MA22019 2025 - Tutorial Week 1

RStudio and R Markdown

Overview

This week's tutorial questions revise some fundamentals of the R programming language. You will further learn how to produce a PDF/HTML file using R Markdown. Note, you will be required to use R Markdown for the coursework.

If you are using your own laptop, you want to first run the code in "InstallPackages.R" (available from Moodle) to install all the packages essential for this course.

Tutorial Question 1 - Creating and editing an R Markdown file

R Markdown allows us to produce a document that includes all the R code, plots and accompanying text of our analysis. After completing the following tasks, you will have produced and edited your first R Markdown file:

a) Create an R Markdown file in RStudio:

   1) Go to **File** -> **New File** and then click **R Markdown**.

   2) Select any output option and provide a title and author (both can be changed later). Click **Continue** at the bottom of the window. This should load an R Markdown file with some material already added to it.

b) Save the file in a directory on your OneDrive. Then click the **Knit** button (the ball of yarn and needle) to compile the document into a PDF/Word/HTML document. After a few seconds, you should see a pop-up window with the produced document. If you are using your own laptop, you may first need to install a version of LaTeX using the following code:

```
install.packages( "tinytex" )
tinytex::install_tinytex()
```

c) Have a look at the document "Editing the R Markdown file" on Moodle. Then edit, and ultimately knit, your R Markdown file such that it includes

   1) A chunk of R code which samples 10,000 observations from a Normal($\mu = 5$, $\sigma^2 = 4$) distribution and stores the samples in a vector $\mathbf{y}$.

   2) A second chunk of R code which extracts the minimum and maximum value in $\mathbf{y}$. You may want to consider using the functions min() and max().

3) A third chunk of R code which creates a histogram of the data in y. Your plot should be centred and be 60% of the text width.

4) Some text before each R chunk to describe what the different chunks of R code are doing.

Tutorial Question 2 - Tuition fees in the United States

We are interested in exploring how tuition fees for private colleges have evolved across the U.S. between 2013 and 2020. For the analysis we will consider data collected by the National Center of Education Statistics (NCES).

The file "Private College Costs.csv" gives per U.S. state (except Wyoming) the average annual tuition fee (in U.S. Dollars) charged to full-time students enrolled in private colleges located in the state for 2013–2020. We can load the data file into R using

```
Fees <- read.csv( "Private College Costs.csv" )
```

Note, the data file needs to be in the same folder as the R Markdown file. Use the data to address the following questions:

a) Create a plot which visualizes the recorded average fee for the years 2013–2020 for the state of Alabama. What can we conclude from the plot?

b) For each state, calculate the relative change in tuition fees when comparing the value for the years 2013 and 2020. Provide a histogram which visualizes the calculated values.

c) Which state has recorded the highest relative increase in average tuition fees for private universities when comparing 2013 and 2020?

d) Discuss whether the data may be used to derive the average increase in tuition fees across all private US colleges from 2013 to 2020.