Coding_Challenge_1(1)

2020年1月8日 星期三　　下午2:00

Coding_Challenge...

ADS2 Coding Challenge 1

MI Stefan

Semester 1, 2019/20

Technical Instructions

You have 2 hours to complete this assignment. There are four 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 result 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 assigment. 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).

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. Side effects of medication

A researcher is testing a new drug for mood disorders. On top of testing the drug itself, it is important to test for possible side effects. One possible side effect is weight gain.

In order to study this, the researcher weighs each mouse in the treatment group before the beginning of the experiment and again at the end. The difference in weight for each mouse is recorded (positive means a gain in weight, negative means weight loss).

The data are available in file `mouse_weight_changes.csv`

Questions

- Plot the data
- Is there weight gain in mice that have been treated with this medication? Choose and conduct an appropriate test. Explain why the test is appropriate, and discuss what the results mean
- Name and explain one way in which the experiment could be improved or one possible direction for future study.

2. Drug use among college students

A survey about drug use among undergraduate students is desgined as follows:

- Participants are shown two questions, A and B
- Participants born between 1 January and 30 June are asked to answer question A
- Participants born between 1 July and 31 December are asked to answer question B
- The person administering the survey does not know which question the participant is answering or when the participant was born

The survey questions are:

A. Have you ever taken illegal drugs?
B. Is your birthday on an even day?

Of 250 students who completed the survey, this was the result

Reply	Number
Yes	108
No	142

Questions

- What proportion of students are likely to have used illegal drugs?
- What is the reason for setting up the survey in such a complicated way?

$250 \times \frac{181.25}{365.25}$ question A

$250 \times \frac{184}{365.25}$ B

$250 \times \left(\frac{181.25}{365.25} (x) + \frac{184}{365.25} \times \frac{90}{184} \right) = 108$

$\frac{108}{250}$

3. Student improvement in a beginning programming course

Professors teaching an introductory programming course want to know whether the course improves students' programming skills. In order to measure skill levels, the professors ask students to self-assess using the following question:

"Please rate your own programming skills on a scale from 0 (unskilled) to 10 (very hihgly skilled)"

They ask students the question twice: Once on the first day of class, before the course starts ("before"), and once after the students have taken their final exams ("after").

The course ran three times from 2015 to 2017. Data for all years is provided in file `student_skills.csv`

Questions

- Visualise the dataset in a useful way
- Explain in a few short sentences why this is a useful way to look at the data
- You may have noticed that for some students, taking the course *decreases* their programming skills. For how many students is this true?
- Why do you think this is the case, and how might it be fixed?

4. Finishing times in the Chicago Marathon

The file `Chicago2013_random_finishers.csv` contains finishing times (in hours) for the 2013 Chicago Marathon for 85 randomly selected women and 85 randomly selected men. It also makes a note of the country of origin of each runner.

A reasonable question to ask is whether there is an effect of gender on finishing time. There may also be an effect of country of origin. In particular, since this is an event held in the USA, it may be the case that participants who travelled from afar to run this marathon are more highly motivated than participants who only travelled from within the US, and this may translate into faster finishing times.

Thus, runners come in four different groups (combinations of gender and finishing time).

Questions

- Plot the data in a useful format
- Is the average difference between two runners in the same group, smaller, bigger, or equal to the average difference between two runners in different groups? What does this tell you?
- Is there a statistical test that you could do to determine whether gender and/or country of origin has an effect on finishing time? What assumptions need to be met in order to conduct that test, and are those assumptions met in this dataset?