ADS2 Coding Challenge 1

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 data sets (.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. Benefits of swimming for long-distance runners

Long-distance runners don't just train for their races by running, they also use a variety of other exercises to improve their performance. One such example is the use of swimming to improve strength and muscle mass. We tested this by selecting 50 runners from a local club which has in total 500 members. In this question, you will be presented with the time taken for these 50 runners to complete a half-marathon before and after they each completed the same swimming training programme. We want to know whether this swimming improved their times.

The data can be accessed from the file `swimming.txt`. Each row contains the results from a single athlete with the total time for the half-marathon before swimming given in minutes and seconds described in 'before_minutes' and 'before_seconds'. The time taken after swimming is described in the same way for each swimmer in 'after_minutes' and 'after_seconds'.

Questions

- What would be a suitable statistical test for these data and why? Note you may need to tidy these data before deciding on which test to use.
- What are your null and alternative hypotheses?
- Is there a statistically significant improvement on runners' times after swimming?

2. Number of emergency room admissions

Dr. Horsey leads a hospital. She wants to better understand the number of people admitted to the hospital's emergency room, in order to be able to plan staff schedules. In particular, she wants to know whether there is a difference in emergency room admissions between weekends and weekdays. For this purpose, she looks at admissions data for the afternoon shift (eight hours from 1pm to 9pm) of every Monday (weekday) and every Sunday (weekend) for an entire year. The data collected is the number of patients admitted during that hour.

The dataset is provided in file `hospital_admissions.csv` It contains columns recording the week (numbered from 1 to 52), day (Monday or Sunday), hour (numbered 1 to 8) and number of patients admitted during that hour.

Questions

- Import the dataset and plot the data in a useful way.
- Is there a difference in patient admission rates between Mondays and Sundays?
- Based on your findings, what advice would you give Dr. Horsey?

3. Spinal cord injury and novel biomaterials

Complete spinal cord injury (SCI) is a severe condition caused by the trauma of the spinal cord that leads to its breakage. Depending on the site and type of injury, it may lead to the loss of motor and sensory functions and disability. The common prognosis is bad, with less than 30% of patients being able to regain at least some functions over the affected body parts spontaneously within 1 year. The commonly accepted clinical scale to describe the patient condition is American Spinal Injury Association Impairment Scale (AIS scale), which has 5 grades: 'A' (the complete motor and sensory functional impairment, complete disability), 'B', 'C', 'D', and 'E' (the above-mentioned functions are normal, healthy subject).

Your team has developed a novel biomaterial that may help to regenerate injured nervous tissue after SCI. Studies on animals showed promising results, and now you perform Phase I-II trial on patients with complete chronic SCI. Your team recruited patients with chronic SCI (at least 1 year since the trauma passed with no signs of recovery), recorded their condition upon admission, and installed implants with the novel material. 15-18 months later, you recorded patient AIS scores again.

The data is in the `SCI_before.csv` and `SCI_after.csv` files. Merge both tables appropriately, analyse the data, draw conclusions, and give further suggestions.

Questions

- Import, arrange the data (merge both pieces of data and make the data possible to analyse), and make it suitable for analysis, e.g. the values. You should perform all the manipulations in R and provide the code.
- Check your data carefully. Identify features of the data and discuss your conclusions. Make illustrative plots.
- Formulate the correct statistical hypothesis to compare the groups, choose the appropriate statistical test, and check assumptions for this test. Explain your choice briefly. Then, perform this test and identify whether the difference between the experimental groups is statistically significant.
- Discuss the results you got. What did you obtain? Are there any flaws in the experimental design and what would you suggest to your colleagues? Support your statements with the appropriate statistics and/or effect size estimates.