

Introduction to Statistics

homework 2

Johan A. Elkink
jos.elkink@ucd.ie

Due 25 October 2017, 5 pm

You will submit two files: one PDF file¹ 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.²

Please send both files to jos.elkink@ucd.ie.

(5%) of the grade is used for an overall evaluation of the presentation of your work (the PDF file) and (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 British Election Survey, using the data from the surveys around the Brexit referendum. While we use a subset of the data, with fewer variables and all missing data removed, you will be able to find the original data on the BES website.³

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/files/data/brexit_subset.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

¹Word files will be sent back—note that newer versions of Word can easily save to PDF format.

²If you only have the HTML version, you can open it in a web browser and then print to PDF.

³<http://www.britishelectionstudy.com/data-object/wave-10-of-the-2014-2017-british-election-study-internet-panel/>

use in this homework. Make sure to include this in the SPSS Syntax file when you do so.

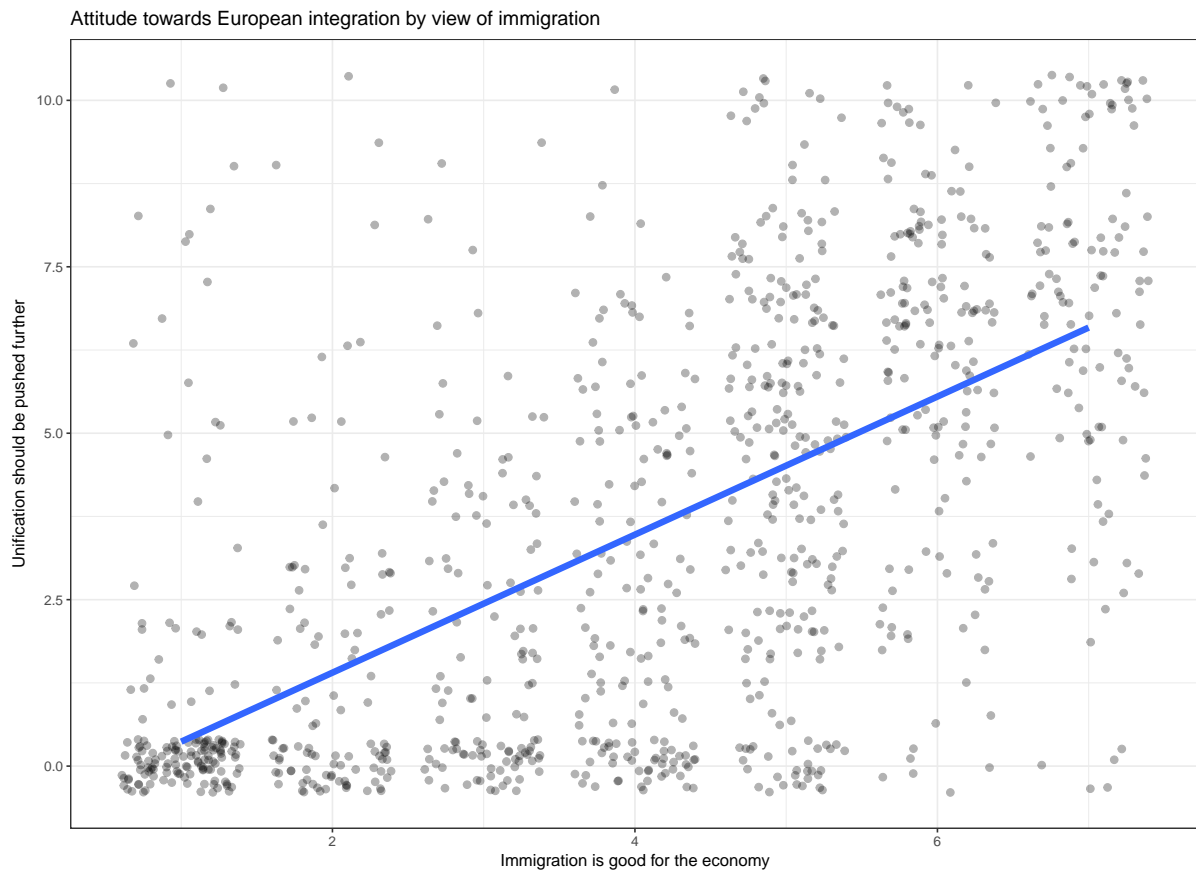


Figure 1: Attitude towards European integration, 0 means “Unification has already gone too far” and 10 means “Unification should be pushed further”, by evaluation of impact of immigration on the economy, 1 means “bad for the economy” and 7 is “good for the economy”.

2. Figure 1 shows the relationship between attitudes towards immigrants (**immigrationEcon**)—more specifically, their impact on the UK economy—and towards European integration (**proIntegration**).
 - (5%) Calculate variances and standard deviations for both variables.
 - (5%) Calculate the covariance and correlation coefficient between those two variables.
 - (5%) Regress the attitude towards European integration on the attitude towards immigrants.
 - (5%) Produce a publishable regression table.⁴
 - (10%) Fully interpret the regression coefficients and the R^2 . What do you conclude about the relationship between the two variables? (approx. 200 words)

⁴See http://www.joselkink.net/wp-content/uploads/2013/01/POL50050_Spring_2013_note_regression_presentation_and_interpretation.pdf.

