LESSON 2. DESCRIPTIVE STATISTICS – PART 2

1.3. Measures of:

- Percentile
- Decile and Quartile
- Outlier

1.4. Data representation

- Frequency distributions and histograms
- Boxplots
- Graphs: Time series, Pie graphs, Scatter plots

1.5. Shapes of frequency distributions

- Skewness
- Kurtosis
- Shapes of distributions

Chapter 1 – Part 2: Assignments

1. 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.

Classroom A	Classroom B
156	185
175	175
189	169
165	182
160	179
154	163
158	191
170	182
171	180
169	174
180	161
175	180
172	176
169	174
162	182
	173

a) Develop an ungrouped frequency table for all 31 students (1 table in total)

To develop a ungroup table, we need to start by taking the number and plotting how many there are.

To develop an ungrouped table, we need to start by writing all the dataset numbers down from 156 to 191. Then we need to count how many numbers there are and then we need to sort the "available" numbers.

We have done this in Ark 2.

b) Construct a grouped frequency table for all 31 students (1 table in total)

It is also done in the Excel file.

The part where we sort the numbers after counting them is the "grouped" part.

c) Plot the frequencies of each class for all 31 students (1 histogram in total)

This is made in Excel, the problem is that it has been made two times but because Excel represents two groups - then it results a not exact answer.

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

The distribution can be seen that the distribution is holding more towards the higher numbers such as 95th and so on. Whereas if you look at the other side, then you can see that the distribution is lower towards the lower percentiles such as 50th and 20th and so on.

Percentile	Score
95th	140
80th	120

65th	101
50th	94
35th	91
20th	87
5th	80

a. Symmetrical bell-shaped

- Out of the dataset, it can be clearly concluded that it is not symmetrical.

b. Skewed left (negatively skewed)

- It is not negatively skewed, because we are not getting negative scores at higher values and so on.

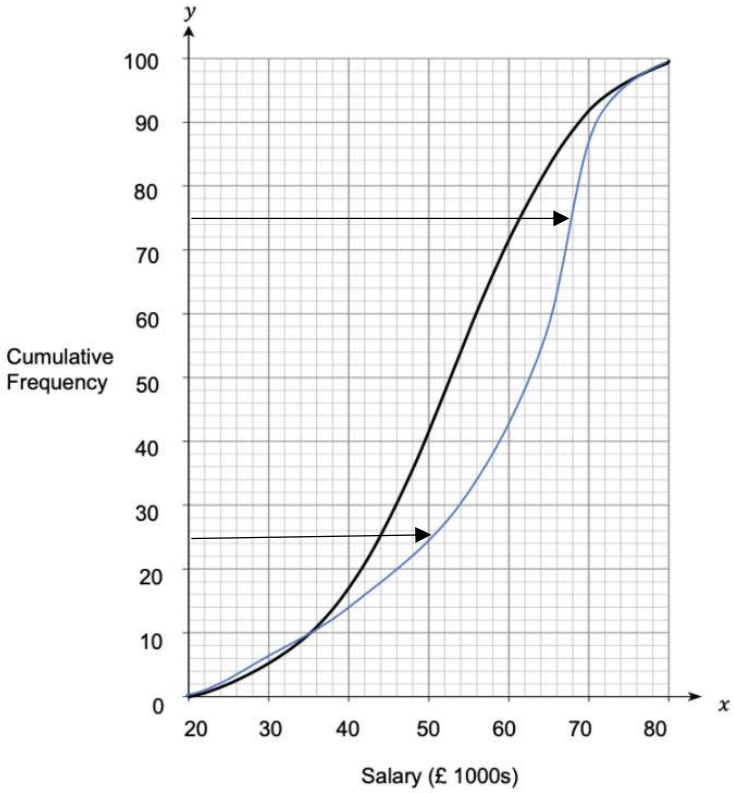
c. Skewed right (positively skewed)

It is most possibly positively skewed and that is because, we have the lower scores at the lower percentiles and the higher scores at the higher percentiles.

d. Impossible to tell from the above

- It is possible to tell 😊

3. 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

- First we will start by looking at the y-axis where we can see that Q3=75%, so we will pick 75. Thereafter we will go in the middle and point the meeting point with the x-axis, for the intersection with the blue graph.

