Computer Exercise 1

This exercise will be graded formally and contribute 10% toward your final mark for this course. Hand in your answers to the departmental office by midday Thursday January 26th 2005.

Learning Outcomes: This exercise is intended to help you become familiar with the fundamental principles underlying OLS and maximum likelihood estimation and the issue of consistency.

Read in the data set *computer_i.dta* from the Ec3327 web site, where the i suffix refers to the number you were allocated.

This contains data on a sample of 800 individuals in a region of the United Kingdom. 800 individuals in each sample is in work and 800 are not in work. Those in work are asked about their wages and various socio-economic characteristics. The variables in each data set are as follows:

sex	= 1 if male, = 2 if female
age	age in years
hourpay	gross hourly pay (in £ an hour)
emp	=1 if employed, =0 otherwise
region	number of region
yearsed	years of education

Keep a log file of all the (correct) commands and output that you have generated. Hand this log file (annotated with some explanatory comments related to what you have done or observed) into the departmental office

1. Write a do loop to calculate the average (mean) hourly pay for men and women in the sample

(check your answers using the "summary" command in stata)

(10 marks)

2. Using the stata matrix algebra commands that you have been taught during the course, program the OLS estimates of the coefficient on the constant and on years of education **and** the OLS estimates of the variance of these estimates based on a regression of the log of the hourly wage on years of education

$$\text{Ln(Hourpay)}_i = b_0 + b_1 * \text{yearsed}_i + u_i \qquad i=1, 2, \ldots N$$

a) for the first 10 observations in your data set
b) all observations with wage information and valid years of education

(check your answer is correct using the "regress" command in stata).

What is the estimated effect of 1 extra year of education on hourly pay in both regressions?

N.B.1 You will have to generate a log hourly wage variable in stata

N.B.2 Not all observations on the years of education variable are valid. You will have to decide which observations are invalid and delete these invalid observations before you do the matrix algebra

N.B.3 To create the X matrix for the OLS estimates you will have to create a column vector of ones and join this to the vector of observations on years of education, (see earlier computer exercises for an example of how to do this)

N.B.4 You may need to adjust stata to allow you to create a matrix with more than 200 rows (the default number of rows in stata). To do this type

set matsize 800

before you create any matrices

(40 marks)

3. Now use the bootstrap command to draw 1000 samples of size a) 10
b) whatever the sample size is in Question 2 part b.

Estimate and obtain a distribution of OLS coefficients from these repeated samples

Plot the distribution of OLS estimates in both cases.

Find the mean and variance of each distribution

Comment on your findings

N.B. use the “set seed 1” command before you run the bootstrap. This will help ensure you can replicate your results

(20 marks)

4. Run a logit model of the probability that an individual is observed in work conditional on years of education and a gender dummy variable
(you will have to generate a dummy variable so that 1=male and 0 = female)

Now
a) write down the likelihood and log likelihood functions for this model
b) program the maximum likelihood estimator of this model using the “ml” maximum likelihood command in stata and the basic syntax for maximum (log) likelihood used in the previous problem sets

Comment on your estimates

N.B. if you have deleted the observations with emp=0, it may be easier to clear the data (type “clear”) and re-read in the original data. If you do this remember there are some invalid observations on the variable yearsed.

(30 marks)