# Regression Analysis
## Fall 2023, Lab 2
## Due September 22nd (Friday)

Prepare a short report with relevant output, your comments, and answers to the questions (this does not need to be exhaustive or polished, but should contain enough to show that you completed all tasks and analyses). Submit the report at cybercampus with one pdf file using R markdown

YOU ONLY NEED TO SUBMIT ONE PDF FILE INCLUDING YOUR CODE AND RESULT!

The dataset *spaghetti.txt* contains the weight (in oz) of 20 spaghetti boxes of a famous pasta brand.

- Load the dataset *spaghetti.txt* in R, using the function *read.table.*

A consumers' association would like to sue the company, affirming that the mean box weight is lower than the nominal one (16 oz). To be sure about their statement, they ask you to perform a suitable test with level 1%.

- First compute the sample mean and the standard deviation of the box weight.
- Perform the test for the consumers' association.
  - What is your null hypothesis H0?
  - What is your alternative hypothesis H1?
  - What distribution can you assume for the test statistic (type of distribution and parameters), and why?
  - Compute the test statistic.
  - Compute the test p-value.
  - Do you reject the null hypothesis, at significance level 1%?
- Compute the 99% two-sided confidence interval for the mean of the box weight.

Consider again the dataset *record.txt* that you used for Lab 1.
This dataset contains running records obtained from athletes from different countries in various types of athletics events (sprints and middle-distance).
We have data about 55 countries (observations) and 6 records (variables): 100 meters, 200 meters, 400 meters, 800 meters, 1500 meters and 3000 meters.

- Load the dataset *record.txt* in R, using the function *read.table* (remember to set `sep=' '`)
- Draw a scatterplot and compute the correlation for all pairs of variables in the dataset. Interpret the results you obtained: what can you observe about the relationship among the variables?

Consider the variables m100 and m400.

- Using the equations(not lm function or matrix equation in r), compute the least square estimators for the coefficients of a single linear regression model, with response m1500 and predictor m800. How do you interpret the slope or the regression line?
- Produce a scatter plot for m1500 vs m800 with the fitted regression line superimposed.
- Re-compute the least square estimators of beta0 and beta1 using the matrix equation.

- Re-compute the least square estimators of beta0 and beta1 using the R function *lm*, and visualize the summary of the regression.
- Which are n and p for the considered regression model?
- Compute the fitted values and the residuals, using the estimated regression line.
- Consider the $25^{th}$ and $75^{th}$ percentiles of the variable m100 on the dataset. Use the estimated regression line to estimate the fitted value of the variable m1500 at each of these two percentiles.