SDA 2019 — Assignment 7

For these exercises you can use the functions[1] `mhuber` and `mlogis` which compute the Huber and logistic M-estimators (see syllabus). Use the function `mad` to compute the MAD estimator for spread.

Make a concise report of *all* your answers in *one single PDF file*, with only *relevant R* code *in an appendix.* It is important to make clear in your answers <u>how</u> 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 7.1 This exercise is about assessing the quality of different estimators of location.

a. Generate 1000 times a sample of size 100 from a t-distribution with 4 degrees of freedom. (You may use the command `rt`.) Compute for each of the 1000 samples the 0, 10, 20, 30, 40 % trimmed means, the median, the logistic M-estimator and the Huber M-estimator (the last two with `scale=1`).

b. Which estimator gives in your case the best estimates? Explain your findings based on your intuition *and* based on theory from the syllabus.

Explanation: for each of the eight estimators, report at least the sample means and variances of the 1000 realized values and also an estimate of a good measure for the quality of the estimates, like the mean squared error[2] *or the mean absolute deviation*[3] *(not the* **median** *absolute deviation). When you generate more than 1000 samples, you may obtain more accurate results, which is even better.*

c. Repeat part a, but now generate samples from the Cauchy distribution distribution instead of the t-distribution. (You may use the command `rcauchy`.)

d. Which estimator gives in your case the best estimates? Explain your findings based on your intuition *and* based on theory from the syllabus.

Hand in: summary of results of parts a and c (see *Explanation*) and your answers to parts b and d.

Find Exercise 7.2 on the next page!

[1] You can find these functions in the file `functions_Ch5.txt`.

[2] mean squared error: $E[(\hat{\theta}-\theta)^2]$ for an estimator $\hat{\theta}$ of θ. If the value of θ is known, the mean squared error can be estimated based on B realizations $\hat{\theta}_1, \ldots, \hat{\theta}_B$ of the estimator via $\frac{1}{B}\sum_{i=1}^{B}(\hat{\theta}_i-\theta)^2$.

[3] mean absolute deviation: $E[|\hat{\theta}-\theta|]$. It can be estimated via $\frac{1}{B}\sum_{i=1}^{B}|\hat{\theta}_i-\theta|$.

Exercise 7.2 With *cloud seeding* a small airplane is used to add a particular substance to clouds in order to change the precipitation properties[4]. In a cloud seeding experiment in 1975 precipitation values of two groups of clouds, *seeded* and *unseeded*, were compared. The precipitation data of this experiment are contained in the file `clouds.txt`.

a. Investigate the two data sets, `seeded` and `unseeded`, graphically as well as numerically to obtain a global impression of their (dis)similarities. Compute, among other things, the sample means. (You do not need to hand in these results.)

b. The sample standard deviation is often used as a measure for the accuracy of the measurements. Determine for both series the corresponding estimate of the accuracy.

c. Determine bootstrap estimates for the standard deviations of the estimators of accuracy used in part b.

d. Deduce from the answers in b and c and supported by reasonable confidence intervals whether one of the types of clouds shows more spread in the precipitation values than the other.

e. Repeat parts b and c, now using the robust estimator for spread MAD as measure for the accuracy of the measurements.

f. Which estimator for the accuracy do you prefer for these data, the sample standard deviation or the MAD? Explain how you reached your conclusion.

Hand in: Results of parts b, c, and e, and your answers to parts d and f.

[4]For more on that see, e.g. `http://en.wikipedia.org/wiki/Cloud_seeding`.