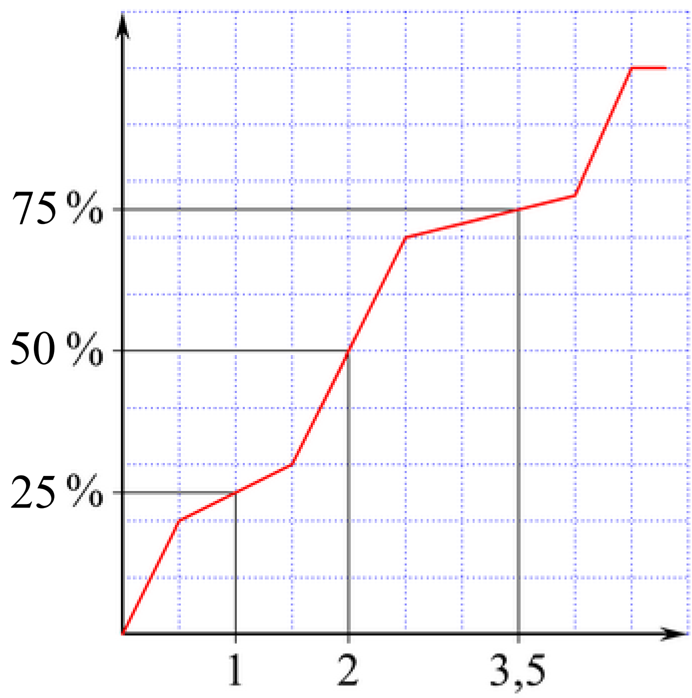
Deskriptiv statistik

for hf-matC



2018 Karsten Juul

Deskriptiv statistik for hf-matC

Hvad er deskriptiv statistik?

Ugrupperede data

Grupperede data

Stikprøver

Deskriptiv statistik for hf-matC
 29/4-2018

Hvad er deskriptiv tatistik?

1.1 Hvad er deskriptiv statistik?

Deskriptiv statistik er metoder til at få overblik over tal vi har indsamlet.
De tal vi har indsamlet, kalder vi data eller observationer.

1.2 Hvad er grupperede og ugrupperede data?

Hvis der er mange forskellige data, så grupperer vi dem i intervaller. (Hvis vi kaster en terning 1000 gange, er der mange data, men kun 6 forskellige, så disse skal ikke grupperes).

1.21 Eksempel på ugrupperede data.

Vi har talt antallet af bær i 15 pakker.
Antal bær i en pakke: 24 24 22 24 23 22 24 23 26 26 23 28 27 22 24

1.22 Eksempel på grupperede data.

Vi har vejet 200 frugter: Mellem 100 og 110 gram: 16 frugter

Mellem 110 og 120 gram: 68 frugter

Mellem 120 og 130 gram: 90 frugter

Mellem 130 og 140 gram: 26 frugter

Ugrupperede data

2.1 Middeltal (middelværdi) for ugrupperede data

Middeltallet for nogle tal er det vi plejer at kalde gennemsnittet.
Vi kan udregne <u>middeltallet</u> (middelværdien) ved at lægge tallene sammen og dividere resultatet med antallet af tal.
Middeltallet betegnes $\overline{x}$.

2.11 Udregne middeltal

I 7 prøver opnåede en elev følgende pointtal: 6 9 8 8 9 7 9

Sådan udregner vi middeltallet:

$$\overline{x} = \frac{6+9+8+8+9+7+8}{7} = 7{,}85714$$

Middeltallet for elevens pointtal er $\underline{\underline{7{,}9}}$

2.2 Median for ugrupperede data

For grupperede data skal vi gøre noget helt andet. Se afsnit 3.5 på side 16.

2.21 Sådan finder vi medianen

En klasse har haft en prøve. De 17 elever fik følgende point:

52 69 70 20 47 71 48 27 27 62 15 48 23 52 49 39 36

Vi ordner disse tal efter størrelse så tallet til venstre er mindst:

$$\overbrace{15\ \ 20\ \ 23\ \ 27\ \ 27\ \ 36\ \ 39\ \ 47}\ \ 48\ \ \overbrace{48\ \ 49\ \ 52\ \ 52\ \ 62\ \ 69\ \ 70\ \ 71}$$

Vi ser at det midterste af tallene er 48. Man siger at tallenes <u>median</u> er 48 .

Antag at der i stedet havde været et lige antal tal:

$$\overbrace{3\ \ 3\ \ 4\ \ 5}\ \overbrace{6\ \ 6\ \ 8\ \ 9}$$

Da der er et lige antal tal, er der ikke et tal der står i midten. I stedet udregner vi gennemsnittet af de to midterste tal:

$$\frac{5+6}{2} = 5{,}5 \ .$$

Man siger at tallenes <u>median</u> er 5,5 .

2.3 Kvartilsættet for ugrupperede data.

For grupperede data skal vi gøre noget helt andet. Se afsnit 3.6 på side 16.

2.31 Hvis der er et midterste tal:

15 20 23 27 27 36 39 47 48 48 49 52 52 62 69 70 71

Medianen for tallene til venstre for det midterste tal kalder vi nedre kvartil.
Dvs. nedre kvartil er 27.

Medianen for tallene til højre for det midterste tal kalder vi øvre kvartil.
Dvs. øvre kvartil er 57.

Når vi taler om **kvartilsættet** for nogle tal, så mener vi de tre tal

nedre kvartil , **median** , **øvre kvartil** ,

dvs. **kvartilsættet** for tallene ovenfor er de tre tal

$$\underline{\underline{27\ ,\ 48\ ,\ 57}}\ .$$

2.32 Hvis der ikke er et midterste tal:

3 3 4 5 6 6 8 9

Medianen for den venstre halvdel af tallene kalder vi nedre kvartil.
Dvs. nedre kvartil er 3,5 .

Medianen for højre halvdel af tallene kalder vi øvre kvartil.
Dvs. øvre kvartil er 7 .

Kvartilsættet er de tre tal $\underline{\underline{3{,}5\ ,\ 5{,}5\ ,\ 7{,}0}}$.

2.33 Det udvidede kvartilsæt

Når vi taler om **det udvidede kvartilsæt** for nogle tal, så mener vi de fem tal

mindste observation , **nedre kvartil** , **median** , **øvre kvartil** , **største observation** .

dvs. det **udvidede kvartilsæt** for tallene i 2.31 ovenfor er de fem tal

$$\underline{\underline{15\ ,\ 27\ ,\ 48\ ,\ 57\ ,\ 71}}\ .$$

2.34 Kvartilbredde

Kvartilbredden er forskellen på øvre og nedre kvartil.

Kvartilbredden for tallene i 2.31 er altså

$$57-27 = \underline{\underline{30}}$$

2.35 Variationsbredde

Variationsbredden er forskellen på det største og det mindste af tallene.

