Introduction to Statistics
homework 4

Johan A. Elkink

jos.elkink@ucd.ie

Due 14 December 2016, 5 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. 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.

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

1. For this homework, just as for the previous one, we will make use of the Latino Barometer data, which is publicly available at `http://www.latinobarometro.org/`. Follow the link to "Data Bank" and download the 2015 data file—see Figure 1. Download the data file prepared for the software package you are using (SPSS, Stata, or R).[2] Note that the data is in ZIP format and you might have to download an unzip program first.

---

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

[2]For R users, this means you will open the data file differently from those to date. Instead of using `read.dta()` to open a Stata file, you use the `load()` function to open an R workspace, the `.rdata` file. The Stata file is written for a newer version of Stata and cannot be opened in R.
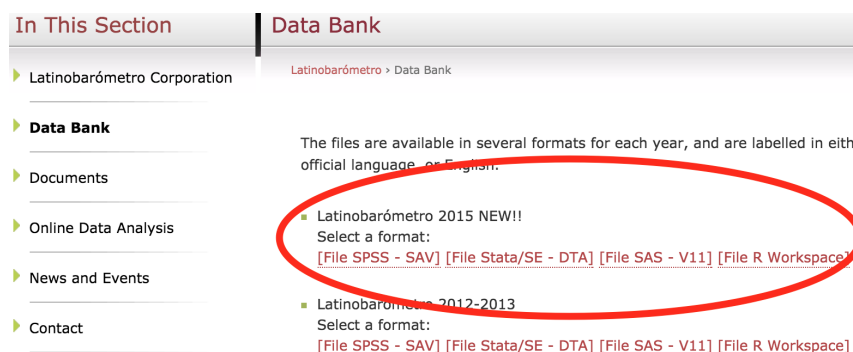

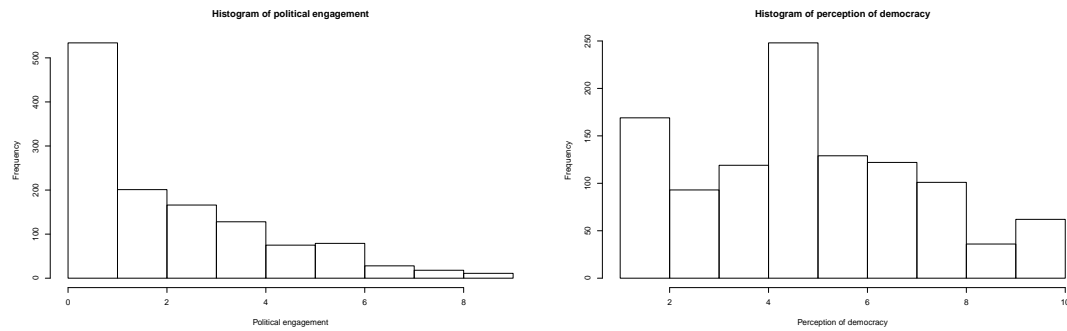Figure 1: Screen shot of data download location.

Figure 2: Distributions of two key variables in the linear regressions.

There are many available for free.[3] Note that in the ZIP file you will also find the questionnaire, which will be crucial for you to interpret the meaning and coding of the different variables.[4]

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

   (b) Select only the Brazilian respondents, which is classified in the data set as the variable **idenpa** being equal to 76. To check, produce a frequency table of variable **S12**, which represents gender, with 1 the males and 2 the females. If you did the download, selection, and table correctly, you should obtain that there are 618 males and 632 females.

      For the remainder of the homework, remember to always look at each variable first before using it, for example through the use of frequency tables or histograms.

2. First we will investigate the relationship between approval of the government and perceived level of democracy in the country.

   (a) (5%) Replicate Model 3 in Table 2.[5] Make sure you check the questionnaire belonging to the data file for exact question wordings. See Table 1 for an overview which variables were used from the data set.

   (b) (5%) Recode the variable for race (**S23**), keeping the "black", "mestizo", "mulatto", and "white" categories, grouping all other categories as "other".

