# 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.

**LATE SUBMISSIONS**

In the case of the late submission of an assignment, if no special consideration has been granted, 10% of the earned mark will be deducted for each day that the assignment is late, up to a maximum of 50%. After 5 days, including weekends and public holidays, a mark of 0% will be awarded for the assignment.

1. In the dataset linked below, some data from a random sample of 201 weightlifters have been collected. The variables recorded for each subject are listed below.

| Variable | Description |
|---|---|
| id | Subject ID |
| gender | Either "female" or "male" |
| bodyweight | The weight of the subject |
| age | The age of the subject in years |
| weightlifted | The maximum weight lifted by the subject in an unsepcified exercise |

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

| ID | gender | bodyweight | age | weightlifted |
|---|---|---|---|---|
| subj1 | male | 113.8 | 31.4 | 177.0 |
| subj2 | male | 110.9 | 24.9 | 175.1 |
| subj3 | female | 66.9 | 23.9 | 106.9 |
| subj4 | male | 115.4 | 23.7 | 179.7 |
| subj5 | female | 84.5 | 32.6 | 113.2 |

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 age of female and male weightlifters?

(b) What is the relation between the body weight of weightlifters and the weight lifted?

In the first question you have to check if the distribution of age has the same mean for females and males. 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 body weight of weightlifters and the response variable weight lifted. In this case you have to take care of the role of gender: can you study females and males 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.