Introduction to Statistics
homework 3

Johan A. Elkink
jos.elkink@ucd.ie

Due 23 November 2018

Note revised submission date—still 5 pm deadline.

Submission instructions

You will submit **two files**: one **PDF** file[a] 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.[b]

Please ensure that the **file names** contain your **UCD** 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`

---

[a]Word files will be sent back—note that newer versions of Word can easily save to PDF format.
[b]If you only have the HTML version, you can open it in a web browser and then print to PDF.

(5*%)[1] 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.

---

[1]Where percentages are marked with an asterisk, positive marking will be applied—in all other cases negative marking.

Data

The data set we will use for this homework is based on the replication data for Ross (2001), which is available through the Harvard Dataverse Network.

Questions

1. Find the replication data for Ross (2001) at `http://thedata.harvard.edu/`:

    (a) Under "Search Studies", search for the title of the article, then click on the link to the replication data.

    (b) Read the description and scope on the first page.

    (c) Click on the "Data & Analysis" tab, then "Download as" and select the Stata binary.

    (d) (5%) Open the data and make sure the command for doing so is in the command file.

    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.

    (e) (5%) Select only the observations in 1995 from this data set (i.e. the variable **year** should equal 1995)—see "Subsetting data" in Lab 3 for instructions. For the remainder of the homework, use only this data from 1995.[2]

2. For the following statements, (2%) formulate the null hypothesis and the alternative hypothesis, (3%) perform the appropriate t-test, and (4%) formulate the conclusion from the test (50–100 words each):

    (a) Countries in the OECD (**oecd**) are more likely to be democratic (higher mean on **regime**) than countries outside the OECD.

    (b) On average, countries are more likely to be autocratic (**regime** less than 6) than democratic.

    (c) The top 25% of oil producers (**oil** greater than 2.22) are less democratic (**regime**) than the other 75%.

    (d) More than a quarter of the population in a country, on average, is illiterate (**illit** greater than 25%).

3. Table 2 provides information on the relationship between Islam and democracy, using transformed versions of the **islam** and **oil** variables.

---

[2]Stata and SPSS users should be able to see variable labels—for R users, this might be of help (in RStudio only): `View(cbind(names(oil), attr(oil, "var.labels")))`.

	Minority Islam	Majority Islam	
Top 50% oil exporter	74	26	100% (39)
Bottom 50% oil exporter	86	14	100% (43)

$$\chi^2 = 1.11; \; p\text{-value} = 0.2916$$

Table 1: Relation between oil export and prevalence of Islam.

(a) (4%) Explain why the percentages are calculated in the direction they are and interpret the relation between oil and Islam, taking into account the χ^2-test (approximately 150 words).

(b) (10*%) Describe in your own words what the p-value represents (approximately 200 words), statistically.

4. In this question we will look somewhat deeper at the relationship between reliance on oil exports and democracy.

(a) (4%) Run a regression whereby you explain the level of democracy (**regime**) by the level of fuel exports (**oil**).

(b) (4%) Repeat, adding the following control variables: **islam**—the percentage of the population that is Islam; **logGDPpc**—the logarithm of the GDP per capita; **illit**—the percentage of the population that is illiterate.

(c) (10%) Present a regression table properly formatted as for a publication, with separate columns for the two regressions, next to each other.

(d) (20*%) Write a one page report (400–500 words) in which you fully interpret the regression output—what does it tell you about the relationship between oil and democracy? Make sure you interpret everything in the regression table, and translate it to substantive context—i.e. about politics, not about statistics only.

References

Ross, Michael L. 2001. "Does oil hinder democracy?" *World politics* 53(3):325–361.

Grade conversion scheme

Score	Grade UCD	Grade TCD	Score	Grade UCD	Grade TCD	Score	Grade UCD	Grade TCD	Score	Grade UCD	Grade 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.