Exam – for 0.5 ECTS credit points

Data: A set of data collected from students at the University of Lausanne is available in the file **etubiol.csv** (courtesy of F. Schütz).

Goals: Get to know the overall structure of the data. Summarize variables numerically and graphically. Model relationships between variables.

Exam is graded as "pass" or "fail".

For a pass:

- Do all exercises, add comments to explain what you do and why
- Copy relevant output from command line into your script file as comments
- Use R functions to import data and export graphics (not GUI buttons)

- Submit analysis by e-mail to the trainer, at the latest 1 week after the course.
- Subject line: Exam FSWR
- Please bundle your script and graphics in a .zip file.

Exam - part I

Important note on variable names in the data set:

height_M: height of mother
height_F: height of father
n_siblings_F: number of female siblings
n_siblings_M: number of male siblings

Let's explore the dataset to see what it contains.

1) Have look at the file in R text editor to get familiar with it.
2) Open a new script file in R studio, comment it and save it.
3) Read the file, assign it to object "df". Examine "df".
   a) How many observations and variables does the dataset have?
   b) What is the structure of the dataset?
   c) What are the names and types of the variables?
   d) Get the summary statistics of "df".
4) Calculate the BMI of each person and add an extra variable "bmi" to a new data frame "df_bmi". Check that df_bmi contains a new column "bmi". Export df_bmi to a csv file. (Google the BMI formula). Typical BMI values should range between 15 and 35. Check if this is the case here.
5) Make a global scatter plot of all pairs of variables in the dataset.

Exam – part II

Investigate the following questions with numerical summaries and visual elements, as well as statistical analysis when possible:

1. Is there a difference in bmi means between males and females?
2. Is there a difference in bmi means between smokers and non-smokers?
3. How strong is the linear (Pearson) correlation between shoe size and height? Is it significant?
4. If you model a linear relationship, how much does the shoe size increase per added cm of height? Is the change significant? What if you do this for males and females separately?
5. Come up with a question of your own that includes one or more variable(s) of your choosing from the etubiol data set and investigate it.

- Make plots as seen in the course to try to give visualization-based answers to these questions. You can use multiple plots per question.
- Test your hypotheses using tests and modeling techniques from the course, based on the type of variables you have. Visually assess normality where appropriate.