

LESSON 1. DESCRIPTIVE STATISTICS – PART 1 (HANDWRITTEN)

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, categorical, ordinal, discrete, continuous

1.3. Measures of:

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

- A. Vi starter allerførst med at finde medianen ved at opstille vores rækkefølge på følgende måde for begge klasser.

Klasse A: 156,175,189,165,160,154,158,170,171,169,180,175,172,169,162

Klasse B: 185,175,169,182,179,163,191,182,180,174,161,180,176,174,182,173

- Nu kan vi se at vi ønsker at sortere selve rækkefølgerne.

Sorteret A: 154,156,158,160,162,165,169,170,171,172,175,175,180,189

Sorteret B: 161,163,169,173,174,174,175,176,179,180,180,182,182,182,185,191

- Fordi vi ønsker at udregne medianen, kan vi anvende følgende formel:

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

- Fordi vi kan se, at vi har 14 numre i Klasse A og 15 numre i Klasse B, kan vi indsætte placering som er 50% på k's plads og derved antallet ved n's plads.

$$A_{50} = \frac{50}{100} * (14 + 1) = \frac{15}{2} = 7,5$$

$$B_{50} = \frac{50}{100} * (15 + 1) = 8$$

Median A: Fordi vi får en værdi som er 7,5, så fortæller den om medianens placering i rækkefølgen. I vores tilfælde er det positionen mellem 169,170, så derfor finder vi middelværdien af disse numre og finder medianen:

$$\text{Median A} = \frac{169 + 170}{2} = 169,5 \approx 169$$

Median B: Fordi vi får en værdi som er 8, så fortæller den om medianens placering i rækkefølgen. I vores tilfælde er det positionen mellem 176,179, så derfor finder vi middelværdien af disse numre og finder medianen:

$$\text{Median B} = \frac{176 + 179}{2} = \frac{355}{2} = 177,5$$

- B. For at udregne middelværdien for begge klasser, skal vi bare lægge alle tallene sammen og dividere med antal tal der findes i rækkefølgen.

Sum A	154+156+158+160+162+165+169+170+171+172+175+175+180+189 = 2525
Sum B	161+163+169+173+174+174+175+176+179+180+180+182+182+182+185+191 = 2826

Mean A	$\frac{2525}{14} \approx 180,3571$
Mean B	$\frac{2826}{15} = \frac{942}{5} = 188,4$

- C. Nu skal vi finde typetallet for begge klasser. Her skal vi huske begreberne som unimodal, bimodal, multimodal. Kig nærmere på den sorterede klasseliste.
- Klasse A: Er kendt som Bimodal Mode, fordi den har 2 værdier som optræder mere end engang. I tilfældet er det 169 og 175.
 - Klasse B: Er kendt som Multimodal Mode, fordi den har 3 værdier som optræder mere end engang. I tilfældet er det 174, 180, 182.

- D. Nu skal vi beregne selve Midrange og dette er muligt ved at tage den største og mindste værdi og derved tage summen af disse to dividere med 2.

Midrange A	$\frac{154 + 189}{2} = \frac{343}{2} = 171,5$
Midrange B	$\frac{161 + 191}{2} = 176$

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

- Vi skal finde middelværdien herhenne, derfor skal vi bare ligesom før finde summen og dividere med antal.

Sum C	20 + 26 + 40 + 36 + 23 + 42 + 35 + 24 + 30 = 276
Mean C	$\frac{276}{9} = \frac{92}{3} \approx 30,66667$

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

- Vi skal finde medianen og her skal vi bare anvende den simple måde.

- Vi kan se, at vi har ulige værdier som vi skal bare sortere i rækkefølge og finde det midterste tal.

Median D	292,300,311,401,595,618,713
----------	-----------------------------

- Det kan ses at medianen er 401 fordi det er den midterste tal i den ulige række.

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

- Vi skal finde medianen og her skal vi bare anvende den simple måde.
- Vi kan se, at vi har ulige værdier som vi skal bare sortere i rækkefølge og finde det midterste tal.

Median E	656,684,702,764,856,1132,1133,1303
----------	------------------------------------

- Vi kan se, at vi har lige antal i rækkefølgen og derfor finder vi de 2 midterste tal og dividere dermed med 2.

$$Median E = \frac{764 + 856}{2} = 810$$

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

- Fordi vi skal finde typetallet, skal vi bare kigge efter hvor mange værdier som optræder mere end en gang.

Mode F	8, 9, 9, 14, 8, 8, 10, 7, 6, 9, 7, 8, 10, 14, 11, 8, 14, 11
Mode F (Sorteret)	6,7,7,8,8,8,8,9,9,9,10,10,11,11,14,14,14

- Vi kan se, at vi har markeret de værdier som optræder mere end en gang, med forskellige farver. Det skal understreges, at der er snak om en multimodal typetal fordi der optrædes mere end 3 værdier, mere end en gang.

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

- Fordi vi skal finde typetallet, skal vi bare kigge efter hvor mange værdier som optræder mere end en gang.

Mode G	110, 731, 1031, 84, 20, 118, 1162, 1977, 103, 752
Mode G (Sorteret)	20,84,103,110,118,731,752,1031,1162,1977

- Vi kan se, at der er ingen typetal fordi vi har ingen af de samme værdier som optræder mere end en gang, eftersom alle værdier er forskellige fra hinanden.

