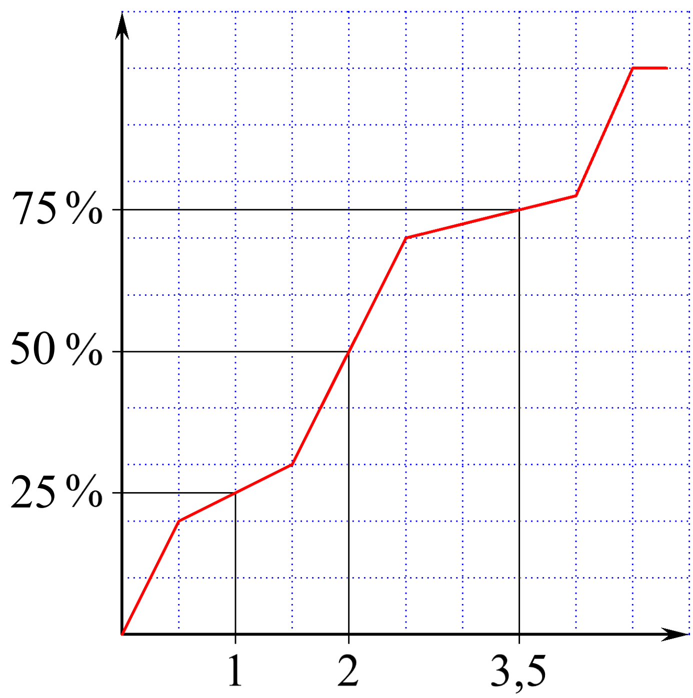
Deskriptiv statistik

for C-niveau i hf



2015 Karsten Juul

DESKRIPTIV STATISTIK

1.1 Hvad er deskriptiv statistik? 1
1.2 Hvad er grupperede og ugrupperede data? 1
  1.21 Eksempel på ugrupperede data 1
  1.22 Eksempel på grupperede data 1

UGRUPPEREDE DATA

2.1 Hvordan udregner vi middeltal for ugrupperede data? 1
2.2 Hvordan finder vi medianen for ugrupperede data? 2
2.3 Hvordan finder vi kvartilsættet for ugrupperede data? 2
  2.31 Hvis der er et midterste tal. 2
  2.32 Hvis der ikke er et midterste tal. 2
2.4 Hvordan tegner vi et boksplot? 2
2.5 Hvordan sammenligner vi boksplot? 3
  2.51 opgave. 3
  2.52 opgave. 3
  2.53 opgave. 3

GRUPPEREDE DATA

3.1 Hvordan tegner vi et histogram? 4
3.2 Et grupperet datasæt er en model af virkeligheden der er meget forenklet. 4
3.3 Hvordan tegner vi en sumkurve? 5
  3.31 Hvis der er oplyst procent for hvert interval 5
  3.32 Hvis der er oplyst antal for hvert interval. 5
  3.33 Hvis start-tal og/eller slut-tal mangler. 5
3.4 Hvordan aflæser vi på en sumkurve? 6
  3.41 Hvor mange procent af rørene er UNDER 3,7 meter? 6
  3.42 Hvor mange procent af rørene er OVER 5,5 meter? 6
  3.43 Hvor mange procent af rørene er MELLEM 3,7 og 5,5 meter? 6
  3.44 Hvor mange procent af rørene er LIG 3,7 meter ELLER DERUNDER ? 6
3.5 Hvordan finder vi medianen for grupperede data? 7
3.6 Hvordan finder vi kvartilsættet for grupperede data? 7
  3.61 Nedre kvartil. 7
  3.62 Øvre kvartil. 7
  3.63 Kvartilsæt. 7
3.7 Middeltal for grupperede data når antal (hyppighed) er oplyst 8
3.8 Middeltal for grupperede data når procent (frekvens) er oplyst 8

4.1 Hvordan grupperer vi data? 8
4.2 Hvor brede skal vi gøre intervallerne når vi grupperer data? 9

5 Sumkurve og lineær sammenhæng. 9

BESVARELSER SKREVET I NSPIRE

6.1-6.9 Besvarelser skrevet i Nspire. 10-15

DESKRIPTIV STATISTIK

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.

