ADS2 Coding Challenge 1 - Mock paper

ADS2

Semester 1 2023/24

Technical Instructions

You have 3 hours to complete this assignment. There are three questions, all of which need to be completed. The instructions and datasets (.csv files) can be downloaded from Blackboard Learn.

Please make an R Markdown file for your response. A template is provided. Please follow the structure set out in the template. Please remember to include your roll number (but not your name) in the author field, as well as in the name of the final document.

The final submission is a pdf knitted from the R Markdown file (if you cannot knit to pdf directly, then knit to Word and convert the outcome to a pdf file using the "Export" function in Word or another text editor).

The submission should contain explanatory text, answers to questions, all results, and all the code used to generate the results. There is one exception: When you read a .csv file and if your name is in the file path, you are allowed to hide that code chunk, so that your anonymity is maintained.

You will be graded not only on your answers to the questions, but also on your ability to compile a well-formatted and readable R Markdown document. It is therefore advisable to knit early and often, and check that your document can be knitted without errors and that the result is in line with your expectations. If you have code chunks that take a long time to run, use the code chunk option cache = TRUE. This means that the results of the code chunk get saved and will be used in the next knit, instead of being computed again (provided the code chunk has not changed).

Please upload your pdf file to the assessment dropbox at the end of the assignment. We are aware that due to increased traffic when everybody uploads their file, your upload may be a few minutes past the deadline. In such cases, we will consult the time at which the pdf document was produced and use this to determine whether or not your submission counts as a late submission. If so, the same penalties apply as for other in-course assessments.

Honour Code

This is an open-book assessment. This means you are allowed to work on your own computer, consult your previous notes, and use your previous code. You are also allowed to look up commands online, if you need to (though the assessment is designed in such a way that you should not need commands or methods beyond what has been taught in this class). If you use code from an online source, please state what the source is (name of site, author if possible, url, date accessed).

You are **not** allowed to work with other students on this assessment. This is why we do not allow mobile phones. Of course, because we are allowing internet access, we cannot completely rule out the possibility of you working together. But we ask that you don't.

We appeal to your sense of honour and integrity. It is wrong to cheat, so don't do it.

By submitting this assignment, you declare that this is the result of your own work and that you did not either get help from, or help, other students.

If, in marking the finished work, we find evidence that students have colluded, this will be treated as a potential violation of academic integrity and brought before the ZAMO.

1. Stroop test

Is your cognitive function better in the morning or the afternoon? A group of students at ZJE tried to find out by making participants take a Stroop test. In this test, participants have to name the colour of a word. Some words are "neutral" (such as "sky" or "car"), while other words are the name of a different colour. The Stroop score is computed from the difference in time it takes participants to name the colour in both conditions. It is an indication of how much automatic reading of colour names interferes with the task of naming the colour.

In order to investigate whether performance on this cognitively demanding test depends on time of day, some participants were asked to take the test in the morning, and some in the afternoon.

The data are provided in file `stroop_test.csv`

Questions

- Import the data and plot them in a useful way.
- Is there a difference in performance on the Stroop task between the morning and afternoon group?
- Name one way in which the study could be improved or followed up on.

2. Marathon finishing times

Long-distance running is a popular activity. But how does the average speed of amateur runners depend on age and gender?

The file `Chicago2013_random_finishers.csv` contains finishing times (in hours) for 170 randomly chosen finishers of the 2013 Chicago marathon (85 male and 85 female).

Questions

- Import the dataset, and plot it in a way that addresses the question we are interested in.
- What are the average finishing times and standard deviation for each gender? What are the average finishing times and standard deviation for each age quartile?
- If you had to suggest a statistical test to determine the effect of age quartile and gender on marathon finishing time, what test would you suggest, and why?

3. Antiviral drug

Prof. Liu studies a new antiviral drug in rats infected with an influenza virus. Half of the rats are given the new drug, and half of the rats (the control group) are given an existing, already approved drug. To measure the efficacy of the treatment, Prof. Liu measures the time it takes animals to fully recover (in days).

Questions

- What are the Null Hypothesis and the Alternative Hypothesis in Prof. Liu's trial?
- For her statistical analysis, Prof. Liu uses the commonly used significance level α of 0.05. What type of error does this relate to, and how?
- After ensuring that the sample size is big enough and that the assumptions of the statistical tests are met, Prof. Liu runs a statistical test and gets a p-value of 0.059. What is a p-value? Based on this result, what should Prof. Liu report as the outcome of this study?