Macquarie University

STAT6170 **Statistical Report**

DUE DATE

See iLearn for this report's due date and time.

SUBMISSION INSTRUCTIONS

This assignment must be created using `R Markdown`, and you must upload to iLearn both your pdf document and the `R Markdown` source file that produced it. Your submission must be exactly two files, in exactly these formats. Marks will be allocated based on the content and the way it is presented. Your work must be neat and clear, the questions clearly numbered and ordered, the included plots complete with labels, title, etc. as appropriate.

Each student receives their own unique data set, and must base their analysis on that data only.

A page limit applies (at the marker's heartfelt request). Your submission should be at most 10 pages in length, including any title pages, appendices et cetera, with font size no smaller than 10 points. This is an upper limit, not a target; the best submissions (from students who understand what's required, do just that, and explain it clearly and concisely) will be substantially shorter.

LATE SUBMISSIONS

Unless a Special Consideration request has been submitted and approved, a 5% penalty (of the total possible mark of the task) will be applied for each day a written report or presentation assessment is not submitted, up until the 7th day (including weekends). After the 7th day, a grade of '0' will be awarded even if the assessment is submitted. The submission time for all uploaded assessments is 11:55 pm. A 1-hour grace period will be provided to students who experience a technical concern. For any late submission of time-sensitive tasks, such as scheduled tests/exams, performance assessments/presentations, and/or scheduled practical assessments/labs, please apply for Special Consideration via ask.mq.edu.au

1. In the dataset linked below, some data from a random sample of 255 stoaches have been collected. The variables recorded for each subject are listed below.

Variable	Description
id	Subject ID
species	Either "Spotted" or "Striped"
weight	Body weight of the stoach
albedo	Proportion of light reflected from animal's surface
territory	The amount of territory 'controlled' by the stoach (unknown how this was defined or measured)

The data set (whose first 5 records are shown below) has been emailed to you along with this document.

ID	species	weight	albedo	territory
subj1	Striped	235.2	40.3	27.8
subj2	Spotted	158.2	50.0	17.0
subj3	Spotted	201.5	35.8	18.4
subj4	Striped	207.6	27.8	25.8
subj5	Spotted	167.6	32.4	17.2

Import the data in Excel and check that they are the same as in the table above. *Please note, the data have been simulated and not based on any real data set.*

The goal of this assignment is to use the methods learnt in the unit (please don't use other methods) and write a **statistical report** that addresses the following research questions:

(a) Is there any difference in the average albedo of Spotted and Striped stoaches?

(b) What is the relation between the weight of stoaches and the area occupied?

In the first question you have to check if the distribution of albedo has the same mean for Spotted stoaches and Striped stoaches. In order to do this, you have to check the assumptions, and in particular that the two samples could come from normal distributions with equal variances.

In the second question we want to study the relation, if any, between the weight of stoaches and the response variable area occupied. In this case you have to take care of the role of species: can you study Spotted stoaches and Striped stoaches together, or do you have to split the data set and perform two regressions? Checking the assumptions will often, but not always, help you with the latter point. The results of all the possible regressions you can run will definitely answer this question.

Finally, your job is to write a statistical report. In addition to the correctness of the analyses, we will also evaluate the presentation: is the report well written, are the numerical and graphical summaries clear, are all the parts exhaustive, is the discussion/conclusion section a logical consequence of the results obtained? There are some resources on iLearn on how to write a statistical report. Please use them and ask questions on the forum if you have any doubts, or just want to comment.

Please use R, within an R Markdown document, to perform all of your analysis, and submit both your source file with extension ".Rmd" and the resulting output file with extension ".pdf". The ".Rmd" file must compile to produce the ".pdf" file; if you need help with this, ask well in advance of the due date.