SDA 2019 — Assignment 2

For these exercises you can use the *R*-functions `summary`, `range` and `IQR`. For QQ-plots you may use `qqnorm`, as well as the local functions `qqt`, `qqlnorm`, `qqchisq`, `qqlogis`, `qqexp`, `qqunif` and `qqcauchy` which can be found on the Canvas page (`functions_Ch3.txt`)[1].

Note: to indicate a normal distribution with expectation 2 and *variance* 25, we use the notation $\mathcal{N}(2{,}25)$, whereas *R* uses the parameters `mean=2`, and `sd=5` to indicate this normal distribution.

Make a concise report of your answers in *one single PDF file*, with only *relevant R* code *in an appendix*. It is important to make clear in your answers how you have solved the questions. Graphs should look neat (label the axes, give titles, use correct dimensions etc.). Multiple graphs can be put into one figure using the command `par(mfrow=c(k,l))`, see `help(par)`. Sometimes there might be additional information on what exactly has to be handed in. **Read the file AssignmentFormat.pdf on Canvas carefully**.

Exercise 2.1

a. Make plots of the quantile functions (that is, make 'true' QQ-plots as in Figure 3.4 of the syllabus or Slide "Quantiles of F and $F_{a,b}$ (2)" of Lecture 2) of the following pairs of distributions:

   I. Cauchy – standard normal.
   II. standard normal – t_4.
   III. uniform(2,4) – log normal with meanlog = 0, sdlog = 0.5.

   Comment on each plot on the heaviness of the tails of the two distributions. The tails of a distribution can be seen as the relative height of its density $f(x)$ for $x \to \pm\infty$. For example, the tails of a normal distribution are heavier than the tails of a uniform distribution (which vanish for finite x), cf. Figure 3.4 in the syllabus. You may create for yourself some density plots to get a better idea of the tail behaviour of the different distributions (without handing these in).

   Note: for this exercise you should not generate random samples. Instead, use the true quantile functions for both the x-axis and the y-axis (for example, the R-function `qnorm` *can be used for computation of the quantiles of a normal distribution). For plotting the function* `plot` *should be used, not the function* `qqplot`.

The functions `qqnorm`, `qqt`, `qqlnorm`, etc., can be used to make QQ-plots for the location-scale families of the normal, t, lognormal distributions, etc., respectively. The argument `df` in `qqt` and `qqchisq` is used to set the number of degrees of freedom in the t-distribution and the χ^2-distribution.

b. To get an idea what QQ-plots may look like, test one or more of these functions by generating one or more samples from a distribution and making QQ-plots on your screen. None of these plots need to be handed in.

[1] These functions can be loaded in exactly the same way as the code from the file `temp.txt`, see Exercise 2.3

c. Investigate the data in `sample2019.txt` with the given functions for making QQ-plots and find an appropriate distribution for this data set. Apart from giving a proper location-scale family (e.g., "normally distributed"), also give values for the location and scale parameters. (e.g., "normally distributed with location 2 and scale 5", or "$\mathcal{N}(2,25)$ distributed").

   Hints:
   1. The location and scale parameters in your location-scale family $F_{a,b}$ are $a \in \mathbb{R}$ and $b > 0$, respectively, and not necessarily equal to the parameter(s) of the underlying parametric distribution family.
   2. Sometimes it is useful to apply a transformation to the data.

Hand in: your plots and comments for part a and plots of relevant graphs of part c, as well as a motivation for your trials and your final conclusion for part c.

Exercise 2.2 The file `temp.txt` contains the mean temperatures in July in 4 Dutch cities: De Bilt, Eelde, Vlissingen and Ter Beek, for the years 1964-1983. You can load these data by performing the following steps:

i) save the file to some directory called `thedirectory` with path `path`[2],

ii) set the working directory to `thedirectory` by using the command `setwd("path")` (instead of this, you obviously need to fill in the correct path on your computer!),

iii) finally run the code in the file `temp.txt` using the command `source("temp.txt")`.

The `i`th component of a list called `listname` can be extracted using `listname[[i]]`. This exercise only concerns the data of Eelde and Ter Beek.

a. Check whether the distributions of the Eelde and Ter Beek temperatures have approximately the same shape.

b. Find suitable distributions for the mean July temperature in Eelde and Ter Beek (see 2.1.c), including the correct location and scale parameters.

c. Make a histogram of the differences in mean July temperature between Eelde and Ter Beek. Does one city clearly have a colder July than the other?

Hand in: the answers to the questions and all plots required to give these answers.

[2] For Windows the path is usually `C:/.../thedirectory`, for Mac the path is usually `/Users/.../thedirectory`.