Introduction to Statistics
homework 3

Johan A. Elkink

jos.elkink@ucd.ie

Due 27 November 2017, 1 pm

You will submit two files: one PDF file[1] 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`.

Percentages with a asterisk indicates that positive rather than negative marking will be applied.

(10*%) 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] Throughout the homework, remember to always look at each variable first before using it, for example through the use of frequency tables or histograms.

---

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

	Very unsure	Quite unsure	Quite sure	Very sure
Abstain	0	3	4	2
Leave	29	53	39	45
Remain	71	44	57	53
	100% (7)	100% (239)	100% (538)	100% (216)

$$\chi^2 = 14.7; \ p\text{-value} = 0.0227$$

Table 1: Relation between certainty about remaining within the EU and vote choice in the Brexit referendum.

Questions

1. First some preliminaries:

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

   (b) (5%) Create a new variable **voteChoice** which is similar to the **vote** variable, but where those who abstained are classified as missing.

2. For each of the following statements, (3%) formulate the null hypothesis and the alternative hypothesis, (3%) perform the appropriate t-test, and (3%) formulate the conclusion from the test (50–100 words each). Note that where the question is clearly formulated in one direction, a one-sided t-test should be performed; where the question is formulated without implying a particular direction, a two-sided test should be performed.

   (a) Voters who voted "remain" (**voteChoice**) are less supportive of Nigel Farage (**likeFarage**) than those who voted "leave".

   (b) Voters are more certain about what would happen when staying in the EU (**certaintyRemain**) than they are when leaving the EU (**certaintyLeave**).

   (c) The certainty of what would happen when the UK leaves the EU (**certaintyLeave**) differs between those who votes "leave" and those who voted "remain" (**voteChoice**).

   (d) Respondents were asked "How much impact do you think that Britain leaving the EU would have on you personally?" with answer categories:
   1 No impact at all
   2 A small impact
   3 A moderate impact
   4 A large impact
   5 A very large impact
   Test whether one can say that, on average, voters rate the impact (**impactSelf**) as 3 or less.

3. Table 1 provides information on the relationship between certainty about remaining within the EU and vote choice in the Brexit referendum.

   (a) (3%) Explain why the percentages are calculated in the direction they are (approximately 50 words).

   (b) (3%) State the null hypothesis and the alternative hypothesis of the χ^2-test.

   (c) (5%) Interpret the relation between certainty about remaining within the EU and vote choice in the Brexit referendum, taking into account the χ^2-test (approximately 150 words).

4. (8*%) Describe in your own words what the p-value represents (approximately 200 words), statistically (i.e. not in the context of any specific test, but in general).

5. See Table 2 for a linear regression explaining certainty about leaving the EU by income levels (on a scale from 1–15, see Figure).
   (20*%) Write a 400–500 word report in which you fully interpret the regression output– what does it tell you about the relationship between income and a sense of certainty about UK's future when it votes to leave the EU? Make sure you interpret everything in the regression tables, and translate it to substantive context—i.e. about politics, not about statistics only.

<div align="center">Table 2:</div>

	Certainty Brexit
Income level	−0.018***
	(0.007)
Constant	2.841***
	(0.051)
N	1000
R-squared	0.008
Adj. R-squared	0.007
Residual Std. Error	0.715 (df = 998)
F Statistic	7.886*** (df = 1; 998)

***p < .01; **p < .05; *p < .1

Linear regression explaining certainty about leaving the EU by income level

What is your gross household income?

1. ○ under 5,000 per year
2. ○ 5,000 to 9,999 per year
3. ○ 10,000 to 14,999 per year
4. ○ 15,000 to 19,999 per year
5. ○ 20,000 to 24,999 per year
6. ○ 25,000 to 29,999 per year
7. ○ 30,000 to 34,999 per year
8. ○ 35,000 to 39,999 per year
9. ○ 40,000 to 44,999 per year
10. ○ 45,000 to 49,999 per year
11. ○ 50,000 to 59,999 per year
12. ○ 60,000 to 69,999 per year
13. ○ 70,000 to 99,999 per year
14. ○ 100,000 to 149,999 per year
15. ○ 150,000 and over
16. ○ Don't know
17. ○ Prefer not to answer

Figure 1: Income levels are recorded in the Brexit BES survey.

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 *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.