Variationsbredden for tallene i 2.31 er altså

$$71-15 = \underline{\underline{56}}$$

2.4 Nogle betegnelser

min = mindste tal

max = største tal

Variationsbredde = $max - min$

m = median

Q_1 = nedre kvartil

Q_3 = øvre kvartil

Kvartilbredde = $Q_3 - Q_1$

Kvartilsæt = $(Q_1,\ m,\ Q_3)$

Udvidet kvartilsæt = $(min,\ Q_1,\ m,\ Q_3,\ max)$

$\bar{x}$ = middeltallet

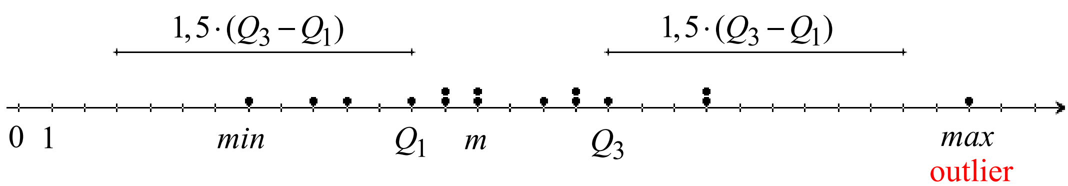
2.5 Outlier

Et tal kaldes en **outlier** hvis

det ligger mere end halvanden kvartilbredde under nedre kvartil

eller

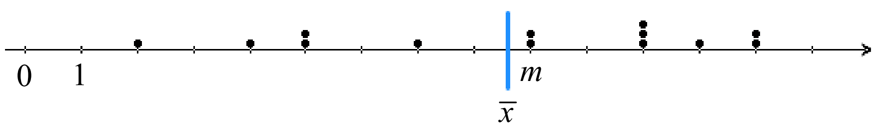
det ligger mere end halvanden kvartilbredde over øvre kvartil .



2.6 Venstreskæv og højreskæv

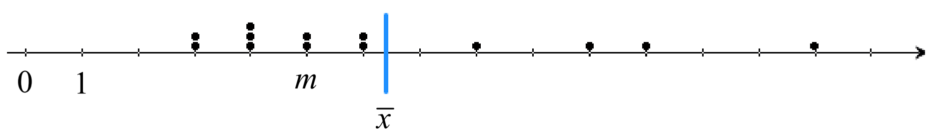
Fordelingen af nogle tal kaldes **venstreskæv**

hvis middeltallet er mindre end medianen: $\bar{x} < m$



Fordelingen af nogle tal kaldes **højreskæv**

hvis middeltallet er større end medianen: $m < \bar{x}$



2.7 Hvordan tegner vi boksplot?

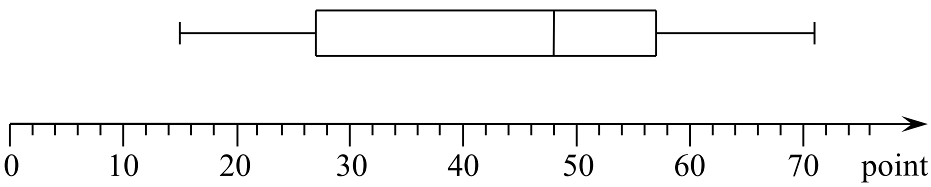
Ved at undersøge datasættet

15 20 23 27 27 36 39 47 48 48 49 52 52 62 69 70 71

kan vi se at

mindste tal	=	15
nedre kvartil	=	27
median	=	48
øvre kvartil	=	57
største tal	=	71

Disse oplysninger har vi vist på figuren. Sådan en figur kaldes et boksplot.



De to små lodrette streger i enderne viser at mindste og største tal er 15 og 71.
De to lodrette streger i hver ende af rektanglet viser at nedre og øvre kvartil er 27 og 57.
Den lodrette streg i midten af rektanglet viser at medianen er 48.

Rektanglet anskueliggør at den midterste halvdel af tallene ligger i intervallet fra 27 til 57.

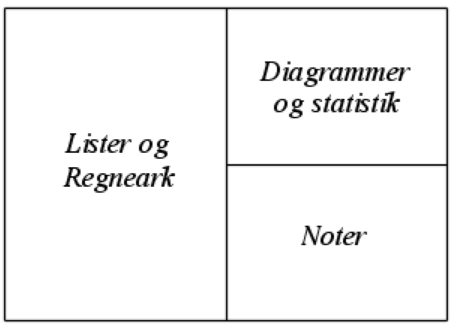
Den vandrette streg til venstre anskueliggør at den fjerdedel af tallene der er mindst, ligger i intervallet fra 15 til 27.

Den vandrette streg til højre anskueliggør at den fjerdedel af tallene der er størst, ligger i intervallet fra 57 til 71.
Når boksplot skal sammenlignes, skal de tegnes over hinande på samme figur.

Rektanglets længde er **kvartilbredden**.
Hele diagrammets længde er **varationsbredden**. .

2.8 Hvordan tegner vi boksplot i Nspire-dokument?

Når du skal tegne boksplot i Nspire,
så start med at dele vinduet op sådan:



2.81 Hvis der kun er ét boksplot

Tast søjlens navn øverst:

Nedenfor er navnet **blå** .

Tast så det **udvidede kvartilsæt**, og

tast medianen to gange .

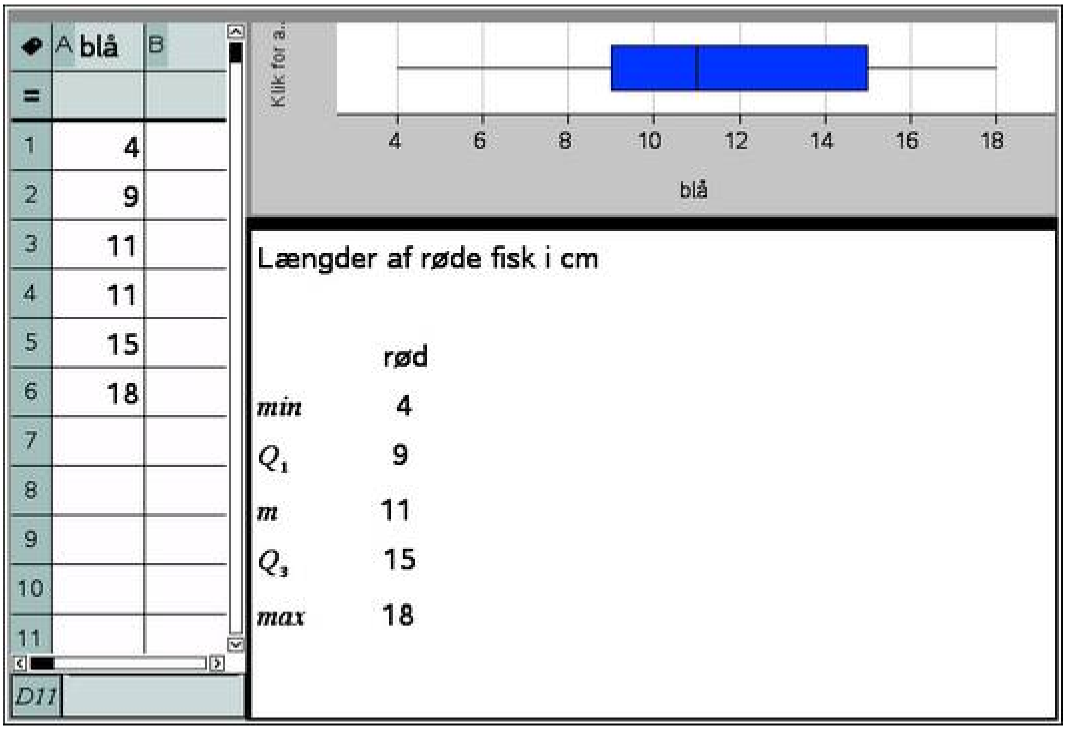
Klik under x-aksen i delvinduet med Diagrammer og statistik og