1.2 Hvad er grupperede og ugrupperede data?

Hvis der er mange forskellige data, så grupperer vi dem i intervaller.

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 Hvordan udregner vi middeltal for ugrupperede data?

For grupperede data skal vi gøre noget andet. Se afsnit 3.7 på side 9.

Middeltallet for nogle tal er det vi plejer at kalde gennemsnittet.
Vi kan udregne middeltallet ved at lægge tallene sammen og dividere resultatet med antallet af tal.

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

Sådan udregner vi middeltallet:

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

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

2.2 Hvordan finder vi medianen for ugrupperede data?

For grupperede data skal vi gøre noget helt andet. Se afsnit 3.5 på side 8.

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:

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

Vi ser at det midterste af tallene er 48. Man siger at tallenes median er 48 .

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

3 3 4 5 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 median er 5,5 .

2.3 Hvordan finder vi kvartilsættet for ugrupperede data?

For grupperede data skal vi gøre noget helt andet. Se afsnit 3.6 på side 8.

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 og øvre kvartil,

dvs. kvartilsættet for tallene ovenfor er de tre tal 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 3,5 , 5,5 , 7,0 .

2.4 Hvordan tegner vi et boksplot?

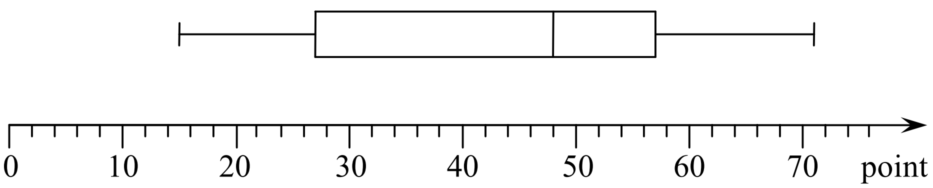
Ved at undersøge datasættet

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 inden i 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.

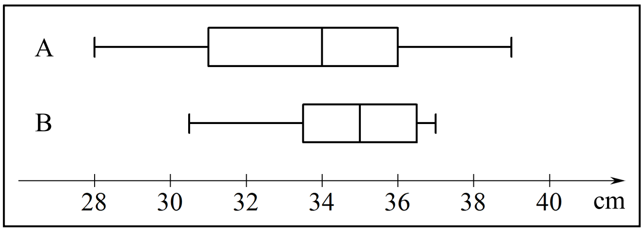
Den vandrette streg til højre anskueliggør at den fjerdedel af tallene der er størst, ligger i intervallet fra 57 til 71.

2.5 Hvordan sammenligner vi boksplot?

2.51 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, ← begrundelse
så højderne er altså overvejende større på B selv om den største højde er på A. ← resultat

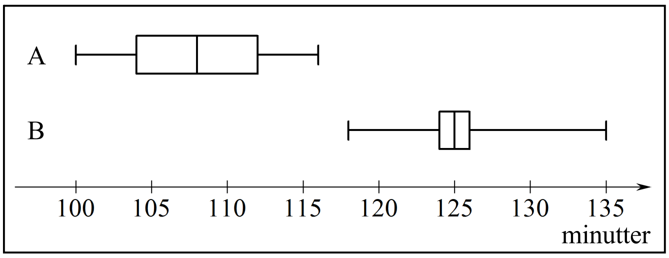
Sammenlign spredning

Både hele diagrammet og kassen er bredere på A's diagram end på B's, ← begrundelse
så højderne fra A er mere spredt end højderne fra B. ← resultat

2.52 Opgave

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

Sammenlign tiderne for A og B.



