Experimental Design and Data Analysis: Assignment 2

This assignment consists of 4 exercises. Throughout this assignment tests should be performed using a level of 0.05.

EXERCISE 1

Test whether the data in `dataexp` stems from the standard exponential distribution by performing a bootstrap test using the test statistic
$mean(X_1, \ldots, X_N) - median(X_1, \ldots, X_N)$.
You can generate a sample from the standard exponential distribution using `rexp(n)` with `n` the sample size you wish.

EXERCISE 2

In 1879 Michelson performed 100 experiments to determine the speed of light and in 1882 another 23 experiments. The measurements minus 299000 are given in the file `light1879` and `light1882`. Please answer the following questions based on the numbers in the file, i.e. without adding 299000.

1. Make a histogram of both data sets.

2. Determine a confidence interval for the population mean of `light1879`.

3. Determine a confidence interval for the population mean of `light1882`.

4. Determine a confidence interval for the population median of `light1879`.

5. Determine a confidence interval for the population median of `light1882`.

6. Comment on the four intervals found (i.e. compare length, values etc.)

Background: Michelson used the method of the French physicist Foucault. Light bounces from a fast rotating mirror to a fixed mirror at a distance and back to the rotating mirror. The speed of the light is calculated from the measured distance between the mirrors and the deflection angle of the emitted and received light on the rotating mirror. In the first series of experiments the distance between the mirrors was 600 m and in the second series 3721 m. Theoretically the observations in the second series should be 24 smaller than those in the first series.

EXERCISE 3

The file `klm` contains the delivery durations (in days) of aircraft parts delivered by Boeing to KLM (use `scan` to read this file). The maximum delivery duration of these parts is 70 days.

1. Test (using an appropriate test) the null hypothesis that the median duration is smaller or equal to 35 days against the alternative hypothesis that this median is greater than 35 days at level 0.05. Motivate your choice of test.

2. KLM is willing to accept that (on average over a long period) at most 10% of the parts arrives after the maximum delivery period of 70 days. Design a test analoguously to the sign test to check whether this criterium is met. Perform this test on the KLM data, at level 0.05.

EXERCISE 4
(For this exercise you need to study the slides of lecture 4)
To study the effect of energy drink a sample of 24 high school pupils were randomized to drinking either a softdrink or an energy drink after running for 60 meters. After half an hour they were asked to run again. For both sprints they were asked to sprint as fast they could, and the sprinting time was measured. The data is given in the file `run.txt`.

1. Study the data and make a few graphical representations.

2. Test separately, for both the softdrink and the energy drink conditions, whether there is a difference in speed in the two running tasks.

3. For each pupil compute the time difference between the two running tasks. Test whether these time differences are effected by the type of drink.

4. Can you think of a plausible objection to the design of the experiment if the main aim was to test whether drinking the energy drink speeds up the running?

5. Is there a similar objection to the design relative to the analysis under 3)?

6. The vector of differences in 3) has 24 elements. Which distributional assumption on these differences is needed for the analysis in 3)? How would you transform this vector into 24 residuals to investigate this assumption in QQ-plots? Make this QQ-plot(s).