Bayesian Statistics: Assignment

Shota Gugushvili

10 October 2024

1 Achtung

Read this file prior to starting to work on the assignment.

2 Summary

Use **R** and build a Baysian model to analyse a small dataset.

3 Description

Since the days of Sir Ronald Fisher, statistics has played a fundamental role in life sciences. In this assignment, you will analyse one such dataset. Data are real data from an undisclosed source.

An experiment was devised to investigate the effect of diet on blood coagulation time. There were 24 subjects participating in the experiment, receiving 4 different diets A, B, C and D. The subjects were randomly allocated to diets and blood samples were taken and tested in a random order. The interest lies in finding out whether there are significant differences in mean coagulation times for 4 different diets.

Data are in the `undisclosedsource.dat` file. The file contains three columns: a variate (`run`) to indicate the order of collecting and testing blood samples; a factor (`diets`, with 4 levels) to identify the treatment group for each subject; and the response variate (`y`) obtained from each subject, i.e. coagulation time (in seconds). You will decide yourself which variates to use in your analysis.

Build a Bayesian model with **brms** to analyse this experiment.

You must produce a report containing your analyses and findings. Some useful steps (but there could be more) are:

- Data exploration and summarisation
- Building a Bayesian model (with suitable priors) in **brms**
- Examining and summarising the fit

- Examining convergence of Markov chains
- Validating model assumptions
- Comparing diets
- Reporting conclusions

Finally, attention points are the following:

- The report must read like a self-contained story
- The **R** script must be well-documented and explain each of the steps that you undertook

4 Rules

You can read and use anything you want in the literature or on the internet. Asking questions in discussion groups is not allowed. You are supposed to work in a team with another student. ChatGPT might not help you with analyses, but may improve your writing once you know what to do and have already done it.

5 Hints

brms has a well-documented website with nice vignettes: `https://paul-buerkner.github.io/brms/`. My study materials might be helpful.

A convenient tool to wrangle data in **R** is **tidyverse**. **ggplot2** is excellent for making pictures. Of course you can use the base **R** too.

6 Submission guidelines

The assignment must be submitted no later than October 17, 2024 (date inclusive). Only one submission per a team of two is required. Indicate names in your report.

Download the Declaration of Originality[1] and sign it[2]. It sets out the rules. Use either the English or the Dutch version.

You can prepare the submission using **R Markdown**[3], which allows direct incorporation of the **R** code. Submit both the compiled pdf (or html) file and the source Markdown file. Include the Declaration of Originality. Zip all the files in a single archive.

[1] The link is: `https://www.medewerkers.universiteitleiden.nl/binaries/content/assets/ul2staff/vr/onderwijs-op-afstand/verklaring-van-originaliteit.pdf`

[2] You can sign pdf files digitally on your mobile device using a specialised app. Adobe Fill & Sign works nicely.

[3] In case you are unfamiliar with R Markdown, this is a good starting point: `https://rmarkdown.rstudio.com/lesson-1.html`

Alternatively, you can type the report in a text editing programme (e.g. Word or LaTeX), paste in the code and figures, and convert the file into a pdf file. Submit the pdf file and the Declaration of Originality in a single zip archive.

7 Grading

Grading is on the $1 - 10$ scale. For a score 6, the statistical analysis must be sufficiently complete and the accompanying script must be functional, with enough details to reconstruct what you did. See also the grading scheme for more details.

Good luck!