ulm university universität
uulm

R Worksheet 5: Statistical Testing

Please document your code for answering the following questions in an R script and check that your code compiles.

For all of the following exercises, use the dataset mgus2.xlsx
As you already did for Worksheet 1, import the dataset into R and make sure factors and numerical variables are appropriately defined.

Exercise 1: Testing continuous variables

a. You hypothesize that the two age groups (1) below 50 years old, and (2) over 50 years old, are significantly different concerning haemoglobin and creatinine. Describe the two variables with descriptive statistics.
b. Write down your hypotheses concerning the differences between the age groups.
c. With the appropriate test for each variable, find out if the differences between the two age groups are statistically significant. Write down your interpretation of the results.
d. You now want to compare the three age groups (1) below 60, (2) 60 to 75, (3) above 75 concerning haemoglobin and creatinine. For each variable, plot three histograms for the three age groups. Each variable should have the three histograms in one plot.
e. Use the appropriate test for comparing the three groups concerning haemoglobin and creatinine, respectively. (You may need to create a new variable in the dataset that splits for the three age groups.)

Exercise 2: Testing categorical variables

a. You want to test the hypothesis that your two age groups from 1.a. have different probabilities for developing plasma cell malignancy (variable pstat). Create a frequency table concerning both variables and run the appropriate test.
b. If you had a smaller sample with less than five patients per cell in your frequency table, what would be the appropriate test? Write down the command.

Exercise 3: Survival data

a. Previous studies suggest that the illness mostly occurs in the elderly, more specific in people over 50 years old. You explore the hypothesis that if the illness occurs in people less than 50, the long-term prognosis is worse than in people over 50. Compare the two age groups concerning follow-up time (futime) and survival (death).