

## MA334 Individual Project, Spring 2022

Deadline: **12 noon, Wednesday 27th April 2022**; 100% of module marks

- **Important:** You may choose a data set of your own choice. The marking will be on the quality of statistical argument alone. The R code should be in an appendix and is not part of the assessment itself, hence you are free to use code snippets etc at will. It is the quality of the unique data analysis you perform which will be assessed, not the R coding. Therefore, be aware that this is not a review article nor is it an R coding exercise and the main body of your project report should not include background discussion on standard statistical methodology or standard R coding details. If your data wrangling required you to produce bespoke coding, then this would merit an explanation whilst introducing your data set within your project's main body. If you choose a large or complex data set, it will be necessary to select limited aspects of the data set for the purposes of this project. Beware of the tendency to want to produce only clear positive results such as very low p values for hypothesis tests etc. There is no need to find a satisfying result, good statistical arguments embrace inconclusive results.

**All the following conditions must be satisfied:**

- Reports must be submitted only as a **pdf** file. Do not submit in any other format, do not zip your file. It must be in a .pdf format when uploaded to Faser. Any file not submitted as a pdf will be allocated a zero mark and possibly re-considered at a later date.
- Your report must a maximum of five A4 pages excluding R code. R code must be in an appendix, not in the body of the report and does not count towards the 5 page maximum allowed length. There will no need for a greater length of report if the content is restricted in the way described above and excessively long reports, not following these requirements, will receive a low mark.
- Use any methodology from the module and any R code provided but do not describe these in your report. The report must be about your data analysis and not include description of standard methodology. References to the set book are optional and not required but the analysis and all text in your report must be your own unique work. It is likely that the only reference you offer will be to the data set itself within the early part of the report.

## The basic structure of the report

1. A careful introduction to the data set chosen and importantly the variables chosen within this data set. Give a reference to the data set. This section in your report is important as people will have chosen a variety of data sets. Data exploration, the usual first step should be in the next section, not here. **[20 marks]**
2. Description of the analysis performed on the chosen data, the results and discussion of these results. Begin this section with a data exploration which usually provides univariate statistical summaries and discussion. Do not include discussion of standard methodology rather only include the data analysis you performed. Do not include the R code here, this must be separated into an appendix. Basic R graphics are sufficient although intelligent use of visualisation may be helpful. The project is about quantitative statistical analysis. The use of graphics and data without associated quantitative analysis will be marked low. **[70 marks]**
3. A minimal conclusion and a brief overview of what your analysis provides. **[ 10 marks]**