- The same method is used for Q1=25%, where we start at 25 in the y-axis and then thereafter go towards the x-axis.
- Because we want to calculate the quartile range, then we can afterwards say the following. 69-51= 18. That tells us, that the finish range is not 60000.

ii. The median for the Welsh Bank is £62000

- No because if we intersect a point between 50% in the y-axis and the x-axis then we can see that we are not getting 62,000, but instead under it

iii. 78000 is an outlier for the Bank Welsh*

- To find the outlier, we need to know the quartiles and the range between them.
- We can see that Q1=43 and Q3=61. We calculate the range. IQR=61-43=18.
- Because we have calculated the range, we can now calculate the outliers.
- O1=Q1-1.5*IQR= 16
- O2=Q3+1.5*IQR= 88

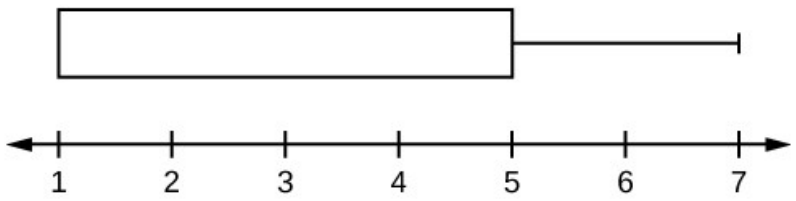
iv. The range for both banks is the same

- We can see, that from the previous task we found the quartile range for the Welsh Bank to be 18.
- If we want to calculate the quartile range, then we will get the answer to be 69-53= 16.
- We have used the same method as before.

4. 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.	positively skewed
d.	symmetrical
e.	bimodal.

5. 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.

6. 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

To calculate the percentile of 12, we need to solve the numbers.
2,3,5,6,8,10,12,15,18,20

To calculate, we count the numbers on the dataset. Thereafter we count the numbers to 12 (12 is excluded). Thereafter we multiply with hundred.

After that, we need to 6/10*100=60

7. A teacher gives a 20-point test to 10 students. Find the value corresponding to the 25th percentile.
   18, 15, 12, 6, 8, 2, 3, 5, 20, 10
   So, we know, that there are 10 numbers. Thereafter we know that the percentile rank is 60.

$$0{,}25 = \frac{x}{10}$$

⇳ *Ligningen løses for x vha. WordMat.*

$$x = 2{,}5$$

It is 2,5 enteries which are needed. In this case, we are saying that we need 2,5 datapoints to locate between two points. For example, we have 2 and 3 in the dataset. So, we can plot 2,5 in between these two.

8. Find Q_1, Q_2, and Q_3 for the data set.
   15, 13, 6, 5, 12, 50, 22, 18
   Just in case, we will start by sorting the dataset: 5,6,12,13,15,18,22,50