vælg søjlen med tallene.

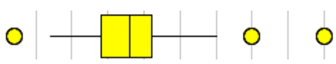
Vælg i værktøjsmenuen

Diagramtyper / **Boxplot**.

Så tegnes boksplottet.



Hvis dit boksplot har prikker som vist nedenfor,



så højreklik og vælg **Udvid boxplotgrænser** .

Hvis du ændrer tallene i regnearket,
så tilpas boksplot-vinduet ved at højreklikke et tomt sted og vælge **Zoom** / **Zoom-Data** .

2.82 Hvis der er to boksplot

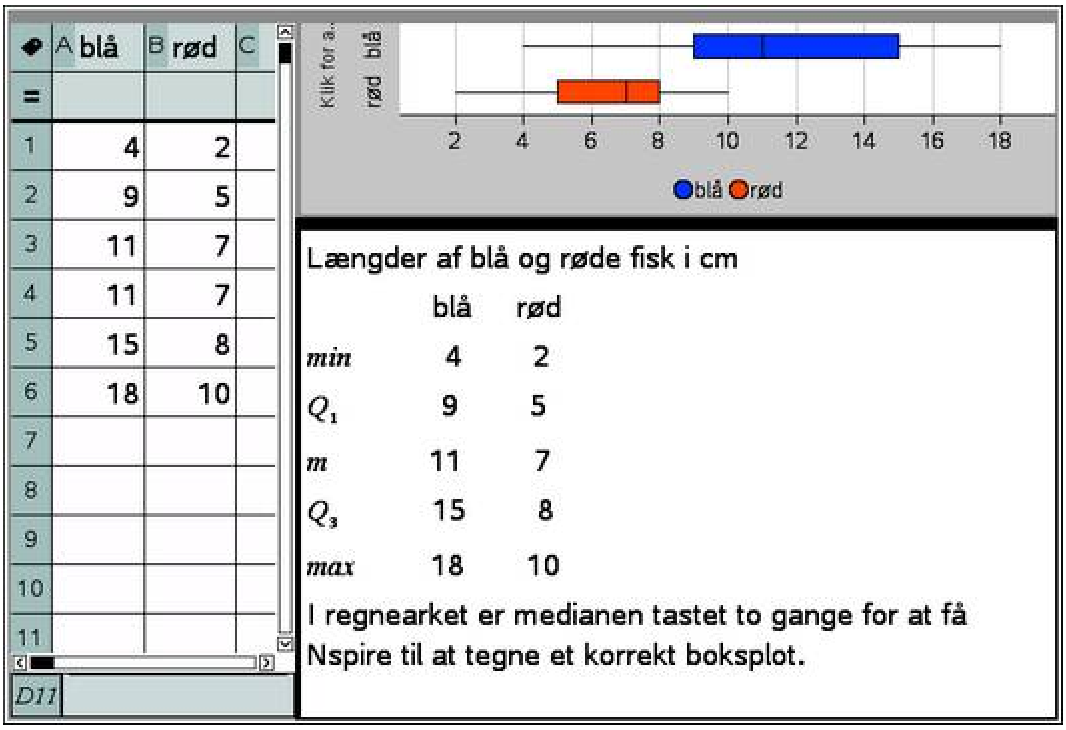
For at tilføje endnu et boksplot skal du gøre sådan:

Tast endnu en søjle i regnearket,
og husk at søjlen skal have et navn (nedenfor er navnet **rød**).
Og husk at **medianen skal tastes to gange** .

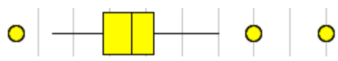
I delvinduet med Diagrammer og statistik:

Højreklik under x-aksen, vælg tilføj x-variabel, og vælg navnet som du har skrevet over den nye søjle.

Så fremkommer et vindue med to boksplot.



Hvis dit boksplot har prikker som vist nedenfor,



så højreklik og vælg **Udvid boxplotgrænser** .

Hvis du ændrer tallene i regnearket,
så tilpas boksplot-vinduet ved at højreklikke et tomt sted og vælge **Zoom** / **Zoom-Data** .

2.45 Ændre farve på boksplot

For at ændre farven på et boksplot skal du gøre følgende:

Højreklik på boksplottet.

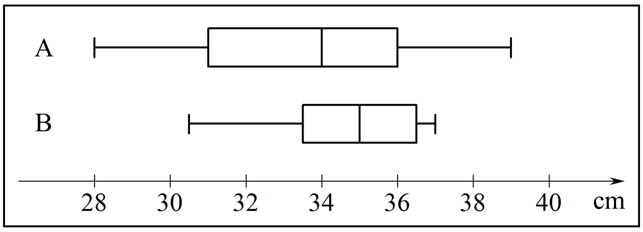
Vælg **Farve** / **Udfyldningsfarve** .

2.9 Hvordan sammenligner vi boksplot?

2.91 Opgave

Diagrammet viser højdefordelingen
for en plante på to marker A og B.

Sammenlign højderne på A og B.