| | |
|-----------------------|--|
| Pro-Integration | -0.66*** (0.03) |
| <i>Intercept</i> | 5.74*** (0.13) |
| <i>N</i> | 1000 |
| <i>R</i> ² | 0.385 |
| <i>Note:</i> | Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01 |

Table 1: OLS regression explaining liking Nigel Farage by attitude towards European integration.

3. Table 1 produces a regression table whereby the **likeFarage** variable is regressed on the **proIntegration** variable.
 - (a) (5%) Replicate the regression analysis.
 - (b) (5%) Standardize both variables.
 - (c) (5%) Repeat the regression analysis using the standardized variables.
 - (d) (8%) Produce a regression table identical to Table 1, but with a column next to it including the regression results using standardized variables.
 - (e) (10%) Fully discuss what you would conclude about the relationship between attitude towards European integration and liking Nigel Farage. In particular fully interpret the coefficients of both regressions and R^2 . (approx. 200 words).
4. 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.⁵ We will investigate two models:

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

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

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

⁵The data is taken from http://en.wikipedia.org/wiki/List_of_legislatures_by_country.

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> ² | 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> ² | 0.84 |

Table 3: Regression based on Eq. (2)

Grade conversion scheme

| Homeworks | UCD | MDP | Homeworks | UCD | MDP |
|-----------|-----|-------|-----------|-----|-------|
| 97-100% | A+ | 78.33 | 74-76% | C- | 51.67 |
| 94-96% | A | 75.00 | 71-73% | D+ | 48.33 |
| 91-93% | A- | 71.67 | 68-70% | D | 45.00 |
| 88-90% | B+ | 68.33 | 65-67% | D- | 41.67 |
| 85-87% | B | 65.00 | 54-64% | E+ | 38.33 |
| 83-84% | B- | 61.67 | 44-53% | E | 35.00 |
| 80-82% | C+ | 58.33 | 33-43% | E- | 31.67 |
| 77-79% | C | 55.00 | 0-32% | F | 25.00 |

Note that the percentage scores will be translated to UCD grades before entering on the system. Overall module grade will be calculated by the system based on the UCD grades. For MDP students, grades will then be translated to TCD marks. Note that TCD marks are

