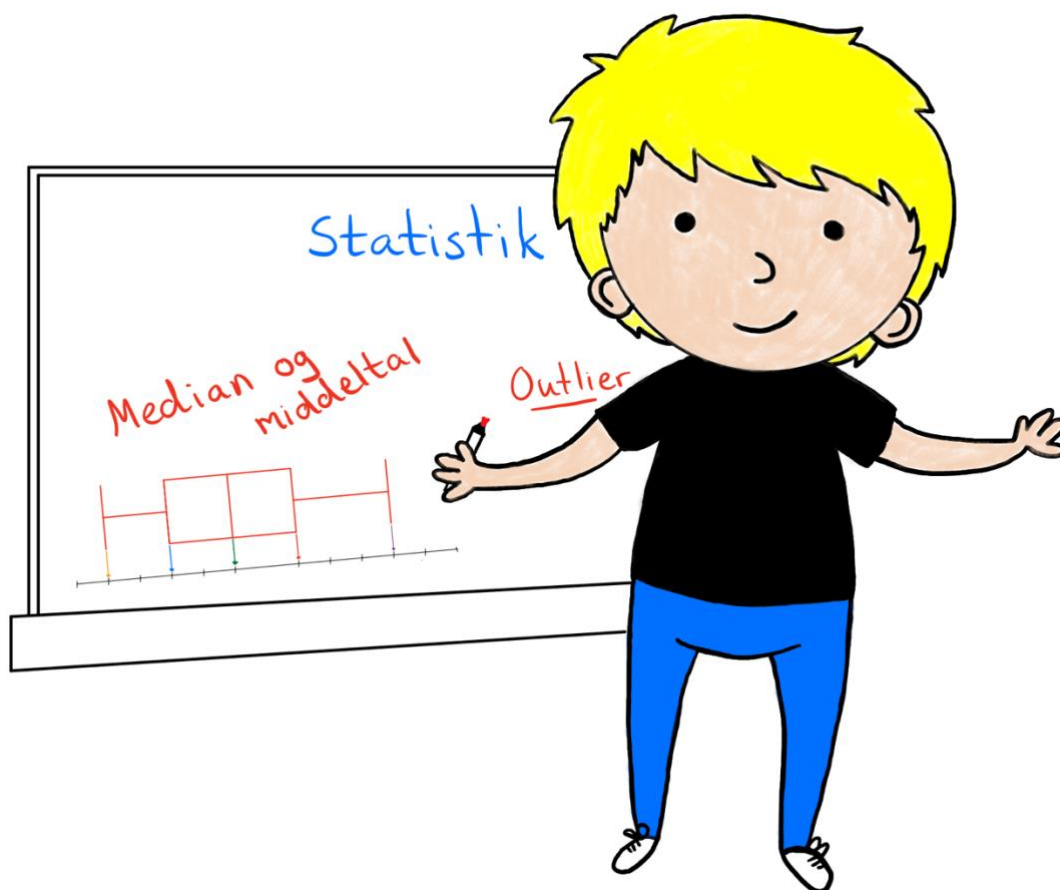


SCHACKS NOTER

KAPITEL 5 - UGRUPPERET STATISTIK



Udarbejdet af Kristian Schack Blom, Århus Akademi

Fejlrettelser samt ris og ros bedes sendes til aaks@aarhusakademi.dk

Den nyeste version af noterne kan altid findes på www.matematikC.dk

Indholdsfortegnelse

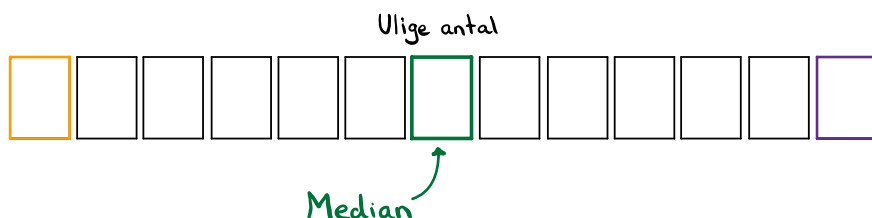
TÆLLEMETODEN	2
BOKSPLOT	4
UGRUPPERET STATISTIK I GEOGEBRA	8
OUTLIER	11

TÆLLEMETODEN

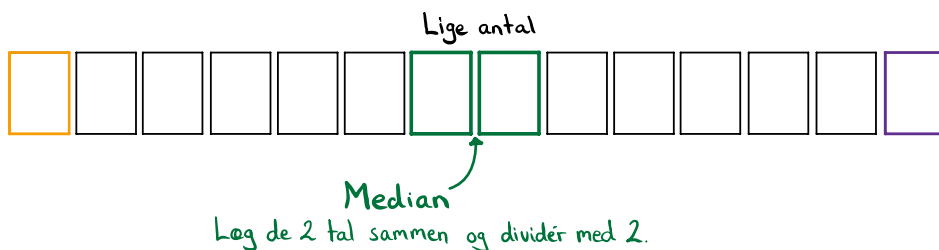
Alle begreber til ugrupperet statistik er defineret og beskrevet i formelsamlingen, så nedenstående er en uddybning og eksemplificering af disse begreber.

