

Introduction to Statistics

homework 4

Johan A. Elkink
jos.elkink@ucd.ie

Due 14 December 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

For this homework we will use the replication data from Tajima (2013), which can be accessed on the course teaching data page³ and is listed as “Indonesian census”.⁴ You will need the *indovill_regency.dta* dataset, which contains data on Indonesian villages, averaged by regency. Make sure you also open the description file, so that you have a list of variables in the data set, with brief descriptions.

Questions

1. 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/indovill_regency.dta.

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.joselkink.net/data.php>

⁴Taken from the PODES DataVerse, <http://hdl.handle.net/1902.1/19477>.

2. Perform a multiple regression whereby you explain the support for Golkar⁵ by the level of identification as Muslim, inequality, poverty, violence and urbanisation.

(a) (15%) Present a regression table properly formatted as for a publication.

(b) (30*%) Write a 400–500 word report in which you fully interpret the regression output—what does it tell you about the relationship between Islam and support for Golkar?

Make sure you include at least:

- Interpretation of regression coefficients and R^2 .
- Interpretation of the size of the main effect (Islam on support for Golkar).
- Interpretation of the t -tests and F -test.
- Discussion of each control variable—are they appropriate controls? Why / why not?
- What control variables might be important to add?

3. Perform a multiple regression whereby you explain the level of interethnic violence by ethnic fractionalisation, levels of inequality, poverty, distance to the nearest police post, relative (“horizontal”) inequality between ethnic groups and urbanisation.

(a) (15%) Present a regression table properly formatted as for a publication.

(b) (30*%) Write a 400–500 word report in which you fully interpret the regression output—what does it tell you about ethnic fractionalisation and the likelihood of ethnic violence in Indonesian villages?

Make sure you include at least:

- Interpretation of regression coefficients and R^2 .
- Interpretation of the size of the main effect (ethnic fractionalisation on ethnic violence).
- Interpretation of the t -tests and F -test.
- Discussion of each control variable—are they appropriate controls? Why / why not?
- What control variables might be important to add?

References

Tajima, Yuhki. 2013. “The institutional basis of intercommunal order: Evidence from Indonesia’s democratic transition.” *American Journal of Political Science* 57(1):104–119.

⁵<http://en.wikipedia.org/wiki/Golkar>

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.