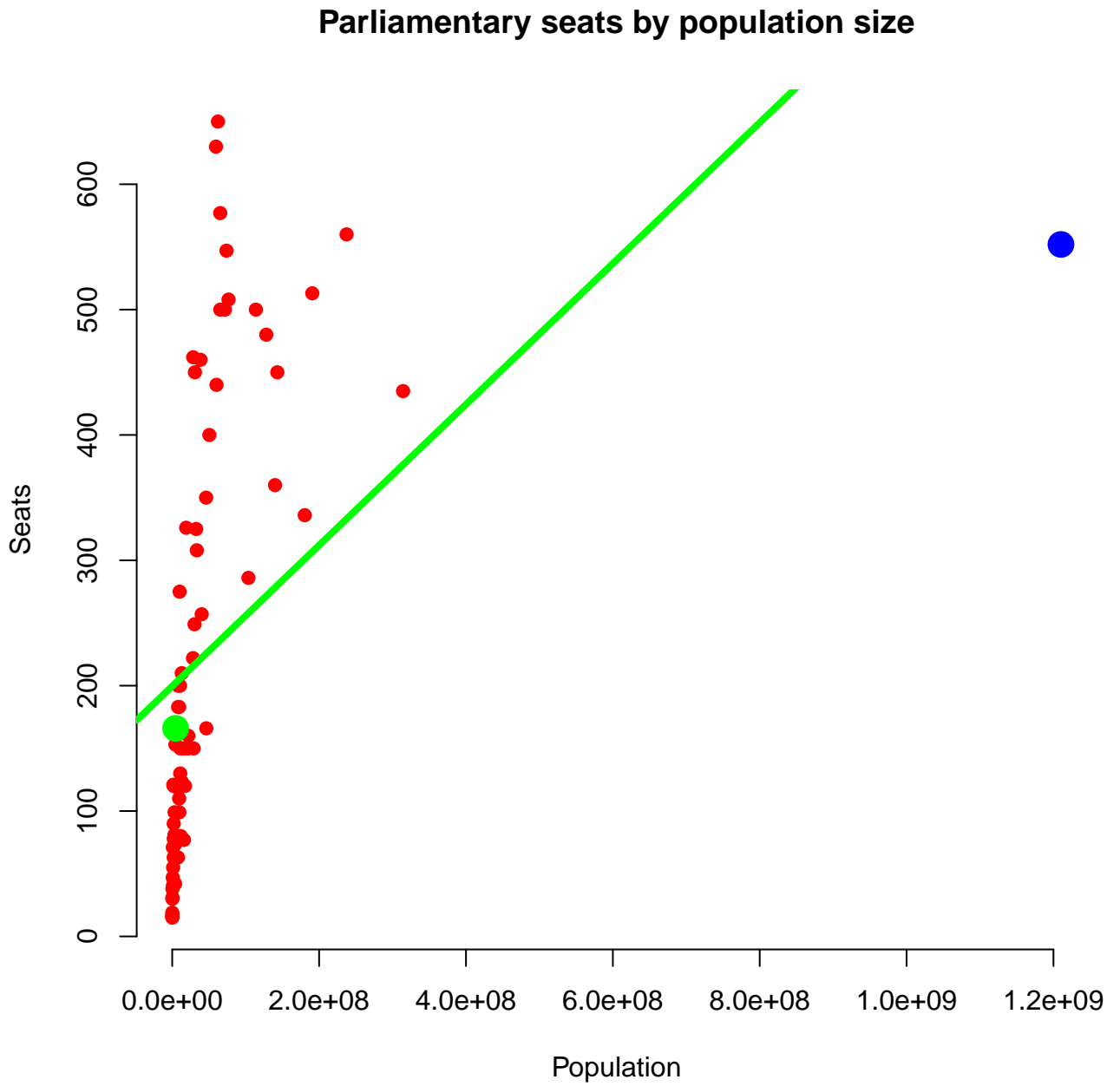


Figure 2: Seats by population and linear regression line. Green dot represents Ireland and blue dot represents India.

