Final Quiz of the Statistical Inference course 2020/2021 Part 2

Olga Klopp

This part of the quiz must be done using R. A report with the command scripts, results, plots and your comments must be submitted on Moodle in pdf, html or word format.

In this exercise, we analyze the ozone dataset available on Moodle. This dataset comes from the Applied Mathematics Laboratory of Agrocampus Ouest which contains 112 data collected in Rennes during the summer of 2001. It contains the following 14 variables:

- `obs` : month-day ;
- `maxO3` : maximum ozone concentration observed during the day;
- `T9, T12, T15` : temperature observed at 9 a.m., 12 p.m. and 3 p.m ;
- `Ne9, 12, Ne15` : cloudiness observed at 9 a.m., 12 p.m. and 3 p.m.;
- `Vx9, Vx12, Vx15` : east-west component of the wind at 9 a.m., 12 p.m. and 3 p.m. ;
- `maxO3y` : maximum ozone concentration observed the day before ;
- `wind` : wind direction at 12 o'clock ;
- `rain` : occurrence or not of precipitation.

1) **Execute the command `getwd()`.**

2) **Import data ozone.txt**

We are interested in the variable `maxO3`: maximum ozone concentration observed during the day. *Attention in the name of the variable "O" is the capital letter O and not zero*

3) (0.5 Pts) **What are the values corresponding to the first Quantile, the Median and the third Quantile?**

4) (0.5 Pts) **Estimate the mean and standard deviation of the variable `maxO3`.**

5) **Execute the command `getwd()`**

6) (2 Pts) **Plot the boxplot of the maximum ozone concentration observed the day before. If your student number is even the boxplot must be in red. If it is odd the boxplot must be in blue. Add your student number as a caption. Does the distribution seems to be symmetric? Explain.**

7) (2 Pts) **It is proposed to model this variable by a Gaussian distribution. Suggest a "visual" method to check if this hypothesis is plausible. Set up this method and comment on your results.**