Statistics for Intelligent Systems 21.03.2022

Exercise 1. The file Chem.csv contains data on starting salaries for chemists, with missing data. Complete the data once using random values from the data, and a second time using the mean values. Save the data to variables A and B respectively.

Exercise 2. The file Per.csv contains the heights of baby brothers and sisters in different families with some data missing. Using one command fill the data with median values and save it as variable C.

Exercise 3. Scale the data in file Norm.csv, so it has the mean 0 and sd equal to 1.

Exercise 4. From the file Fam.csv use the function transform to change the number column to factors.

Exercise 5. Use the variable C from exercise 2 to find the range of heights of the brothers, the interquartile range for sisters and the correlation between the two variables.

Exercise 6. Use the function summary to compare the vectors A and B from exercise 1.

Exercise 7. Draw the points of form (x,x) where x is an integral from 1 to 10, using the function plot().

Exercise 8. Draw a histogram for the data in the file Chem.csv.

Exercise 9. The file Gal.csv contains the velocities of 83 observed galaxies. Draw a histogram which divides the data into 12 buckets, and describes relative frequency.

Exercise 10. Draw a density plot for the examples from exercises 8 and 9.

Exercise 11. Make a boxplot of the data in Chem.csv, omit the outliners.

Exercise 12. Draw a scaterplot for the date in file Per.csv.