Svar

Sammenlign størrelser

Alle dele af diagrammet bortset fra højre endepunkt ligger længere mod højre på B's diagram end på A's, dvs. mindste tal, nedre kvartil, median og øvre kvartil er mindre for A end for B mens største tal er større for A end for B, begrundelse
så højderne er altså overvejende større på B selv om den største højde er på A. resultat

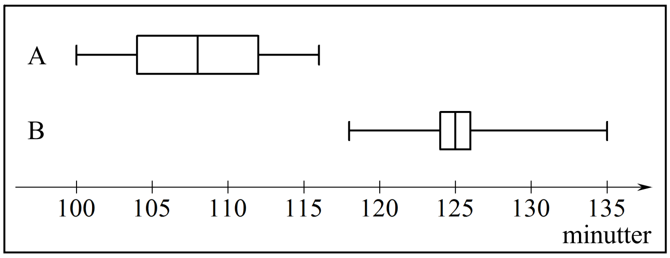
Sammenlign spredning

Både hele diagrammet og kassen er bredere på A's diagram end på B's,
dvs. variationsbredde og kvartilbredde er større for A end for B, begrundelse
så højderne fra A er mere spredt end højderne fra B. resultat

2.92 Opgave

Diagrammet viser fordelingen af tider
for to løbere A og B.

Sammenlign tiderne for A og B.



Svar

Sammenlign størrelser

Venstre endepunkt for B-diagrammet ligger til højre for højre endepunkt for A-diagrammet, begrundelse
så B's mindste tid er større end A's største tid. resultat

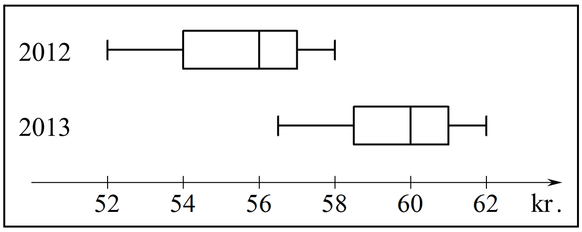
Sammenlign spredning

Kvartilbredden (kassens længde) er meget større for A-tider end for B-tider, begrundelse
så midterste halvdel af tiderne er meget mere spredt for A end for B. resultat

Variationsbredden (hele diagrammets længde) er ca. den samme for A og B, begrundelse
så forskellen på største og mindste tid er ca. den samme for A og B. resultat

2.93 Opgave

Diagrammet viser hvordan priserne på en vare er fordelt i 2012 og i 2013.



(a) Sammenlign priserne i 2012 og 2013.

(b) I 2012 betalte en person 53,50 kr. for varen. Hvordan ligger denne pris i forhold til alle 2012-priserne for varen?

(c) En person betalte et beløb i den laveste halvdel af den højeste halvdel af 2012-priserne. Hvad fortæller dette om størrelsen af beløbet.

Svar på (a)

Sammenlign størrelser

Hele 2013-diagrammet ligger til højre for venstre halvdel af
2012-diagrammet
dvs 2013-mindsteværdien er større end 2012-medianen, ← begrundelse
så alle 2013-priserne er over laveste halvdel af 2012-priserne. ← resultat

Hele 2012-diagrammet ligger til venstre for kassen i 2013-diagrammet,
dvs. 2012-størsteværdien er mindre end nedre kvartil for 2013, ← begrundelse
så alle 2012-priserne er lavere end de 75 % højeste 2013-priser. ← resultat

Sammenlign spredning

Hverken for kassen eller hele diagrammet er længden ændret væsentligt fra
2012 til 2013,
dvs. variationsbredde og kvartilbredde er ikke ændret væsentligt fra
2012 til 2013, ← begrundelse
så der er ikke meget forskel på hvor spredt priserne er i 2012 og 2013. ← resultat

Svar på (b)

53,50 ligger på diagrammets venstre linjestykke,
dvs. 53,50 er mindre end nedre kvartil, ← begrundelse
så 53,50 kr. er i den nederste fjerdedel af 2012-priserne. ← resultat

Svar på (c)

Når et beløb er i den laveste halvdel af den højeste halvdel, er det i højre del af kassen,
dvs. mellem median og øvre kvartil, ← begrundelse
dvs. mellem 56,00 kr. og 57,00 kr. ← resultat

Grupperede data

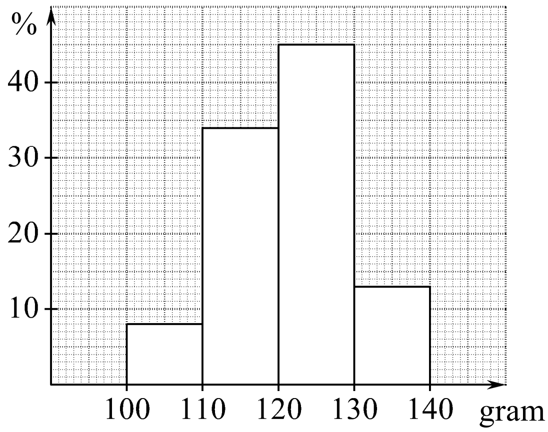
Dette er kun mundtligt pensum

3.1 Hvordan tegner vi et histogram?

Tabellen viser fordelingen af nogle frugters vægt.

Vægt i gram	100-110	110-120	120-130	130-140
Procent	8	34	45	13

Histogrammet nedenfor viser oplysningerne i tabellen.



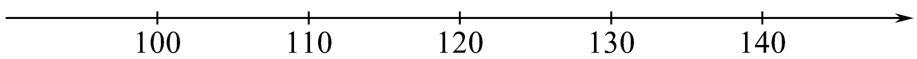
Rektanglet over intervallet 100-110 har højden 8 %.

Dette viser at 8 % af frugterne vejer mellem 100 og 110 gram.

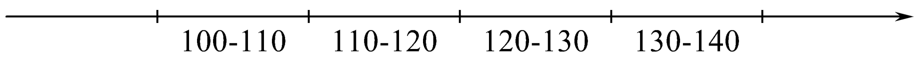
Bemærk: Denne måde at tegne et histogram på
kan kun bruges fordi intervallerne 100-110,
110-120 osv. er lige lange.

Advarsel: Den vandrette akse skal tegnes som en sædvanlig tallinje.

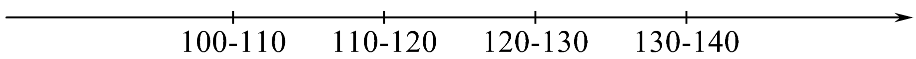
RIGTIGT:



FORKERT:



FORKERT:



3.2 Hvordan tegner vi et histogram i Nspire-dokument?

OPGAVE

Vægt af nogle sten er fordelt sådan:

Vægt i g:	60-70	70-80	80-90	90-100	100-110	110-120
Frekvens:	9 %	28 %	20 %	11 %	32 %	20 %

Tegn et histogram for denne fordeling.

BESVARELSE

er vist i den røde ramme nedenfor. Gør sådan:

-Vælg **Indsæt** / **Opgave** .

-Klik på og del siden op i tre vinduer: .

-Klik i vinduer og tilføj som vist i den røde ramme:

Noter, **Lister og Regneark**, **Diagrammer og statistik** .

