Experimental Design and Data Analysis: Assignment 3

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

EXERCISE 1
The file `peruvians` 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
The table `clouds` (to be read from the file `clouds` using `clouds=read.table(...)`) contains the precipitation values of seeded and unseeded clouds. This concerns two independent samples. We are interested in possible differences between the two populations.

1. Test whether the two populations are equal by performing three tests: the two samples t-test, the Mann-Whitney test and the Kolmogorov-Smirnov test. Indicate whether these tests are applicable for our research question. What is your conclusion?

2. Repeat the same procedure on the square root of the values in `clouds`. What is your conclusion here?

3. Repeat the same procedure on the square root of the square root of the values in `clouds`. What is your conclusion here?

EXERCISE 3
One of the parameters of an evolutionary algorithm is the mutation probability. In order to find the minimal value of a target function on a domain, the target function is evaluated repeatedly on sets of points in its domain. A current set of points is replaced by a new set that is formed by applying several "genetic" operators, and hopefully contains points with a smaller target value. Several of these operators have a chance character, so that the output of the algorithm can be considered random. For instance, the mutation operator exchanges the digits 0 and 1 in a binary representation of a point with a given probability.

An evoluationary algorithm for minimizing a particular function was stopped after a fixed number of iterations, and the minimum value over the last set of points was returned. The seven columns in the file `genal.txt` give the minimum values reached in 100 independent runs of the algorithm with the mutation probability set equal to 0.01, 0.02, 0.03, 0.04, 0.05, 0.06 and 0.07, respectively. The other genetic operators were fixed to standard settings.

1. Make a (single) boxplot of the 7 samples.
2. Make QQ-plots of the 7 samples against the normal distribution (a separate plot for each value of the mutation probability). Is it reasonable to assume that these samples were taken from normal populations?
3. Transform the data by the square root function (type `sqgenal=sqrt(genal)` if `genal` is the data.frame containing the data). Make QQ-plots of the 7 transformed samples against the normal distribution. Is it reasonable to assume that the samples were taken from normal populations?
4. Test the null hypothesis that the mutation probability has no influence on the genetic algorithm.
5. Which mutation probability gives the best result?
6. Make a QQ-plot of the residuals of the analysis.

EXERCISE 4

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.
2. 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?
3. 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.
4. Perform the Kruskal-Wallis test for the same null hypothesis. What is the conclusion here?
5. Explain possible differences between the conclusions under parts 3 and 4.