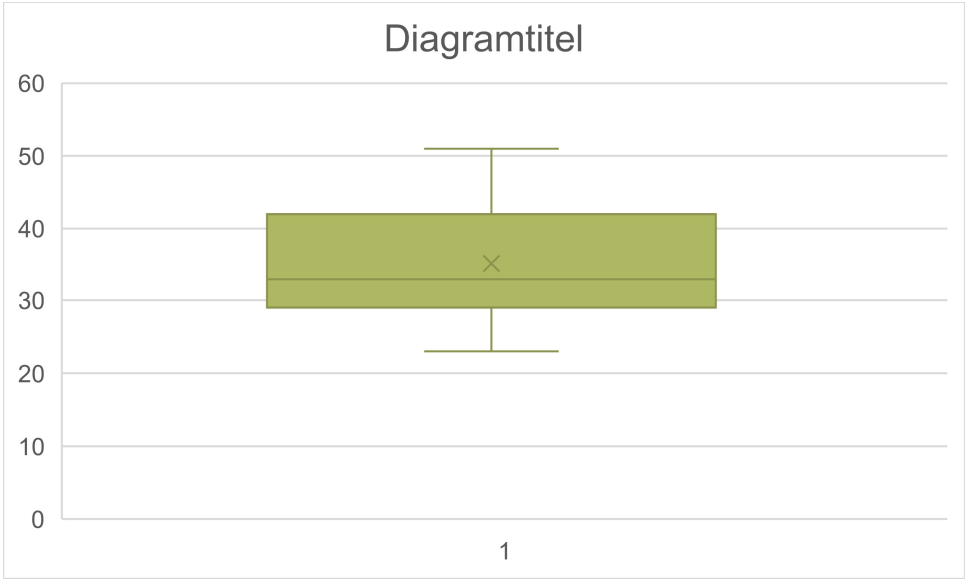
Q2= 13+15=28/2=14
Q3=18+22=40=40/2=20
Q1=12+6=18=18/2=9

9. 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.

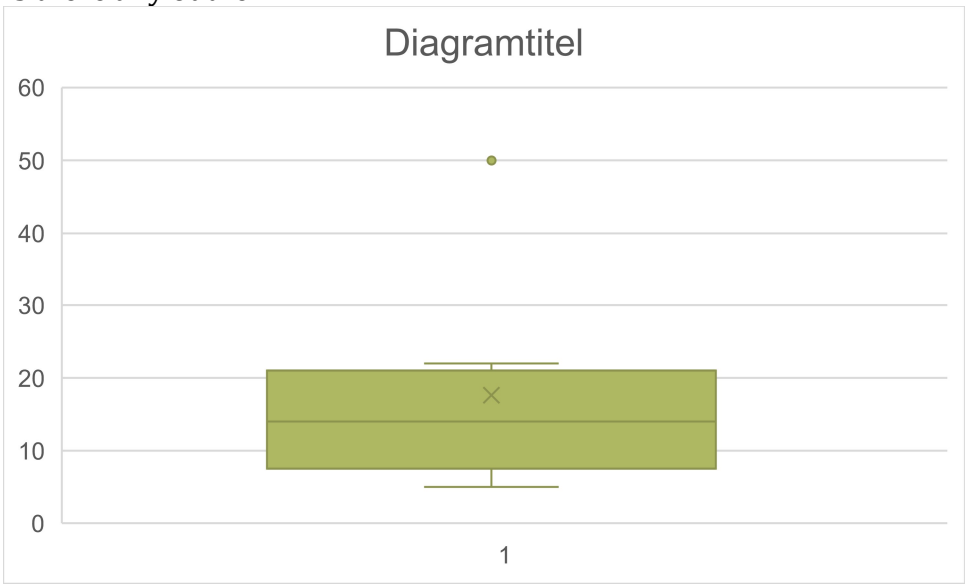
10. A percentile score of 40 indicates that a person:
    a. answered 40% of the questions correctly on the test.
    b. knows 40% of the material covered by the examination.
    c. has earned a score equal to or better than 40 persons in his class.
    d. has earned a score equal to or better than 40% of the persons in his class.

11. Construct two boxplots (“Type 1” and “Type 2”) for the data.

33, 38, 43, 30, 29, 40, 51, 27, 42, 23, 31



12. In this data set:
    15, 13, 6, 5, 12, 50, 22, 18
    Is there any outlier?



Yes, there is an outlier, because of the cross.

13. 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 A,B,A,O,B,A
    a) Can you construct a histogram? Can you construct a bar graph?

- If you define the different letters with numbers, then you can create a histogram and then thereafter a bar graph.

    b) Considering your reply in item a, construct the correct graph

- X