-Tast intervalendepunkterne i første søjle. I anden søjle: Tast 0 ud for mindste endepunkt. Ud for hvert intervals højre endepunkt skal du taste intervallet frekvens som vist.

-Klik under x-aksen og vælg **vægt** fordi det er det navn som vi har givet første søjle.

-**Højre**klik til venstre for y-akse, vælg **Tilføj y-værdiliste** (**IKKE** "Tilføj variabel"!) og vælg **frekv** fordi det er det navn vi har givet den anden søjle.

-Klik på (Værktøjsmenuen) og vælg

Diagramegenskaber / **Egenskaber for histogram** / **Søjleindstillinger** / **Lige store intervaller** ,

og tast Bredde som 10 og Søjlestart som 60,01 . (Søjlestart skal være en anelse større end første endepunkt).

-Hvis du ændrer på tallene, så tilpas histogram-vinduet ved at højreklikke et tomt sted og vælge **Zoom** / **Zoom-Data** .

Tabellen til højre viser fordel-ingen af vægtene af nogle sten.

Første søjle indeholder intervalendepunkter (gram).

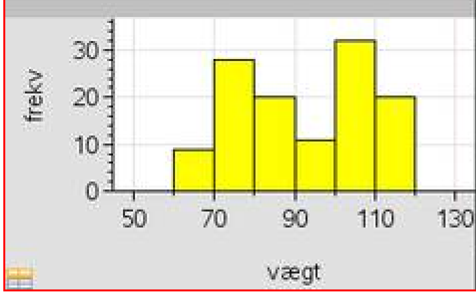
Ud for højre endepunkt står intervallets frevens i procent, f.eks. har 70-80 frekvens 28%

Nedenfor er tegnet et histogram for denne fordeling.

Søjlehøjde angiver frekvens.

	A vægt	B frekv
=		
1	60	0
2	70	9
3	80	28
4	90	20
5	100	11
6	110	32
7	120	20
8		
9		
10		
11		
12		

D12



3.3 Et grupperet datasæt er en model af virkeligheden der er meget forenklet.

Ovenfor har vi set på følgende grupperede datasæt:

Vægt i gram	100-110	110-120	120-130	130-140
Procent	8	34	45	13

Da dette datasæt er grupperet, skal vi regne som om

de 8 % i første interval er helt jævnt fordelt i dette interval

de 34 % er helt jævnt fordelt i andet interval

osv.

Dette betyder bl.a. et vi f.eks. skal regne som om

0 % af dataene er præcis lig 110.

Dette er ikke i modstrid med virkeligheden, for når vi siger at noget vejer 110 g, mener vi ca. 110 g. Hvis vi hermed mener "mellem 109 g og 111 g", så er der ifølge tabellen 4,2 % der vejer ca. 110 g. Til eksamen plejer man ikke at spørge om sådan noget.

Der gælder altså:

Den procentdel af dataene der er 110 eller mindre,
er lig den procentdel der er mindre end 110.

Det giver ingen mening at spørge om 110 er talt med i intervallet 100-110 eller i intervallet 110-120. Dette spørgsmål giver mening i en opgave hvor du selv skal gruppere nogle data.

4 Hvordan tegner vi en sumkurve?

4.1 Kumuleret frekvens og sumkurve

Et intervals frekvens, er den procentdel af dataene som intervallet indeholder. Ordet ”kumuleret” betyder ophobet.

Den **kumulerede frekvens** af et tal t er den procentdel af dataene der er af størrelse t eller derunder.

Sumkurven er grafen for den kumulerede frekvens.

4.2 Hvis der er oplyst procent for hvert interval.

Vægt i gram	100-110	110-120	120-130	130-140
Frekvens	8 %	34 %	45 %	13 %

For at tegne en sumkurve, udregner vi kumulerede frekvenser.
Vi har skrevet dem i tabellen, og vi har udregnet dem sådan:

$8\% + 34\% = 42\%$, $42\% + 45\% = 87\%$, osv.

Vægt i gram	100	110	120	130	140
Kumuleret frekvens	0 %	8 %	42 %	87 %	100 %

For at tegne sumkurven gør vi sådan:

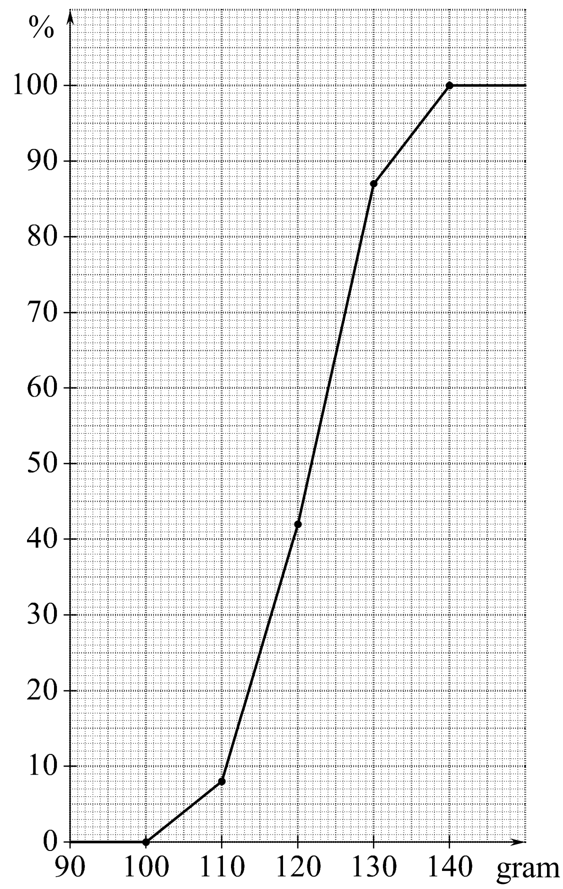
0 % er mindre end 100, så ved $x = 100$ afsætter vi et punkt ud for 0 % på y-aksen.

8 % er mindre end 110, så ved $x = 110$ afsætter vi et punkt ud for 8 % på y-aksen.

42 % er mindre end 120, så ved $x = 120$ afsætter vi et punkt ud for 42 % på y-aksen.

Osv.

Da dataene er jævnt fordelt i hvert interval, skal vi forbinde punkterne med rette linjestykker.



4.3 Hvis der er oplyst antal for hvert interval.

I tabellen står antal i stedet for procent. Så må vi omregne til procent for at kunne tegne sumkurven.

Længde (m)	0,5-2	2-3	3-4	4-5	5-8
Antal rør	34	58	91	72	27

I tabellen kan vi skrive "hyppighed" i stedet for "antal rør". Det har vi gjort i tabellen nederst.

Nedenfor lægger vi sammen før vi omregner til procent. Det er for at undgå mellemfacitter med mange cifre.

Antal data er $34+58++91+72+27 \; = \; 282$.

