

# Statistisk Dataanalyse 2023

Instruktør: Vivek Misra

# Exercise Class NR2

Solutions to the Tasks

# Task 1 - Description

- The table below shows the height of students in classroom A (total of 15 students) and classroom B (total of 16 students), measured in centimeters.
  - Solution: We have solved the task in the following manner:
    - 10A for Ungrouped Frequency Table
    - 10B for Grouped Frequency Table
    - 10C for Histogram

# Task 1A - Solution

**Class A – Sorted Ungrouped Frequency Table**

154	156	158	160	162	165	169	170	171	172	175	180	189
1	1	1	1	1	1	2	1	1	1	2	1	1

**Class B – Sorted Ungrouped Frequency Table**

161	163	169	173	174	175	176	179	180	182	185	191
1	1	1	1	2	1	2	1	2	3	1	1

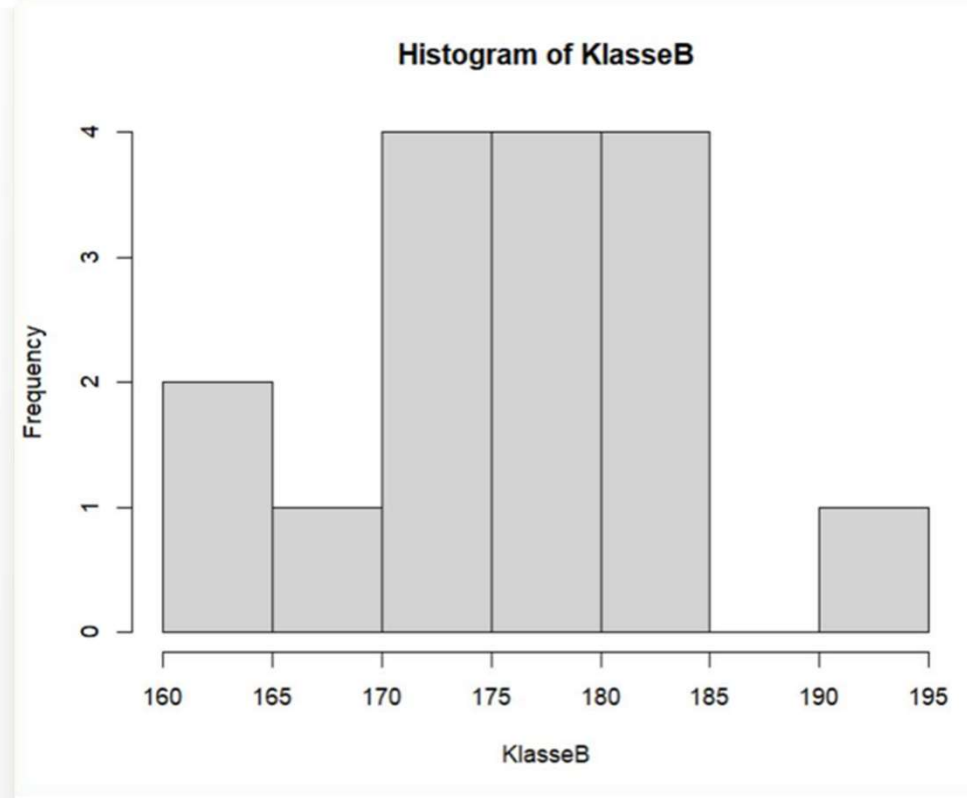
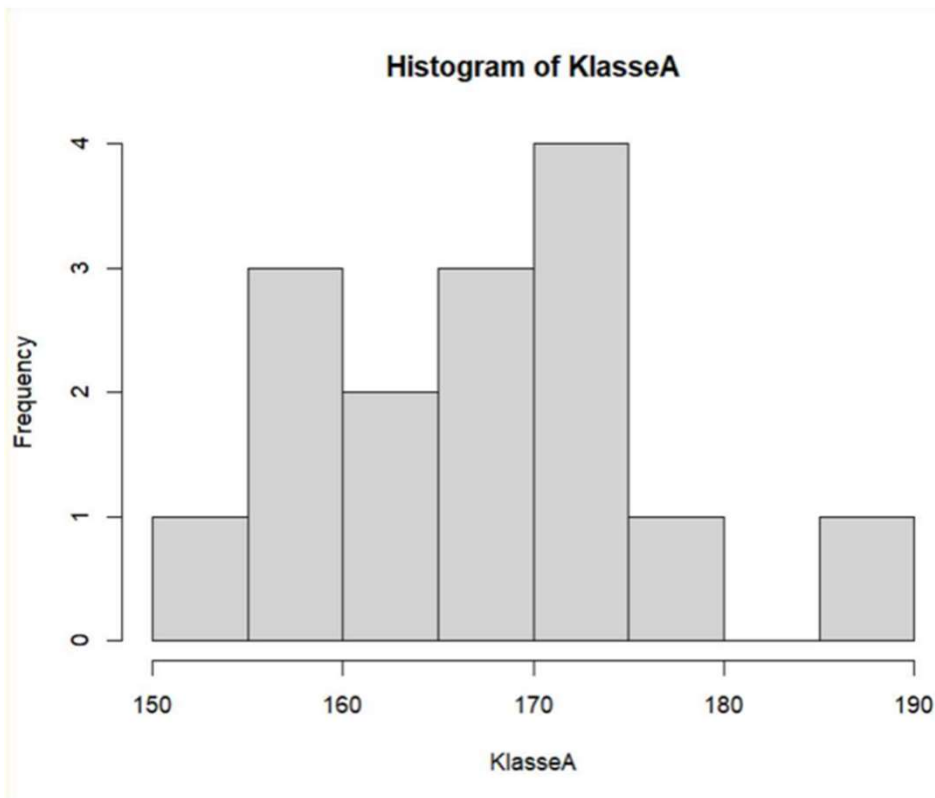
# Task 1B - Solution

