UNIVERSITY OF ESSEX
2017–18

DEPARTMENT OF ECONOMICS
Dr. Abhimanyu Gupta

EC114 INTRODUCTION TO QUANTITATIVE ECONOMICS
Spring Term Assignment 2018

NOTE: The deadline for handing in this assignment is Monday, 19 March 2018, 12:00 mid-day (online via FASER). It is acceptable to copy/paste output from Stata, but to get full marks you are required to comment appropriately on the output.

Important: You must submit your answers online on FASER.

Question 1 is theoretical and carries 40 marks.

Questions 2, 3, 4, 5, 6, 7 and 8 are based on your individualized data set, and carry 60 marks. Your data set is available on FASER under feedback for the EC114 Spring Term Assignment, as explained in an earlier email. If you have not obtained the data set, please contact Julie Oliver (joliver@essex.ac.uk) **as a matter of urgency.**

Your data set is unique to you. You **must not** use the data set of any other student for your exercise. Answer the questions using your data set and Stata.

1 Theory

QUESTION 1

Consider the simple regression model $Y = \alpha + \beta X + \epsilon$. The ordinary least squares (OLS) estimates of α and β are denoted a and b respectively.

(a) [5 marks] Classify each of $Y, X, \alpha, \beta, \epsilon$ as observable or unobservable.

(b) [2 marks] Give the formulae for a and b.

(c) [9 marks] A friend tells you that having the disturbance term ϵ in the model is silly because it makes the relationship between Y and X uncertain. You have to explain to him that he is wrong. Explain clearly the role of ϵ, the types of things it can capture and why it is an essential feature of an economic relationship. Then explain how we can minimize the error that ϵ captures in the relationship between Y and X. Finally, relate the last explanation to the OLS estimates a and b.

(d) [4 marks] State the assumptions under which a and b are unbiased estimates of α and β respectively.

(e) [10 marks] Under the assumptions stated in (d) above, prove that a and b are unbiased estimates of α and β respectively.

(f) [10 marks] Suppose now that there is no intercept in the model, so that $Y = \beta X + \epsilon$. Derive the OLS estimate of β. Simply stating the formula will fetch zero marks.

2 Data analysis

The variables in your data set are: *health* = (self-reported) health status, *private* = indicator which takes the value 1 if the individual has private health insurance, and 0 otherwise, *age* = individual's age in years, *educ* = years of education, *inc* = annual household income (in EUR), *female* = indicator which takes the value 1 if the individual is female, and 0 otherwise, *bluec* = variable which takes value 1 for blue-collar work, and 0 for white-collar work, and *docvis* is the number of visits to the doctor in the past three months.

QUESTION 2

[10 marks] For each of the variables *health*, *private*, *inc*, and *docvis*, what is the average, the standard deviation, the minimum value, and the maximum value in your data set? What is the number of observations in your data set?

QUESTION 3

(a) [3 marks] Generate a new variable *loginc* as the (natural) logarithm of *inc*. Label this variable as "logarithm of income".

(b) [2 marks] Label *private* as "indicator for private health insurance".

(c) [4 marks] Label each value of the variable *private* appropriately.

(d) [6 marks] What percentage of individuals has private health insurance? What is the average level of health status for individuals with private health insurance who are at least 40 and at most 49 years old?

(e) [6 marks] Create a new variable, *incgroup*, which takes value 1 if household income is less than 10,000 EUR, value 2 for income group 10,000-19,999 EUR, 3 for income group 20,000-29,999 EUR, 4 for income group 30,000-39,999 EUR, 5 for income group 40,000-49,999 EUR, and 6 for income group 50,000 EUR and above. Tabulate the new variable. What fraction of households is in the income group 40,000-49,999 EUR?

QUESTION 4

[4 marks] Plot a histogram of health status, separately for males and females.

QUESTION 5

[4 marks] Calculate the correlation coefficient between health status and years of education. Is there a strong relationship between these two variables? Answer the question using a statistical test.

QUESTION 6

(a) [4 marks] Run a regression of health status on *age*, *female*, and years of schooling and report the regression results.

(b) [2 marks] If education changes from 8 to 11 years, what is the effect on predicted health status?

QUESTION 7

(a) [4 marks] Run a regression of health status on *age*, age^2, age^3, *female*, and years of schooling, and report the regression results.

(b) [3 marks] Test if the three coefficients on the age variables are jointly significant (at a significance level of 5%).

QUESTION 8

1. [4 marks] Do you prefer the regression in question 7 (a) to the regression in question 6 (a)? Explain.

2. [4 marks] What might be the use of *bluec* in explaining health status? Feel free to run any regressions you think may help in answering the question, but this is not necessary.

*****