# **Assignment Instructions [Taken from handbook]**

Your individual project report will be based on a dataset provided on the Assessment page of the course Learn site. Your main submission will be a pdf document generated from an R Markdown script, with a length limit of 15 pages. Please make sure that you keep all your assessment work in an R Markdown and you knit to pdf to obtain a file for submission.

In your Markdown script, please set the options up so that all your code and the output is visible in the knitted pdf document (echo = TRUE). If you see warnings or messages that take up a lot of space (for example when loading the tidyverse package, or when uploading data), you can switch those off with message = FALSE and/or warning = FALSE. Please knit your document and check it before submitting – to make sure that you meet the page limit, and to check that all graphs and tables have rendered as expected. While we will not be marking you on the quality of your Markdown code, if certain elements don’t appear in the pdf document, we will not be able to mark them.

# Dataset description

The dataset includes data from fictional two care homes, and a range of factors relating to each person within the homes. The key purpose of the data was to record the average physical activity of each of the individuals. All fictional participants agreed to their data being recorded. All participants had their physical activity recorded and an average amount of time for each category delivered back to the study team. All participants had the measured trial endpoint of death.

The dataset includes the following variables:

|  |  |
| --- | --- |
| **Variable** | **Description** |
| participant\_id | The study ID of the participant |
| sex | Male (M) or Female (F) |
| age\_at\_recording | The age of the participant at the time of data (in years) |
| age\_at\_death | The age of the participant at their time of death (in years) |
| physical\_activity | A physical activity variable indicating the most common type of exercise undertaken by each individual: sitting, low, light, moderate. |
| moderate\_activity | Percentage of time each individual spent undertaking activity classified as moderate. |
| BMI | Each patients BMI value |
| carehome\_id | The carehome ID for each patient (1 or 2) |

The Scottish Government is now considering creating an intervention to increase the amount of exercise undertaken by those in care homes. The Government officials have asked you to write a short report describing your analysis of the raw data, and your recommendations as to whether this would be useful. The report will be read by people who are not expert users of statistics, so they need you to provide clear explanations of any graphs, tables, and statistical results that you present. Your report is supposed to help them understand what the data are showing and decide whether an intervention should be considered for implementation in Scotland. The Scottish Government officials have asked specific question and expect to find separately answers to following questions:

1. What were the demographic and baseline characteristics of individuals in the trial? Please present one table, at least one graph and interpretation of any statistical tests you conduct. [20 marks]
2. Is there an association between the care home that an individual was in and the type of physical activity undertaken? Additionally, was a longer amount of moderate activity observed depending on which care home the individual resided in? Make sure to explain the analysis you picked, why you picked it, and what the results show. [35 marks]
3. Does the data provide any evidence that those who are more physically active live longer? Provide a formal recommendation on whether an intervention should be provided based on these results, as part of your answer. [20 marks]
4. It is thought that BMI tends to decrease as the proportion of moderate activity (undertaken) increases? Please use appropriate visualisation and a linear regression model to test this hypothesis and to quantify the relationship. [25 marks]

IMPORTANT: Please ensure your R code is included throughout your report.

# Grading criteria

Your answer to each question will be marked separately. The marks listed next to each question indicate its weight towards the final mark. In each question, we will be looking for the following:

* Statistical analyses that are appropriate to the questions
* Statistical analyses that are clearly described and explained
* Appropriate, readable and well-labelled graphs (where appropriate)
* Tidy presentation and lay-friendly language

**Additional guidance**

1. Your written assignment should be submitted as a single pdf file, generated from R Markdown. It should have a header which states the following information only: the assignment title and your exam number (starting with B, for example “B12345”). Your name should not appear anywhere in your assignment. If you do not know your exam number, please email pgt.dshsc@ed.ac.uk.
2. When you save your assignment, please give it your exam number as its name - for example “B12345.pdf”. Please do not give it any other title.
3. Assignments must be submitted no later than the specified date/time. Marks will be deducted for assignments which are handed in late at a rate of 5% per day (except where there are acceptable extenuating circumstances, and the student has submitted a special circumstances form in advance). This applies for up to seven days, after which a mark of zero will be given.
4. Markers will only look at the first 15 pages of your submission. Assignments longer than this will likely lose marks, so please do stay within the page limit.
5. Please submit your assignment electronically to the Turnitin drop-box on the course Learn page.
6. You are not expected to include any references to scientific literature – this is solely a statistics assignment.