Assignment
Tasks
Translated!

1. Read the dataset "leistungsdaten.dat". Get an overview of its structure. Sort the dataset by schools and classes.
    a. Convert the variable for gender into a factor, in which afterwards "0" denotes previous male test participants and "1" denotes female test participants.
    b. Which variables have missing values?

2. Which school did the most students attend? Determine the scale level of the school variables. Report (appropriate) statistical characteristics.
3. Create a clear table of the gender distribution in the three class levels. How many girls are there in the seventh grade? What percentage of all test takers does this group represent?
    a. Describe the structure of the table you created.
    b. Graphically represent the percentage frequency of boys and girls in the different grades and describe the structure of your graphic.
4. Compare the German performance of the eighth graders between the five schools in a descriptive way. Report the median and mean values for the overall sample. Create a box plot of the distribution of German performance per school and additionally (and clearly) draw in the respective mean value.
    a. One student from school 3 received a high score in German (47 points). You consider this student to be an outlier and would like to remove his score from the data set - of course after a long and extremely careful weighing. Mark the value as Missing (NA) in the record.
    b. Will the median and mean change after you remove the outlier value? If so, why? If not, why not?
5. Form a sum score for the general school performance level from the measured results in the German, Maths and English tests. Describe the distribution of these new variables using appropriate descriptive statistical parameters. Create a histogram of the distribution. Make at least two changes to this graphic which distinguishes it from the standard output.
6. Now consider the relationship between motivation to test processing and math performance. Calculate a suitable correlation measure and justify your choice. What does the coefficient mean in terms of content? In addition, create a scatter diagram for graphical illustration.
7. Did you know that R can also import spreadsheets from Excel (.csv files)? Look for a suitable command and load (and check) a dataset.