For at bestemme **mindste observation** ($\min$), **nedre kvartil** (Q_1), **median** (m), **øvre kvartil** (Q_3) og **største observation** ($\max$) benyttes **tællemetoden**, når observationerne er opstillet i rækkefølge.

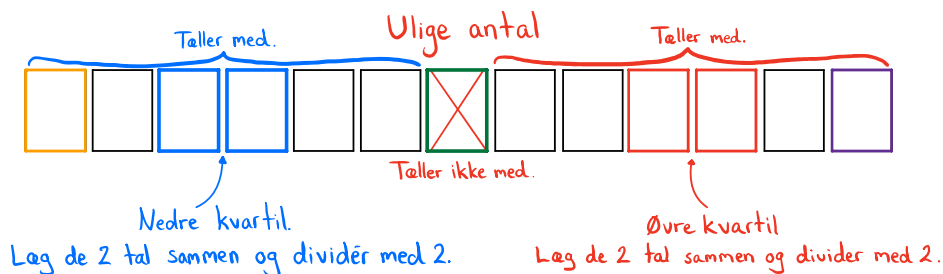
Ved et **ulige** antal observationer, er medianen den midterste observation:



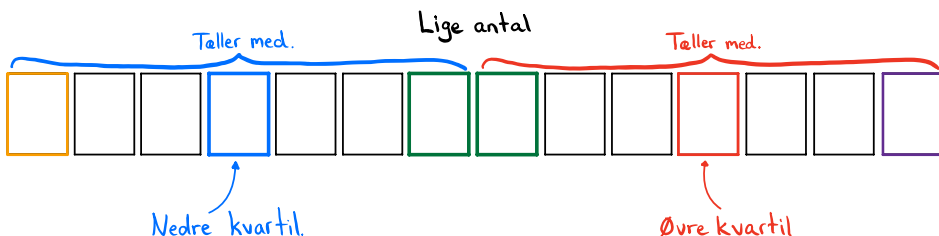
Ved et **lige** antal observationer, er medianen midt mellem de to midterste observationer:



Hvis medianen er den midterste observation, tæller denne ikke med, når nedre og øvre kvartil skal bestemmes:



Hvis medianen er midt mellem de to midterste observationer, tæller tallene med, når henholdsvis nedre og øvre kvartil skal bestemmes:

