Assignment 1

Due date: 10 September 2023

This assignment is worth 10% of final mark for STAT 2402.

This assignment involves analysing one dataset. The task is described below. Page 2 provides information on the requirements of the assignment and page 3 contains details of the dataset.

Task The data frame `abalone` is available in the file `abalone.txt`.

Aims of Analysis: To determine how the number of rings on the shell of the abalone depends on the other variables in the data.

Please turn over

Assignment Submission Information

Submission: Assignment 1 *must be submitted the LMS submission link* by *11:59 pm on Sunday 10 September 2023*. A pdf document is preferred, but Word document or html will also be accepted. You are strongly encouraged to use `R, knitr` and LaTeX to prepare your submission.

Please be sure to include your name and student number on the assignment. Late submissions will attract a penalty (5% per day late) unless the unit lecturer is advised as soon as possible of any extenuating circumstances and an extension is granted.

Requirements of the assignment:

The assignment task involves the analysis of data. If you have difficulty accessing the data set you should contact the unit co-ordinator *immediately*.

You should hand in a report based on your analysis. The report should be *no longer than four single A4 pages (excluding any relevant Figures, Tables) in no smaller than 10 pt font.* The page margins should *not be less than 2 cm. Submissions will be marked down for exceeding this page limit.* The aim of the report is to convey the aims, methodology and results of your data analysis in a concise, readable fashion.

Style of report

This is a *technical* report, so you should include model equations.

It is strongly recommended that you structure your report into sections, along the following lines. I have suggested the length of each section, but the total length **must not exceed 4 pages**.

- *Executive summary:* Briefly introduce the data, the context, the aims of the analysis and the main findings. This should not exceed half a page.
- *Introduction:* Begin by setting the context of estimating or predicting the number of rings based on the other measurements. Search for and find literature relating to this, and summarise the findings. Then provide a more detailed description of your data and context, and the aims of the analysis. Between 1 and 1.5 pages.
- *Methodology:* Describe the statistical methods that you use. Roughly speaking, you should give enough information to enable a professional statistician to reconstruct what you did. Half a page.
- *Results:* Describe the results of your analysis. You will include appropriate exploration of data (tables and graphs), and describe your findings. You will also describe the model fitting procedure (such as steps in arriving at the final model). **No `R` outputs or code should be given here.** Note that this section contains results only, not interpretation.

  Select your graphs and tables carefully. You will only include graphs and tables that highlight a specific aspect of the data or analysis that cannot be equally simply described in words. Any graphs or tables that are included in the report *must be discussed.*

  About one page.
- *Discussion:* Draw conclusions **based on your results**. Discuss any issues in the analysis that may affect the conclusions (e.g. potential weaknesses of your analysis, alternative analyses). About half a page.
- *References:* A list of references that you have used and cited in your report. Any standard form of referencing and citing may be used, but be consistent in the format that you use.
- *Appendix:* This is optional. You may include tables of figures that you deem are worth showing but do not belong in the main report. Think carefully whether you really need an appendix. If you choose to include an appendix, select carefully what you put in it.

Assessment:

Your report will be assessed as follows.

- Exposition (40%): The report should be well organised, readable, and concise. The report should be readable English, not computer language. The writing should be concise and to-the-point. Graphics should be selected for appropriateness and relevance. Marks will be deducted for exceeding the page limit. Marks will be deducted for writing that is vague, circumlocutious, illogical or self-contradictory.
- Statistical content (60%): marks will be awarded for the correct use of appropriate statistical techniques, and for the correct interpretation of results from these techniques. Marks may also be awarded for the 'completeness' or thoroughness of the analysis (e.g. careful inspection of the data at the beginning). Marks may be deducted for grossly inappropriate selection of technique, gross misinterpretation of the results, failure to notice important features of the data, and grossly unjustifiable conclusions (e.g. speculation, exaggeration).

Warning: The work that you submit must be your sole effort (i.e. not copied from anywhere or anyone else). Penalties for plagiarism will apply according to faculty policy.

Information on the Data

The dataset is on 4177 abalones. This data can be used to obtain a model to predict the age of abalone from physical measurements. The age of abalone is determined by cutting the shell through the cone, staining it and counting the number of rings through a microscope. Other physical measurements are easier to obtain, and may be useful in predicting the number of rings, which determines the age.

Analyse the data to determine the relationship between the number of rings and the other variables in the data.

The variables are as follows.

- `Sex`: M = Male, F = Female, I = Infant
- `Length`: Longest shell measurement (mm)
- `Diameter`: Shell diameter (mm)
- `Height`: Shell height (mm)
- `Wholewt`: Weight of the abalone (g)
- `Shuckedwt`: Weight of meat without the shell (g)
- `Viscerawt`: Gut weight (after bleeding) (g)
- `Shellwt`: After being dried (g)
- `Rings`: Number of rings, +1.5 gives the age in years