

# Introduction to Statistics

## homework 3

Johan A. Elkind  
jos.elkind@ucd.ie

Due 23 November 2016, 5 pm

You will submit two files: one PDF file<sup>1</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. Please send both files to jos.elkind@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, 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).<sup>2</sup> Note that the data is in ZIP format and you might have to download an unzip program first. There are many

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

<sup>2</sup>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.

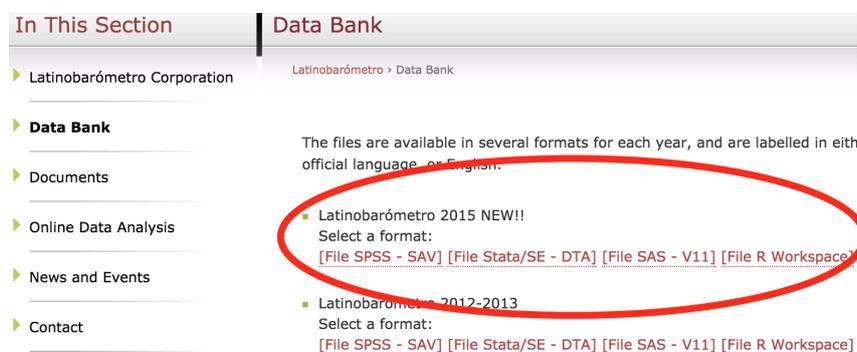


Figure 1: Screen shot of data download location.

	Agree	Disagree	
Male	38	62	100% (632)
Female	47	54	100% (618)

$\chi^2 = 9$ ;  $p$ -value = 0.0022

Table 1: Relation between gender and support for gay marriage.

available for free.<sup>3</sup> 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.<sup>4</sup>

- (a) (3%) Open the data and make sure the command for doing so is in the command file.
- (b) (3%) 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. For each of the following statements, (3%) 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) Brazilian women (**S12**) are more left-wing (**P27ST**) than the men.
  - (b) On average, respondents rate Brazil slightly more non-democratic than democratic (i.e. **P17STGBS** is less than 5.5).
  - (c) Less than a third of the population is satisfied with the government's performance (**P48STGBS**).
  - (d) Whites (use **S23**, but (5%) recode is required) in the population perceive a lower level of democracy (**P17STGBS**) than other racial groups.
3. Table 2 provides information on the relationship between gender and support for gay marriage.
  - (a) (4%) 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  $\chi^2$ -test.
  - (c) (5%) Interpret the relation between gender and support for gay marriage, taking into account the  $\chi^2$ -test (approximately 150 words).

<sup>3</sup>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](https://en.wikipedia.org/wiki/Comparison_of_file_archivers).

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

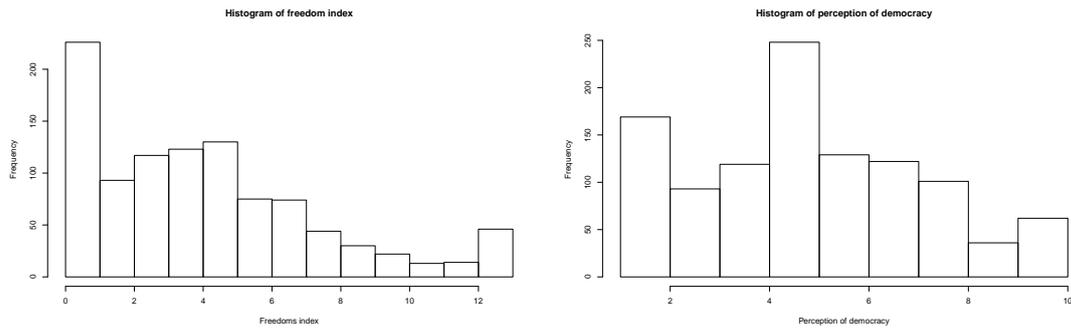


Figure 2: Distributions of two key variables in the linear regressions.

- (d) (8\*%) Describe in your own words what the  $p$ -value represents (approximately 200 words), statistically.
4. In this question we will look a bit closer at what variables might explain the perceived level of democracy in Brazil, looking at perceived freedoms and age as explanatory variables.
- We will make use of a constructed index variable, the Freedoms Index, which is constructed by counting the number of items in Question Q50ST where the answer is “fully” or “fairly generally”. A histogram of this constructed variable is presented in Figure 2.
- (a) (4%) Run a regression whereby you explain the perceived level of democracy (**P17STGBS**) by age (**S13**). The output should match Model 2 in Table 2.
- (b) (15\*%) Write a 400–500 word report in which you fully interpret the regression output—what does it tell you about the relationship between perceived freedoms and perceived democracy? And between age and perceived democracy? Make sure you interpret everything in the regression tables, and translate it to substantive context—i.e. about politics, not about statistics only.

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

	Perception of democracy	
	OLS	
	(1)	(2)
Freedoms index	0.199*** (0.022)	
Age		0.008* (0.005)
Constant	4.382*** (0.126)	4.881*** (0.199)
N	938	1,079
R <sup>2</sup>	0.077	0.003

\*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.