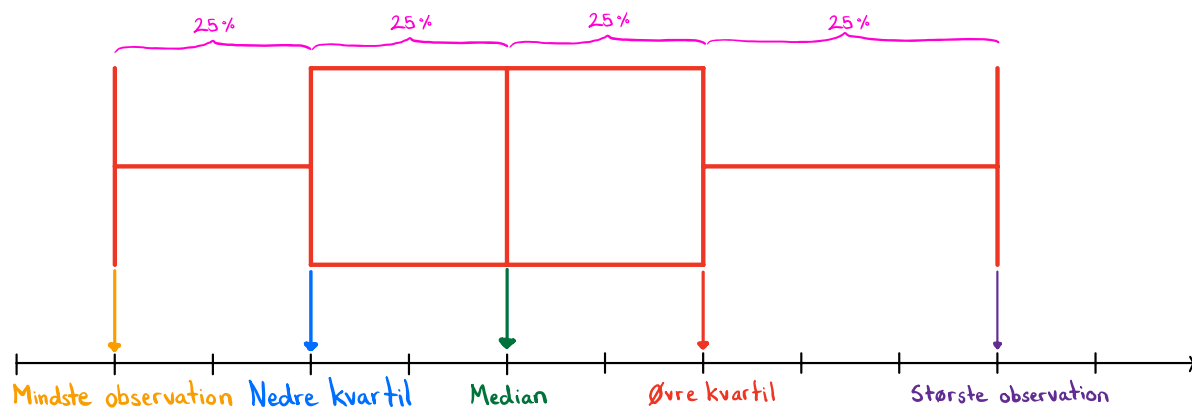


 [Kort video om bestemmelse af median, nedre kvartil og øvre kvartil med tællemetoden.](#)

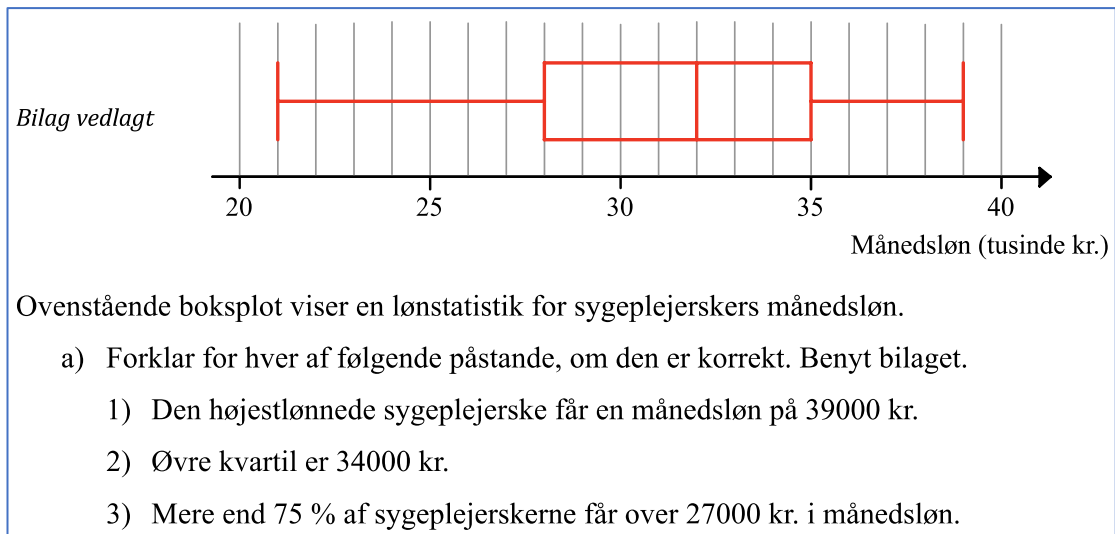
BOKSPLOT

På boksplottet kan vi aflæse mindste observation, nedre kvartil, median, øvre kvartil og største observation.

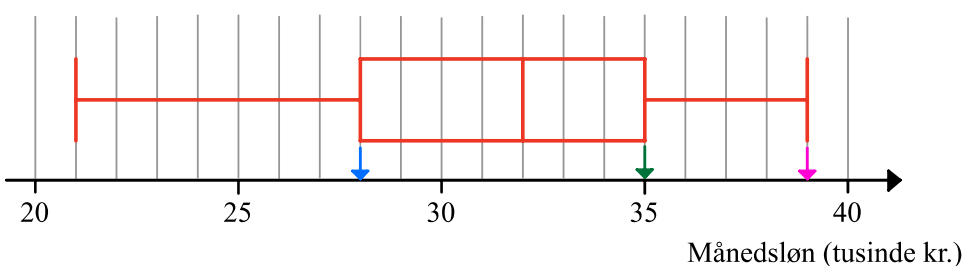
Mellem hver af disse værdier er der 25% af observationerne.



