (0.042)
<i>Intercept</i>	2.930*** (0.232)
Observations	479
R <sup>2</sup>	0.035

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 1: OLS regression explaining government satisfaction by attitude towards immigration.

- Table 1 produces a regression table whereby the level of satisfaction with the government (**satisf\_gov**) is regressed on the attitude towards immigration **immigrants** variable.
  - (5%) Replicate the regression analysis.
  - (5%) Standardize both variables.
  - (5%) Repeat the regression analysis using the standardized variables.
  - (8%) Produce a regression table identical to Table 1, but with a column next to it including the regression results using standardized variables.
  - (15%) Fully discuss what you would conclude about the relationship between attitude towards immigrants and towards the government. In particular fully interpret the coefficients of both regressions and  $R^2$ . (approx. 300 words).
- For this question you will not need any software. We will look at the relationship between the number of seats in parliament and the size of the population. Only those parliaments classified as “lower house” in the table have been included.<sup>4</sup> We will investigate two models:

$$Seats_i = \beta_1 + \beta_2 Population_i \quad (1)$$

$$\log(Seats_i) = \beta_1 + \beta_2 \log(Population_i) \quad (2)$$

<sup>3</sup>See [http://www.joselkink.net/wp-content/uploads/2013/01/POL50050\\_Spring\\_2013\\_note\\_regression\\_presentation\\_and\\_interpretation.pdf](http://www.joselkink.net/wp-content/uploads/2013/01/POL50050_Spring_2013_note_regression_presentation_and_interpretation.pdf).

<sup>4</sup>The data is taken from [http://en.wikipedia.org/wiki/List\\_of\\_legislatures\\_by\\_country](http://en.wikipedia.org/wiki/List_of_legislatures_by_country).

For Model (1), the regression coefficients can be found in Table 2 and the scatter plot with regression line can be found in Figure 2. For Model (2), the regression coefficients can be found in Table 3 and the regression line in Figure 3.

- (a) (3%) Based on Model (1) in Table 2, what do you conclude about the relationship between a country's population and the size of the lower house?
- (b) (5%) Based on Model (1) in Table 2, if the population size of country A is one million people larger than the population size of country B, how much would you expect their size of their lower houses to differ? Which one will be larger? Include the reasoning and/or calculation.
- (c) (3%) Both visually in Figure 2 and based on Table 2, how well do you think this linear regression summarizes the data? Explain why you think so.
- (d) (3%) For Model (2) in Table 3, both variables are transformed before performing the linear regression. Both visually and based on the information in the table, how does this affect the extent to which the model describes the data? Explain why you think so.
- (e) (3%) Based on both regressions, tables 2 and 3, what would you conclude about the Irish and Indian cases: Are they typical or unusual compared to other countries? Explain why you think so.
- (f) (5%) Based on Model (2) in Table 3, if country C has a population of 30 million, how many seats would you expect the lower house in parliament to have? Include the reasoning and/or calculation.

Population	0.000000562
<i>Intercept</i>	199.6
<i>N</i>	77
<i>R</i> <sup>2</sup>	0.20

Table 2: Regression based on Eq. (1)

Log(Population)	0.436
<i>Intercept</i>	-2.03
<i>N</i>	77
<i>R</i> <sup>2</sup>	0.84

Table 3: Regression based on Eq. (2)

