Problem Sheet

BIOM4025 - Statistical Modelling - 2024

Introduction

For this *Problem Sheet*, you will be analysing some of the data provided through the *Model seeks Data* link on the ELE page under Task 1.3.1).

I have exported the data for you and I have also done a bit of cleaning-up (but not too much!). You can find the data as a `Comma seperated values (.csv)` file in the Assessments section of the ELE page.

You will see that I have kept the description of the tasks general. This means that you have a lot of freedom in how you choose to approach each one of them. What I expect from you is to show off your statistical modelling and `R` skills by applying what you have learned during the lectures and especially the practicals. For more information, see the marking criteria on the ELE page.

Note that during the lecture and practical Q&A sessions we will NOT answer questions that are *directly* related to this problem sheet.

What to submit

You are required to perform all of the tasks listed on the next page. Remember that there are often many ways of achieving the same thing, but simpler is usually better. You are limited to *3 pages* (including your code and figures), so you will have to be both selective and concise, and you have to think carefully about what you present.

You will need to provide both the `R` code (but only the code that worked and that is relevant to the tasks) and the relevant output. There are two ways of doing this:

1. Copy-paste everything into a Word document. Use a fixed-width font like `Courier New` or `Menlo` for both your `R` code and the output. Make sure that you annotate your code so that someone else (*i.e.* me) can follow why you did what, and to show that you have understood what you did. To copy-paste a figure, go to the `Plots` tab in the bottom-right panel and click on `Export` → `Copy to Clipboard...`. Alternatively, you can use `Save as Image...` or `Save as PDF...` and import this file into Word. There are no specific requirements with respect to font size, line spacing or margins, but please use common sense and keep it clear and easy to read. Save the final document as a PDF.
2. You can use `R Markdown`, which provides a powerful tool to combine `R` code, output and any other text, images, equations, etc., into a single document. As a matter of fact, most of the documents for this module (including this one) are written using `R Markdown`! Although I think that on the long-term learning `R Markdown` might be worth it, it is yet another new thing to learn and you probably have plenty of other things to do as well. It is therefore completely up to you whether you use `R Markdown` or not. If you would like to know more, have a look at the `R Markdown` website.

How to submit

Submit a PDF via the submission point on the ELE page. The deadline is **December 3 at 12:00 (noon)**.

Tasks

We have data for a diverse set of variables, including both continuous and categorical variables, and while some are expected to be normally distributed, others probably are not.

Task 1

Choose a dependent variable and one or more predictor variables and formulate (in words) the hypothesis that you would like to test. Tip: Choose at least one numerical variable, either as dependent or predictor variable as plotting models with only categorical variables can be tricky.

Task 2

Use linear, generalised linear, linear mixed and/or generalised linear mixed modelling to test this hypothesis. Provide the code that you used, as well as any relevant output. Add comments (using `#`) to your code that allow me to understand why you did what.

Task 3

Write one or more sentences for the *Results* section of a scientific article in which you summarise your most important result(s). It does not matter if your P-value is smaller or larger than 0.05!

Task 4

Provide a publication-ready figure that illustrates this result, including the code used to generate the figure.