Intro to R Programming: Class Test 1

During the test you are not allowed to talk to or otherwise communicate with other students (email, instant messaging, etc.), or access the internet/course material. The only material you can use is the R reference manual provided as well as the R help within RStudio.

Only at the end of the test when specified by the invigilators, please log on to Moodle and upload your script by clicking on the link *Upload your class test script*. Please upload one R script only including your code and comments. You do not need to upload R output. Please do not upload Word documents or Zip files. The name of the file submitted **must be** your student ID number.

Please make sure you regularly save your R script, just in case RStudio crashes.

You have to attempt all questions within the time of one hour

Task 1

The file `cats.csv` contains data on the body and brain weights of cats. It has three variables.

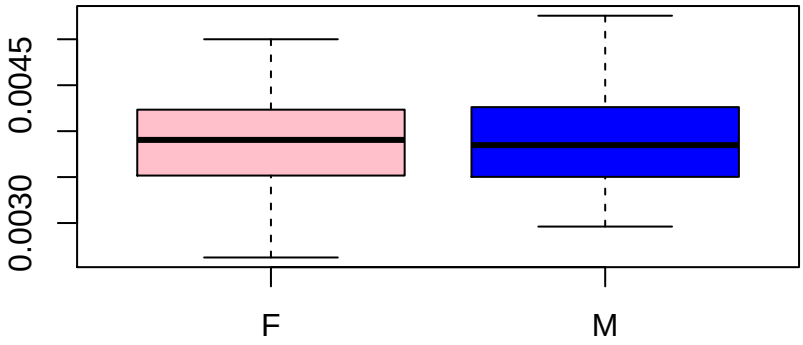
Sex	Gender of the cat (M/F)
Bwt	Weight of the cat in kg
Hwt	Weight of the hearth of the cat in g

Give the R code to perform the following tasks.

1. **[1 mark]** Read the data into R and store it in the data frame `cats`.
2. **[2 marks]** Create two data frames `cats.male` and `cats.female` which contain the data of male and female cats, respectively. Compute the number of male and female cats.
3. **[1 mark]** Compute the average hearth weight for male cats weighting more than 3kg.
4. **[1 mark]** Identify the gender and the hearth weight of the cat with the largest body weight.
5. **[1 mark]** Create a new variable `Hearth.percent` which contains the relative weight of the hearth, i.e.

$$\texttt{Hearth.percent} = \frac{\texttt{Hwt}}{\texttt{Bwt} \times 1000}$$

6. **[2 marks]** Create boxplots of `hearth.percent` by cat's gender. Colour the boxplot for female in pink and the boxplot for male in blue. Your plot should look similar to the one below.



7. **[2 marks]** Create a scatterplot of `Bwt` (on the x-axis) against `Hwt` (on the y-axis). Use a different symbol to denote the gender of the cat. The label of the x-axis should be `Body weight` and the label of the y-axis `Hearth weight`.
8. **[2 marks]** Adapt the code from part 7 to colour the observations for female cats in pink and for male cats in blue. Add a legend to your plot.

9. **[1 mark]** A linear regression fitted to the dataset `cats` gives the following estimated regression equation:

$$\texttt{Hwt} = -0.36 + 4.03\texttt{Bwt}.$$

Add the regression line to the plot.

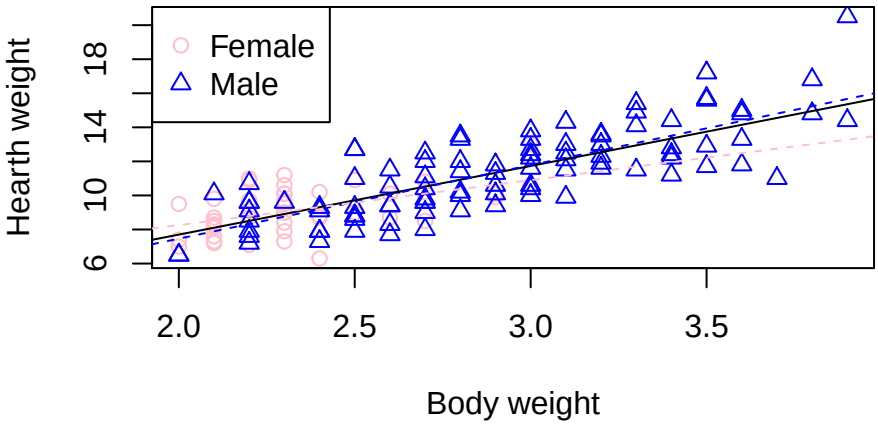
10. **[2 marks]** A linear regression fitted to `male.cats` gives the estimated regression:

