# Problem set 9: week 13

This problem set is to be completed prior to your Stata class. The data set used, wage2.dta, is available on the EC203 website, please copy it on to a memory stick or your H: drive. I will highlight the main commands, use either: **help**, **findit** or the **search** function for guidance on each.

**1.** Open a do file in Stata. All the commands we use in this problem set will be copied into here. This is so you can recall what we have done, and the analysis can be repeated. It will also be useful for you to annotate the do file as you go along. Load the wage2.dta dataset into Stata. Open a log file to record the output from Stata. This should be written into either your memory stick or H: drive.

**2.** Use a t-test to test for a difference in mean wages between males and females.

**3.** Use a regression model to carry out the same test.

**4.** Run the following regression: $ln(wage_i) = \alpha + \beta_1 school_i + \beta_2 male_i + \epsilon_i$. Interpret the regression coefficients. Plot the predicted wages from your regression.

**5.** Generate a new dummy variable indicating whether an individual lives in the north or the south. Run the following regression: $ln(wage_i) = \alpha + \beta_1 school_i + \beta_2 exper_i + \beta_3 exper_i^2 + \beta_4 male_i + \beta_5 North_i + \beta_6 (Male_i * North_i) + \epsilon_i$. Interpret the regression coefficients. Test if there is a difference in gender wage gaps between the north and south.

**6.** For the following regression you decide to only include male observations. Explain why this may be a good idea.

**7.** Generate a set of regional dummies, one for each category of the **region** variable. Run the regression: $ln(wage_i) = \alpha + \beta_1 school_i + \beta_2 exper_i + \beta_3 exper_i^2 + \sum_{j=2}^{8} \delta_j R_{ji} + \epsilon_i$, where the omitted dummy is $R_1$ the London dummy. Interpret two of the coefficients on the regional dummies. Which region has the lowest earnings? Test the hypothesis that the regional dummies are jointly significant: $H_0 : \delta_2 = \delta_3 = ... = \delta_8 = 0$.