Advanced Quantitative Methods
Homework 1: Math Review & Linear Regression

Johan A. Elkink
jos.elkink@ucd.ie

Due February 20, 2017, 5pm

Please submit by email in PDF format. Add R code in a separate `.R` *file, or SPSS code in a separate* `.sps` *file, or Stata code in a separate* `.do` *file, or the code for any other package you use separately. Note that if you do not use Latex, there is an "Equation" entry in Microsoft Word under "Insert" that will allow you to include matrices and other mathematical equations.*

Percentages with an asterisk indicate that positive rather than negative marking will be applied.

(10*%) of the grade is used for an overall evaluation of the clarity and presentation of your code.

Data

The last part of this homework is based on the replication data for Ross (2004).[1] While Ross (2004) uses panel data – a set of countries observed over a number of years – we will be using a cross-section – only one year – for 1997. The data is available as `ross_1997.dta` ("taxation and representation") on the teaching data page.

Note that this homework assumes you know how to deal with categorical independent variables and you should identify for yourself where this is the case in the exercises below. Make sure you look at the distribution of each variable such that no incorrect data are included and make corrections if and when necessary.

[1] `http://hdl.handle.net/1902.1/14305 UNF:5:8Ymx2XXoRulfSgFb5Xq1uQ==.`

Questions

1. Consider a matrix $\mathbf{X}$ containing the test scores of $n = 5$ students on $k = 2$ tests and $\mathbf{u}$ a vector of n ones. $\mathbf{D}$ contains the inverse of the standard deviations of $\mathbf{X}$.

$$\mathbf{X} = \begin{bmatrix} 8 & 6 \\ 3 & 6 \\ 5 & 7 \\ 7 & 1 \\ 7 & 2 \end{bmatrix} \qquad \mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{D} = \begin{bmatrix} \frac{1}{n}\sigma_1 & 0 \\ 0 & \frac{1}{n}\sigma_2 \end{bmatrix} = \begin{bmatrix} 0.5000 & 0 \\ 0 & 0.3701 \end{bmatrix}$$

   Write down the result of

   (a) (4%) $\mathbf{u}'\mathbf{X}$
   (b) (4%) $\mathbf{M} = \frac{1}{n}\mathbf{u}\mathbf{u}'\mathbf{X}$
   (c) (4%) $\mathbf{A} = \mathbf{X} - \mathbf{M}$
   (d) (4%) $\mathbf{Z} = \mathbf{A}\mathbf{D}$ (you will see that $\mathbf{Z}$ contains the standardized columns of $\mathbf{X}$)

2. Consider the following matrices:

$$\mathbf{X} = \begin{bmatrix} 1 & -0.01 & 13 \\ 1 & 0.03 & 15 \\ 1 & -0.09 & 11 \\ 1 & -0.20 & -2 \\ 1 & -0.04 & 10 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 4.38 \\ 5.26 \\ 3.32 \\ -0.60 \\ 3.12 \end{bmatrix} \qquad \hat{\beta} = \begin{bmatrix} 1.13 \\ 6.49 \\ 0.25 \end{bmatrix}$$

   Write down the result of

   (a) (4%) $\mathbf{X}\hat{\beta}$
   (b) (4%) $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\beta}$
   (c) (4%) $\mathbf{e}'\mathbf{e}$
   (d) (4%) $\frac{1}{5}\mathbf{X}'\mathbf{X}$

3. When we run an OLS regression, we estimate the coefficients using $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. We can subsequently calculate the residuals as

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} = \mathbf{M}\mathbf{y}.$$

   (a) (6%) Show that $\mathbf{M}$ is symmetric ($\mathbf{M}' = \mathbf{M}$).
   (b) (6%) Show that $\mathbf{M}$ is idempotent ($\mathbf{M}'\mathbf{M} = \mathbf{M}$).

4. This exercise is meant as a revision of multiple regression. You do not need to use matrix algebra in this exercise. Taking corruption as the dependent variable and the level of democracy as the key independent variable:

   (a) (4%) Perform a regression analysis with only those two variables.
   (b) (4%) Repeat the regression analysis adding at least four relevant control variables (not including region).

(c) (4%) Repeat, including the same control variables and adding the `region` variable to the model.

(d) (14%) Present a regression table with three columns, one for each model, properly formatted as for a publication. (See also the note for Week 3 on the course website.)

(e) (20*%) Write a one page report in which you fully interpret the regression output (coefficients, standard errors, R^2)—what does it tell you about the relation between democracy and corruption?

Grade conversion scheme

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

Ross, Michael. 2004. "Does taxation lead to representation?" *British Journal of Political Science* 34:229–249.