Grouped Frequency Table for Class A				
(154-160(	(160-165(	(165-170(	(170-175(	(175-189)
3	2	3	3	4

Grouped Frequency Table for Class A				
(160-165(	(165-170(	(170-175(	(175-180(	(180-191)
2	1	3	3	7

# Task 1C - Solution

- Solution: The following diagram has been made in R.



## Task 2 - Description

- The distribution of entrance test scores of freshmen in a particular university has the following percentile scores. How may the distribution be described.

- A. Symmetrical bell-shaped
- B. Skewed left (negatively skewed)
- C. Skewed right (positively skewed)
- D. Impossible to tell from the above.

Percentile	Score
95th	140
80th	120
65th	101
50th	94
35th	91
20th	87
5th	80

## Task 2 - Solution

- Solution: Compare the mean with the median
- We already know that the median is 94.
  - And the mean is 101,85
- Mean > Median = Positive Skewed.
  - The score-numbers are rising along with Percentile.

Percentile	Score
95th	140
80th	120
65th	101
50th	94
35th	91
20th	87
5th	80
Mean	$\frac{713}{7} = 101,85$



## Task 3 – Description

- The cumulative frequency graph below shows the salaries of 100 employees who work for Welsh Bank (black) and 100 employees who work for the Bank of Finland (blue).
- Based on the given graph, evaluate the following sentences as TRUE or FALSE:
  - i. The interquartile range of the data for the Bank of Finland is 60000
  - ii. The median for the Welsh Bank is £62000
  - iii. 78000 is an outlier for the Bank Welsh
  - iv. The range for both banks is the same

