EXERCISE - STATISTICS FOR AI
Summer Semester 2025 (Mag. Thomas Forstner)

366.591 366.592 366.593 366.594 366.595

12. Suppose 100 people were asked whether they suffer from high cholesterol and/or obesity. The relationship between these two conditions is shown in the cross table below. Calculate the missing percentages and add them into the gray fields.

High Cholesterol x Obesity					
			Obesity		TOTAL
			NO	YES	
High Cholesterol	NO	Count	65	2	67
		% within high cholesterol	97.00%	3.00%	100.00%
		% within obesity	76.47%	13.33%	100%
		% of total	65.00%	2.00%	67.00%
	YES	Count	20	13	33
		% within high cholesterol	60.60%	30.40%	100.00%
		% within obesity	23.50%	86.70%	33.00%
		% of total	20.00%	13.00%	33.00%
TOTAL		Count	85	15	100
		% within high cholesterol	85.00%	15.00%	100.00%
		% within obesity	100.00%	100.00%	100.00%
		% of total	85%	15.00%	100.00%

13. The table below presents the income distribution of some people.

income in 1,000 Euros	number of people	class mark x_i	class length d_i	relative frequencies
up to 10	47,996	5	10	8.457%
10 up to 20	191,492	15	10	33.741%
20 up to 30	124,498	25	10	21.937%
30 up to 50	104,428	40	20	18.399%
50 up to 100	67,988	75	50	11.979%
100 and over	31,125	N.A.	N.A.	5.484%

a) Add the class mark and the class length for each class directly in the table above.
b) Add the relative frequencies of each class directly in the table above.
c) Calculate the percentage of people with an income of less than 20,000 Euro. 42.198%
d) Calculate the percentage of people with an income between 30,000 Euro and 50,000 Euro. 18.399%
e) Calculate the percentage of people with an income of more than 30,000 Euro. 35.862%

14. In a study, the following class marks were defined for a continuous attribute:

class mark x_i for each class: 200, 220, 240, 260, 280, 300, 320

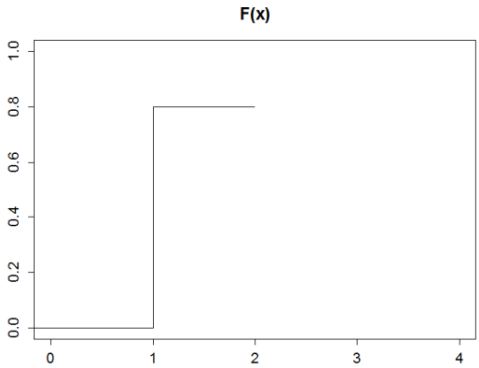
Determine the class length and the class boundaries for each class.

a) class lengths: 20 *Ambiguous Pharasing. It may be [20, 20,..., 20]
b) class boundaries: [190, 210), [210, 230), [230, 250), [250, 270), [270, 290), [290, 310), [310, 330)

15. While preparing for a statistics exam, a student found a partial solution to an exam from the previous semester on the Internet. From this partial solution, she was able to obtain the following information.

Example 1: The attribute "number of books read per person in the last week" is given for a population of 500 people. The attribute has the attribute values 0, 1, 2, and 3. It is known that 50 people have read exactly 3 books.

The student has also found an incomplete sketch of the empirical cumulative distribution function F(x).

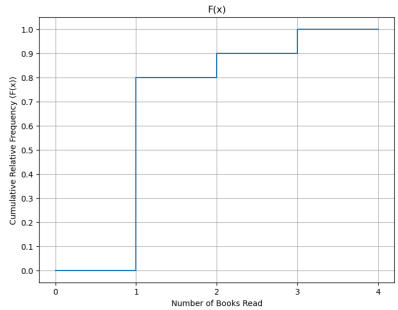


Construct a frequency table for the relative frequencies of the books read per person in the last week, and complete the sketch of the empirical cumulative distribution function F(x).