 **YouTube** [Kort video om boksplots.](#)

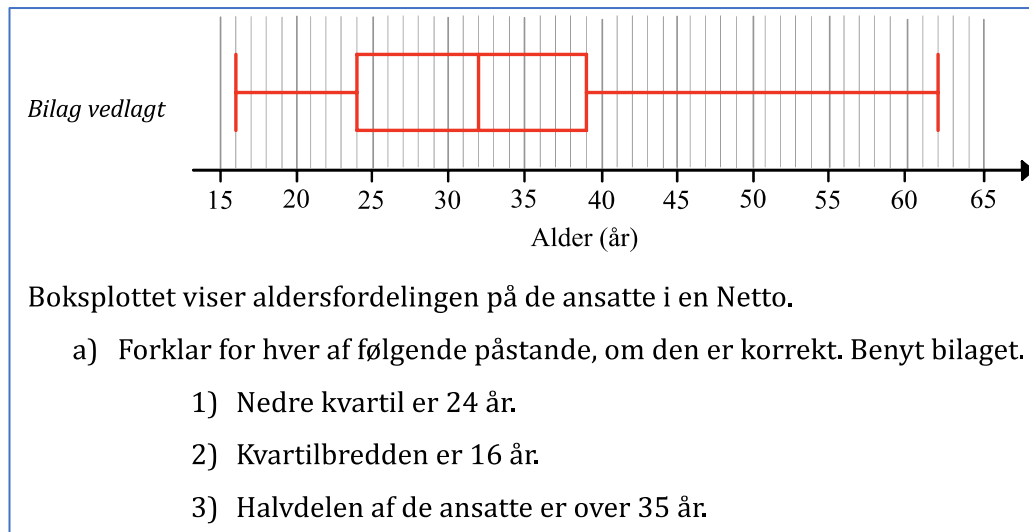


- a) Den højestlønnede sygeplejerske får en månedsløn på 39000 kr., hvilket ses på den pinke pil på bilaget. Så første påstand er korrekt.
- Øvre kvartil er ikke 34000 kr., da den grønne pil på bilaget viser, at øvre kvartil er 35000 kr. Så anden påstand er ikke korrekt.
- Mere end 75 % af sygeplejerskerne får over 27000 kr., da den blå pil på bilaget viser, at nedre kvartil er større end 27000 kr., nemlig 28000 kr. Så tredje påstand er korrekt.



YouTube [Kort gennemgang af opgaven.](#)

GENNEMGANG AF EKSAMENSLIGNENDE OPGAVE #5.3 (DELPRØVE 1)

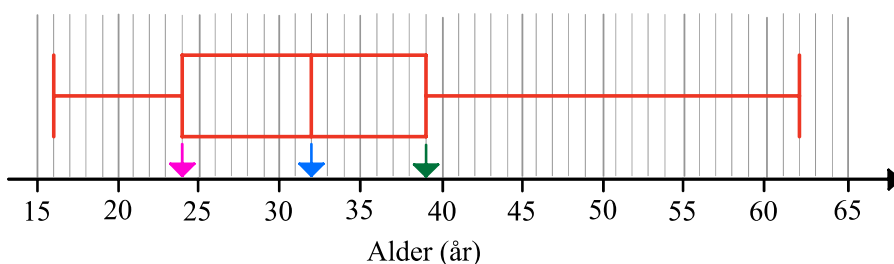


a) Nedre kvartil er 24 år, hvilket ses på den pinke pil på bilaget.
Så første påstand er korrekt.

