Applied Economics
OLS in `gretl`

It is necessary to generate a script (.inp file) that allows you to answer each of the questions.

1. (Based on Wooldridge, pg. 237) Use WAGE1.GDT to estimate the following model:

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \epsilon$$

   i) Estimate the model (using robust standard errors) and comment the results.

   ii) Write the null hypothesis that experience has constant marginal effects on wages. Test, if possible, the previous hypothesis (using a two tail test at a 95% confidence level).

   iii) Is experience relevant to explain wages? Write the hypothesis that allows you to answer this question and perform the adequate test.

   iv) Compute the expected return for the fifth year of experience.

   v) For which value of *exper* more years of experience imply a lower predicted wage? For how many individuals in the sample this is the case?

   vi) Do you think that education and experience help to explain an important percentage of wages' variation?

   vii) Write a model in which the effect of experience is different for men and women. Estimate that model and discuss if you find in fact a different effect by gender.

2. In this exercise you will investigate the relationship between the number of completed years of education for young adults and the distance from their high school to the nearest four-year college. Proximity to college lowers the cost of education, so that students who live closer to a four-year college should, on average, complete more years of schooling. You will use data on a random sample of high school seniors, contained in the data file *dist.dta*. These same students were interviewed as high school seniors in 1980, then re-interviewed in 1986 to determine how many years of education they had completed. Use these data to answer the following questions.

   i) Estimate a simple regression model of years of completed education (*yrsed*) on distance to the nearest college (*dist*), where *dist* is measured in 10's of miles (for example, *dist* = 2 means that the distance is 20 miles).

   ii) Report the estimated coefficients and the standard errors. What's the interpretation of the slope?

   iii) Is the slope coefficient significant? At what significance level?

   iv) Construct a 95% confidence interval for the slope.

   v) Ana's high school was 20 miles from the nearest college. Predict Ana's years of completed education using your estimation.

vi) Does distance to college explain a large fraction of the variance in educational attainment across individuals? Explain.

vii) Repeat question (i) using only the data on females.

viii) Repeat question (i) using only the data on males.

ix) Is there any difference between the effect of distance on men and women? Write a model that allows you to test if there are significant differences.

x) Generate a new variable, $dist2$, representing distance in miles rather than in tens of miles (i.e., multiply $dist$ by 10). How do the regression results change?