Exercises for the course

"An introduction to R"

Sheet 03

Exercise 10: *Skewness is a measure of the asymmetry of a probability distribution. The skewness is negative if the mass of the distribution density concentrates towards the right and is positive if the mass of the distribution density concentrates towards the left. This exercise is to practice working with vectors; the concept of skewness itself is not too important.*
Let the mean of the binomial distribution with 60 trials and success probability $\frac{1}{6}$ be denoted as `m`. Recall that `m` is not related to the sample mean and that you can not calculate `m` using the command `mean()`. Let the third centralized moment be denoted as

$$m3 <- \sum_{i=0}^{60} (i-m)^3 \cdot \text{dbinom}(i, 60, 1/6).$$

Calculate the skewness of the binomial distribution at hand as `skew` $<-$ $\frac{m3}{\sigma^3}$ where

$$\sigma^2 <- \sum_{i=0}^{60} (i-m)^2 * \text{dbinom(i}, 60, 1/6)$$

is the variance of the binomial distribution at hand. Check your result with `skew`$= 0.2\ldots$ Next repeat all calculations with $\frac{1}{6}$ replaced by $\frac{5}{6}$. Then plot the densities `dbinom(,60,1/6)` and `dbinom(,60,5/6)` and explain the signs of the two skewness values. (4 points)

Exercise 11: *Data frames are the typical R representation of data sets. Here we create a data frame "by hand" to become familiar with data frames.*
Use the command `data.frame()` to create a data frame `students` with the following entries

name	degree	mat.nr
Leonie	Master	1111
Luka	Master	1112
Lea	Bachelor	1113
Leon	Bachelor	1114
Laura	Bachelor	1115
Luis	Bachelor	1116

- Get an overview of `students` with the commands `names()`, `str()` and `summary()`.
- Which command returns the fifth element of the vector 'mat.nr'?
- Check existence of the variable `degree` by entering it into the R command line. Now copy `students` into the search path with the command `attach()`. Check again whether `degree` is a known variable.
- Define a new data frame named 'ba.students' which consists of all students with degree `Bachelor` (without using the command `data.frame()`). As all students in `ba.students` have degree `Bachelor` the variable `degree` is not needed in `ba.students`.
- Write the data frame `students` into the file 'studentsfile.txt'. Then read the data frame from this file into the new variable `students2`. If you used the right command, then `students` and `students2` are identical. Check this using the command `all()`.

- Laura and Luis just received their Masters degree. Use `fix()` or `edit()` to update `students` accordingly. Note that these two commands do not work on all systems.
- We wish to change 'degree' into 'deg' to save typing work. Use the command `names()` to accomplish this change. You might need to consult the help page `?names` to find out how to do this. (5 points)

Exercise 12: *The role of the heartbeat in the relation between mother and infant.*
Download the file 'heartbeats.txt' from the course web page (the file 'heartbeats.readme.txt' on the web page contains more information on the data). Read the data from 'heartbeats.txt' into a data frame named `heartbeats`. Get an overview of the structure of the data frame. Attach `heartbeats` to save typing work. The group of newborns which had the heartbeat treatment is consequently referred to as 'heartbeat group'. Calculate

- the mean of the increase in weight of all newborns,
- the mean of the weight increase of the control group,
- the mean of the weight increase in the heartbeat group,
- the mean of the weight increase in each weight class of the control group,
- the mean of the weight increase in each weight class of the heartbeat group,
- the standard deviation of the weight increase of all newborns.

Are the means in the heartbeat group higher than in the control group?

Newborns often loose weight in the first days. Let's have a closer look at that with the data at hand. Define a new vector `negwghtincr` as the elements of `wghtincr` which are ≤ 0.

- Boxplot `negwghtincr` both for the heartbeat group and for the control group.
- Plot an histogram of `negwghtincr` of the first weight class of the heartbeat group.
- Plot an histogram of `negwghtincr` of the first weight class of the control group. (5 points)

Exercise 13: Before reading data, you need to carefully prepare the data file. Consider the following self-generated data which is supposed to be obtained from measuring mice of the two species 'mus musculus' and 'mus spretus'. The following table contains the tail length measured in centimeter.

mus musculus	mus spretus
11.3	5.4
10.6	5.6
12.1	6.1
9.5	4.9
11.7	5.3
10.8	4.6

Prepare a data file as described in Subsection 3.5 of the script using Excel or a text editor. Then read the data file into a data frame.

- Calculate the mean tail length of all mice. (If you have problems with this, then check that you prepared the data file 'the right way').
- Calculate the mean tail length of all mice of the species mus musculus.
- Calculate the mean tail length of all mice of the species mus spretus. (4 points)

Exercise 14:
Create the following matrices:

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Hints: The first matrix is a submatrix of `diag(1,5,4)`. The antidiagonal matrix is can be obtained from the diagonal matrix `diag(4)`. This exercise is more difficult than other exercises.

(3 points)