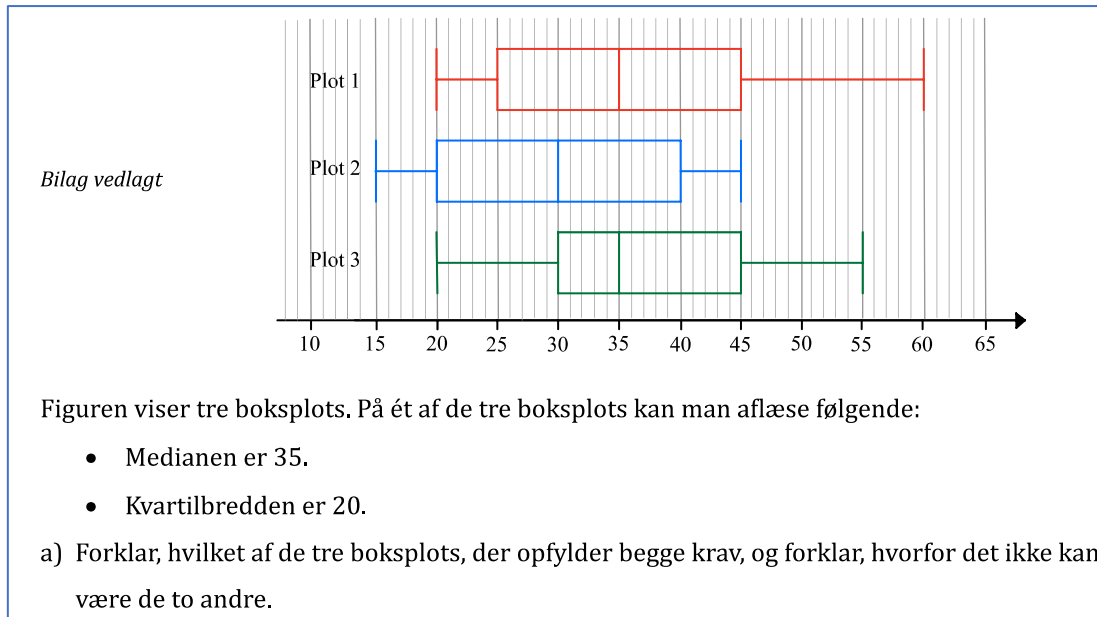
Øvre kvartil er 39 år (se den grønne pil på bilaget). Da kvartilbredden udregnes med $Q_3 - Q_1$, er kvartilbredden $39 - 24 = 15$ år.

Så anden påstand er ikke korrekt, da den siger, at kvartilbredden er 16 år.

Tredje påstand er heller ikke korrekt, da medianen er 32 år - og ikke 35 år (se den blå pil på bilaget).



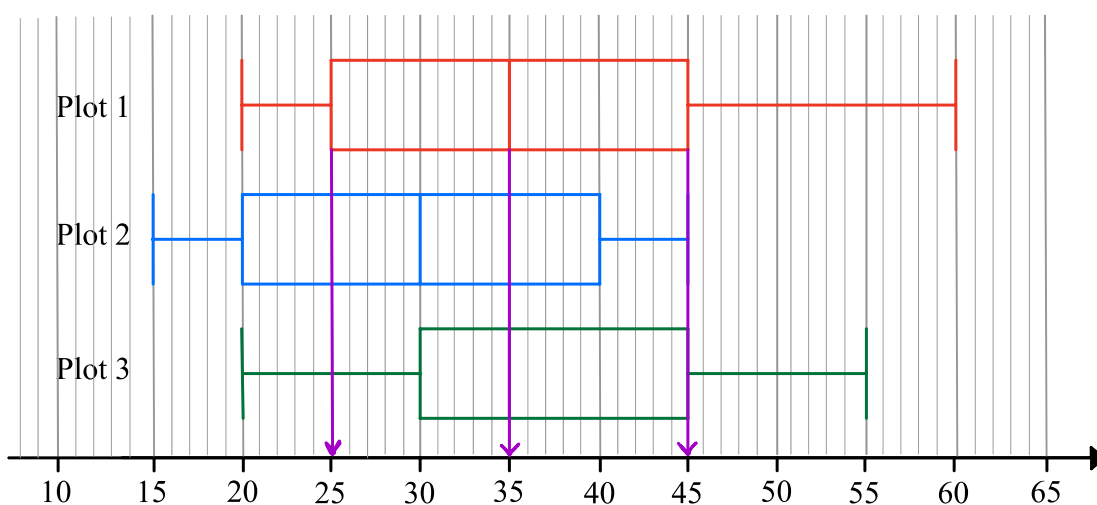
[Kort gennemgang af opgaven.](#)



a) Jeg kan se, at det er Plot 1, der opfylder begge krav, da jeg aflæser nedre kvartil til 25, medianen til 35 og øvre kvartil til 45 (se lilla pile på bilaget). Kvartilbredden er derfor $45 - 25 = 20$. Derfor er begge krav opfyldte.

Ved Plot 2 er kvartilbredden 20, men medianen er 30.

Ved Plot 3 er medianen 35, men kvartilbredden er 15.



