Data Analysis
Week 10: Class Test 2

Introduction

This week is the second of two class tests for Data Analysis and is worth 35% of your final grade. The class test consists of 3 tasks worth a total of **40 MARKS** broken down as follows:

- A report on a statistical analysis of a given data set: **25 MARKS**;
- Further question 1: **7 MARKS**;
- Further question 2: **6 MARKS**;
- Successful upload of `.pdf/html` document: **2 MARKS**

All tasks will be completed within the same R Markdown document. The written report should include:

- An appropriate **Title** and **Introduction** detailing the data and question of interest; **2 MARKS**
- An **Exploratory Analysis** of the data; **7 MARKS**
- A **Formal Analysis** of the data; **12 MARKS**
- Finish with your **Conclusions**; and **2 MARKS**
- Have an appropriate report layout. **2 MARKS**

Instructions

1. **Do NOT** open RStudio until you have downloaded the required files described in Instructions 2. and 3.
2. Go to the **Class Test 2 Files** folder in the **Week 10: Class Test 2** section of the **Data Analysis Moodle page**.
3. Download the files in the **Class Test 2 Files** folder into the **same folder** on your computer:
    - `.csv` files contain the required data sets; and
    - `ClassTest2Template.Rmd` - an R Markdown template for this class test. It loads the R packages necessary to complete the set tasks.
4. Open RStudio and open `ClassTest2Template.Rmd` then save it as `ClassTest2YourStudentNumber.Rmd` in the **same folder** as the `.csv` files are saved on your computer.
5. **Before you start to work**, compile `ClassTest2YourStudentNumber.Rmd` (using `Knit`) and check that the `ClassTest2YourStudentNumber.pdf/html` file is compiled as expected. It is wise to periodically compile and check the `.pdf/html` file as you work through the tasks so you can more easily debug your code as you go. You will **NOT** receive any assistance with compiling your document.
6. For the report part of the class test you **are NOT required** to **include** your R code in the `.pdf/html` file, hence `echo=FALSE` is set as the default in the `.Rmd` template. However, for the further questions you will need to provide your R code in the `.pdf/html` file, and hence should include `echo=TRUE` in any corresponding R code chunks relating to the further questions.
7. When you are ready to submit your class test document, click on the **Class Test 2 .pdf/html Upload** link under **Data Analysis > Week 10: Class Test 2** and upload and submit the file `ClassTest2YourStudentNumber.pdf/html`. **1 MARK** will be deducted if the document is not named as instructed.

8. Also, upload and submit the R Markdown file `ClassTest2YourStudentNumber.Rmd` using the **Class Test 2 .Rmd Upload** link. Again, **1 MARK** will be deducted if the document is not named as instructed. Please note that only the `.pdf\html` file will be marked. The `.Rmd` file will only be considered if there was a problem compiling the `.pdf/html` file. **Note**, the `.pdf/html` file uploaded to Moodle will be considered as your **complete** class test, and as such any partial working files **should not** be uploaded in an attempt to obtain **2 MARKS**.

Examination Conditions

- You have 24 hours to complete the class test and can submit your completed tasks anytime within that time.
- You must work on your own - **NO communication** by any means with anyone is permissible.
- You may consult ANY resources (hardcopy or online), e.g. `tidyverse` "cheat sheets" and/or the online tutorials from the course.

Class Test Tasks

Report: Housing Prices

Housing prices are good indicators of economic status. The house price (in pounds) of 300 houses collected at the time of sale by a realtor within the last six months along with the number of bathrooms and the type of parking for each house was collected. The data is contained within the `HousingPrices.csv` file. Use what you have learned to produce a report on the following questions of interest:

What is the relationship between house price and the number of bathrooms? If a relationship exists, does it differ by type of parking available?

25 MARKS

Further Question 1

Information on 7060 products sold in grocery stores is available in the `GroceryPromotion.csv` file. Here we are interesting in examining whether the number of sales of a product can be related to whether there was a promotional event associated with the product.

(a) Using the grocery sales promotion data set, fit a logistic regression model with `Promotion` as the response variable and `Sales` as the explanatory variable. Output the summary table of the results.

2 MARKS

(b) Produce a 95% confidence interval for sales on the odds scale. Interpret this result.

3 MARKS

(c) Produce an estimate of the odds of an item being on promotion given total sales of 2100. Interpret this result.

2 MARKS

Further Question 2

A random walk process $\{X_t \mid t \in T\}$ is defined by $X_0 = 0$ and

$$X_t = X_{t-1} + Z_t,$$

where Z_t is a purely random process with mean 0 and variance σ^2.

(a) Simulate 100 random walk processes each of length $t = 1, \ldots, 100$ where $Z_t \sim N(0, 1)$.

Hint: you may find the `rnorm` and `cumsum` functions useful.

4 MARKS

(b) Create a time series plot with all 100 simulated random walk processes superimposed. In addition, superimpose the overall mean of the 100 random walk processes in black. Comment on what you see in the plot.

2 MARKS

Total: 38 MARKS (+ 2 for pdf upload)