ADS2 Mock Coding Challenge 2

Semester 2, 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 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. Vitamin C and tooth growth

Lack of vitamin C leads to severe health issues. It is not produced in the human body and must be supplied with food. At the same time, personnel that have limited access to fresh vegetables (sailors, spacemen, travelers, etc) may suffer from the insufficiency of this compound in their food. Thus, a vitamin C formulation that can preserve its properties for a long time is of great need.

Researchers developed such a formulation. *In vitro* tests showed its efficiency. Now, they performed an *in vivo* trial. Guinea pigs received the newly developed formulation of Vitamin C or fresh orange juice (normalized according to the concentration of vitamin C) in addition to their standard diet (`supp`). Each type of additives included three concentrations (`dose`) of vitamin C: 0.5, 1, and 2 mg/ml. The measured outcome is the tooth length (`len`) in mm (stem cells that become teeth are sensitive to vitamin C).

Questions

- Import, check, and organize the data appropriately. Reformat columns if needed.
- Plot the data in a useful way.
- Choose, justify, state the statistical hypotheses, and carry out an appropriate test to answer whether the vitamin C formula is useful.
- Present and discuss your results. Is this novel formula useful? What would you suggest doing next?

2. Mutation and survival

You work on the mutation of a certain gene (Gene_X) that likely causes developmental abnormalities in humans but is quite rare, and the precise role of the mutation is not known. You created a mouse model by introducing a similar mutation in a similar location within the murine genome.

You set several breeding pairs and crossed mice as Gene_$X^{WT/mut}$ × Gene_$X^{WT/mut}$. You recorded the genotype of the newborn mice. Your genotyping record (`genotype.csv`) includes `mouse_ID`, birth date (`BD`), `sex`, and `genotype`.

Answer the questions below, provide your analysis, and explain your results. Given the genotyping records you got, what can you say about the studied mutation?

Questions

- Import and organize the data.
- *Describe* the data in a useful way.
- What would you expect under Mendelian inheritance?
- Choose and justify the appropriate statistical test, state the statistical hypotheses, and carry the test out an appropriate test on whether the mutation affects the survival of mice.
- Present and discuss your results. What would you suggest doing next?

3. Coffee shop opening hours

A new coffee shop has opened on campus. Hooray! Coffee shops are normally open from 6am-5pm but the owners are aware that students often sleep later than other members of the society. After being open for one month, they run a month-long trial opening 10am-9pm to see if students prefer these times. They leave an iPad at the serving counter where customers can record if they are 'satisfied' or 'unsatisfied' with the opening times.

During the 6am-5pm opening times, the iPad records 864 presses of the 'satisfied' button by customers and 714 presses of the 'unsatisfied' button. When they change these times to 10am-9pm, they receive 980 'satisfied' pressed and 473 'unsatisfied'.

Questions

- What would be a suitable statistical test for these data and why? (6 points)
- What are your null and alternative hypotheses? (4 points)
- Are students more satisfied with the early or later opening times? (15 points)

Originally created by Dmytro Shytikov and Robert Young in 2024.