## References

### Grade conversion scheme

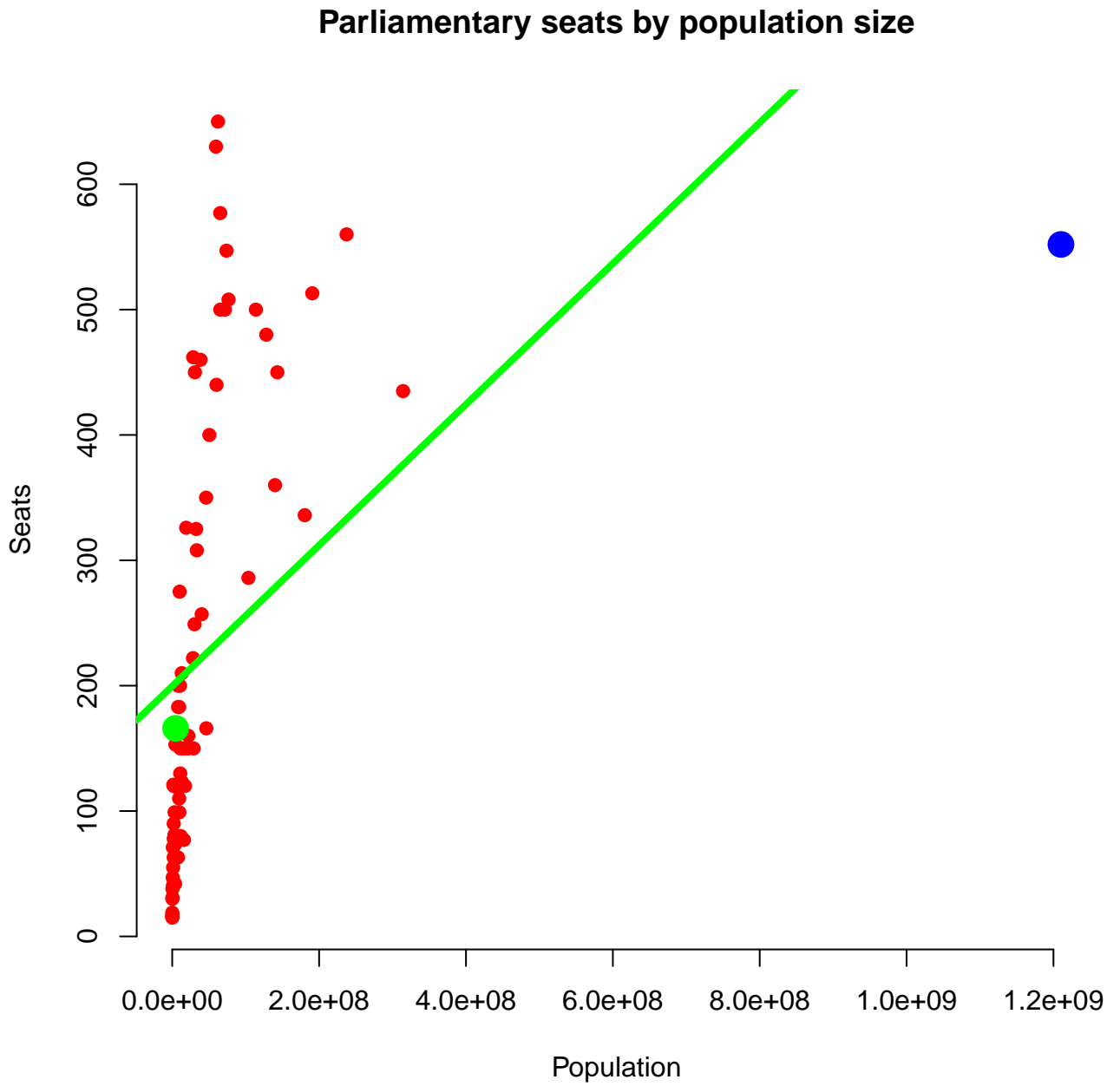


Figure 2: Seats by population and linear regression line. Green dot represents Ireland and blue dot represents India.

