Exercise 1

Let's look at the concepts of the cumulative distribution function (cdf) and the empirical cumulative distribution function (ecdf) on the example of a discrete random variable.

(a) Roll a die 10 times or equivalently run in R:

```
R> sample(1:6, 10, replace = TRUE)
```

Compute the ecdf for your data. Plot the ecdf in R.

(b) Now roll the die 10'000 times. Compare the ecdf to the cdf.

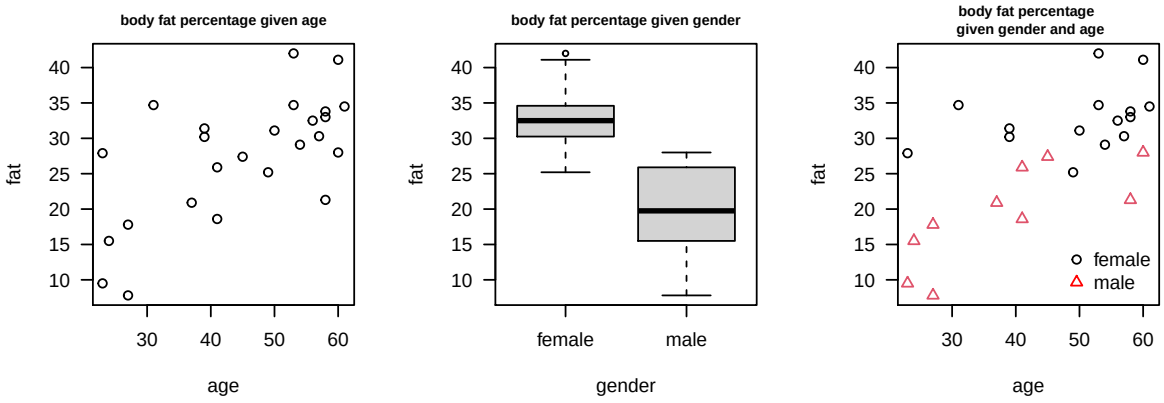
(c) Consider the sum of two dice rolls. Give the probability mass function and the cdf.

Exercise 2

The `agefat` data set from package **HSAUR2** contains 25 observations with the following three variables:

- `age`: the age of the subject
- `fat`: the body fat percentage.
- `gender`: a factor with levels female and male

What hypothesis can you state based on the three following plots? What do you consider as pros and cons of the graphical representations?



Exercise 3

The `CHFLS` data set from package **HSAUR2** is a subset of the Chinese Health and Family Life Survey. Each row in the data frame contains information for the wife (variables starting with `R`) and her husband (variables with `A`). Use graphical methods to investigate

(a) the conditional distribution of the wife's income (`R_income`) given her education (`R_edu`).

(b) the conditional distribution of the wife's self-reported happiness given her education (`R_edu`).

(c) the conditional distribution of the wife's self-reported happiness (`R_happy`) given her income (`R_income`).

(d) the conditional distribution of the wife's age (`R_age`) given her income (`R_income`).

(e) the conditional joint distribution of the couples's income (`R_income` and `A_income`) given the husband's education (`A_edu`).

luisa.barbanti@uzh.ch
lucasheinrich.kook@uzh.ch

Results due 2020-09-22