Stochastic Methods for Material Science (Winter term 2022)

Programming Project – Part 1

(Deadline for submission: January 8th 2023 at 11:59 pm)

Notice: You have to submit an R script including comments and the commands for solving the subsequent tasks. The evaluation and interpretation of the results and data analysis in R can be done via comments within the R script.

1. Task

In an investigation of the Cd contamination of trout in a river, ten trout were caught at each of two locations and their Cd content (in mg/g fresh weight) was determined. The measured values were as follows:

Location A	76.8	72.3	74.0	73.2	46.1	76.5	61.9	62.4	65.9	62.4
Location B	64.4	60.0	59.4	61.2	52.0	58.1	62.0	57.8	57.2	

a) Draw parallel box plots. What do you observe?

b) Test for the level $\alpha = 0.05$ whether the Cd contents measured at **both** locations can be regarded as realizations of a normally distributed random variable.

c) Test for the level $\alpha = 0.05$ whether the variances of the Cd contents are equal or significantly different from each other.

d) Test for the level $\alpha = 0.05$ whether the expected Cd content at location A is significantly greater than that at location B.

2. Task

The measured temperatures (in ^oC) before and 3 hours after taking a drug can be found for ten different patients in the following table:

before	38.4	39.6	39.4	40.1	39.2	38.5	39.3	39.1	38.4	39.5
after	37.6	37.9	39.1	39.4	38.6	38.9	38.7	38.7	38.9	38.7

a) Describe which sample situation (one, two, paired, multiple, ...) we have here.

b) Generate a scatter plot of the data. What do you observe?

c) Estimate the correlation between the temperature values before and after taking the drug.

d) Test for level $\alpha = 0.05$ if the correlation is significant (you can decide in one- or two-sided alternative). To this end, research (in literature or online) which test is suitable for this task and verify that its assumptions are satisfied for the data at hand.

3. Task

In the text file *coal_data.txt* you find the time intervals in days between disasters in British coal mines between 1850 and 1965.

a) Load the data into R and visualize the empirical distribution via a histogram. Also draw a kernel density estimte for the underlying distribution. What do you observe regarding the (shape of the) distribution?

b) Compute a confidence interval for the mean value of days between two disasters for confidence level of 95%.