Exercise sheet 2

Exercise 1: In crossbow contest each contestant has 4 arrows to hit the target. A total of 100 participants registered to the contest, and you are in charge of collecting the data of each contestant. Assuming each contestant is very good and hits the target 1 every 2 tries, the total number of points per contestant could be simulated as follows:

points <- sample(0:4, 100, replace = TRUE, prob = c(0.06, 0.25, 0.38, 0.25, 0.06)).

Using those data, make the following exercises

- Calculate the following

$$mu = \frac{(\sum_{i=1}^{100} points_i)}{100}, \quad a = \sum_{i=1}^{100} (points_i - mu)^2, \qquad s = \sqrt{\frac{a}{100-1}}$$

- Do you recognize to what do the above formulas correspond to? Is there a more efficient way to calculate this in R?
- How many contestants got 4 points, and how many did not get any points?
- Both women and men registered to the contest, use the function sample to generate at random the gender of the 100 contestants (do not set the parameter "prob"). Calculate the mean points for women and men. How many men and how many women hit the target 4 times?
- Within the command to generate the data, some probabilities were fixed. Can you explain how were those probabilities obtained?

Exercise 2: Create in R the following table:

Name	Gender	Biology	Physics	Maths
Andrea	Female	5	5.5	6
Jacob	Male	4.5	5.5	5.5
Camille	Female	5.5	4	3.5
Anne	Female	4	4.5	4
Samuel	Male	6	5.5	5.5
Patrick	Male	3.5	3	3.5
Sabrina	Female	5	3	5.5

- Calculate the *mean* grade by gender for each lecture.
- Calculate the *mean* grade for each student. Which student had the lowest and which one the highest *mean* grade?
- Export the table as the file *student_grades.txt*. Load this new file into a new object called *grades2*, which should be identical to *grades*. Confirm whether the two data frames are identical using the function *all()*.
- Change the column name from *Maths* to *Mathematics*. Use the command re*name()*.

Exercise 3: Download the files co2_uptake.txt, iris_1.csv, and iris_2.csv

- Load the file co2_uptake. It contains the CO2 uptake of different plants at two sites and under two treatments. Additionally, it contains the environmental concentration of CO2. Summarize the *mean* CO2 uptake per site and treatment. Then, create a subset with plants that took at least 10% of the environmental CO2, and export that table as a comma-separated file.
- Load the files iris_1.csv and iris_2.csv. Then, from iris_1.csv create a new file containing column names of your choice.