ulm university universität
uulm

R Worksheet 2: Descriptive Statistics

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: Descriptive statistics for categorical variables

a. Create a frequency table for each categorical variable in the dataset. Also, create tables that display the relative frequencies. Round values to two decimal places.
b. For each variable, also create a bar plot using the relative frequencies. Ideally, display all bar plots in one plot. (Tip: par())
c. Create a contingency table that compares gender (female/male) and death (yes/no).
d. Use your result from c. to answer the question whether more men or women die.

Exercise 2: Descriptive statistics for continuous variables

a. Calculate the arithmetic mean and the median for the following vector z and interpret your result regarding the distribution of the data.

```
z <- c(4, 5, 6, 6, 7, 8, 8, 8, 9, 10, 100)
```

b. Calculate descriptive statistics for the metrical variables concerning haemoglobin (hgb) and creatinine (creat) in the dataset.
c. Us the describeBy() function from the package “psych” to calculate descriptive statistics regarding haemoglobin and creatinine but separating by gender.
d. Create a histogram for both metrical variables. Both histograms should be in one plot and the axis labels should be easy to understand.
e. Add a line indicating a normal distribution to both histograms.
f. Look at your results from b., c. and d. Do the variables follow a normal distribution?