ADS2 Coding Challenge 2

Semester 2, 2022-23

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 text explaining your approach and your working, 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 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. Neural activity patterning and osmotic challenge

Dr Ion Chloride is studying the spiking properties of vasopressin neurons of the hypothalamus and how they respond to a prolonged osmotic challenge, using a salt loading experimental protocol and electrophysiology to record and measure the neurons' spiking activity. There are two overall types of neuron, continuous firing, and phasic firing. The continuous firing neurons can be easily fit by a model. The phasic neurons haven't yet been fit with the model.

Dr Chloride wants to know whether the neurons' intrinsic spiking properties, classified by the choosing the best fit of a model to the recorded spike patterning, are different in the salt loaded neurons, compared to neurons recorded in control conditions. To find the best fit, the model was tested with each neuron using different combinations of afterpotentials (HAP, DAP, and AHP), labelling the best fit as 'HAP' (HAP only), 'DAP' (HAP + DAP), 'AHP' (HAP + AHP), or 'All' (HAP + DAP + AHP). The data from the model fitting is provided in the file `vasotypes.csv`.

Questions

- Import and organise the data set.
- Plot the data in a useful way.
- Choose, justify, state the statistical hypotheses, and carry out an appropriate test to answer Dr. Chloride's question.
- Discuss the result, have the neurons' intrinsic properties changed? Are there any problems or limitations with the current study?

2. Myoinositol and gestational diabetes mellitus

Gestational diabetes mellitus (GDM) is the most common metabolic disorder of pregnancy. Usual treatment includes dietary supplementation with folic acid but it is insufficient to normalize insulin sensitivity by itself. Inositol and its metabolites are reported to alleviate insulin resistance symptoms and may help to treat GDM.

You work in a team of clinicians that want to compare the effect of different dosages of inositol stereoisomers and folic acid supplementation on insulin resistance levels and maternal-fetal outcomes in GDM women. Your team enrolled 80 patients (pregnant women with GDM and impaired glucose tolerance) around the country and randomly allocated them to receive daily for 8 weeks:

- 400 mcg folic acid (control treatment),
- 4000 mg myo-inositol plus 400 mcg folic acid (MI treatment),
- 500 mg d-chiro-inositol plus 400 mcg folic acid (DCI treatment),
- 4000/500 mg myo/d-chiro-inositol plus 400 mcg folic acid (MI plus DCI treatment).

The researchers determined insulin-free time (days of pregnancy without insulin) and body weight at the beginning of pregnancy and 1 week before delivery.

The data are recorded in several separate files, each from the respective hospital (site1.csv, site2.csv, site3.csv, site4.csv, site5.csv, site6.csv, site7.csv, site8.csv). The files include information about patients: patients' names, sites of the clinical trial, treatment group, days of pregnancy spent without insulin supplementation

(insulin-free days), and body weight measurements before treatment and at the endpoint, and comments if required.

Your colleagues wonder whether the suggested treatments improve insulin resistance and prevent excessive weight gain in pregnant women. You may compare the difference in the analyzed characteristics between both time points.

Questions

- Import the data and "clean" it.
- Choose the appropriate method of statistical analysis, check assumptions for this analysis, and explain your choice briefly.
- Formulate the correct statistical hypotheses, conduct the analysis, identify the effect of treatment, and show which factor has a higher effect on the tested characteristics. Plot your data.
- What could be done in the future to follow up on this study? Give some suggestions to the researchers.

3. Classifying neuron types from electrophysiological recordings

You are working on a project studying a region of the brain involved in appetite regulation. Your colleague has recorded the electrical activity of 25 neurons and used their own analysis to classify the recorded neurons into 5 types based on the patterning of their electrical activity.

You have performed your own independent analysis using a model to fit the 25 recordings, varying two fit parameters, *hap1* which represents a half-life or time constant, and *hap2* which represents a magnitude. The results are in the file `vmndata.csv`
It is sometimes better to use the natural log of a half-life value in visualisation and analysis.

Questions

- Import the original data and plot it in a useful format to show the original classifications (type)
- Use clustering of the model fit data to make your own classification of the recordings (You are allowed to use the `kmeans` function in R)
- Plot the outcome of this clustering (e.g. by assigning colours by cluster).
- Test the clustering using different subsets of the fit parameters. Describe in words how your clustering compares to the original classifications.