Data Analysis Skills: Practice Class Test

This is a practice test to help you prepare for the class test which will contribute 25% to your final grade. This practice test does not count towards your final grade but has the same instructions as the actual class test so you know what to expect (and therefore certain information, like the "you have four hours to complete" instruction, does not apply to the practice test).

Introduction

This practice class test consists of three tasks worth a total of **40 MARKS** broken down as follows:

Task 1. A report on a statistical analysis of a given data set: **25 MARKS**;

Task 2. A Further Task: **13 MARKS**;

Task 3. Successful upload of appropriate `.pdf` and `.Rmd` files: **2 MARKS**.

Tasks 1. and 2. are to be completed within the same R Markdown document (and uploaded in Task 3).

Task 1, the written report on a statistical analysis, should include:

- An appropriate Title, the author's student number and Introduction detailing the data set and question of interest **4 MARKS**;

- An Exploratory Analysis of the data **7 MARKS**;

- A Formal Analysis of the data **10 MARKS** and

- Conclusions **2 MARKS**.

2 additional MARKS will be awarded for an appropriate report layout and good English.

Instructions

1. Go to the **Practice Class Test Files** folder in the **Week 4: Practice Class Test** section of the Data Analysis Skills Moodle page.

2. Download the files in the **Practice Class Test Files** folder to the **same folder**:

    - The two `.csv` files containing the required data sets;
    - `PracticeClassTestTemplate.Rmd` - a R Markdown template for the practice class test submission document. It includes the R packages necessary to complete the tasks.

3. Open **RStudio**

4. From within RStudio open `Practice ClassTestTemplate.Rmd` then save it as `YourStudentNumber_PracticeClassTest.Rmd` in the **same folder** as the `.csv` files are saved.

5. **Before you start to work**, compile `YourStudentNumber_PracticeClassTest.Rmd` (using `Knit`) and check that the `YourStudentNumber_PracticeClassTest.pdf` file is produced as you expected. **It is strongly recommended** to periodically compile and check the `.pdf` file as you create your document so you can fix any bugs in your code as you go.

6. For the report part of the class test (Task 1) you are **NOT** required to include your R code in the `.pdf` file, hence `echo=FALSE` is set as the default in the `.Rmd` template.

However, for **Task 2: Further Task** you **need to provide your R code** in the `.pdf` file, and hence should include `echo=TRUE` in any R code chunks relating to the Further Task.

7. When you are ready to submit your document, click on the **Practice Class Test .pdf Upload** link in the **Week 4: Practice Class Test** section and upload and submit the file `YourStudentNumber_PracticeClassTest.pdf`.

8. Also upload and submit the R Markdown file `YourStudentNumber_PracticeClassTest.Rmd` using the **Practice Class Test .Rmd Upload** link. Please note that only the `.pdf` file will be marked. The `.Rmd` file will only be considered if there was a problem when compiling the `.pdf` file.

Examination Conditions

- You have four hours to complete the class test and you can submit your `.pdf` and `.Rmd` files anytime within that time.

- You are required to use `tidyverse` functions (including `ggplot2`) for the analysis and `RMarkdown` to produce your document.

- You may consult resources (hard copy or online), e.g. `tidyverse` "cheat sheets" and/or the online labs from the course.

- You **must not** communicate or correspond with anyone about the class test during the time that submissions are open. You **must not** submit material you have discussed with or copied from others. If your work is similar to those of any other candidate(s) you will both/all be suspected of collusion and referred to Student Conduct.

- The work you submit must be entirely your own effort and must demonstrate your understanding rather than reproduce text from notes, slides, books, or online sources (which is plagiarism). We may conduct a further oral examination to check your knowledge and establish that the submitted files are your own work.

Task 1. Anatomical Data from Domestic Cats

Digitalis is a heart medicine similar to toxins found in a genus of plants commonly known as foxglove. These plants can be poisonous towards humans and pets. As part of digitalis experiments, 144 domestic male and female adult cats had their heart weight in grams (`Hwt`) and body weight in kilograms (`Bwt`) measured. The data are stored in `cats.csv`.

Using a linear model, describe the heart weights of male and female adult cats. What does the model say about the difference, on average, between male and female cats? 25 MARKS

NB: For the purposes of Task 1, you are NOT required to include diagnostics (i.e. checks of the residuals)

Task 2. Further Task

Every four years the Scottish government publishes data called the Scottish Index of Multiple Deprivation (SIMD) which ranks 6,976 "datazones", effectively small postcode areas, across Scotland based on levels of income, employment, health, education, housing, access to services and crime. The ranking of '1' is the most deprived datazone with the higher the ranking the less deprivied an area is. As well as an overall ranking based on the SIMD, rankings are also given for each of the seven criteria which determin the overall ranking.

The rankings for the datazones in the Glasgow and Edinburgh City areas are available in the file `Glasgow_Edinburgh_SIMD2020.csv`. Import this data into R and answer the following questions.

    a. In what ways is the data in `Glasgow_Edinburgh_SIMD2020.csv` **NOT** in `tidy` format (as defined in Week 2's lab)? Write the R code that converts the data to a `tidy` format and include the code in your submitted `.pdf` document. **5 MARKS**

    b. Produce a single scatterplot of the percentage of the population that is of working age against the rank of the overall Scottish index of multipe deprivation for all datazones in Glasgow City and the City of Edingburh, distinguishing between the two cities in your plot. Include the R code used to produce the plot and the plot in your submitted `.pdf` document. **8 MARKS**

Task 3. Upload appropriate `.pdf` and `.Rmd` files. 2 MARKS