## Task 3I - Solution

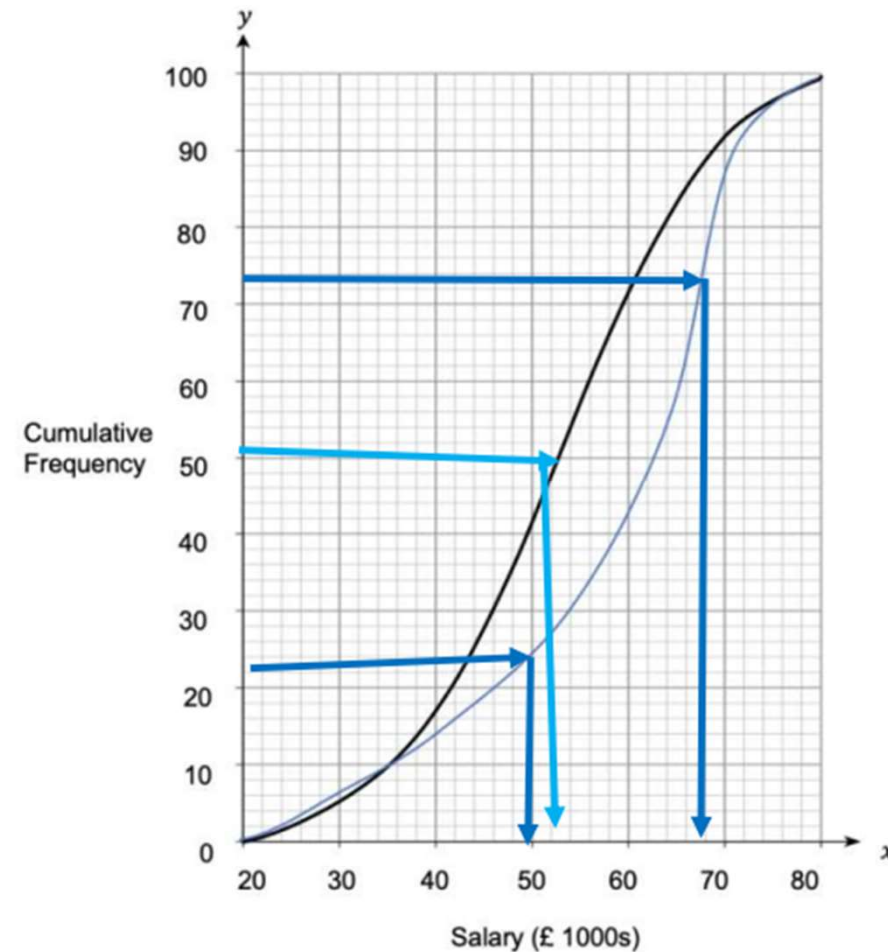
- First we will calculate the Interquartile Range.

$$IQR = 69 - 50 = 19$$

- We can see, that the interquartile range is 19 and therefore the statement is false.

# Task 3II - Solution

- Solution: We will find the median for the welsh bank by starting at 50 from the y-axis and then going towards the x-axis.
  - We can see, that when we hit the x-axis then we will get a value which is 51.
  - Therefore the statement is false.



## Task 3III - Solution

- Solution: To find the outlier, we need to use the formula interquartile formula and insert the values which we found in the graph.

$$Q3 + 1,5 \cdot IQR$$
$$Q1 - 1,5 \cdot IQR$$

- Now we will insert the values inside the IQR-formula.

$$Q3 + 1,5 \cdot 15 = 82,5$$
$$81 - 1,5 \cdot 15 = 22,5$$

- Now we will try and insert 78 between the 82,5 and 22,5 in a interval.

$$22,5 < 78 < 82,5$$

- We can say, that 3.III is false, and there is no Outlier.

## Task 3IV - Solution

- Solution: Because the smallest value and biggest value are the same for both banks. Therefore we just need to subtract them from each other.

$$80 - 20 = 60$$

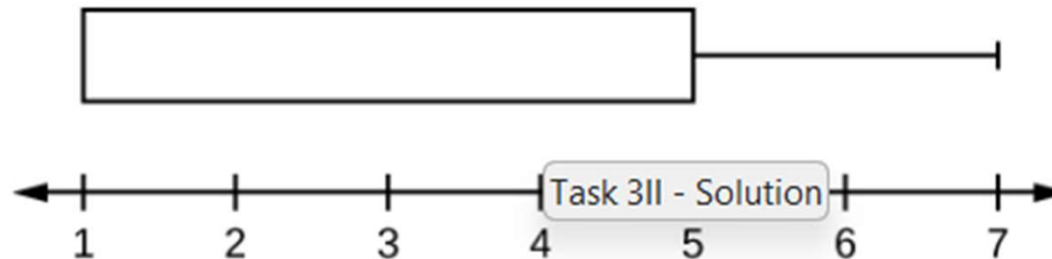
- We can say, that statement 3IV is true and both banks have the same range!

