Statistics and Exploratory Data Analysis

Marcin Chlebus, Ewa Cukrowska-Torzewska

Laboratory 1: Measures of location

Exercises for students

Exercise 1:

Create 1000 random sampling from normal distributions with parameters μ=2 and $\delta^2 = 0.5$.
Can you guess what will be the mean, the median and mode for this sample?
Check your guess with R functions.

Exercise 2:

Use R dataset called "anorexia" from MASS library. Calculate the average change in patients' weight for different kinds of treatments. Which type of mean will you use and why?

Exercise 3:

Use data on wages from the NLSY dataset for the US.

a) How would you summarize wage data? Use various measures such as mean, median, trimmed mean and Winsorized mean. What can you say about the wage distribution?
b) Calculate the mean, the median, quartiles, 1st and 9th deciles, as well as 5th and 95th percentiles of hours by sex and by race. Interpret the data.
c) Do the same for experience and tenure. What can you say about these two variables?

Exercise 4:

Use R dataset called "Animals" from MASS library.

- What is the mean, median and mode of body and brain weight of the 28 species?
- Are the distributions of body and brain weight symmetric? Answers by using mean, median and mode values.
- Dropping two extreme values - one from the top and one from the bottom of body weight distribution - what is the mean body weight?
- What is the mean share of brain weight in the total body weight (have in mind that body weight is in kg and brain weight in g)?
- What is the body weight below which there is 27% of the sample?