Parliamentary seats by population size

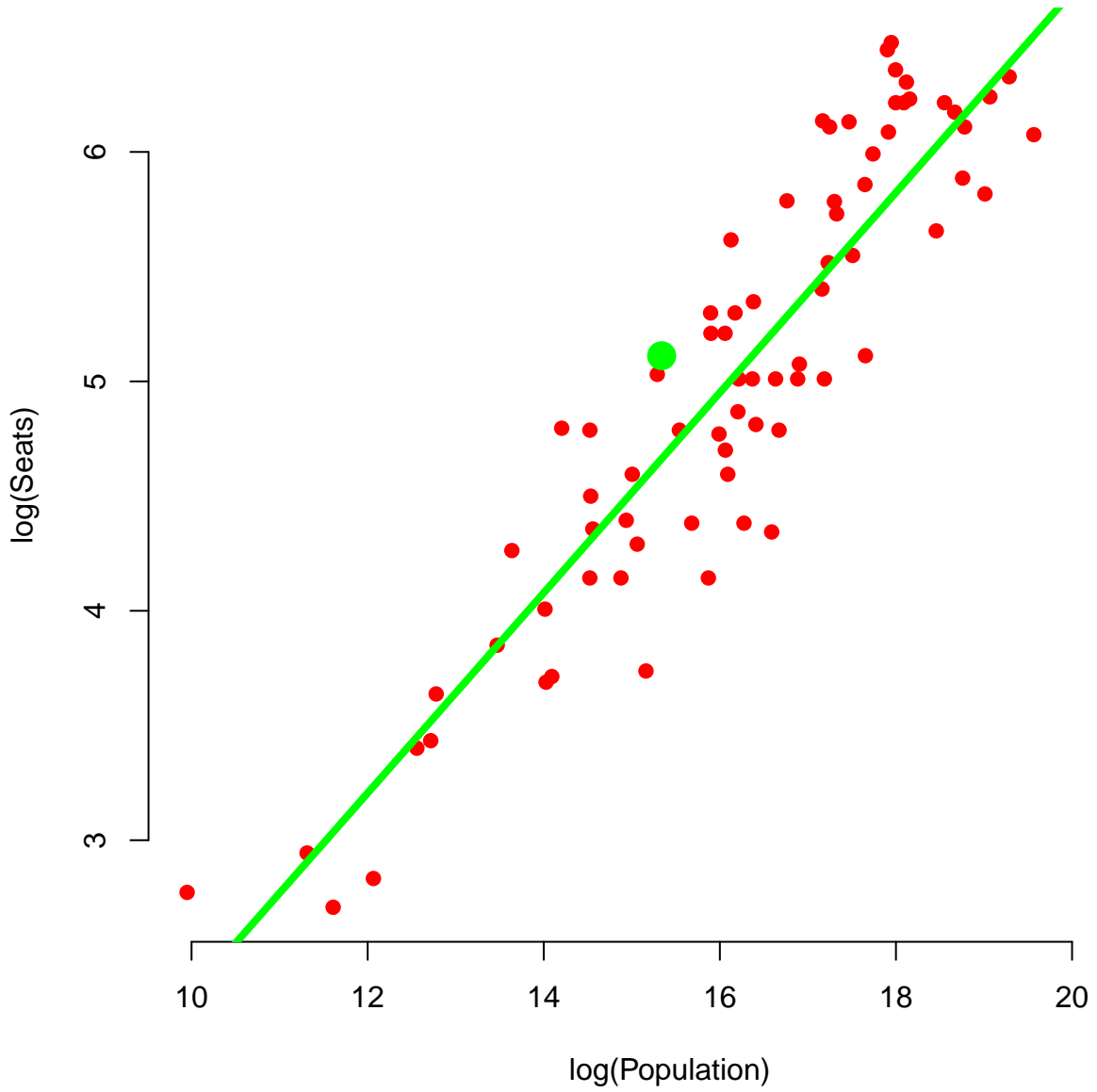


Figure 3: Seats (logged) by population (logged) and linear regression line. Green dot represents Ireland and blue dot represents India.

not percentages and will therefore reflect the above scale. Thus, a 95% score on all your homeworks will generate an A grade on the UCD system, and a 75 mark on the TCD system.