## Task 4 – Description / Solution

- If the mean, median and mode of a distribution are 8,7,6 respectively, then the distribution is:
  - A: Negatively Skewed
  - B: Not Skewed
  - C: Postively Skewed
  - D: Symmetrical
  - E: Bimodal
- Solution: We can see, that mean is 8, and it is bigger than the median which is 7. Therefore it is a positive skewed distribution.

## Task 5 – Description

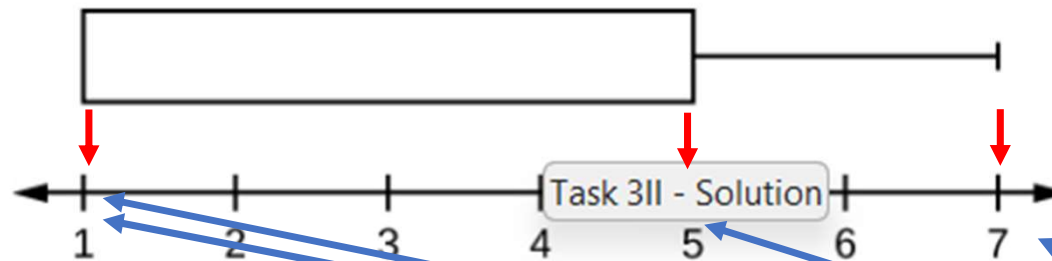
- The following Type 1 boxplot was drawn using a list of numbers.



- What is the incorrect statement regarding this boxplot?
  - A. The number 1 must be in the list of numbers from which the plot was drawn.
  - B. The dataset has no median.
  - C. The boxplot could be derived from the following dataset: 1,1,5,5,7.**
  - D. More than half of the data falls between 1 and 5.
  - E. The Range is 6.

## Task 5 – Solution

- The following Type 1 boxplot was drawn using a list of numbers.



**C. The boxplot could be derived from the following dataset: 1,1,5,5,7.**



## Task 6 – Description / Solution

- A teacher gives a 20-point test to 10 students. Find the percentile rank of a score of 12.
  - 18,15,12,6,8,2,3,5,20,10

Solution: First sort the numbers in the dataset and then use the percentage formula shown on the next page with Lk:

Sorted Dataset									
2	3	5	6	8	10	12	15	18	20

## Task 6 - Solution

- Solution: I have used the following formula to solve this question.

$$L_k = \frac{k}{100} \cdot (n + 1)$$

Note: The k-value defines the percentile and the n-value defines the amount of numbers.

$$L_{25} = \frac{25}{100} \cdot (10 + 1) = \frac{11}{4} = 2,75 \approx 3$$

We can see that 2,75 is a position between 3 and 5. And therefore we can conclude that the 25th percentile is 5.

Next slide shows how the placement order works!



## Task 7 – Description / Solution

- Find Q1, Q2 and Q3 for the dataset.
  - 15, 13, 6, 5, 12, 50, 22, 18
- Solution: First sort the dataset, so that we are able to track the numbers in quartile-arrangement.

Sorted Dataset							
5	6	12	13	15	18	22	50

- Solution: Now proceed further and find the quartile values.

## Task 7 - Solution

- First we will find the Median from the even-amount of numbers.

$$Q2 = \frac{13 + 15}{2} = 14$$

- Now we will calculate the values which are for 1st Quartile.

$$Q1 = \frac{6 + 12}{2} = 9$$

- Now we will calculate the values which are for 2nd Quartile.

$$Q2 = \frac{18 + 22}{2} = 20$$

## Task 8 - Description

- The mean of the population of ten scores, 78, 91, 91, 94, 74, 23, 63, 22, 78, 89 is 70.3, and the modes are 78 and 91. The skewness of the population is:
  - A. Negative
  - B. Zero
  - C. Positive
  - D. Not Determined
  - E. Positive or Negative Depending on the Score

## Task 8 - Solution

- Solution: We need to compare the mean and the median with each other. But first let's find the median, by sorting the list **as marked**.

