

# Matematik i Biologi A

Jette Vestergaard

FIP biologi

Silkeborg Gymnasium onsdag d. 15. januar 2020

Hvidovre Gymnasium torsdag d. 16. januar 2020

# Matematik i Biologi A

*Matematik i de vejledende sæt 1 og 2, specielt*

- ▶ Boksplot (vejledende sæt 2 opgave 4)
- ▶ Middelværdi og spredning (vejledende sæt 2 opgave 2)
- ▶ ”Foretag relevant databehandling” (vejledende sæt 1 opgave 1)

# Boksplot

- hvad er det og hvad kan det bruges til?

- ▶ Et boksplot er et diagram, som illustrerer de 5 værdier, der indgår i *det udvidede kvartilsæt* for en række observationer.
- ▶ De 5 værdier er: minimum , nedre kvartil (Q1) , median (Med) , øvre kvartil (Q3), maksimum.

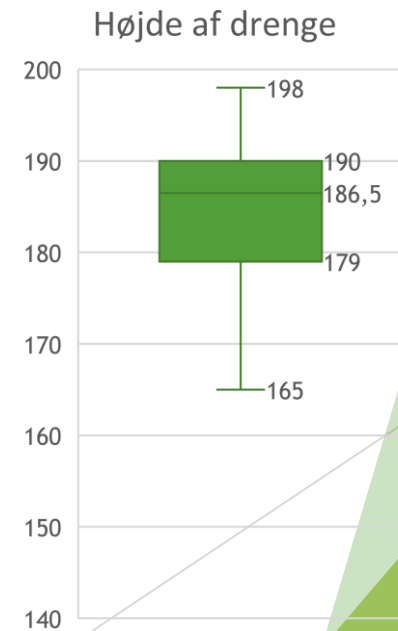
## ▶ *Eksempel*

- ▶ I en klasse måles højden af alle drengene.

▶ 165 , 172 , 179 , 180 , 185 , 188 , 188 , 190 , 191 , 198



- ▶  $Q1=179$  ,  $Med=186,5$  ,  $Q3=190$
- ▶ Det udvidede kvartilsæt: ( 165 , 179 , 186,5 , 190 , 198 )



► *Boksplots er smarte hvis man skal sammenligne flere datasæt.*

► **Eksempel**

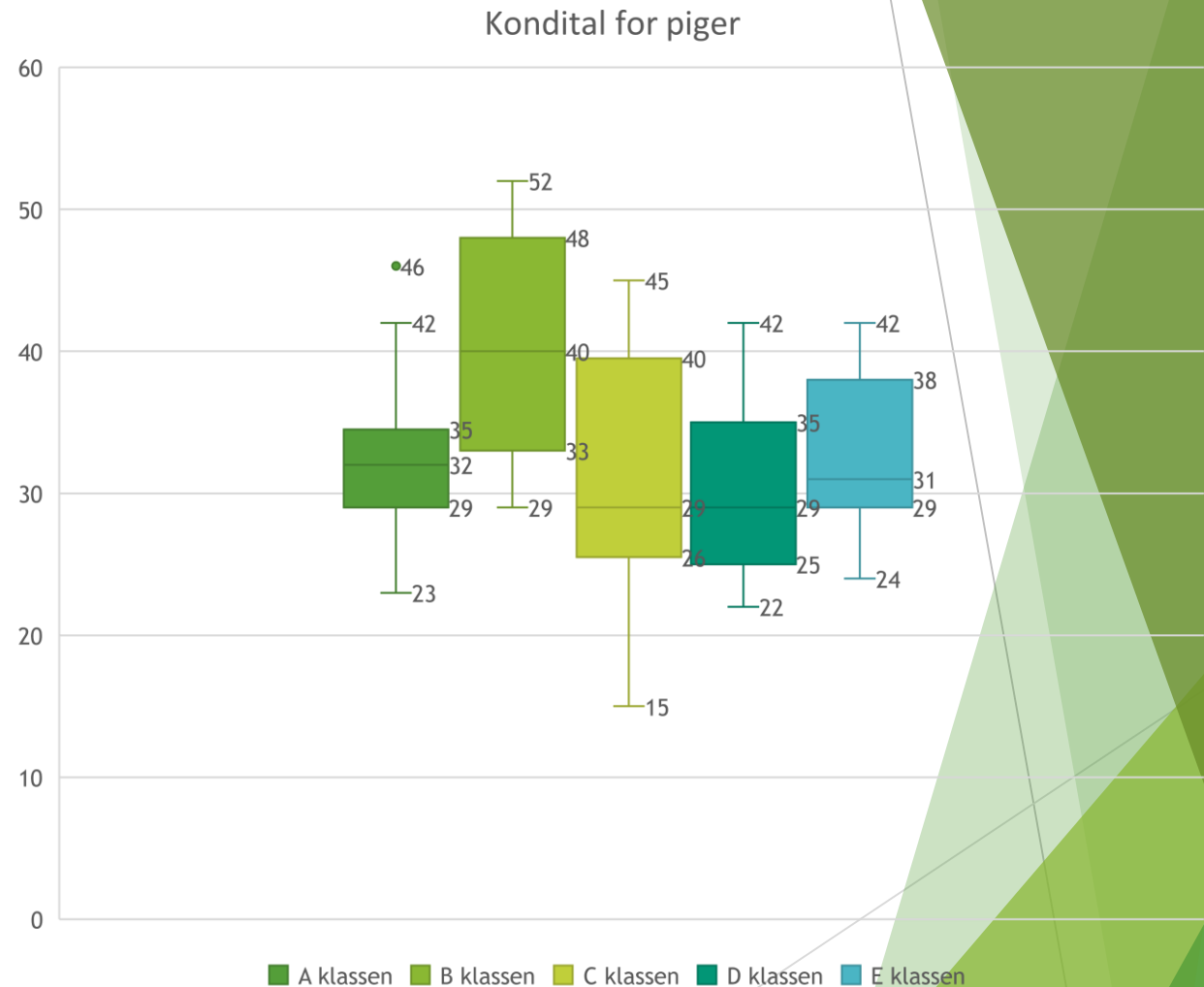
► På et gymnasium testes de nye 1.g elever, og deres kondital noteres. Da drenges og pigers kondital vurderes forskelligt, opdeles data efter køn. Tabellen viser det udvidede kvartilsæt for **pigerne** i 5 klasser.

