Introduction

For the third piece of CA, you will produce your **Data analysis** documentation. In this CA you will pre-process your data, examine it using descriptive statistics techniques, describe each of the variables you will use, and evaluate the most appropriate statistical method to test your hypothesis assumptions.

Use the following headings to create your CA document.

Abstract

Complete your abstract when you have all other sections finished. Outline in a few sentences what your research question will answer and what problem(s) are you examining. Then briefly describe the statistical method you applied to your datasets. Finish the abstract by briefly describing your results and outline what these results mean and how current approaches could be changed because of your findings. Approximately 250 words.

Data Description

Prepare your data by cleaning and suitably restructuring. Then analyse your data using descriptive statistics techniques. Create new variables if required. Choose the optimum data variables that will allow you to perform your analysis and answer your research question. Describe each step that you complete throughout this process, and display your findings using relevant data and charts.

Once your dataset is optimised, define and evaluate the statistical method(s) you will use on your data variables. You should justify your selected statistical method by examining each of the following:

- the structure of your data variables
- type of response variables under scrutiny
- number of groups being studied
- how these enable you to answer your research question
- assumptions you are making about your data variables

Hypothesis testing and sample size determination

Once you have identified your proposed statistical method, you should investigate the hypothesis you will assume about your population parameters.

Using standard notation for the null and alternative hypothesis, describe the relationships between your variables of interest, and describe how you plan to implement your hypothesis tests. Refer back to your original research question when defining your null and alternative hypothesis. Rephrase the original research question if required.

Using R, you should examine the effects of sample size, effect size, desired significance level, and desired power. Critique their impact on each other to select the most appropriate and relevant values of each criteria for your data under consideration so that your analysis will yield meaningful results. Output the results from R into your CA document.

Additional unforeseen relationships

During the examination process of your data, you may include new datasets and variables previously unconnected to your data that will help answer your hypothesised questions. If such relationships are identified, describe this new data, and iterate through each process described above.

Results

Using R, apply the statistical method to your datasets and answer the null and alternative hypothesis tests. Justify in your results why you decided on a particular outcome for each hypothesis.

Conclusion

Conclude your data analysis by describing the processes required to pre-process, evaluate and describe your dataset. And critique your findings within your data.

Important Information

Plagiarism will not be accepted and will result in an automatic mark of zero. If you use references, the Harvard referencing must be adopted. Please use the following link which might help you create the references required: http://www.neilstoolbox.com/bibliography-creator/.

Any deviation from the above project specification must be approved by myself before submission.

Note: Submit your work as a pdf file. Save all of your R scripts on a public repo on GitHub. **Provide a link to this repo in your submitted document.**

Due Date: Sunday 10th May 2020 before 23:59. You must submit your work through Blackboard. A cover sheet must be the first page of your submitted document.