Exercise 1 - Data collection: fundamental terms

Specify population, sample (size), statistical unit, statistical variable (feature), the variable's type (discrete, dichotomous, ...) and, if provided, the variable's realizations in the following situations.

a) Locations of lightning strikes in Germany for 2020 are collected in order to visualize them.

b) A Student asks his two seatmates whether they like their new teacher or not in order to get an impression of the new teacher's popularity in the class. One likes the new teacher, the other does not.

c) Hourly wind direction data $w_i \in [0, 360), i \in \{1, ..., 24\}$, is collected at Zugspitze at a given day in order to get an impression of the air movement on that day.

Exercise 2 - Levels of measurement

Describe the following data by specifying which level of measurement they have (nominal, ordinal, dichotomous etc.):

a) Party preference at an election

b) Level of difficulty in a video game

c) Age of zoo animals

d) Calendar time with the birth of Christ at point zero

e) Enrollment number of a student

f) Grades in school

Exercise 3 - Measures of central tendency and variation

Measuring the age of $n = 10$ randomly selected people on the street one obtains the following sample:

Observational unit i	1	2	3	4	5	6	7	8	9	10
x_i	39	44	29	51	58	54	54	49	48	43

Calculate the following measures based on the sample and judge whether the measures are meaningful for the given situations:

a) Arithmetic mean $\bar{x}$

b) Median $\tilde{x}_{0.5}$
Additional question: Why do the median and the arithmetic mean differ for this sample?

c) Mode x_{mod}

d) Sample standard deviation s

Exercise 4 - Descriptive Statistics

The following table gives the amount of rain (in litres per square metre), measured at the volcano Merapi (Indonesia) between January 1st and January 20th, 1995.

Rain	Date	Rain	Date
2	01.01.1995	50	11.01.1995
9	02.01.1995	12	12.01.1995
18	03.01.1995	0	13.01.1995
2	04.01.1995	0	14.01.1995
23	05.01.1995	0	15.01.1995
42	06.01.1995	0	16.01.1995
11	07.01.1995	3	17.01.1995
13	08.01.1995	3	18.01.1995
40	09.01.1995	40	19.01.1995
12	10.01.1995	48	20.01.1995

a) What is the level of measurement of the variable `rain`?

b) Draw a histogram using the intervals [0,10),[10,20),[20,30),[30,40),[40,50].

c) Calculate the the following measures of location and dispersion based on the sample: mode, median, arithmetic mean, lower quartile, upper quartile, sample variance, sample standard deviation, and coefficient of variation.

d) Use the results obtained in (c) to draw a boxplot of the empirical distribution of `rain`, and interpret it.

Exercise 5 - Association of categorical variables

In the following, we will analyze data about all Titanic passengers that were on board during its sinking. Use the information given in the contingency tables to calculate and interpret Odds Ratios to quantify the difference in the chance to survive (Yes: Person survived, No: Person did not survive) between the following groups:

a) Passengers from first and third class

	1st	2nd	3rd	Crew
No	122	167	528	673
Yes	203	118	178	212

b) Male and female passengers

	Male	Female
No	1364	126
Yes	367	344

c) Male crew members and male passengers from third class

Contingency table only based on males:

	1st	2nd	3rd	Crew
No	118	154	422	670
Yes	62	25	88	192