	A klassen	B klassen	C klassen	D klassen	E klassen
Minimum	23	29	15	22	24
Nedre kvartil, Q1	29	33	25,5	25	29
Median, Med	32	40	29	29	31
Øvre kvartil, Q3	34,5	48	39,5	35	38
Maksimum	46	52	45	42	42

► Det er svært at danne sig et overblik over data, når de præsenteres på denne måde.

- ▶ Det ses tydeligt, at det giver et meget bedre overblik over data, når de præsenteres på denne måde.
- ▶ Ved en sammenligning kan være en god ide at komme ind på følgende 3 ting:
- ▶ Interessante kvartiler
- ▶ Variation i klasserne
- ▶ Konklusion
- ▶ Det ses, at B klassen generelt har et bedre kondital end de andre klasser. De har en "top" der ligger over de andre klasser, mens deres svageste elev er bedre end "bunden" i de andre klasser.
- ▶ Des ses desuden, at eleverne i E klassen er mest ens mht kondital, mens eleverne i C klassen er mest forskellige mht kondital.

- ▶ Boksplots for pigerne i de 5 klasser:



# Middelværdi og spredning

- hvad er det og hvad kan det bruges til?

- ▶ Middelværdien for en observationsrække er gennemsnittet af observationerne.

- ▶ Formel:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

- ▶ Spredningen for en observationsrække er et mål for variationen i observationsrækken.

- ▶ Formel:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}}$$

► *Eksempel*

- Vi betragter eksemplet med pigernes kondital. Middelværdien og spredningen beregnes for pigerne i hver klasse.

	A klassen	B klassen	C klassen	D klassen	E klassen
Middelværdi	33	40	32	30	32
Spredning	6,2	8,0	8,9	6,3	5,6

- Det ses, at B-klassen klart har den højeste middelværdi, altså har de som gennemsnit det højeste kondital.
- Det ses desuden, at spredningen er størst i C-klassen mens den er mindst i E-klassen. Det betyder, at variationen er størst i C-klassen mens variationen er mindst i E-klassen. Altså er eleverne i C-klassen mest forskellige mht kondital, mens eleverne i E-klassen er mest ens mht kondital.

# ”Relevant databehandling”

- hvad er det?

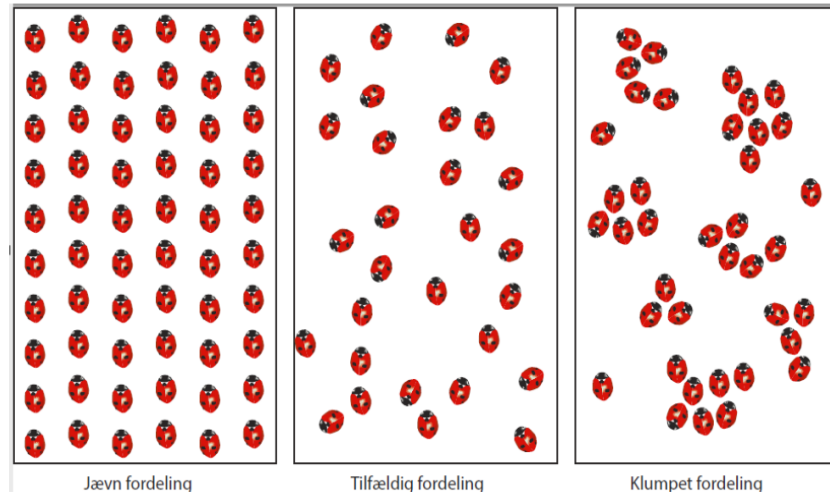
Afhænger af data.