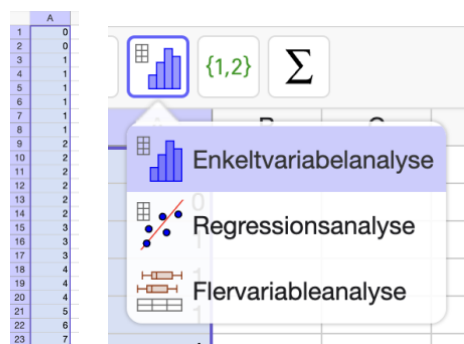
YouTube [Kort gennemgang af opgaven.](#)

UGRUPPERET STATISTIK I GEOGEBRA

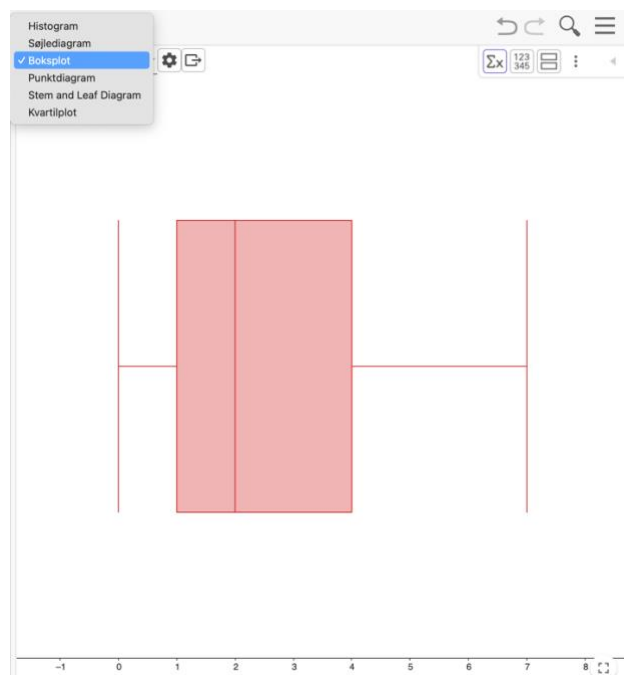
Det er meget nemt at arbejde med ugrupperet statistik i Geogebra. Først skal vi vælge "Regneark", hvor vi skal indsætte vores data. I dette eksempel bruger vi observationerne fra den eksamenslignende opgave #5.1, som vi selv indtaster. Til eksamen ligger data nogle gange i et bilag (se eksamenslignende opgave #5.5):

	A
1	0
2	0
3	1
4	1
5	1
6	1
7	1
8	1
9	2
10	2
11	2
12	2
13	2
14	2
15	3
16	3
17	3
18	4
19	4
20	4
21	5
22	6
23	7

Vi skal derefter markere data og vælge "Enkeltvariabelanalyse".



Der dukker nu et nyt vindue op, hvor vi blandt andet kan få Geogebra til at tegne et boksplot ved hjælp af menuen til venstre:



Vi kan også få Geogebra til at angive alle de ting, vi har brug for, ved at trykke på dette ikon:



Så kommer følgende vindue frem:

n	23
Middel	2.4783
σ	1.7905
s	1.8308
Σx	57
Σx^2	215
Min	0
Q1	1
Median	2
Q3	4
Maks	7

 **YouTube** [Kort video om ugrupperet statistik i Geogebra.](#)

Tabellen viser gennemsnitsalderen i landets kommuner.

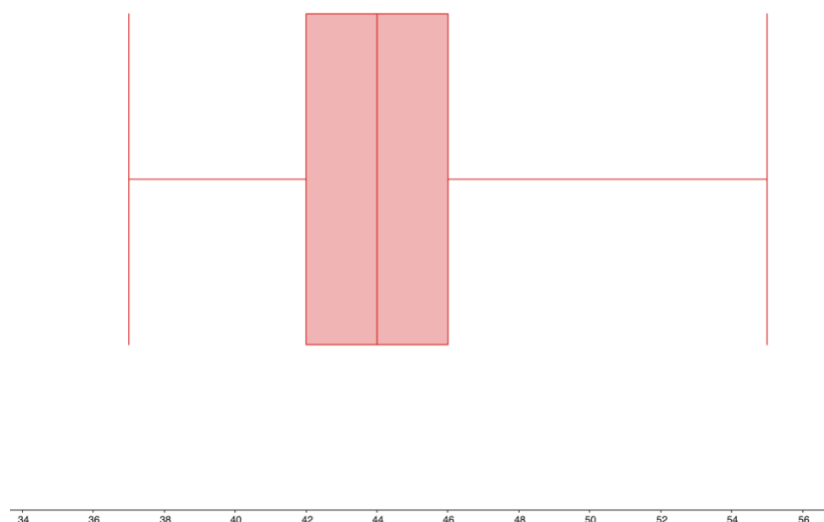
Gennemsnitsalder	41	47	...	45	49
------------------	----	----	-----	----	----

