LESSON 1. DESCRIPTIVE STATISTICS – PART 1

1.1. Statistics

- What is statistics?
- Definitions: Variable, data, population and sample
- Descriptive and Inferential statistics

1.2. Variables and Types of Data

Qualitative, quantitative, categorial, ordinal, discrete, continuous

1.3. Measures of:

- Central Tendency (Location):
    - Mean
    - Median
    - Mode
    - Midrange
- Variation (Dispersion)
    - Range
    - Variance
    - Standard Deviation
    - Coefficient of Variation

Chapter 1: 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. For each of the classroom, calculate the following:
    a. Median
    b. Mean
    c. Mode
    d. Midrange

Display the results on a table.

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

2. Find the mean of the following data:
    20, 26, 40, 36, 23, 42, 35, 24, 30

First you have to start by sorting the numbers.
20, 23, 24, 26, 30, 35, 36, 40, 42.
You can see, that there are 4 numbers on the both of sides of 30. Therefore because 30 is a even number, you will get the answer 30.

3. Find the median of the following measurements:
    713, 300, 618, 595, 311, 401, and 292

Now we will sort the numbers.
292, 300, 311, 401, 595, 618, 713
After that, we can see that there are 3 numbers on the left side and 3 on the right side. Therefore, our answer will be 401.

4. Find the median of the following measurements:
    684, 764, 656, 702, 856, 1133, 1132, 1303

Now we will sort the numbers.
656, 684, 702, 764, 856, 1132, 1133, 1303
After that, we can see that there are to even numbers in the middle. Here in this case is 764, 856 and we will add them together and divide by 2, to find the median.

(702+856)/2=810

5. Find the mode of the following measurements:
   8, 9, 9, 14, 8, 8, 10, 7, 6, 9, 7, 8, 10, 14, 11, 8, 14, 11

To find the mode, you have to count how many times you have got the same number. The number which comes the most, is the mode and in this case 8.

6. Find the mode of the following measurements:
   110, 731, 1031, 84, 20, 118, 1162, 1977, 103, 752

Because you don't have the same number, then there is no mode.

7. Find the midrange of these data:
   2, 3, 6, 8, 4, 1

The midrange is found by subtracting the high number with the lower number. Here in this case, if we sort the numbers, we get the following.
1,2,3,4,6,8
Because 8 is the highest and 1 is the lowest, 8-1=7. The following number is the result.

8. A researcher wants to collect data on 100 inhabitants living in one specific town. Classify the following collected variables according to their type (Nominal, ordinal, discrete, or continuous)
   i. Occupation ("blue collar", "white collar", "unemployed")

- In this case, it is norminal.

   ii. Highest attained education ("low", "medium", high")

- The second one is ordinal.

   iii. Monthly salary

- This can be argumented because it is either discrete or continuous.

   iv. Civil status ("single", "married", "widow")

- This is ordinal.

   v. Number of children

- This is discrete.

9. Evaluate the following statements as true or false:
   a. In statistics, a population always refers to humans.

- Yes, but it can also be puppets or things.

   b. A sample is a subset of the study population.

- A sample is a fraction of a large part of a population, which is used for drawing conclusions to the bigger conclusions. The reason for using sample is, to make big conclusions from smaller fractions of data.

   c. Inferential statistics are statistical techniques used to draw conclusions about one specific sample.

- Here comes the tricky part. Inferential statistics is used to draw conclusions to the bigger populations.

   d. A survey will be given to 100 students randomly selected from the freshmen class at Odense High School. The sample is all the freshmen at Odense High School.

10. Find the range, variance and standard deviation for the data set for the samples of Brand A and Brand B paint.

- If we want to find the range, then we will have to start by subtracting the big number value with the smaller number value. In this case we get the answer 50, because: 60-10=50
- In the other section of Brand B, the range will be 15.

Now comes the tricky part, and here in this case we can start by finding the mean of both sides.
10+60+50+30+40+20=210. Now 210, will be divided with 6 because we have 6 values, and we will get mean to become 35.
The same method is used for Brand B, which results in the following answer. It is also 35.

Kig venligst på selve Excel-Filen.

Brand A	Brand B
10	35
60	45
50	30
30	35
40	40
20	25

11. If the variance of a distribution is 9, the standard deviation is:
    a. 3

- The reason behind, that it is 3 because if you take the square root of 9 that will give you 3. You just have to follow the same rule of square root, as if it was a real calculative question.

    b. 6
    c. 9
    d. 81
    e. impossible to determine without knowing n.

12. The standard deviation of a dataset is 10. If 5 were subtracted from each measurement, the standard deviation of the new dataset would be:

a) 2
b) 10/25
c) 5
d) none of these.

If we start by measuring 10 numbers, then we can in this case say 1,2,3,4,5,6,7,8,9,10.

If subtract the numbers from each number, then the result will become this.

10-5=5

9-5=4

8-5=3

7-5=2

6-5=1

5-5=0

4-5=-1

3-5=-2

2-5=-3

1-5=-4

If we put all the sum numbers together, then we can see that we get the number 5.

Standardafvigelse ændrer ikke (HUSK DET TIL EKSAMEN!)