Advanced Quantitative Methods
Homework 1

Johan A. Elkink
jos.elkink@ucd.ie

Due January 31, 2012, 9am

Please submit by email in PDF format. Add R code in a separate R file.

1. Consider the following three matrices:

$$\mathbf{A} = \begin{bmatrix} 6 & 7 & 6 \\ 3 & 65 & 9 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 1 & 9 \\ 7 & 1 \\ 8 & 3 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

   Write down the result of

   (a) (5%) $\mathbf{A} + \mathbf{B}'$
   (b) (5%) $\frac{1}{2}\mathbf{AB}$
   (c) (5%) $\mathbf{A}'\mathbf{CA}$

2. (10%) For matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix},$$

   write down what $\mathbf{X}'\mathbf{X}$ looks like.

3. Suppose $P(A) = \frac{1}{4}$ and $P(B) = \frac{2}{7}$ are independent events, what are

   (a) (3%) $P(A|B)$
   (b) (3%) $P(A \cap B)$
   (c) (3%) $P(A \cup B)$

4. (10%) Suppose a medical test provides a false positive or false negative in 5% of the tests. The disease is rare and occurs for only six in a thousand in the overall population. The test is taken for all citizens at a particular age and does not depend on observed symptoms or particularly risky behaviour by the patient. What is the probability that a patient has, in fact, the disease, when the test results in a diagnosis that the patient does? You should be using Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}$$

5. This exercise is meant as a revision of regression in R. You will need the *dpes.sav* dataset, which contains data from the 2002 Dutch Parliamentary Election Study.

```
library(foreign)
dpes <- read.spss(file.choose(), to.data.frame = TRUE)
```

The variables we will use are:

SYMP_CDA	Sympathy score for the CDA (Christian democrats)
EUTHANASIA	Attitude towards euthanasia (1 = should be forbidden)
INCOME	Attitude towards income redistribution (1 = differences should be increased)
AGE	Age of the respondent

Make sure you look at the distribution of each variable such that no incorrect data are included. For example, many survey data sets contain numerical codes for missing data, which should not be included as if it is regular data.

Perform the following regression: $CDA_i = \beta_0 + \beta_1 EUTHANASIA_i + \beta_2 INCOME_i + \beta_3 AGE_i + \varepsilon_i$

(a) (16%) Present a regression table properly formatted as for a publication.

(b) (25%) 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 support for the CDA?

(c) (15%) Plot, based on the previous regression, SYMP_CDA as a function of AGE, including the estimated regression line.