Alle tabellens 104 data findes i vedhæftede fil.

- Tegn boksplot for gennemsnitsalderen i landets kommuner.
- Bestem median og middeltal.
- Hvad fortæller disse to tal om gennemsnitsalderen i landets kommuner?

a)

Jeg har importeret alle observationerne ind i et regneark i Geogebra, hvorefter jeg har lavet enkeltvariabelanalyse. Jeg får derefter Geogebra til at tegne et boksplot:



b)

Jeg skal nu bestemme medianen og middeltallet. Jeg får Geogebra til at vise statistikken:

n	104
Middel	44.1058
σ	3.0411
s	3.0558
Σx	4587
Σx^2	203275
Min	37
Q1	42
Median	44
Q3	46
Maks	55

Her kan jeg aflæse, at medianen er 44 år, og middeltallet er 44,1 år.

c)

Medianen fortæller, at i 50% af landets kommuner er gennemsnitsalderen 44 år eller derunder.

Middeltallet fortæller, at gennemsnitsalderen i gennemsnit er 44,1 år i landets kommuner.

[Kort gennemgang af opgaven.](#)

OUTLIER

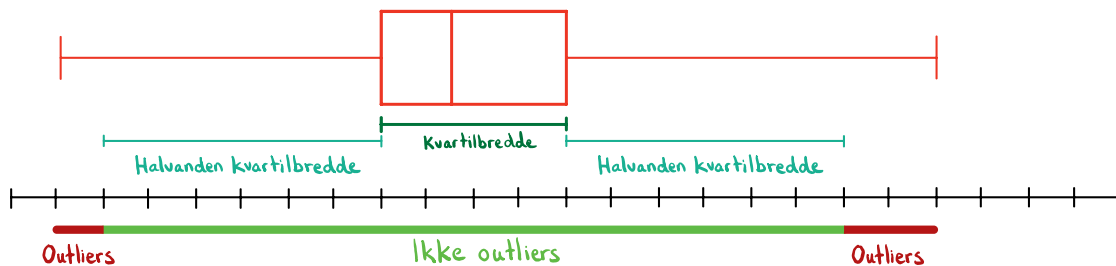
Hvis der er observationer, der ligger mere end halvanden kvartilbredde under nedre kvartil eller mere end halvanden kvartilbredde over øvre kvartil, så kaldes de outliers.

Vi skal derfor først udregne kvartilbredden ved at trække nedre kvartil fra øvre kvartil ($Q_3 - Q_1$).

Derefter skal vi gange kvartilbredden med 1,5.

Til sidst skal vi trække halvanden kvartilbredde fra nedre kvartil og lægge tallet til øvre kvartil. Vi har nu fundet **nedre grænse og øvre grænse for, hvornår en observation er en outlier**. Hvis en observation er under nedre grænse eller over øvre grænse, er der tale om en outlier.

Det kan illustreres således:



EKSEMPEL

I datasættet 1, 2, 3, 4, 5, 5, 5, 6, 6, 7, 7, 8, 9, 14, 16 er nedre kvartil og øvre kvartil markeret.

Kvartilbredden er $Q_3 - Q_1 = 8 - 4 = 4$, og halvanden kvartilbredde er $1,5 \cdot 4 = 6$.

