BCB743 Basic Statistics Assignment

Due date: Tuesday 19 April 2022

1. Preamble

A report comprises of an Introduction, Methods, Results, Discussion, Conclusion, and References. For this assignment each student is required to write a mini-report of no longer than 8-pages.

Find, select and use a dataset of your choice. R has many built-in datasets. One example of a built in dataset used throughout this course is the *ChickWeight* dataset. Information for this dataset (or any built in R dataset) may the obtained by running ??Chickweight in the console. But you are not allowed to use the *ChickWeight* or *Iris* datasets, as they have been used *ad infinitum* in the course. Neither may you use any of the datasets included with this course.

Various datasets may be found online, and here are some examples to get you going:

https://vincentarelbundock.github.io/Rdatasets/datasets.html

https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/00Index.html

https://hofmann.public.iastate.edu/data_in_r_sortable.html

https://www.r-bloggers.com/datasets-to-practice-your-data-mining/

https://r-dir.com/reference/datasets.html

OR

The R Datasets Package—there are around 90 datasets available in the package. Most of them are small and easy to feed into functions in R. See a list of data with the statement below:

```
> library(help = "datasets")
```

OR

Be a motivated scientist and collect your own data.

2. Instructions

Produce a document with the components below. Next to each is indicated the mark allocation for the section.

Introduction **[10%]**

- Background information
- Importance
- Limitations
- Aims and objectives
- Clearly state the hypothesis being tested

Methods **[25%]**

- Materials and/or Site selection, etc., incl.
    - Map of the site location were data were obtained (if relevant)
    - Study design
    - Procedure
    - Mention the statistical analyses that will be used, as well as justifications for these methods (these must relate to the hypotheses being tested)
    - Mention which programs and libraries were used to do these analyses
    - All relevant R code submitted separately as an R script **[30%]**

Results **[25%]**

- Visualisations of descriptive stats, group differences, correlation or regression, etc.—i.e. graphs (these graphs should relate to the question being addressed)

- Statistical tests—all statistical output must be written in full as appropriate for the method, e.g. the outcome of *t*-tests and ANOVAs is written like this (from Dyer et al., 2019):

"The chl-*a* concentrations were significantly different under upwelling and downwelling conditions ($t = -7.173$, $p < 0.001$), with higher concentrations during downwelling conditions (Fig. 3B). There was, however, no difference with distance from the kelp bed ($F = 2.881$, $p = 0.096$, $R^2 = 0.195$) with concentrations remaining constant along the length of the transects (Table 2)."

- Explanation (no interpretation) of the outcome of the graphs and statistical tests, focussing only on the things that needed testing according to the Aims/Hypotheses
Be sure to show all the tests (exploring the data, normality tests, homoscedasticity) before doing any inferential statistical analyses

Tip: Read some papers to see exactly how the Methods sections and Results are constructed.

Discussion **[8%]**

- State the major findings
- Explain the meaning of the findings and why they are important
- Relate the findings to those of similar studies
- Mention/ explain alternative explanations of the findings
- State the relevance of the findings
- Acknowledge the study limitations

Conclusion **[2%]**

- Include the summary of the main points

Bonus marks

- Using Rmarkdown
- Creating a theme for your plots
- Creating graphs or statistical tests with interpretations that has not already been done in class

Final submission

Create a folder titled with your name and surname e.g."amieroh_abrahams". In this folder you will have the following:
- A final word and PDF document
- A script with all of the necessary code used
- All of the datasets used within this research assignment
- If Rmarkdown was used please attach the Rmarkdown scripts as well as the Word and/or PDF document produced from it.
- Compress (e.g. zip) all the files, name it properly, and send it to Jesse and myself by email.

Tips

- Make comments throughout the R script
- Clean scripts will only have a positive influence on the person marking which is important

3. References

Dyer DC, Butler MJ, Smit AJ, Anderson RJ, Bolton JJ (2019) Kelp forest POM during upwelling and downwelling conditions: using stable isotopes to differentiate between detritus and phytoplankton. Mar Ecol Prog Ser 619:17-34. https://doi.org/10.3354/meps12941