### Parliamentary seats by population size

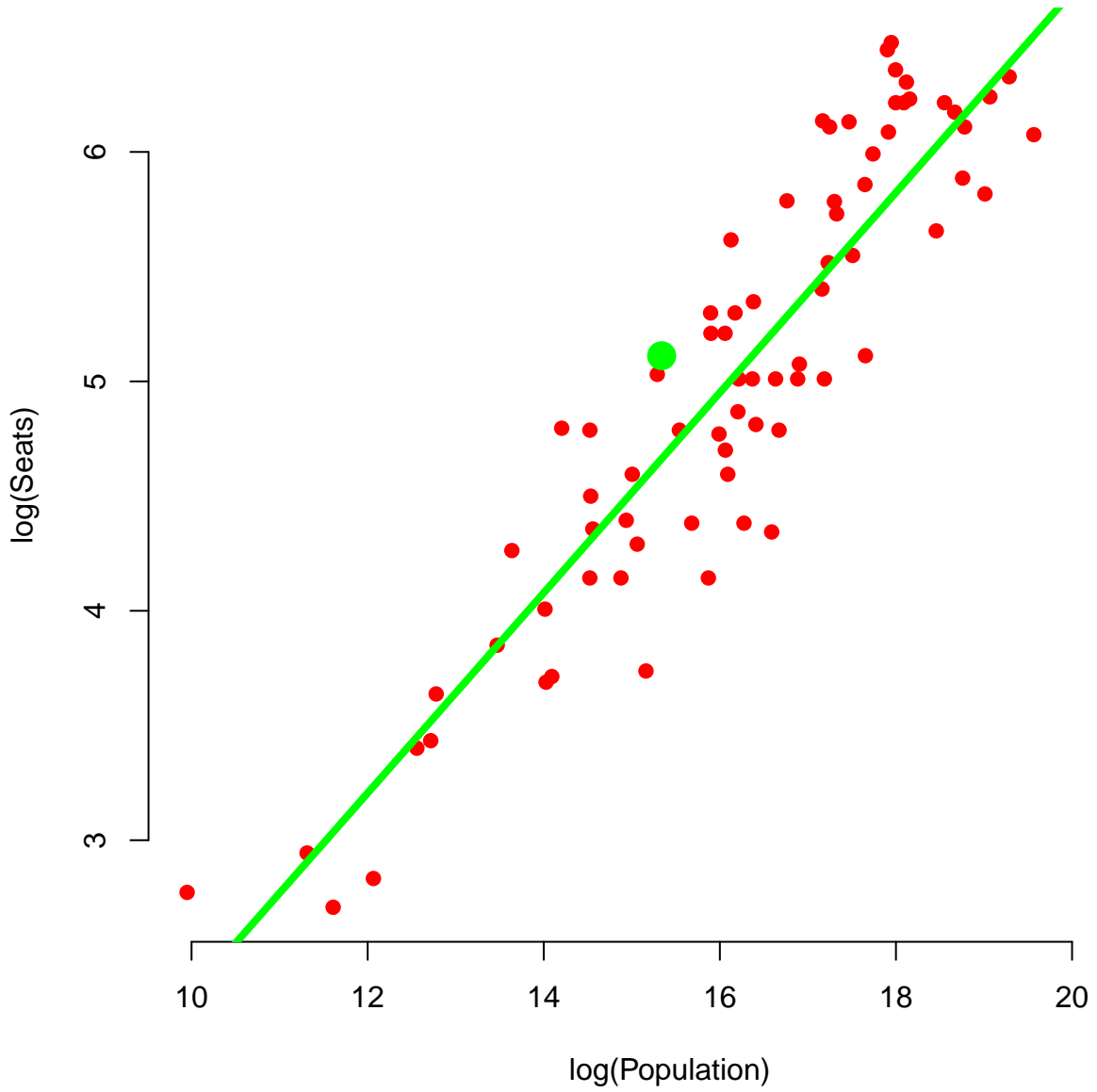


Figure 3: Seats (logged) by population (logged) and linear regression line. Green dot represents Ireland and blue dot represents India.

Score	Grade		Score	Grade		Score	Grade		Score	Grade	
	UCD	TCD		UCD	TCD		UCD	TCD		UCD	TCD
97-100%	A+	A+	85-87%	B	B	74-76%	C-	C	54-64%	E+	D
94-96%	A	A	83-84%	B-	B	71-73%	D+	C	44-53%	E	D
91-93%	A-	A	80-82%	C+	C+	68-70%	D	C	33-43%	E-	D
88-90%	B+	B+	77-79%	C	C	65-67%	D-	C	0-32%	F	F

Note that all results are entered in grades, not percentages, and the average score at the end of the module is therefore also based on the grades, not the percentages.