Bayesian Statistics: Assignment

Shota Gugushvili

17 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 foundational work of Sir Ronald Fisher, analysis of designed experiments through statistics has penetrated numerous branches of science and became indispensable in industry. In this assignment, you will analyse one such dataset.

Blocking is a popular technique to deal with inhomogeneity of experimental units, thereby helping to better uncover the effects of applied treatments. A possible way to analyse experiments with blocks is through a mixed effects modelling approach. An example is given in Section 10.6 of J. J. Faraway's textbook *Extending the Linear Model with R*. This textbook is referred to in the *Linear and Generalized Linear Models* course[1], as well as the *Essentials of Mixed and Longitudinal Modelling* course[2].

In detail, the production of penicillin uses a raw material, corn steep liquor. The penicillin yield is the primary response variable and the researchers wanted to try 4 processes, A, B, C and D, for the production. However, the liquor is quite variable in quality, and it was believed that this alone could cause substantial differences in yield. Luckily, it was found that for experimental purposes blends of liquor could be made that were sufficient for 4 runs. This gave an opportunity to perform an experiment with 4 treatments (A, B, C, D) and 5 blends (blocks). By randomly assigning the order in which the 4 treatments were applied within each blend (block), the blend differences were largely eliminated and

[1] `https://studiegids.universiteitleiden.nl/index.php/en/courses/122105/linear-and-generalized-linear-models`

[2] `https://studiegids.universiteitleiden.nl/index.php/en/courses/121709/essentials-of-mixed-and-longitudinal-modelling`

valid conclusions could be drawn on the treatments. Possibility of interaction between treatments and blocks was not considered in the analysis, as there is hardly any evidence for it. The corresponding experimental design is called the randomised complete block design (RCBD).

Data are available as part of the **R** package **faraway** under the name `penicillin`.

Build a Bayesian model with **brms** to analyse this experiment. Specifically, consider the `treat` variable as a fixed effect, and `blend` as a random effect. The latter is justified, because blends that were used can be considered as being selected from a notional population of all the blends.

You must produce a report containing your analyses and findings. For inspiration and ideas you can refer to Faraway's textbook, who does a frequentist analysis of the data. Bayesian analysis has its own twists, though. 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
- Drawing conclusions on treatments
- 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 25, 2024 (date inclusive). Only one submission per a team of two is required. Indicate names in your report.

Download the Declaration of Originality[3] and sign it[4]. It sets out the rules. Use either the English or the Dutch version.

You can prepare the submission using **R Markdown**[5], 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.

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!

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

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

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