ETW2001 Foundations of Data Analysis and Modelling

Assignment 2

Semester 1 2021

A. How it works

- This assignment is an individual work, involving analysis of a set of data using the techniques you learned in this unit.
- The final product will be a **report** (in word or pdf) using the given data. Also, you will submit R-script that you worked for the analysis.
- The assignment worth 25% of your final mark.
- The assignment is assessed based on the quality and content of your report.

B. The Topic

You are to undertake a small investigation of some data from households in Africa, to learn something about the social and economic conditions these households face. The data set comprises the results of a household survey undertaken in 2014 of more than 1,700 poor, rural households in Kenya, Uganda and Tanzania. These households are all in rural villages and small farms, relying on their small piece of land for most of their food, income, and daily needs.

There are 3 topics, and you will write a report for **ONE** topic.

Students whose student ID ends with 1,2,3 will do Topic 1.

Students whose student ID ends with 4,5,6 will do Topic 2.

Students whose student ID ends with 7,8,9,0 will do Topic 3.

e.g.) if your student ID is <u>3XXXXXX4</u>, your last digit is 4, hence do Topic 2.

Topics

Topic 1: Household Crop production

Poor, rural households rely on production of food crops for most of their food needs, and also to earn a little income to buy more food and other essentials. You are to investigate the total production of crops among these households, looking at the following questions:

- What are the characteristics of the households in this study? What indicators are there of poverty among households?
- What are the characteristics of the farms these households live on? What do they produce? How much land do they have?
- Are households producing enough crops to meet their daily food needs?
- Does having more land mean households produce more food?
- What is the effect of household composition on crop production? Does having more children mean less is produced? Or more elderly? Or working age population?
- What are the implications of your findings for priorities in programs to improve the livelihoods of poor, rural households in Africa?

Topic 2: The experience of food shortages

- What are the characteristics of the households in this study? What indicators are there of the level of poverty among households?
- What are the characteristics of the farms these households live on? What do they produce? How much land do they have?
- Are households experiencing food shortages? How often?
- Do households that are producing more crops experience less food shortages?
- Are wealthier households (with more land, better houses) less vulnerable to food shortages?
- What is the effect of household composition on the experience of food shortages? Does having more children mean there are more shortages? Or more elderly? Or working age population?
- What are the implications of your findings for priorities in programs to improve the food security of poor, rural households in Africa?

Topic 3: Children and Schooling

- What are the characteristics of the households in this study? What indicators are there of the level of poverty among households?
- What are the characteristics of the farms these households live on? What do they produce? How much land do they have?
- Are school-aged children attending school?
- Are households experiencing food shortages? How often?
- How does the education level of the household head affect attendance at school?
- Are children from wealthier households (with more land, better houses) attending school more?
- What is the effect of household composition on attendance at school? Does having more children mean attendance is better? Or more elderly? Or working age population?
- What are the implications of your findings for priorities in programs to improve education for poor, rural households in Africa?

C. What to do?

- You should only use R-studio for any analysis conducted for the report.
- Download the spreadsheet file of the data. Import the data into R-studio and randomly select **1,000 households** as a sample to use in your analysis.
  - For consistent output, use the last **4-digits** of your student ID to extract random sample.
    <u>Example</u>
  - set.seed(3854) if your student ID is let say 32733854
  - data <- as.data.frame(sample_n(excel data name,1000))
  - This should generate the identical random sample even though you re-run the code.

- Looking at the list of variables in the data set, choose some variables that describe the characteristics of the households and their farms. Prepare a brief summary using percentages, summary measures, histograms, etc.

- Given your topic, identify the relevant "outcome" variable(s). Describe the characteristics of the sample for this variable.
- Perform an adequate technique for the outcome variable in relation with some of the other important variables that might be relevant to the questions given above.
- Estimate appropriate multiple linear regression models to address some of the questions.
- After interpreting results, draw brief implication for development programs.

D. The Final Product

1. A report with 10 pages limit (excluding Appendix). Figures and estimated output (in equation form) should be included within the report, and the corresponding R-Studio output should be included in the appendix (the appendix will not be marked. It will be used only to check if the results are reported correctly).
2. Font: either Arial or Times New Roman, font size 12, spacing 1.5.
3. Your report should include the following:
   1. Introduction and review of literature on situation in African countries.
   2. Analysis
      - Descriptive analytics.
      - Predictive analytics.
   3. Policy implication / recommendation from your analysis.
   4. Conclusion and limitation of the study.

E. Assessment Criteria

There are three components to how we will assess your report.

1. Content:
- Your analysis should include a range of techniques that have been covered in this unit.
- We will look for a good understanding and interpretation of the techniques that you use. The choice of techniques needs to be appropriate to answer the questions of interest.
- We will look for quality, not quantity. Do not report many, many results. Concentrate on reporting a few things clearly and well, but showing a variety of techniques.

2. Structure of report:
- The report should follow a logical flow from introducing the issues, some research on the background, description of important variables using good visualization, explanation of the analysis, discussing the results and drawing implications.

3. Marking criteria
- Your report will be evaluated based on the marking guide.
- You must submit a R-script and we will run your R-script to see whether your outputs are obtained smoothly without errors.
- Your coding in R-script should work elsewhere. If we are unable to run the model using the given code, we will go through investigation for plagiarism. All students should get different result as you are using random observations from the given data.