POLS0012 Causal Analysis: Practical Session 1

In a famous paper titled "Islam and Authoritarianism", Steven Fish asks whether Muslim societies are less democratic.[1] To find out, he runs a series of cross-sectional regressions of countries' Freedom House scores (an indicator of the level of a country's democracy) on characteristics of the countries, including whether they are predominantly Muslim.

The paper's dataset is in the spreadsheet "fishdata.csv", which you should load using the `read.csv()` command. It contains (amongst others) the following variables:

- *FHREVERS* - Freedom House scores, a measure of democracy where higher values indicate that a country is more democratic and lower values indicate greater authoritarianism
- *MUSLIM* - =1 if a country is predominantly Muslim, 0 otherwise
- *GDP90LGN* - the country's GDP in 1990
- *GRW7598P* - the country's average annual economic growth from 1975-98, in percent
- *BRITCOL* - =1 if the country was a British colony, 0 otherwise
- *OPEC* - =1 if the country is a member of the OPEC group of oil-exporting countries, 0 otherwise

a) First, we'll practice taking subsets and summarising variables:

   i) How many countries are predominantly Muslim?
   ii) What percentage of countries are predominantly Muslim?
   iii) How many countries have GDP in 1990 of above 3.0?
   iv) How many countries are both Muslim and a former British colony?
   v) How many countries have either average economic growth from 1975-98 of above 0.6% or GDP in 1990 of above 2.5?
   vi) Create a new dataset consisting only of countries that are both Muslim and a member of OPEC

   Code Hints:
   - Use square brackets to denote subsets of a variable or dataset. You'll also need the `length()` function

b) What is the difference in mean Freedom House score between Muslim and Non-Muslim countries? Calculate it both by hand and using a regression, verifying that your answers are identical.

[1] M. Steven Fish (2002). "Islam and Authoritarianism." *World Politics* 55 (1): 4-37

c) Is the difference in means in (b) likely to be biased? If so, in which direction and why?

d) Conduct a t-test for the difference in means in (b) using R's `t.test()` function. Is the difference statistically significant?

e) Conduct the t-test again, this time coding it by hand. Confirm that your answer is identical to (d)

f) Calculate:

   i) The percentage of Muslim countries that are former British colonies

   ii) The percentage of non-Muslim countries that are former British colonies

   iii) The correlation between being a former British colony and Freedom House score

   Use these results to predict what impact controlling for *BRITCOL* will have on the estimated difference in means

g) Now estimate a regression of *FHREVERS* on *MUSLIM* and *BRITCOL*. Do the results make sense?

h) Repeat (f) for *OPEC*, *GRW7598P* and *GDP90LGN*. For the latter two variables, simply calculate the correlation between each one and *MUSLIM* instead of the percentage for steps (i) and (ii)

i) Now estimate a regression of *FHREVRERS* on *MUSLIM*, *BRITCOL*, *OPEC* and *GRW7598P*. Again, do the results make sense?