State your solution in the box below:

Frequency Table:

Books Read	Absolute Freq.	Relative Freq.
0	0	0.000
1	400	0.800
2	50	0.100
3	50	0.100

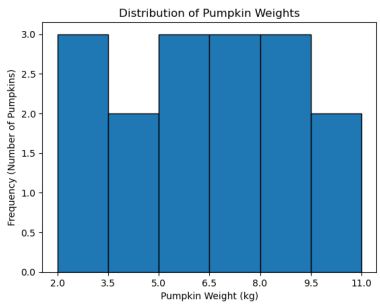


16. The weights of pumpkins (in kg) from a farm are given:

2.5 , 3.0 , 3.2 , 4.1 , 4.5 , 5.0 , 5.5 , 6.0 , 6.5 , 7.0 , 7.5 , 8.0 , 8.5 , 9.0 , 9.5, 10.0

Present this distribution using a histogram. Use appropriate self-defined class limits for this histogram.

State your solution in the box below:



17. A simple story was read to some children. Their ability to understand this story was tested by six questions related to the story. For each child, the number of questions answered incorrectly is given below:

1 2 0 6 0 0 1 2 0 0 2 1 6 2 2 1 6 6 1
0 2 0 2 0 0 1 0 0 1 0 0 1 0 0 1 1 0 3

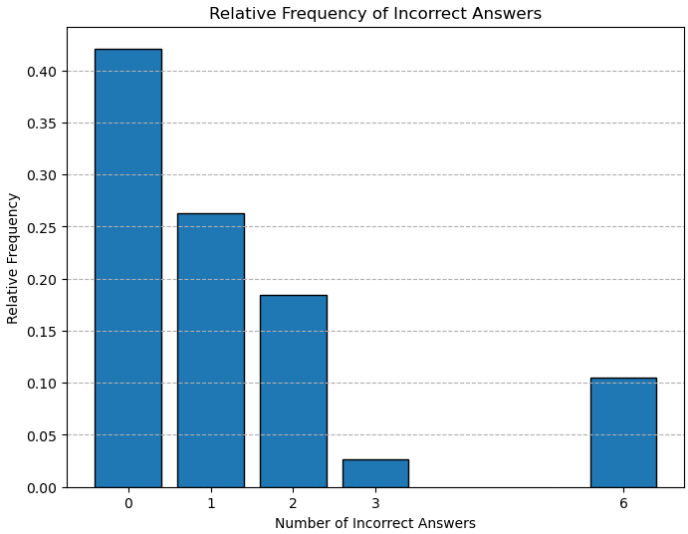
a) Construct a frequency table for the absolute and relative frequencies of the given data.
b) Construct an appropriate graphical representation of the relative frequencies.
c) Construct an appropriate graphical representation of the cumulated relative frequencies.

State your solution in the box below:

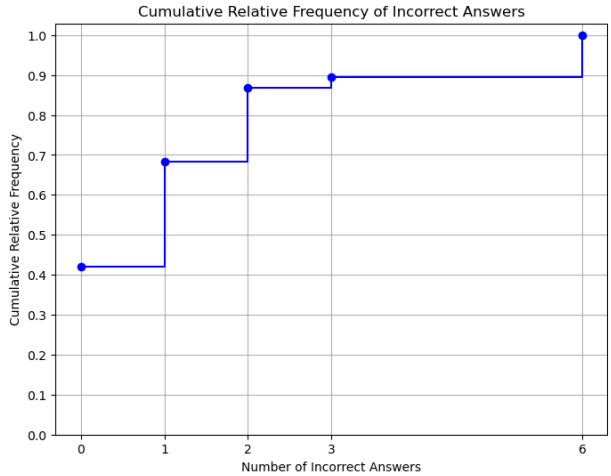
Frequency Table:

Incorrect Answers	Absolute Freq.	Relative Freq.
0	16	0.421
1	10	0.263
2	7	0.184
3	1	0.026
6	4	0.105

17a



17b



17c

18. A company has 18 subsidiary companies that each achieved the following revenues (rounded in millions of Euros) in a given year:

2	4	6	4	7	5	7	5	3
8	3	10	9	4	3	7	4	6

a) Construct a frequency table for the absolute and relative frequencies of the given data.
b) Construct an appropriate graphical representation of the cumulated relative frequencies.
c) Construct a histogram using the revenue-class definition I below.
revenue-class definition I: [1.5 , 3.5), [3.5 , 5.5), [5.5 , 7.5), [7.5 , 9.5), [9.5 , 11.5)
d) Construct a histogram using the revenue-class definition II below.
revenue-class definition II: [1.5 , 3.5), [3.5 , 6.5), [6.5 , 10.5)

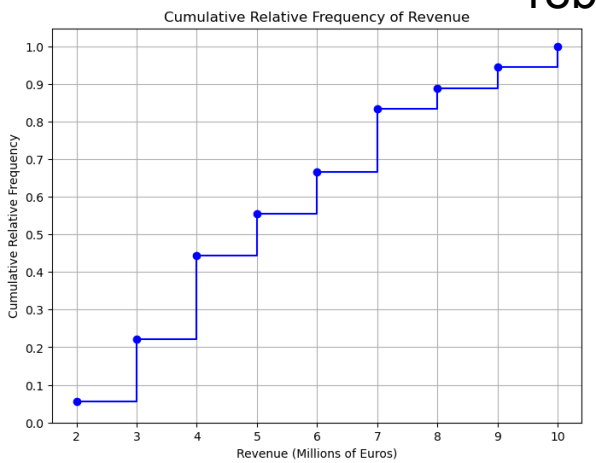
State your solutions in the box below:

18a

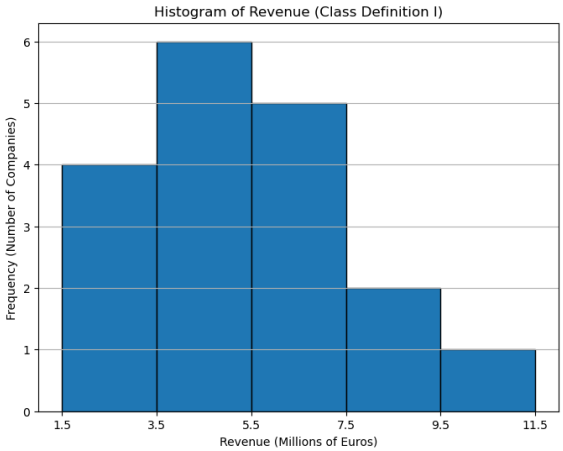
Frequency Table:

Revenue (Millions)	Absolute Freq.	Relative Freq.
2	1	0.056
3	3	0.167
4	4	0.222
5	2	0.111
6	2	0.111
7	3	0.167
8	1	0.056
9	1	0.056
10	1	0.056

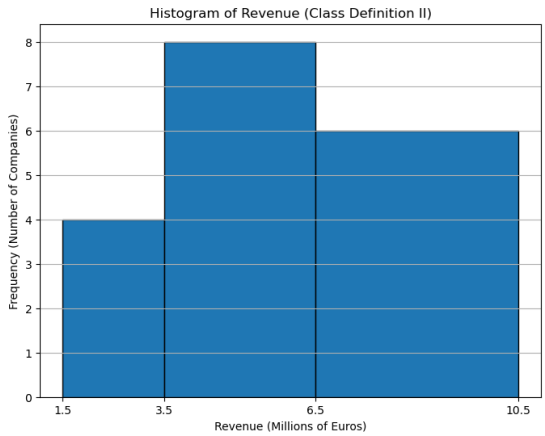
18b



18c



18d



19. Suppose over 200 days, the three elevators in an office building were observed. On 90 days, at most two elevators worked, on 10 days, only one elevator worked. On two days, all three elevators did not work. Calculate the absolute and relative frequencies of the number of working elevators.

absolute and relative frequencies of the number of working elevators:

```
Frequency Table:
-----------------------------------------
Working Elevators | Absolute Freq. | Relative Freq.
-----------------------------------------
0                 | 2              | 0.010
1                 | 10             | 0.050
2                 | 78             | 0.390
3                 | 110            | 0.550
-----------------------------------------
```

Please keep the formal guidelines for submitting the homework assignments in mind to avoid losing points unnecessarily.