

Afsluttende STATISTIK forløb for 10K over tre lektioner

HUSK PAUSER!

I har allerede et godt tag i statistiske undersøgelser; håber jeg! I skal nu afslutte emnet over tre lektioner. **Alle opgaverne skal afleveres** (sendes) inden mandag d. 23.3.2020.

Repetition (slå op i [Formler og Fagord](#), hvis I har glemt noget!):

1. Repetition (afleveres)

- Nævn de seks deskriptorer, og hvad de betyder
- Lav en hyppighedstabel over flg. data, hvor man har spurgt 50 elever om antal fag i skolen:

Person	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Antal fag	6	7	8	9	12	6	8	8	9	8	6	7	9	5	9	7	5	9	7	8	9	5	7	6	8
Person	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Antal fag	9	7	9	8	7	10	11	12	9	10	5	9	8	7	8	9	7	10	8	10	8	10	10	8	8

- Find de seks deskriptorer, MEGET gerne ved brug af Excel
- Tegn et søjlediagram over fordelingen
- Tegn et cirkeldiagram over fordelingen

2. Grupperede data (intervalldata)

I denne opgave skal I bruge/lære om **grupperede data**, **kumuleret hyppighed**, **kumuleret frekvens** og **sumkurver**.

Grupperede data er blot data, som er placerede i intervaller. Hvis man fx spørger en lang række personer om deres alder, vil det være praktisk at lave en hyppighedstabel med intervaller, fx 10-års intervaller (210):

Men først lidt information:

- Intervaller** skrives normalt via kantede parenteser, fx: [20, 30]. Dette betyder fra 20 til 30. Vejen som parenteserne vender, angiver om endepunkterne er med: i [20, 30] er både 20 og 30 er med, mens]20, 30[hverken har 20 eller 30 med.

Hyppighedstabel over alder					
Alder	Hyppighed	Frekvens		Kumuleret hyppighed	Kumuleret frekvens
		I brøk	I procent		
[0;10]	21	21/210	10,0%	21	10,0%
]10;20]	33	33/210	15,7%	54	25,7%
]20;30]	44	44/210	21,0%	98	46,7%
]30;40]	44	44/210	21,0%	142	67,6%
]40;50]	35	35/210	16,7%	177	84,3%
]50;60]	18	18/210	8,6%	195	92,9%
]60;70]	9	9/210	4,3%	204	97,1%
]70;80]	4	4/210	1,9%	208	99,0%
]80;90]	2	2/210	1,0%	210	100,0%
I alt	210	1	100,0%		

- De røde kolonner, **kumuleret hyppighed** og **kumuleret frekvens**, ser man sjældnere i en hyppighedstabel. Kumuleret betyder sammenlagt; man lægger altså de ovenstående værdier sammen. Hvis man indtegner de kumulerede frekvenser og/eller hyppigheder, kalder man kurven for en **sumkurve**. Hvis man er i tvivl om begreberne, kan man, som altid, se i [Formler og Fagord](#). Man kan også se en video fra webstedet: [FriViden matematik](#). Video 04 er om sumkurver. Nogle af de andre videoer er også gode! Desuden har jeg lavet denne PowerPoint til jer, som I måske også kan bruge! Men nu til opgaven →

Nedenstående tabel viser aldersfordelingen i en mindre dansk kommune:

Alder	Antal personer
0 - 9	1.851
10 - 19	1.561
20 - 29	1.995
30 - 39	2.226
40 - 49	2.009
50 - 59	1.966
60 - 69	1.287
70 - 79	984
80 - 89	492
90 - 99	80
100 - 109	7
I alt	14.458

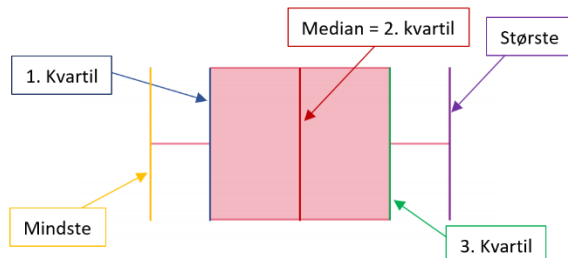
Afleveres:

- Indtast den samlede hyppighedstabel i Excel. Den skal indeholde:
Aldersinterval (obs.), hyppighed, frekvens, kumuleret hyppighed og kumuleret frekvens
- Find de seks deskriptorer (gerne via Excel). **Jeg har tidligere lagt en note om Excel på OneDrive!**
Husk det er i intervaller. Det hedder altså ikke størsteværdi, men største interval.
- Tegn et søjlediagram over for fordelingen
- Tegn et cirkeldiagram over fordelingen
- Tegn en sumkurve over fordeling

3. Boksplot

Nu er I næsten verdensmestre i statistik. I mangler kun boksplot. Hvad boksplot er, hvad man bruger det til, og en opgave i brugen af boksplot følger nedenfor. Et boksplot er MEGET brugt inden for forskningens verden, da det på en god og hurtig måde at beskrive en statistisk undersøgelses fordeling på.

For at tegne et **boksplot**, skal man bruge et datasæts **mindsteværdi**, **kvartilsæt** og **størsteværdi**. Men heldigvis, er det noget som GeoGebra gør automatisk se senere. Selve boksplottet ser således ud:



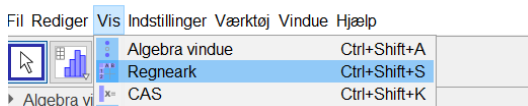
I kender medianen, også kaldet 2. kvartil, men hvad er, og hvordan finder man, 1.- og 3. kvartil?

Kvart betyder jo en fjerdedel, så 1. kvartil er "nummer 25 % observation", mens 3. kvartil er "nummer 75 % observation". Hvis man fx har 23 observationer: 1,2,4,5,7,8,9,4,6,7,9,3,2,5,7,8,9,3,4,5,3,4,5 og sætter dem i rækkefølge: 1,2,2,3,3,3,4,4,4,4,5,5,5,5,6,6,6,6,8,8,9,9,9 ses **1.kvartil** - **2. kvartil** og **3. kvartil**

Hvis observationsantallet havde været et lige antal, kunne man selv vælge kvartilsiden, ligesom I er vant til for medianens vedkommende (man kan også bare spørge Excel eller GeoGebra).

I disse videoer, kan I se: [Hvad et boksplot er](#) og [hvordan man laver et boksplot](#).

VIGTIGT: For at få regnearket frem i GeoGebra skal I klikke i menuen vis regneark:



Nu skal I selv prøve. Som hjælp kan I selvfølgelig stadigvæk benytte [Formler og Fagord!](#)

To 10. klassers forbrug af Facebook pr. dag i minutter																					
10. A																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	62	55	110	40	0	55	14	158	25	81	62	15	51	94	26	75	52	32	59	55	12
10. B																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17				
	26	45	48	74	52	16	42	65	72	46	51	61	43	43	53	33	18				

Afleveres:

- Find de seks deskriptorer for hver klasse (gennemsnit, maksimum, typetal, etc.)
- Hvad er 1. kvartil for de to klasser?
- Hvad er 3. kvartil for de to klasser?
- Tegn boksplot for hver klasse
- Sammenlign de to klassers Facebookforbrug