Assignment 2

December 2023

Course:	KZ4016 Kemisk DataAnalys
Tutor:	Maksim Posysoev

- For this assignment you need only NumPy and Matplotlib

- Also you need function **np.where** to select data

- Load data into Jupyter Notebook and try to create a model or models for predicting foot size based on height

  - Make scatter plot with different colors for different groups
  - Make transparent histograms for different groups
  - Write the text in the Markdown cell to explain what you see
  - Compute mean, median, and standard deviation for those distributions
  - Make two linear models and two models quadratic models
  - Draw confidence intervals for all models
  - Compute Mean Squared Error for all models (MSE)

- All plots should have named axis

- Save all plots as pdf file

- Use these models to predict your foot size. Does the result fit the confidence interval?

- Explain how to improve those models (Write a markdown cell)

- What is better: two models for different groups or one model for everyone? Prove it with MSE