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, unless otherwise specified.

EXERCISE 1

A telecommunication company has entered the market for mobile phones in a new country. The companys marketing manager conducts a survey of 200 new subscribers for mobile phones. The data of her survey are in the data set `telephone.txt`, which contains the first months bills.

1. Test whether the data in `telephone.txt` stems from the exponential distribution Exp(λ) with $\lambda = 0.035$, by performing a bootstrap test using the test statistic $T(X_1, \ldots, X_N) = \text{median}(X_1, \ldots, X_N)$.

2. Make an appropriate plot of this data set. What information can be extracted from these data? What marketing advice would you give to the marketing manager?

EXERCISE 2

In 1849 and 1850, the French physicists Fizeau and Foucault had separately devised methods of measuring the velocity of light. Foucault's method, as refined and improved by Newcomb and Michelson, was the source of the more accurate subsequent determinations. Foucault's method consists in essence of passing light from a source off a rapidly rotating mirror to a distant fixed mirror, and back to the rotating mirror. The velocity of light is then determined by measuring the distance involved, the speed of the rotating mirror and the angular displacement of the received image from its source.

In 1879 Michelson performed 100 experiments to determine the speed of light and in 1882 another 23 experiments. The measurements in *km/sec* minus 299000 are given in the file `light1879.txt` and `light1882.txt`, respectively. The file `light.txt` contains Newcomb's measurements (made in 1882 on three different days: the first 20 measurements on day 1, the next 20 onday 2, the last 26 on day 3) of the passage time it took light to travel the distance 7.442 *km*, The coding of Newcomb's measurements is as follows: from the original times in *microseconds* measured by Newcomb first 24.8 was subtracted, after which the results were multiplied with 1000.

1. Make histograms and box plots of the data sets. What do you observe?

2. Determine confidence intervals for the speed of light in *km/sec* for all three data sets (use population means and medians).

3. Comment on the intervals found.

4. Find on the internet the currently most accurate value for the speed of light. Is it consistent with the measurements of Michelson and Newcomb?

EXERCISE 3
The file `klm.txt` contains the delivery durations (in days) of aircraft parts delivered by Boeing to KLM (you can 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 μ is smaller or equal to 32 days against the alternative hypothesis that this median is greater than 32 days. 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 analogously to the sign test to check whether this criterium is met. Perform this test on the KLM data.

EXERCISE 4
To improve rain fall in dry areas, an experiment was carried out with 52 clouds. Scientists investigated whether the addition of silver nitrate has an effect on rainfall. They chose 26 out of a sample of 52 clouds and seeded it with silver nitrate. The file `clouds.txt` contains the precipitation values (records the rainfall in feet per acre) of seeded and unseeded clouds.

1. Test whether silver nitrate has an effect by performing three tests: the two samples t-test (argue whether the data are paired or not), the Mann-Whitney test and the Kolmogorov-Smirnov test. Indicate whether these tests are applicable for our research question. Comment on your findings.

2. Repeat the same procedure on the square root of the values in `clouds.txt`. Comment on your findings.

3. Repeat the same procedure on the square root of the square root of the values in `clouds`. Comment on your findings.

Experimental Design and Data Analysis: Assignment 3

This assignment consists of 3 exercises. Throughout this assignment tests should be performed using a level of 0.05, unless otherwise specified.

EXERCISE 1
The file `peruvians.txt` contains the data of Peruvian men after migrating to a modern society (see lecture 3). The column `migration` displays the years since migration and column `wrist` contains the heart rate. Neglect the columns `chin`, `arm` and `calf` for this exercise (use e.g. `peruvians[,-c(5,6,7)]`). The meaning of the remaining columns is given by their names. In this exercise we want to investigate which variables are related to the years since migration.

1. Use `pairs` to make plots of each pair of two variables. Based on this picture, which variables do you expect to correlate (in rank) with `migration`?

2. Perform a test for each of the variables to test the rank correlation between that variable and `migration`. Give your conclusions of each test separately.

EXERCISE 2

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).

EXERCISE 3

The concentrations (in nanograms per millimeter) of plasma epinephrine were measured for 10 dogs under isofluorane, halothane, and cyclopropane anesthesia. The measurements are given in `dogs.txt`. We are interested in differences in the concentration for the different drugs.

1. Make boxplots of the 3 samples. Make QQ-plots of the 3 samples against the normal distribution (a separate plot for each of the drugs). Is it reasonable to assume that these samples were taken from normal populations?
2. Test the null hypothesis that the concentration is the same under the different drugs using normal theory. Give the estimated concentration of plasma epinephrine for each of the three anesthesia drugs.
3. Perform the Kruskal-Wallis test for the same null hypothesis. What is the conclusion here? Explain possible differences between this conclusion and the conclusion under part 2.