   (c) (5%) Extend Model 3 in Table 2 by adding the race variable as recoded.
      (10%) Produce a regression table for this model.

   (d) (15*%) Write a 400–500 word report, based on Table 2 and the new regression table, with a particular focus on the relation between approval of the government and perceived level of democracy in the country.

3. The second investigation asks where younger people are more likely to engage in politics

---

   [3]See for a list of applications, including indication which are free and which can handle ZIP files: `https://en.wikipedia.org/wiki/Comparison_of_file_archivers`.

   [4]Note that most question numbers start with "Q", while the corresponding variable names starts with "P", but otherwise they appear identically named.

   [5]No need to report in PDF, just include in code file.

Perception of democracy	P17STGBS	
Political engagement	P20TGB.A	constructed as 12 minus the sum of these three variables
	P20TGB.B	
	P20TGB.C	
Protest for democratic rights	P22ST.D	
Government approval	P48STGBS	dummy variable, 1 if approve
Female	S12	dummy variable, 1 if female
White	S23	dummy variable, 1 if white
Left-right	P27ST	
Age	S13	

Table 1: Overview of variables used in regressions

than older people. We construct a new political engagement variable which simply sums the three items in question **Q20STGBS** in the survey and then reversing the index. The result is a scale from 0 to 9, where 0 is the lowest level of engagement ("never" to all questions) and 9 the highest level ("very frequently" to all questions). A histogram is presented in Figure 2.

(15*%) Write a 400–500 word report, based on Table 3, with a particular focus on the relationship between age and active engagement with politics.

4. The third investigation looks into the relation between perceptions of national democracy and willingness to protest for democratic rights.

   (a) (15%) Find three control variables in the data set, using the questionnaire provided with the data, that seem appropriate when evaluating the relationship between perceptions of national democracy and willingness to protest for democratic rights. In approximately 150 words each, explain why these are appropriate control variables.

   (b) (10%) Extend Model 1 by adding the three control variables and produce an appropriate regression table.

   (c) (10*%) Write a 200–300 word report, based on the extended regresion, with a particular focus on the relation between perceptions of national democracy and willingness to protest for democratic rights.

Table 2: Linear regressions explaining the perceived level of democracy in the country.

| | Perception of democracy | | |
| | OLS | | |
	(1)	(2)	(3)
Government approval	1.025***	1.006***	0.878***
	(0.165)	(0.166)	(0.219)
Age		0.005	0.001
		(0.005)	(0.007)
Female		−0.070	−0.073
		(0.151)	(0.199)
Left-right			0.084**
			(0.035)
Constant	4.882***	4.707***	4.680***
	(0.090)	(0.218)	(0.328)
N	1,024	1,024	617
R^2	0.036	0.038	0.038
Adjusted R^2	0.035	0.035	0.032

$^*p < .1$; $^{**}p < .05$; $^{***}p < .01$

Table 3: Linear regressions explaining the level of political engagement.

| | | Political engagement | | |
| | | OLS | | |
	(1)	(2)	(3)	(4)
Age	−0.012***	−0.012***	−0.012***	−0.007
	(0.004)	(0.004)	(0.004)	(0.006)
Female		−0.350***	−0.356***	−0.336*
		(0.125)	(0.133)	(0.171)
White		−0.046	−0.043	−0.045
		(0.127)	(0.136)	(0.174)
Perceived democracy			0.090***	0.058*
			(0.028)	(0.035)
Left-right				0.036
				(0.030)
Constant	2.823***	2.988***	2.615***	2.544***
	(0.166)	(0.185)	(0.245)	(0.337)
N	1,240	1,215	1,055	633
R^2	0.008	0.015	0.025	0.016
Adjusted R^2	0.007	0.012	0.021	0.008

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

Table 4: Linear regression explaining willingness to protest for democratic rights.

	Willing to protest for democratic rights
	OLS
Perception of democracy	0.172***
	(0.043)
Constant	5.344***
	(0.248)
N	1,014
R^2	0.015

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

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.