

Statistisk Dataanalyse 2023

Instruktør: Vivek Misra

VELKOMMEN TIL SYDDANSK UNIVERSITET ODENSE!

Studiestartsprøven



- Skal være lavet færdigt inden 7. September.
- Man har 2 forsøg.
 - 1. Forsøg gælder til 7. September 2023.
 - 2. Forsøg gælder til 14. September 2023.
- Vælger man ikke at lave Prøven efter 2. forsøg!
 - **Udskrivning fra Skolen!!**
- Fordele
 - Studiestartsprøve giver indblik i at eleven følger med på Universitet!
 - Vælger man at skifte uddannelse på SDU året efter, er Studiestartsprøven betragtet sammen med Karaktersnittet fra Gymnasiet.



Biografi

- Fuldenavn: Vivek Misra
- Langua Franca: Hindi, Engelsk, Dansk og lidt Tysk
- Uddannelse: Software Engineering 3. Semester
- Email: vimis22@student.sdu.dk
- Fritid: Arbejder i Bilka Horsens + Youtube + Småprojekter

Statistik Øvelsestimer

- Jeg vil være tilstede i de 12 Øvelsestimer.
 - Skulle jeg være fraværende, skal studerende henvises til andre instruktører.
- Jeg lægger facit op (kun) for øvelsesopgaverne efter kl.16.
 - Min egen webside: vimis22.github.io.
- **Note: Pointgivende- og Eksamensopgaver bliver IKKE offentliggjort!**
- **I Statistik har i 2 Pointgivende Aktiviteter (frivillig).**
 - *Hvert Pointgivende Aktivitet bidrager 5%+5%=10% maks.*
 - *Pointgivende Aktiviteter giver overblik i mangler og forbedringsområder.*
 - *Mist ikke håbet, hvis du klarer dig dårligt men se det som en mulighed!*
- **MAN KAN FÅ 12 UDEN & MED POINTGIVENDE AKTIVITET!**
 - *Mister man Pointgivende Aktivitet er der ingen Re-Prøver for Pointgivende Aktiviteter!*

Exercise Class NR1

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. For each of the classroom, calculate the following:
 - A. Median
 - B. Mean
 - C. Mode
 - D. Midrange

Task 1 - Description

Here is the table:

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

Task 1A - Løsning

- To find the median, we need to sort the data given from the table.

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

- Classroom A: There are uneven numbers, so the middle numbers are selected.
- Classroom B: There are even numbers, so the average of the two numbers marked with green in the middle.

$$\frac{179 + 179}{2} = 179$$

Task 1B - Løsning

- To find the mean, we need will find the total sum and divide by amount.

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

$$\text{mean } A = \frac{2525}{15} = 168,3$$

$$\text{mean } B = \frac{2826}{16} = 176,6$$

Task 1C - Løsning

- To find the mode, we need to count the numbers which repeats itself.
- First, we need to sort the numbers so that we can get a clearer view.

Classroom A	154	156	158	160	162	165	169	169	170	171	172	175	175	180	189	
	155	157	159	161	163	166	170	170	171	172	173	176	176	181	190	
	156	158	160	162	164	167	171	171	172	173	174	177	177	182	191	
	157	159	161	163	165	168	172	172	173	174	175	178	178	183	192	
Classroom B	161	163	169	173	174	174	175	176	179	180	180	182	182	182	185	191
	162	164	170	174	175	175	176	177	180	181	181	183	183	183	186	192
	163	165	171	175	176	176	177	178	181	182	182	184	184	184	187	193
	164	166	172	176	177	177	178	179	182	183	183	185	185	185	188	194

- Classroom A: Here we have an Bimodal, because 169 and 175 are two numbers which repeat themselves.
- Classroom B: Here we have a Multimodal, because 174, 180 and 182 are the numbers which repeat themselves.

Task 1D - Løsning

- Find the average of largest and smallest value from the dataset.

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

$$\text{Midrange } A = \frac{154 + 189}{2} = 171,5$$

$$\text{Midrange } B = \frac{161 + 191}{2} = 176$$

Task 2 – Description / Solution

- Find the mean of the following data:
 - 26,26,40,36,23,42,35,24,30
- Solution, we need to sort the dataset, and then find the average.

Dataset (Sorted)	20	23	24	26	30	35	36	40	42
------------------	----	----	----	----	----	----	----	----	----

Mean	$20 + 23 + 24 + 26 + 30 + 35 + 36 + 40 + 42 = 276$ $\frac{276}{9} = 30,66$
------	--

Task 3 - Description / Solution

- Find the median of the following measurements:
 - 713,300,618,595,311,401,292
- Solution: Sort the dataset, and find the middle number based on amount of amount.

Dataset (Sorted)	292	300	311	401	595	618	713
-----------------------------	------------	------------	------------	------------	------------	------------	------------

Dataset (Median)	292	300	311	401	595	618	713
-----------------------------	------------	------------	------------	------------	------------	------------	------------

- **Note: We have uneven amount of numbers, because of total 7 numbers. Therefore there is only one number in the middle and that is 401.**

Task 4 - Description / Solution

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

Dataset (Sorted)	656	684	702	764	856	1132	1133	1303
-----------------------------	------------	------------	------------	------------	------------	-------------	-------------	-------------

Dataset Median	656	684	702	$\frac{764 + 856}{2} = 810$	1132	1133	1303
---------------------------	------------	------------	------------	-----------------------------	-------------	-------------	-------------

Note: That we have even amount of numbers which is now 8, and therefore we need to find the mean of the two middle numbers.

