Exercise Session 1.1

We will use the TBAD data to familiarize with the basics of a data analysis in R through obtaining the results from the first theory session.

Before we start, download the data file (TBAD.txt) and the R script (Rcode session 1-1.R) and save them to your hard drive.

Open R and the R script. Follow the instructions in the script.

Exercise Session 1.2

Analyze the BP data, based on the questions below. The data set is a reduced table from a large scale measurement campaign for calibrating different techniques to measure the blood pressure. In this reduced version the classical Korotkoff technique used by a physician only is shown together with the patient’s sex. Prior to the campaign expert physicians rated each patient as a function of cardiovascular risks. The main question was which of the measurement devices of the BP was able to reflect the expert risk assessment the best. In this setting, the main question is whether the Korotkoff technique reflects the risk assessment properly.

1. Analyze the frequency table for the variable sex to see that both groups are well represented.
2. How are the patients distributed over the risk categories?
3. Explore the boxplot of the Korotkoff scores to check differences among sex.
4. Explore the boxplot of the Korotkoff scores over the risk groups for a visual analysis.
5. Formalize the visual conclusions of the boxplot w.r.t. sex differences by means of the appropriate statistical tests supported by descriptive statistics.
6. Based on the boxplot over the risk categories, do you agree with the risk groups assessment of the physicians? Where would you define the boundaries for a good blood pressure?
7. Do you agree that the patients do not deviate significantly from their long term mean blood pressure?
8. **Extra**. Use the TBAD data to solve the following exercises:

* make a frequency table and a bar chart for the variable ‘certification level’
* obtain the descriptive measures for the variable ‘total number of patients per month’
* make a histogram and a boxplot for the variable ‘total number of patients per month’
* obtain the descriptive measures and create a boxplot for the variable ‘years of experience’, for males and females separately
* Are the years of experience different between doctors who followed medical school in the USA or in a foreign country?
* Is the total average cost per patient per month less than $ 60 for specialty pediatrics?