# ECS4002: DATA ANALYSIS

# CASE STUDY

**DUE DATE: Thursday 26th November 2020, 11:59pm**

**Please submit on the Turnitin link provided on the course Unihub page. All answers should be written on a single Word document. Please save your file as .doc as this format works better when uploading into turnitin. Copy a paste all the Stata commands in an appendix, placed at the end of your document.**

**Answer all questions.**

**Exercise 1**

For this exercise open the Excel file titled ‘**HH Energy Data’** in UniHub.

This data comes from a study of annual household energy consumption in the UK (measured in Kilowatt hours per year). This data was collected from a random sample of houses from across the UK in 2018. The aim was to establish a baseline of energy consumption prior to the introduction of new energy efficiency regulations.

\*Kindly generate ‘largehh’ variable in stata as it is missing by using command –

gen large hh=hhsize>2

The dataset contains the following variables:

|  |  |
| --- | --- |
| **Variable name** | **Description of variable** |
| **id** | Identifies each household in the study |
| **hhsize** | Number of people living in the household |
| **largehh** | 1=household with 3 or more people living in it, 0=all other |
| **rural** | Whether a household is defined as rural (1=rural, 0=urban or suburban) |
| **kwh** | Household energy consumption per year (in Kilowatt hours, kWh) |

**Questions:**

1. Please provide a brief description of the data including: (**10** marks)
   1. A table with summary statistics
   2. A histogram of energy consumption
   3. A short textual description of the sample statistics
2. The national energy regulator claims that the average household in the UK consumes on average about 4000 kWh of energy per year. (**15 marks**)
   1. Test this claim using a two-tailed t-test.
   2. Report the 95% confidence interval for mean energy consumption.
   3. Discuss your results. Clarify whether your hypothesis test confirms your confidence interval estimation.
3. Perform appropriate tests to identify: (**10 marks**)
   1. Whether rural households consume more energy than urban/suburban ones.
   2. Whether large households consume more energy than smaller ones
   3. Whether rural households are larger than urban ones.
4. A pilot study of the impact of proposed new energy efficiency technologies is being designed to be implemented next year. Based on results of the above study, do you have any suggestions for this pilot study might be most effective? Write a short paragraph explaining your reasoning. (1**0 marks)**

Exercise 2

The dataset ‘London\_data.dat’ contains information on individuals’ budget share on clothing and food, their total expenditure, age and number of children, for a total of 850 people.

1. Provide a table of descriptive statistics for your data and discuss what this initial analysis tells you about the characteristics of your sample (**9 marks**)
2. Produce two scatter plots, the first one using food expenditure (Y) and total expenditure (X) and the second one using clothing expenditure (Y) and total expenditure (x). Include the fitted line on your scatter plot. What can you say about the relationship of the two budget shares with respect of total expenditure? **(10 marks)**
3. Estimate the following budget share equation for the food budget share and the clothing budget share:

Report and discuss your results. In your discussion comment on how total expenditure, age and number of children influence the various budget proportions. Interpret you R-squared and comment on the statistical significance of your coefficient estimates, using the 5% and the 1% significance level. **(12 marks)**

1. Re-estimate the food budget share but this time express TOTEXP in logarithms. Interpret your results following the instruction in question b) **(12 marks)**
2. Re-estimate the clothing budget share expressing the both the share and TOTEXP in logarithms. Interpret your results following the instruction in question b) **(12 marks)**

Produce professional looking tables (using the outreg2 command) and figures.