$$\texttt{Hwt} = -1.18 + 4.32\texttt{Bwt},$$

whilst for `female.cats` is

$$\texttt{Hwt} = 2.98 + 2.64\texttt{Bwt}.$$

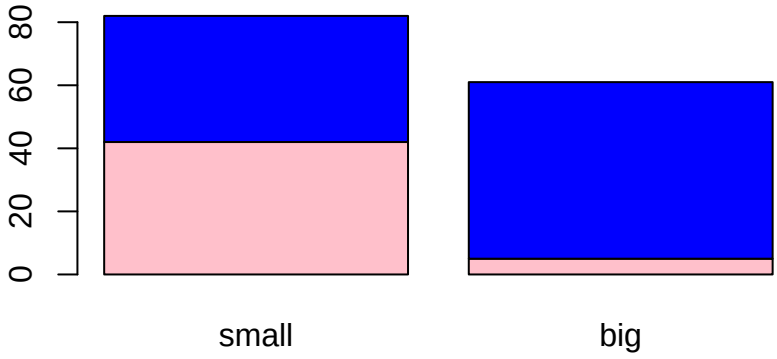
Add these two regression lines to the plot from part 9. Both lines should be dashed. The line for male should be blue and the line for female should be pink. Your plot should look similar to the one below.



11. **[2 marks]** Find the median of the variable `Bwt`. For an odd number of observations the median is the value separating the higher half from the lower half of the data. Do not use the function `median`.

12. **[1 mark]** Create a new variable called `Bwt.discrete` which splits `Bwt` into two groups: `small` corresponding to body weights less than the median computed in part 11, and `big` otherwise.

13. **[2 marks]** Create a bar plot illustrating the number of cats of small and big body weight by gender. Colour the bar for female in pink and the one for male in blue. Your plot should look similar to the one below.



Task 2

Consider the matrix $\mathbf{A}$ and the vector $\mathbf{b}$ defined as

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 3 \\ 0 & 4 & 2 & 0 \\ 0 & 2 & 17 & 8 \\ 3 & 0 & 8 & 49 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 4 \\ 5 \end{pmatrix}$$

1. **[1 mark]** Give the code to construct the matrix $\mathbf{A}$.
2. **[1 mark]** Give the code to construct the vector $\mathbf{b}$. Do not manually assign the entries of the vector.
3. **[1 mark]** Solve the linear system of equations $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{x}$ is the unknown.
4. **[3 marks]** Use the function `eigen` to compute the eigenvalues and the eigenvectors of $\mathbf{A}$. Letting $\mathbf{\Gamma}$ be the matrix of eigenvectors and $\mathbf{\Lambda}$ be the square matrix with eigenvalues on the diagonal, verify that $\mathbf{A} = \mathbf{\Gamma\Lambda\Gamma}^{-1}$.

Task 3

Two friends want to meet up for coffee. Both arrive at the café at a random time between 9 and 11am, independently of each other, with each arrival time between 9 and 11am being equally likely.

Suppose each of them stays at the café for 20 minutes and then leaves, irrespective of whether they have actually met their friend or not.

Use simulation to find the probability that the friends manage to meet each other. Simulate $n = 1000$ such scenarios and find for which proportion the two friends manage to meet up. To do this follow the steps below:

1. **[1 mark]** Create a 10000×2 data frame `arrival` with columns representing the arrival times of each friend *Hint: the function* `runif(n,a,b)` *simulates equally likely numbers between* `a` *and* `b`. *You should think of the interval between 9 and 11am as a time frame of 120 minutes.*}
2. **[2 marks]** Add the column `meet` to the data frame `arrival` which should be `TRUE` if the absolute difference between the arrival time of the two friends is less or equal to 20 mins.
3. **[1 mark]** Use the column `meet` to compute the proportion of times the friends meet.