

Assignment 2 - Data Programming with R

Instructions

This assignment is due on Sunday 20th November 2022 at 11:59pm.

- You should submit it to the ‘Assignment 2’ assignment object in Brightspace.
- You should submit two files only:
 1. Rmd file detailing the commented code you used to obtain your answers.
 2. final document in pdf format which should contain answers to the questions below.
{If you created an HTML file, please convert it to pdf. You can use Google Chrome: File > Print > Destination [Change...] > select Save as PDF. If you have created a word document, please convert it to pdf by saving it as a .pdf file}
- You may submit it multiple times before the deadline, but only the last version will be marked.
- There is a maximum of 19 marks for this assignment. This assignment is worth 19% of your final grade.
- The marks available for each question are shown in brackets.
- Late submissions will score 0, unless a “Late Submission of Coursework” form is submitted.
- You may have to discover and learn some new functions. Use `help()` and `help.search()` to find what you need.
- Some tips on using R Markdown are given at the end of this document.
- Complete your assignment using R Markdown, check that all the output and code are correctly and nicely shown in your final document. Knit your document frequently to fix errors. Once completed, submit the Rmd file and the resulting pdf or word document which shows all your code. **[0.5 marks]**

Plagiarism

While you are encouraged to ask about the module material, this assignment should be completed individually. Any student who plagiarises will receive a 0 mark. If you are unsure whether a question about the project would be considered as plagiarism, please email the question to the lecturer rather than posting on the discussion forums. The UCD Plagiarism Policy applies to all students. This can be consulted at the following *link*.

Description of the Dataset

The data file `s50_1995.txt` contains information on substance use and sporting behaviour for a cohort of 50 pupils aged 13 in 1995 in a school in the West of Scotland.

alcohol - Alcohol consumption: 1 (not), 2 (once or twice a year), 3 (once a month), 4 (once a week) and 5 (more than once a week).

drugs - Cannabis use: 1 (not), 2 (tried once), 3 (occasional) and 4 (regular).

smoke - Smoking status: 1 (not), 2 (occasional) and 3 (regular, i.e. more than once per week).

sport - Sport participation: 1 (not regular) and 2 (regular)

Questions

1. Load in the data. Convert each column to an ordered factor with appropriate labels [Hint: look at the arguments of the function `factor`, in particular levels and labels]. Display the structure of the dataset. [2.5 marks]
2. Using base R, create two suitable graphs, with labels, colours etc., one illustrating the variable `smoke` and the other illustrating the variable `sport`. Put the two plots next to each other on the same page. Comment on the resulting plots. [3 marks]
3. Produce some code to answer the following questions:
 - What is the proportion of pupils who smoke at least occasionally? [1.5 marks]
 - What is the proportion of pupils who regularly practiced sport and smoke at least occasionally? [1.5 marks]
4. We would like to be able to summarise such data sets as new data arrive. For this reason, we want to turn the object containing the data into an S3 class called `s50survey` and write a summary method that will show the **proportion of students for every level of each variable**. Test your function on the `s50_1995.txt` data. [5 marks]
5. What is the proportion of pupils who did not use cannabis? [1 marks]
6. Follow up data on the same students has been collected also in 1997. Read in the file `s50_1997.txt`, convert each column to an ordered factor, and assign the class `s50survey` to this dataset as well. Test the summary S3 method on this new dataset. [3 marks]
7. Did the proportion of students practising sport regularly increased or decreased with respect to the 1995 data? [1 mark]

Tips for R Markdown

- Ensure that you use R markdown to its full potential - there should be some/more free-flowing text outside code blocks and headings, in order to have a more comprehensive and readable report.
- It's important to learn how to use sentences and text in markdown files, so your knitted document is not just headings and code and code output
- Be aware that a common error is to give the same label to two different code chunks!

```
```{r cars}
summary(cars)
```
```

```
```{r cars}
plot(cars)
```
```

You can fix this by changing the label to one of them:

```
```{r cars2}
plot(cars)
```
```

- If you want to improve the appearance of your plot in your knitted document you can set up the dimension of your figure:

```
```{r, fig.height = 10, fig.width = 7, fig.align = "center"}
plot(Nile)
```
```

- In case of an error in your code, add the option `error = TRUE` into the R chunk to run the code, show the error message on the knitted file. For example:

```
```{r, error = TRUE}
x <- "a"
sum(a)
```
```

- For all the available options for the R chunk, you can see here: <https://yihui.name/knitr/options/>
- R Markdown website: <https://rmarkdown.rstudio.com/>
- R Markdown cheatsheet is available here: <https://www.rstudio.com/resources/cheatsheets/#rmarkdown>