Kumuleret hyppighed udregner vi sådan:

$$34+58 \; = \; 92 \;\; , \;\; 92+91 \; = \; 183 \;\; , \;\; \text{osv.}$$

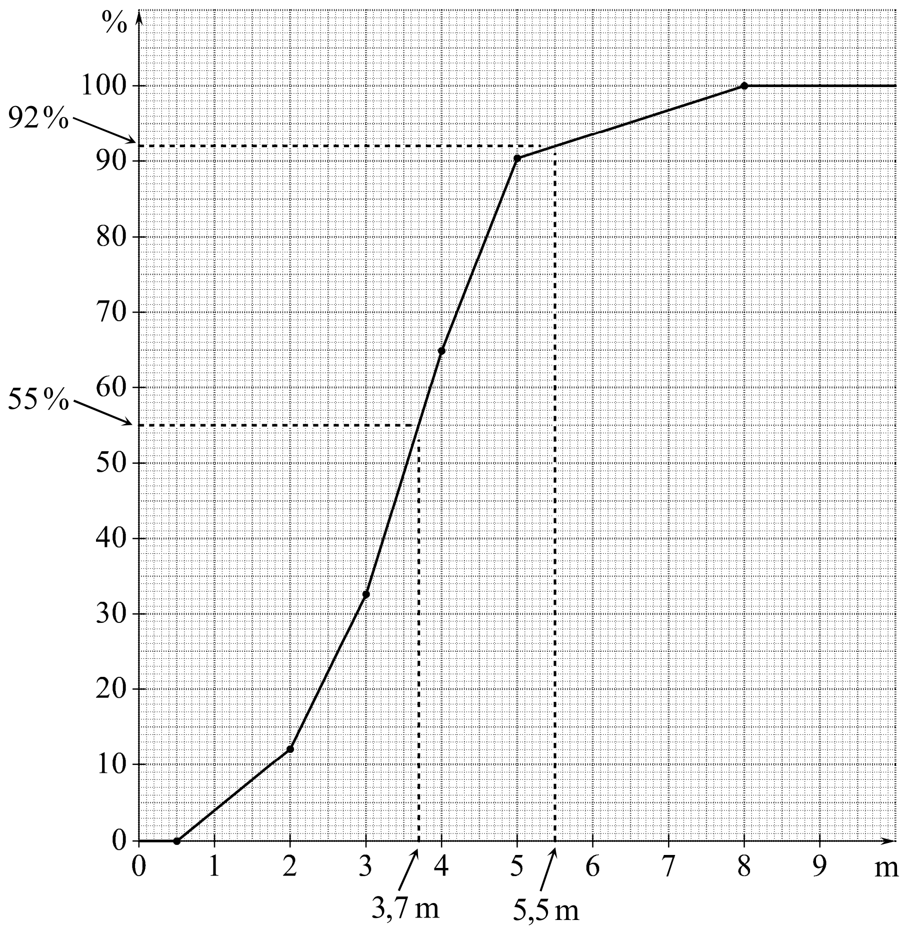
Kumuleret frekvens udregner vi sådan:

$$\frac{34}{282} \; = \; 0{,}120567 \, , \;\; \frac{92}{282} \; = \; 0{,}326241 \, , \;\; \text{osv.}$$

Længde i meter	0,5	2	3	4	5	8
Kumuleret hyppighed	0	34	92	183	255	282
Kumuleret frekvens	0 %	12,1 %	32,6 %	64,9 %	90,4 %	100,0 %

5 Hvordan aflæser vi på en sumkurve?

Figuren viser sumkurven for rørene fra tabellen på foregående side.



5.1 Hvor mange procent af rørene er UNDER 3,7 meter?

Svar: Som vist på figuren aflæser vi at $\underline{\underline{55\,\%}}$ af rørene er under 3,7 meter.

Dette kan udtrykkes ved at sige at 55%-**fraktilen** er 3,7 .

5.2 Hvor mange procent af rørene er OVER 5,5 meter?

Svar: Som vist på figuren aflæser vi at 92 % af rørene er under 5,5 meter.
Da $100\% - 92\% = 8\%$, er $\underline{\underline{8\%}}$ af rørene over 5,5 meter.

5.3 Hvor mange procent af rørene er MELLEM 3,7 og 5,5 meter?

Svar: Fra de 92 % der er under 5,5 meter, skal fraregnes de 55 % der er under 3,7 meter.
Da $92\% - 55\% = 37\%$, er $\underline{\underline{37\%}}$ af rørene mellem 3,7 og 5,5 meter.

5.4 Hvor mange procent af rørene er LIG 3,7 meter ELLER DERUNDER ?

Svar: Det er samme spørgsmål som spørgsmålet 3.41 ovenfor da
0 % af rørene er præcis lig $3{,}70000\cdots$ meter.
Det at der på sumkurven er 0 % der er lig 3,7 meter, er ikke i modstrid med at nogle af rørene er målt til 3,7 meter. (Læs evt. forklaringen på dette i afsnit 3.3 på side 12).

6 Hvordan finder vi medianen for grupperede data?

For ugrupperede data skal vi gøre noget helt andet. Se afsnit 2.2 på side 2.

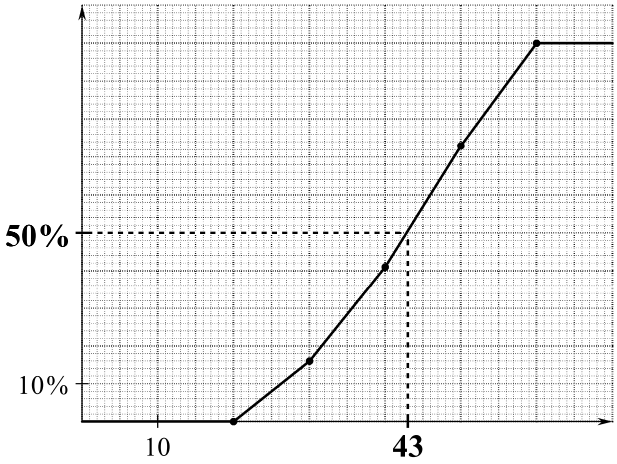
For at finde medianen skal vi bruge sumkurven når det er grupperede data.

Vi starter i 50 % på y-aksen,
går vandret hen til sumkurven,
går lodret ned på x-aksen,
og aflæser x-værdien.

Denne x-værdi er medianen.

At et tal er **median**, betyder altså at
50 % af dataene er mindre end dette tal
og 50 % af dataene er større end dette tal.

På figuren er medianen 43.