- ▶ Hvis data består af flere observationsrækker, hvor der hver gang er målt på det samme men i flere undergrupper, vil det ofte være relevant at lave noget sammenlignende deskriptiv statistik.
- ▶ Hvis data består af to observationsrækker med sammenhørende værdier (altså nogle målepunkter med to variabler) vil det ofte være relevant at plote data og lave forskellige regressioner for at undersøge data.



# Eksempel

## - Vejledende Sæt 1 opgave 1



Figur 1  
Tre eksempler på fordeling af individer i en biotop.

Kast nr.	Diameter 3 cm	Diameter 6 cm	Diameter 9 cm
1	0	4	0
2	2	3	8
3	1	2	3
4	0	4	6
5	1	1	5
6	1	3	7
7	2	3	4
8	2	3	5
9	2	3	3
10	1	3	8
11	1	3	4
12	2	3	4
13	1	2	5
14	0	2	8
15	0	4	6

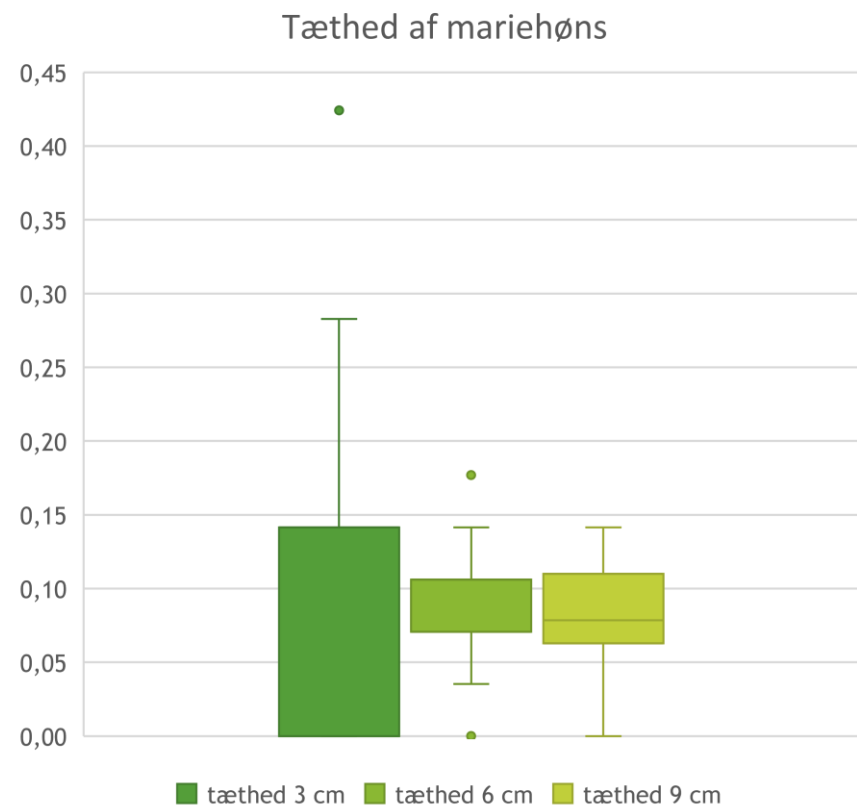
- ▶ En gruppe elever udfører tre serier á 50 stikprøver. I første serie anvendes en plastikskive med et areal på  $7,07 \text{ cm}^2$ , i anden serie en skive med et areal på  $28,27 \text{ cm}^2$  og i tredje serie en skive med et areal på  $63,62 \text{ cm}^2$ .
- ▶ Den faktiske tæthed af mariehøns på papiret er  $0,064$  pr.  $\text{cm}^2$ .
- ▶ *Foretag relevant databehandling, der kan vise, hvilken af de tre stikprøvestørrelser, der kommer nærmest på den faktiske tæthed af mariehøns på papiret.*

	Diameter 3 cm	Diameter 6 cm	Diameter 9 cm	Populationstæthed (antal/ $\text{cm}^2$ )
Middelværdi af antal mariehøns	0,92	2,76	5,34	
areal af skive ( $\text{cm}^2$ )	7,07	28,27	63,62	
Gennemsnitligt antal mariehøns pr $\text{cm}^2$	0,130	0,098	0,084	0,064

- ▶ Det ses, at den stikprøve, der kommer tættest på den faktiske tæthed af mariehøns på papiret, er stikprøven, hvor diameteren er 9 cm.

- ▶ Man kunne også lave boksplots for at undersøge de tre observationsrækker nærmere.
- ▶ Rækkerne er imidlertid ikke direkte sammenlignelige.
- ▶ For hver observation beregnes estimatet for tætheden ved at dividere observationen med arealet af skiven.

F	G	H
tæthed 3 cm	tæthed 6 cm	tæthed 9 cm
0,00	0,07	0,00
0,28	0,14	0,13
0,14	0,04	0,05
0,00	0,11	0,09
0,14	0,11	0,08
0,14	0,11	0,11
0,28	0,11	0,06



- ▶ Det ses, at der er størst variation i de estimerede tætheder ved brug af den mindste skive.
- ▶ Men svaret på opgaven er ikke boksplottene. Det er de estimerede tætheder fra forrige side.