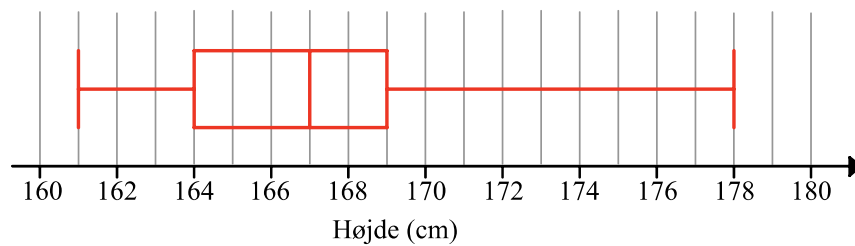
Trækkes halvanden kvartilbredde fra nedre kvartil, får vi en nedre grænse på $4 - 6 = -2$.

Lægges halvanden kvartilbredde til øvre kvartil, får vi en øvre grænse på $8 + 6 = 14$.

Observationen 16 er derfor en outlier, da observationen er større end den øvre grænse.

 [Kort video om outliers og eksemplet.](#)

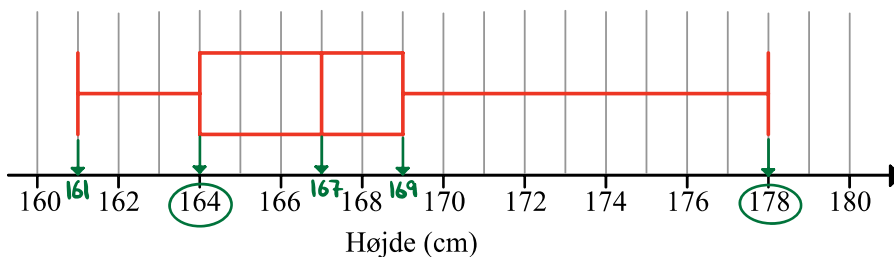
Bilag vedlagt



Figuren viser et boksplot for fordelingen af pigernes højde på et håndboldhold.

- Aflæs det udvidede kvartilsæt for pigernes højde. Brug bilaget.
- Bestem kvartilbredden, og benyt denne til at undersøge, om den højeste pige på håndboldholdet er en outlier.

- Jeg har markeret det udvidede kvartilsæt med 5 grønne pile på bilaget.
 Altså er mindste observation 161 cm, nedre kvartil 164 cm, medianen 167 cm, øvre kvartil 169 cm og største observation 178 cm.
- Jeg bestemmer kvartilbredden ved $Q_3 - Q_1$, så kvartilbredden er $169 - 164 = 5$ cm.
 Jeg udregner nu $Q_3 + 1,5 \cdot \text{kvartilbredde} = 169 + 1,5 \cdot 5 = 169 + 7,5 = 176,5$ cm.
 Alle observationer større end 176,5 cm er outliers, så den højeste pige på 178 cm er en outlier.



[Kort gennemgang af opgaven.](#)

Tabellen viser de 20 fodboldholds antal point i 2019-2020-sæsonen af den engelske fodboldliga, Premier League.

Antal point	21	34	...	66	81	99
-------------	----	----	-----	----	----	----

Alle tabellens 20 data findes i vedhæftede fil.

a) Bestem kvartilsættet for de 20 fodboldholds antal point.

b) Bestem kvartilbredden.

Holdet med flest point fik 99 point.

c) Benyt kvartilbredden til at gøre rede for, at dette antal point er en outlier.

a)

Jeg har importeret alle observationerne ind i et regneark i Geogebra, hvorefter jeg har lavet enkeltvariabelanalyse. Jeg får derefter Geogebra til at vise følgende:

n	20
Middel	52.4
σ	17.3245
s	17.7746
Σx	1048
Σx^2	60918
Min	21
Q1	40
Median	53
Q3	60.5
Maks	99

Altså er nedre kvartil 40 point, medianen 53 point og øvre kvartil 60,5 point.

b)

Jeg skal bestemme kvartilbredden, hvilket jeg gør med formlen $kvartilbredde = Q_3 - Q_1$. Jeg indsætter tallene:

$$kvartilbredde = 60,5 - 40 = 20,5$$

Altså er kvartilbredden 20,5 point.

c)

Jeg skal gøre rede for, at 99 point er en outlier.

Jeg udregner først $Q_3 + 1,5 \cdot kvartilbredde = 60,5 + 1,5 \cdot 20,5 = 91,25$

Alle observationer større end 91,25 point vil være outliers, og da 99 point er større end 91,25 point, er 99 point en outlier.

 [YouTube Kort gennemgang af opgaven.](#)