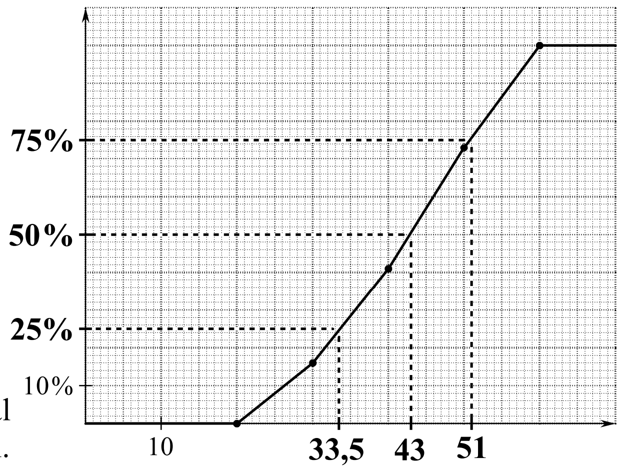


Dette har du brug for at vide når du har fundet medianen og skal svare på hvad dette tal fortæller. I dit svar skal du i stedet for "data" bruge det ord der står i opgaven, f.eks. "længde", og i stedet for "dette tal" skal du skrive det tal du har fundet, f.eks. 43.

7 Hvordan finder vi kvartilsættet for grupperede data?

For ugrupperede data skal vi gøre noget helt andet. Se afsnit 2.3 på side 3.

For at finde kvartilsættet skal vi bruge sumkurven når det er grupperede data.



7.1 Nedre kvartil.

Vi starter i 25 % på y-aksen,
går vandret hen til sumkurven,
går lodret ned på x-aksen,
og aflæser x-værdien.

Denne x-værdi er nedre kvartil.

At et tal er **nedre kvartil**, betyder altså
at 25 % af dataene er mindre end dette tal
og 75 % af dataene er større end dette tal.

På figuren er nedre kvartil 33,5 .

Dette har du brug for at vide når du har fundet nedre kvartil og skal svare på hvad dette tal fortæller. I dit svar skal du i stedet for "data" skrive det ord der står i opgaven, f.eks. "længde", og i stedet for "dette tal" skal du skrive det tal du har fundet, f.eks. 33,5.

7.2 Øvre kvartil.

Vi starter i 75 % på y-aksen,
går vandret hen til sumkurven,
går lodret ned på x-aksen,
og aflæser x-værdien.

Denne x-værdi er øvre kvartil.

Dette har du brug for at vide når du har fundet øvre kvartil og skal svare på hvad dette tal fortæller. I dit svar skal du i stedet for "data" skrive det ord der står i opgaven, f.eks. "længde", og i stedet for "dette tal" skal du skrive det tal du har fundet, f.eks. 51.

At et tal er **øvre kvartil**, betyder altså at
75 % af dataene er mindre end dette tal
og 25 % af dataene er større end dette tal.

På figuren er øvre kvartil 51 .

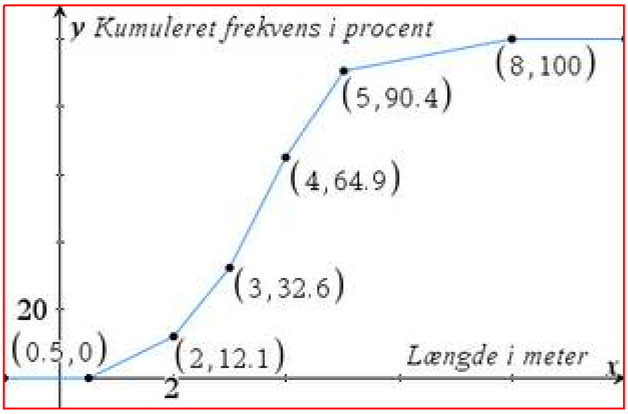
7.3 Kvartilsæt.

Når vi taler om kvartilsættet for nogle tal, så mener vi de tre tal

nedre kvartil , median , øvre kvartil,

dvs. kvartilsættet er de tre tal 33,5 , 43 , 51 .

8 Tegne sumkurve i Nspire.



I afsnit 4.3 er de kumulerede frekvenser udregnet. Ud fra disse procenttal kan vi tegne sumkurven på skærmen.

Gør sådan:

- Vælg **Indsæt** / **Side** .
- Klik i vinduet og vælg **Tilføj Grafer** .
- Klik på (Værktøjsmenuen) og vælg
  Vindue/Zoom / **Indstillinger for vindue**
- Giv **XMin** og **XMax** værdier så der kommer mere med end tabellens *x*-værdier (I afsnit 3.33 er det længderne 0,5 , 2 osv. der er *x*-værdier).
- Sæt **YMin** til –10 og **YMax** til 110 (da en sumkurves *y*-værdier altid går fra 0 % til 100 %)

- Klik på (Værktøjsmenuen) og vælg
  Geometri / **Punkter og linjer** / **Punkt** .
- Klik et tomt sted for at afsætte et punkt. Afsæt 8 punkter da der er 6 *x*-værdier og vi skal bruge to ekstra punkter.
- Tryk på **Esc** for at fjerne ikonen når du er færdig med at afsætte punkter.
- Højreklik på et af de afsatte punkter, og vælg **Koordinater og ligninger**. Så fremkommer punktets koordinatsæt.

  Dobbeltklik på punktets *x*-koordinat (tallet før kommaet), ret den til den første *x*-værdi (altså 0,5), og tryk på **enter** .

  Dobbeltklik på punktets *y*-koordinat (tallet efter kommaet), ret den til den første *y*-værdi (altså 0 som står i tabellens tredje række), og tryk på **enter** .

  Fortsæt med at sætte et af de andre punkters koordinater til (2 , 12,1) osv. til tabellens 6 punkter er afsat.
- Det første af de to ekstra punkter afsætter du lidt til venstre for tabellens første punkt (altså med *y*-koordinat 0). Det andet ekstra punkt afsætter du lidt til højre for tabellens sidste punkt (altså med *y*-koordinat 100).

- Klik på (Værktøjsmenuen) og vælg
  Geometri / **Punkter og linjer** / **Linjestykke** .

  Så fremkommer linjestykke-ikonen. Før markøren hen til det første punkt så prikken bliver større, og klik. Før markøren hen til det andet punkt så prikken bliver større, og klik. Fortsæt til hele sumkurven er tegnet.

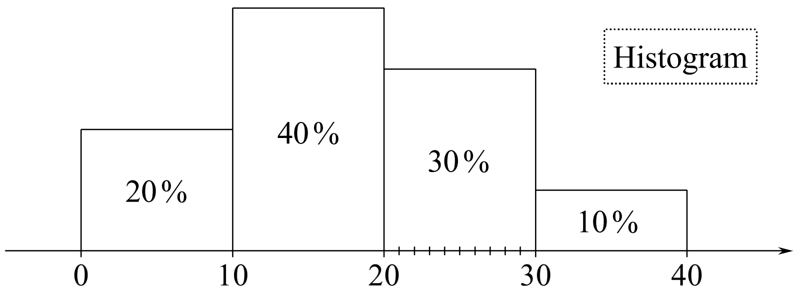
  HUSK at trykke på **Esc** for at fjerne ikonen når du er færdig med at tegne linjestykker.
- Skriv tekst på akser.