- **22,23,63,74,78,78,89,91,91,94**

$$\text{Median} = \frac{78 + 78}{2} = 78$$

- Now we already know that the mean is 70,3 and the median is now 78.
- Because the median is higher than the mean, the result is negative skewed.

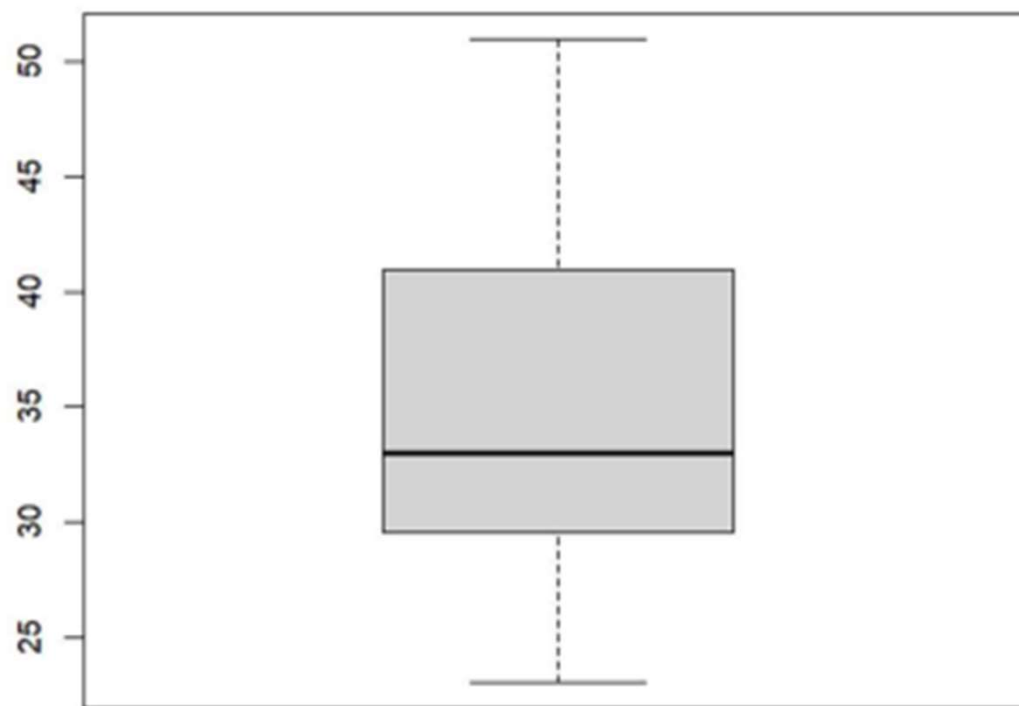
## Task 9 – Description / Solution

- A percentile score of 40 indicates that a person:
  - a. answered 40% of the questions correctly on the test.
  - b. knows 40% of the material covered by the examination.
  - c. has earned a score equal to or better than 40 persons in his class.
  - **d. has earned a score equal to or better than 40% of the persons in his class.**
- Solution: We will use the “selfish-concept” where we say that the person has either earned equal or better than the rest of the class.
- So the correct answer is D.