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

- Vi skal finde midrange og her kan det ses at vi finder største og mindsteværdien og derved dividerer med 2.

$$Midrange = \frac{8 - 1}{2} = \frac{7}{2} = 3,5$$

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)
- Occupation ("blue collar", "white collar", "unemployed")
 - Highest attained education ("low", "medium", "high")
 - Monthly salary
 - Civil status ("single", "married", "widow")
 - Number of children

- Vi skal klassificere her, hvilken kategori selve variablerne tilhører under:
- Dette er Kvalitativ Nominal, fordi det er farver som ikke er rankeret i hierarkisk orden, men er på lige fod med andre 😊
 - Dette er Kvalitativ Ordinal, fordi det er rankeret i hierarkisk orden som gør at der er forskel i niveau med andre.
 - Dette er Kvantitativ Kontinueret, fordi det kan både stige og falde men at det fortsætter i evighed - så længe man har et job.
 - Dette er Kvalitativ Nominal, fordi om man er i ægteskab eller ej har ingen hierarkisk betydning.
 - Dette er Kvantitativ Diskret, fordi dette kan være meget begrænset i kvantitet afhængigt af forholdet.

9. Evaluate the following statements as true or false:
- In statistics, a population always refers to humans.
 - A sample is a subset of the study population.
 - Inferential statistics are statistical techniques used to draw conclusions about one specific sample.
 - 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.

- Her skal vi svare på spørgsmål, som blandt andet står i selve sliden.
- Nej, det gør det ikke. En Population kan også være andre ting, som en hel masse genstande som stole som man ønsker at lave inferential statistik på.
 - Ja, det er rigtigt. Et sample er også kendt som en håndfuld gruppe af data som man opsamler fra en Population til at drage konklusioner til den store Populationsgruppe.
 - Nej, det er ikke rigtigt men det omvendte svar. Inferential statistik er der hvor vi anvender en sample gruppe til at drage konklusioner imod den store Populationsgruppe.
 - Nej, det er ikke rigtigt men det omvendte svar. Det er faktisk de 100 tilfældige elever fra freshmen klassen fra Odense Gymnasium, som er sample gruppe.

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

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

- Vi starter med at udregne range og her er det sådan at vi bare trækker største og mindsteværdien fra hinanden i de begge Brands.

$$\text{Range A: } 60 - 10 = 50$$

$$\text{Range B: } 45 - 25 = 20$$

- Nu skal vi udregne Variansen, og her i tilfældet er der lavet en tabel nedenfor hvor vi udøver forskellige steps.
- Først starter vi med at finde middelværdien.

Mean A	$10 + 60 + 50 + 30 + 40 + 20 = 210$	$\frac{210}{6} = 35$
Mean B	$35 + 45 + 30 + 35 + 40 + 25 = 210$	$\frac{210}{6} = 35$

- Nu trækker vi middelværdien fra de enkelte værdier fra datasættet og derved kvadrerer dem. Først starter vi med Brand A og derefter Brand B.

MinusA	35-10	60-35	50-35	35-30	40-35	35-20
Svar A	25	25	15	5	5	15
Kvadrat A	$25*25=625$	$25*25=625$	$15*15=225$	$5*5=25$	$5*5=25$	$15*15=225$

MinusB	35-35	45-35	35-30	35-35	40-35	35-25
Svar B	10	10	5	0	5	10
Kvadrat B	$10*10=100$	$10*10=100$	$5*5=25$	$0*0=0$	$5*5=25$	$10*10=100$

- Vi kan se, at vi har fundet vores kvadrerede værdier og derfor skal summen lægges sammen af dem.

Sum A	$625+625+225+25+25+225=1750$
Sum B	$100+100+25+0+25+100=350$

- Nu skal summen divideres med antal-1 og derved findes variansen.

Variansen A	$\frac{1750}{6-1} = 350$
Variansen B	$\frac{350}{6-1} = 70$

- Nu skal vi udregne Standardafvigelsen og dette er samme procedure, men det eneste vi skal gøre her, er at vi skal bare tage kvadratroden af de udregnede Varianser - dermed kan vi finde Standardafvigelsen.

Standardafvigelsen A	$\sqrt{350} = 5 \cdot \sqrt{14} \approx 18,70829$
Standardafvigelsen B	$\sqrt{70} = \sqrt{70} \approx 8,3666$

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

- a. **3**
- b. 6
- c. 9
- d. 81
- e. impossible to determine without knowing n.

- Vi skal her beregne Standardafvigelsen og derfor skal vi anvende den samme metode som tidligere fra Opgave 10, hvor vi bare tager kvadratroden af 9.

$$sd = \sqrt{9} = 3$$

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.

- Vi skal her trække 5 fra enhver tal der starter fra 10 og nedad imod 0.

$$10-5=5,$$

$$9-5=4,$$

$$8-5=3,$$

$$7-5=2,$$

$$6-5=1,$$

$$5-5=0,$$

$$4-5=-1,$$

$$3-5=-2,$$

$$2-5=-3,$$

$$1-5=-4,$$

$$0-5=-5$$

- Vi kan se, at det ender med at være 5.