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

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).
5) Make a global scatter plot of all pairs of variables in the dataset.

Exam, Part II

Assume that you have been given the following questions.

1. Is there a significant difference in bmi means between males and females?
2. Is there a significant 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 for hypothesis testing of your own that includes one or more variable(s) of your choosing from the etubiol data set.

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

Note on variable names in the data set:

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

Exam – proposed workflow

1. Specify your biological question and your experimental design very clearly, then collect your data.
2. Save your data into a csv format in a dedicated folder.
3. Start up R , open a new script file and save it where you save your data. Don't forget to annotate it and save it regularly.
4. Import your data into R. Check everything in your data. Make sure it is what you expect it to be.
5. Explore your data, first with R's plotting functions. Make an hypothesis. Try to guess the answer that your statistical test should give you.
6. Perform your test to confirm your answer.
7. Communicate your findings.
8. Make sure your files (data, scripts, figures, reports) are well organized in your folder.