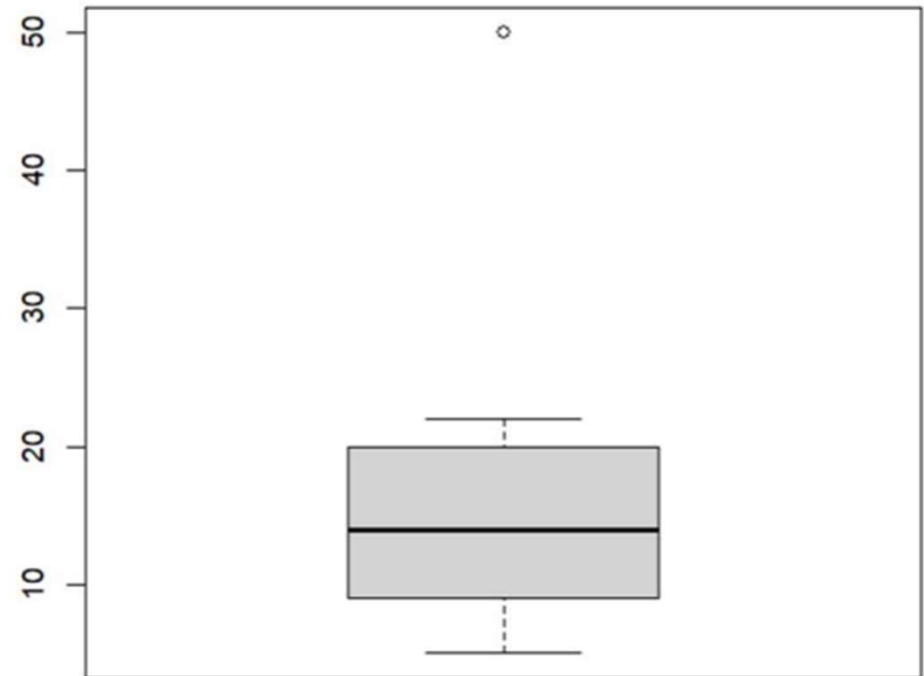
# Task 10 – Description / Solution

- Construct two boxplots (“Type 1” and “Type 2”) for the data.
  - 33, 38, 43, 30, 29, 40, 51, 27, 42, 23, 31
- Solution: In this task it is only one boxplot and the second is for the outlier!
- I have solved this question in R.



# Task 11 – Description / Solution

- In this dataset
  - 15,13,6,5,12,50,22,18
- Solution: We will use R to construct a Boxplot.

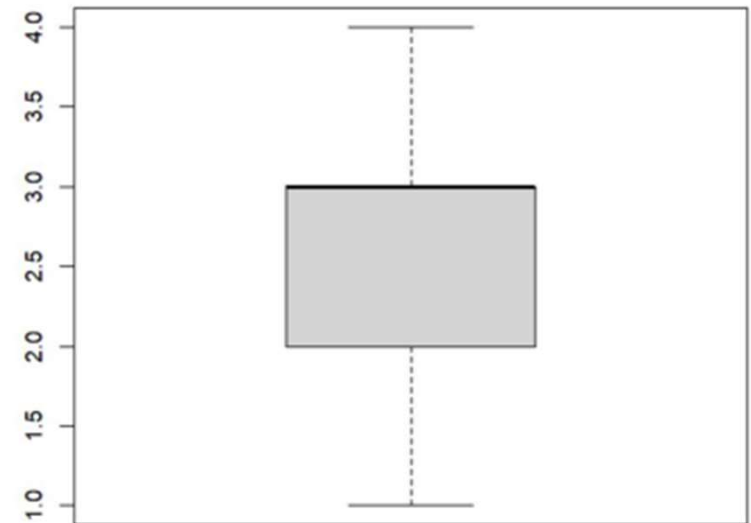


## Task 12 - Description

- Twenty-five people were given a blood test to determine their blood type.
  - Raw Data: A,B,B,AB,O O,O,B,AB,B B,B,O,A,O A,O,O,O,AB AB,A,O,B,A
    - a) Can you construct a histogram? Can you construct a bar graph?
    - b) Considering your reply in item a, construct the correct graph

# Task 12A & Task 12B - Solution

- Solution A: Yes, you can define different factorical values to numerical values.
- Solution B: We will show it here, where  $A=1$ ,  $B=2$ ,  $O=3$  and  $AB=4$ .
  - The following dataset I as a student made:  
1,2,2,4,3,3,3,2,4,2,2,2,3,1,3,1,3,3,3,4,4,1,3,2,1



# Tak for i dag!

Instruktør: Vivek Misra