

02403: Assignment 1

Summary statistics and the multivariate normal distribution

This is the first assignment in 02403, the main topics are summary statistics and the multivariate normal distribution.

The assignment assumes a basic understanding of programming in Python, in particular that you can write simple functions that perform some mathematical tasks and return relevant output.

Data:

A data set from measurements in Skive fjord is used to exemplify the methods you will implement in this assignment. The data set contains the following variables

year: The year of the measurement.

month: The month within the year.

chl_a: Average chlorophyll-a concentration in the month and year.

temp: Average water temperature in the month and year.

gr: Average global radiation (i.e. sunlight) in the month and year.

vmp: Water framework directive, these are different legislations that are effective (given as integers from 0 to 3).

lchl_a: natural logarithm of chl_a.

The data set is also used in the lectures to exemplify different parts of the theory.

Hand in:

The answers should be kept short and concise, and may include Python code and output, but should be handed in as a single pdf-file. Some answers involve writing Python functions and in the case it should be explained what the function do.

Questions:

- a) **Read data:** Read the data and make a very short description of the data, i.e. what period does it cover and are the missing values.
- b) **Present data:** Make boxplots of `lchla` and `temp`, and further present a scatter plot of the two, comment briefly on the result.
- c) **Summary statistics:** Make a Python function that takes a data matrix (or data frame) as input and return the columnwise averages, and the diagonal elements of $\hat{\sigma}$ and the matrix R , in the decomposition

$$S = \hat{\sigma} R \hat{\sigma}.$$

Use the function to report averages, standard deviations and correlation between the variables `lchla` and `temp`.

- d) **"Square root" of variance-covariance matrix:** Using the estimated covariance matrix from the last question, calculate a matrix $\Sigma^{\frac{1}{2}}$, such that $\Sigma = \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}T}$, check that is is correct
- e) **Simulation of multivariate normal random variables:** Using the previous question write a function that simulate reslization from a multivariate normal with some mean and variance, based on standard normal realizations

f) **Standardization of normal random variables:** Define

$$Q_i = (\mathbf{Y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}),$$

using the simulation result from the previous question find $\bar{Q}$ and S_Q^2 , and compare with what you would expect.