Instructions :

1. The dataset you will use is in the Excel spreadsheet EAWE.xlsx
2. This assignment entails the use of econometric models and statistical tools in eco nomic applications. You will use statistical software, Stata, to analyse the educational attainment and wage equations data.
3. Answer all the questions. Show all numerical answers to 2 decimal places if necessary. When you are asked to 'perform a test', you should write explicitly the null hypothesis of the test, and state clearly how you make testing decisions and make conclusions. Please carry out all tests using a 5% level of significance.
4. Please submit your answers in .smcl file along with a pdf file containing the explanations of each question answer.

Data Description :

You will use a subset consisting of 500 observations of the Educational Attainment and Wage Equations (EAWE) dataset to answer the questions. The description of the data set and contained variables can be found in ECMT1020_ASN_Data_set_explanation.pdf. In particular, note that EXP and TENURE in your dataset are, respectively, the number of years at work and the number of years spent working with the current employer. We define a new variable PREVEXP = EXP−TENURE. PREVEXP, thus defined, is the total work experience with previous employers, and will be used in some of the questions. We use LGEARN to denote the logarithm of EARNINGS.

Questions :

1. Fit an educational attainment function using your data set. Regress S on ASVABC, SM and SF, and interpret the regression results. Perform t-tests on the coefficients of the variables in the education attainment function.

2. Use the Box and Cox procedure to evaluate whether the dependent variable of a wage regression of EARNINGS on S and EXP should be linear or logarithmic based on your dataset.

3. Following up on the previous question, now define demeaned variables $S^* = S - \bar{S}$ and $EXP^* = EXP - \bar{EXP}$,

   where $\bar{S}$ and $\bar{EXP}$ denote the sample mean of S and EXP. Regress EARNINGS or LGEARN, depending on your result in the previous question, on S^* and EXP^* and interpret the regression output. In particular, how is the interpretation of the two slope coefficients in this regression different from the regression where regressors are S and EXP instead?

4. Consider the following two regressions:

   (a) Regress LGEARN on S and PREVEXP;

   (b) Regress LGEARN on S, TENURE and PREVEXP.

   Before running any regressions in your software, what is your expectation of the relative magnitude of the coefficients of PREVEXP in the two fitted regressions? Explain why you expect so. Then, run these two regressions in Stata and see if the result confirms your expectation.

5. Explain how you could get the same OLS estimate of the coefficient of PREVEXP in the multiple regression in Question 4(a) using "purged regressions". Implement your procedure in Stata and show the results are matched.

6. Consider a regression of LGEARN on S, EXP, TENURE, PREVEXP, ASVABC, ETHBLACK and ETHHISP. Can you run this regression in Stata? If yes, please report your output. If not, please explain why this is the case.

7. Regress LGEARN on S, PREVEXP, TENURE, ASVABC, ETHBLACK and ETHHISP. Explain how you would conduct an F test for testing the coefficients in front of PREVEXP and TENURE being equal. Please perform the test and interpret your result.

8. Regress LGEARN on S, EXP, TENURE, ASVABC, ETHBLACK and ETHHISP. Perform a t test on the coefficient of TENURE. Please explain why such a t test is a test of the same restriction described in Question 7. Verify that the same result is obtained from the previous F test and the t test here.

9. How do you interpret the coefficients of ETHBLACK and ETHHISP in the fitted regression in Question 8?

10. Add an intercept dummy MALE and a slope dummy defined as the product of MALE and S in the regression in Question 8, and then run the regression again. Interpret the coefficients before these two dummy variables. Is the effect of education on earnings different for males and females?