Task 5 – Description / Solution

- 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
- Solution: Sort the list, and find the repeating number.

Data	6	7	7	8	8	8	8	8	9	9	9	10	10	11	11	14
Data	6	7	7	8	8	8	8	8	9	9	9	10	10	11	11	14

NOTE: Because we have 5 types of numbers which are 7,8,9,10,11 – therefore we can say that there is a multimodal in this dataset.

Task 6 – Description / Solution

- Find the mode of the following measurements:
 - 110,731,1031,84,20,118,1162,1977,103,752
- Solution: Sort the list, and find the repeating number.

Data	20	84	103	110	118	731	752	1031	1162	1977
------	----	----	-----	-----	-----	-----	-----	------	------	------

- **NOTE: Because there is not any repeating number in the dataset, we need to conclude that there is no mode in the list!**

Task 7 – Description / Solution

- Find the midrange of these data:
 - 2,3,6,8,4,1
- Solution: Sort the data, and find the mean of the biggest and smallest value in the dataset.

Dataset (Sorted)	1	2	3	4	6	8
-----------------------------	----------	----------	----------	----------	----------	----------

$$\text{midrange} = \frac{8 + 1}{2} = 4,5$$

Task 8 – Description / Solution

- A searcher wants to collect data on 100 inhabitants living in one specific town. Classify the following collected variables according to their type. (Norminal, Ordinal, Discrete or Continous).
- Solution: Here the task is solved in the table according to the Powerpoint.

Description	Norminal	Ordinal	Discrete	Continous
Occupation (blue collar, white collar, unemployed)	X			
Highest attained education (low, medium, high)		X		
Montly salary				X
Civil status (single, married, widow)	X			
Number of children			X	

Task 9 – Description / Solution

- Evaluate the following statements are true or false:
 - Solution: Look at the table below.

Description	True	False
In statistics, a population always refers to humans		X
A sample is a subset of the study population.	X	
Inferential statistics are statistical techniques used to draw conclusions about one specific sample.		X
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.		X

Task 10 – Description

- Find the range, variance and standard deviation for the dataset for the samples of Brand A and Brand B - Paint.
 - Solution: Will be shown on the next page, where calculation have been made in the following order:
 - 10A for Range
 - 10B for Variance
 - 10C for Standard Deviation

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

Task 10A - Solution

- Solution: Sort the dataset in the ascending order, and subtract the smallest and biggest value from each other.

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

Brand A (Range)	$60 - 10 = 50$					
Brand B (Range)	$45 - 25 = 20$					

Task 10B - Solution

- Solution: Find the mean, and subtract each data from the mean. Thereafter square it and find the num where the division will give the Variance. **NOTE: Both of the Brands have mean calculated to 35.**

Brand A	Brand B
$10 - 35 = -25$	$35 - 35 = 0$
$20 - 35 = -15$	$45 - 35 = 10$
$30 - 35 = -5$	$30 - 35 = -5$
$40 - 35 = 5$	$35 - 35 = 0$
$50 - 35 = 15$	$40 - 35 = 5$
$60 - 35 = 25$	$25 - 35 = -10$
<i>Mean = 35</i>	<i>Mean = 35</i>

10B - Solution

Brand A	Brand B
$-25^2 = 625$	$0^2 = 0$
$-15^2 = 225$	$-10^2 = 100$
$-5^2 = 25$	$-5^2 = 25$
$5^2 = 25$	$0^2 = 0$
$15^2 = 225$	$5^2 = 25$
$25^2 = 625$	$10^2 = 100$
<i>Sum</i> = 1750	<i>Sum</i> = 250

$$\text{Variance } A = \frac{1750}{6 - 1} = 350$$

$$\text{Variance } B = \frac{250}{6 - 1} = 50$$

10C - Solution

Solution: To find the Standard Deviation, we just need to take the Square Root of the Variance.

$$\text{Variance } A = \frac{1750}{6 - 1} = 350$$

$$\text{Variance } B = \frac{250}{6 - 1} = 50$$

$$\text{Standard Deviation } A = \sqrt{350} = 18,70$$

$$\text{Standard Deviation } B = \sqrt{50} = 7,07$$

Task 11 – Description / Solution

- 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.

Solution: Just take the squareroot of the variance!

$$\textbf{Standard Deviation} = \sqrt{9} = 3$$

Result: A is the right answer!

Task 12 – Description

- 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**

Task 12 – Solution

- Solution: To solve this task, we will subtract 5 from each number going from 10 and down to 0.

Solution	Sum of all subtractions
$10 - 5 = 5$	$5 + 4 + 3 + 2 + 1 + 0 + (-1) + (-2) + (-3) + (-4) + (-5) = 0$
$9 - 5 = 4$	
$8 - 5 = 3$	
$7 - 5 = 2$	
$6 - 5 = 1$	
$5 - 5 = 0$	
$4 - 5 = -1$	
$3 - 5 = -2$	
$2 - 5 = -3$	
$1 - 5 = -4$	
$0 - 5 = -5$	

- **The answer is D, which is "none of these".**
- The reason is clear and that is because the distribution remains the same if you subtract all the numbers from the standard deviation.
- Dansk: Hvis du trækker alle værdierne fra bliver spredningen den samme ved standardafvigelsen og gør blandt andet at middelværdien ændres.

Tak for i dag!

Instruktør: Vivek Misra