9 Aflæse sumkurve i Nspire

- Klik på (Værktøjsmenuen) og vælg
  Geometri / **Punkter og linjer** / **Punkt på** .

  Så fremkommer Punkt-på-ikonen.
  I nærheden af det sted på kurven som du vil aflæse, skal du klikke, og klikke igen.
  Så fremkommer et punktpå kurven.
- Højreklik på punktet, vælg **Koordinater og ligninger**, og ret den af koordinaterne du kender.
  Den anden koordinat fremkommer. Denne koordinat er din aflæsning.

10 Sumkurve og lineær sammenhæng.



Histogrammet viser et grupperet datasæt:

Intervallet 20-30 deler vi op i 10 lige store dele (se figur).
Hver af disse små intervaller må indeholde en tiendedel af hele intervallets observationer, dvs. de indeholder hver 3 % af samtlige data.

Vi lader (x, y) være et punkt på sumkurven, dvs.

y er den procentdel af observationerne der har størrelse x eller derunder.

Af histogrammet ovenfor ser vi:

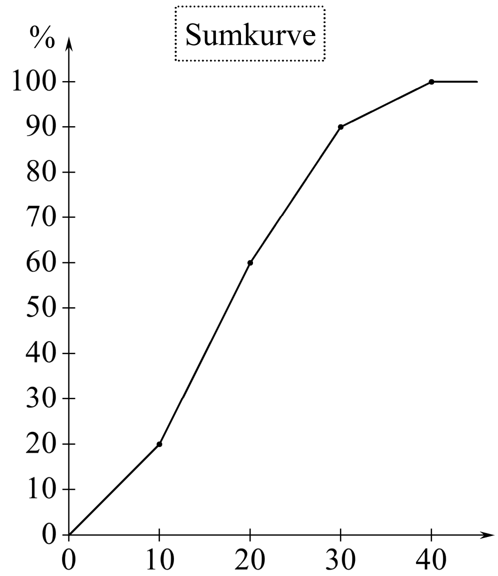
Når $x = 20$ er $y = 0{,}20 + 0{,}40 = 0{,}60$

Når $x = 21$ er $y = 0{,}60 + 0{,}03 = 0{,}63$

Når $x = 22$ er $y = 0{,}63 + 0{,}03 = 0{,}66$

Hver gang x bliver 1 større, vil y blive 0,03 enheder større, så y vokser lineært i intervallet fra $x = 20$ til $x = 30$. Derfor er grafen en ret linje i dette interval, og ligningen er

$y = 0{,}03x + b$.



Vi udregner b :

Når $x = 20$ er $y = 0{,}60$ så

$0{,}60 = 0{,}03 \cdot 20 + b$.

Heraf ser vi at $b = 0$, så ligningen er

$y = 0{,}03x$.

For de fire intervaller er ligningerne:

0-10: $y = 0{,}02x$
10-20: $y = 0{,}04x - 0{,}2$
20-30: $y = 0{,}03x$
30-40: $y = 0{,}01x + 0{,}6$

Hvor mange procent af observationerne har størrelse 27 eller derunder?

Vi ser at vi skal bruge ligningen fra tredje interval:

$y = 0{,}03 \cdot 27 = 0{,}81$

dvs. 81 % af observationerne er 27 eller derunder.

Hvor stor er nedre kvartil?

Vi skal gå ud fra 25 % på y-aksen. Vi ser at vi skal bruge ligningen fra andet interval:

$0{,}25 = 0{,}04x - 0{,}2$.

Vi løser denne ligning mht. x og får $11{,}25$, dvs. nedre kvartil er $11{,}25$.

Stikprøver

11 Stikprøver.

Nogen på et gymnasium mener at der er forskel på hvad piger og drenge mener om et bestemt spørgsmål. For at undesøge denne hypotese, spørger vi nogle piger og drenge.

11.1 Hvad er populationen?

De ting eller personer som vi vil påstå noget om, kaldes populationen.

Er det alle personer i europa som nu er mellem 10 og 20 år?
Er det alle elever på vores gymnasium?
Eller?

Når vi laver en statistisk undersøgelse, skal vi **skrive** en præcisering af

hvad det er for en population vi vil påstå noget om.

11.2 Hvad er stikprøven?

Vi undersøger kun en lille del af hele populationen.
De personer vi får et svar fra (eller de ting vi undersøger), kaldes stikprøven.

Når vi laver en statistisk undersøgelse, skal vi **skrive** en præcisering af

hvordan vi har valgt stikprøven.

Det er **IKKE nok at skrive**:

”Vi har spurgt 47 elever på vores gymnasium.”

Det er **nok at skrive**

”Den 20. februar mellem kl. 8:50 og 9:10 spurgte vi de 47 elever der sad på gangen, og vi fik svar fra dem alle. 10 af drengene og 8 af pigerne var fra 3g FY, 13 af drengene og 16 af pigerne var fra 3g Fy.”

eller

”Den 20. februar kl. 8:50 sendte vi en besked til alle elever på skolen. Stikprøven er de 47 elever der svarede inden kl. 10:00 den 22. februar.”

Disse to beskrivelser af en indsamling af stikprøve er så grundige at læseren kan se om der er grund til tro at der kan være systematiske fejl.

11.3 Systematiske fejl ved valg af stikprøven.

Eksempel 1

Population: Eleverne på vores gymnasium.

Stikprøve: Eleverne i en sproglig klase.

Her kan vi have lavet en systematisk fejl ved valg af stikprøven, for det kan være at en bestemt holdning oftere er blandt sproglige end blandt andre.

Eksempel 2

Hvis vi spørger elever pr. e-mail, og mange ikke svarer, så kan vi have lavet en systematisk fejl, for det er måske især elever med en bestemt holdning der svarer.

11.4 Tilfældige fejl ved valg af stikprøven.

Selv om vi vælger stikprøven tilfældigt blandt hele populationen,
er det ikke helt sikkert at den ligner populationen.

Det kan f.eks. være at vi tilfældigt har fået for mange ja-sigere med i stikprøven.

11.5 Er der skjulte variable?

En skjult variabel er noget der kan ødelægge resultatet selv om stikprøven er udvalgt tilfældigt blandt hele populationen.

Eksempel: Der er flere der overlever på hospital A end på hospital B. Man slutter at behandlingen er bedre på A end på B. Men forskellen skyldes at B har flere ældre patienter. Patienternes alder er en skjult variabel der påvirker resultatet.

Stikordsregister