

# Class Test 2

## *Data Analysis*

### Instructions

You are expected to complete a series of tasks described below. Please follow these instructions carefully before you start your work.

1. **Do NOT** open RStudio until you have downloaded the two 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 these files in the **Class Test 2 Files** folder to the **same folder** on your **M: drive**:
  - `.csv` - which contains the data set;
  - `ClassTest2Template.Rmd` - a R Markdown template for this class test which includes the tasks described below. It loads the R packages necessary to complete the task and also reads the data set into R (assuming you've saved the `.csv` files in the **same** folder as the `.Rmd` file).
4. Open RStudio and open `ClassTest2Template.Rmd` then save it as `YourStudentNumberClassTest2.Rmd` in the **same folder** as the `.csv` files are saved on your **M: drive**.
5. **Before you start to work**, compile `YourStudentNumberClassTest2.Rmd` (using Knit) and check that the `YourStudentNumberClassTest2.pdf` file is produced as you expected. It is wise to periodically compile and check the `.pdf` file as you work through the tasks so you can fix any bugs in your code as you go.
6. Unlike in reports, you **are required** to **include** your R code in the `.pdf` file, hence `echo=TRUE` is set as the default in the `.Rmd` template.
7. Before answering each question, ensure that the data is in **tidy** format (and if it isn't modify it so that it is) and produce graphical summaries relevant to the question(s) asked of the data. Note that you **don't** need to label plots in this class test (unlike in reports) but you should comment briefly about how the graphical summaries informally inform your subsequent analysis.
8. When you are ready to submit your completed tasks, click on the **Class Test 2 .pdf Upload** link under **Data Analysis > Week 10: Class Test 2** and upload and submit the file `YourStudentNumberClassTest2.pdf`.
9. Also upload and submit the R Markdown file `YourStudentNumberClassTest2.Rmd` using the **Class Test 2 .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 compiling the `.pdf` file.

### Examination Conditions

- You have the full two hours to complete the class test and you 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 are required to use **tidyverse** and **infer** functions for the analysis and **RMarkdown** to produce your answers
- You may consult ANY resources (hardcopy or online), e.g. **tidyverse** “cheat sheets” and/or the online tutorials from the course.

## Question 1

A farmer is interested in comparing the effect of different fertilizers on crop yield, and decides to undertake an experiment. He wants to compare three different fertilizers, labelled A, B and C, respectively, against a control group with no fertilizer, labelled D. He partitions his field of potatoes into 40 plots, and applies each of the four treatments A, B, C and D to 10 plots at random. At the end of the experiment he measures the total weight (in kilograms) of potatoes grown in each of the 40 plots. Use bootstrap confidence intervals to compare the effects of the fertilizers on the yield of potatoes harvested.

The results of this experiment are stored in `test1.csv`.

10 MARKS

## Question 2

A social scientist is interested in exploring what people consider the “ideal height” of a life partner. She sampled 100 male and 100 female adults and asked what they considered the ideal height (in centimetres) for a partner to be. The social scientist also recorded if the subject was a man or a woman and their height (in centimetres), as she was interested to see if there were differences between men and women and if ideal partner height can be predicted given an individual’s height.

The data set is stored in the file `test2.csv`.

- (a) Generate numerical and graphical summaries that are appropriate for this data and the research questions. Comment briefly on the summaries with respect to the research questions.

5 MARKS

- (b) Starting with the model with the most parameters, use a model selection method of your choosing to find the model which is most appropriate for this data (you do **not** need to check the model assumptions in this class test). Write down the equation(s) for your chosen fitted model and produce a plot which shows how your chosen model relates to the original data. Use your chosen fitted model to answer the research questions in clear, non-technical English.

6 MARKS

- (c) Suppose someone else analysed this data but ignored the effect of gender. What model would they choose to best describe the data? Produce a graph showing their preferred model and compare its interpretation with your preferred model in part (b). What name is given to this apparent contradiction?

4 MARKS