Svar

Sammenlign størrelser

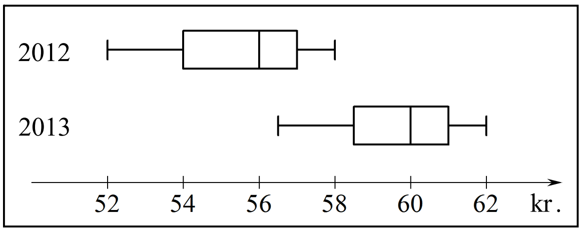
Venstre endepunkt for B-diagrammet ligger til højre for højre endepunkt for A-diagrammet, ← begrundelse
så B's mindste tid er større end A's største tid. ← resultat

Sammenlign spredning

A-kassen er meget længere end B-kassen, ← begrundelse
så midterste halvdel af tiderne er meget mere spredt for A end for B. ← resultat
Hele diagrammet har ca. samme længde for A og B, ← begrundelse
så forskellen på største og mindste tid er ca. den samme for A og B. ← resultat

2.53 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, ← begrundelse
så alle 2013-priserne er over laveste halvdel af 2012-priserne. ← resultat

Hele 2012-diagrammet ligger til venstre for kassen i 2013-diagrammet, ← 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, ← 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, ← begrundelse
dvs. 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, ← begrundelse
dvs. mellem 56,00 kr. og 57,00 kr. ← resultat

GRUPPEREDE DATA

3.1 Hvordan tegner vi et histogram?

Tabellen viser fordelingen af nogle frugters vægt.

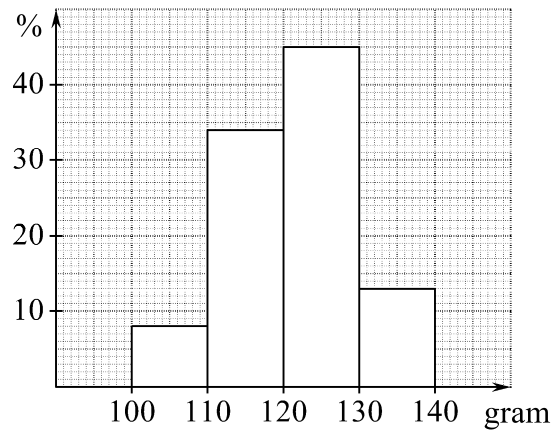
Vægt i gram	100-110	110-120	120-130	130-140
Procent	8	34	45	13

Histogrammet til højre viser oplysningerne i tabellen.

Rektanglet over intervallet 100-110 har højden højden 8 %.

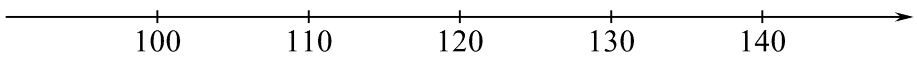
Dette viser at 8 % af frugterne vejer mellem 100 og 110 gram.

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

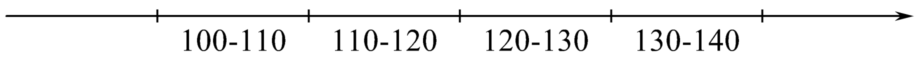


Advarsel: Den vandrette akse skal tegnes som en sædvanlig tallinje.

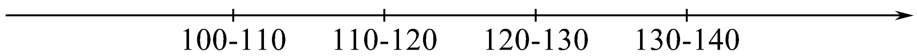
RIGTIGT:



FORKERT:



FORKERT:



3.2 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 opgav hvor du selv skal gruppere nogle data. Se afsnit 4.1 side 8.

3.3 Hvordan tegner vi en sumkurve?

3.31 Hvis der er oplyst procent for hvert interval.

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

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

$$8\% + 34\% = 42\% \ , \quad 8\% + 34\% + 45\% = 87\% \ , \quad \text{osv.}$$

Vægt i gram	100-110	110-120	120-130	130-140
Frekvens	8 %	34 %	45 %	13 %
Kumuleret frekvens	8 %	42 %	87 %	100 %

I andet interval står 42 %. Det betyder at i de to første intervaller er der 42 % af dataene, dvs. 42 % af dataene er under 120.
Sumkurven skal bruges til at aflæse hvor mange procent af dataene der er mindre end et tal.

For at tegne sumkurven gør vi sådan:

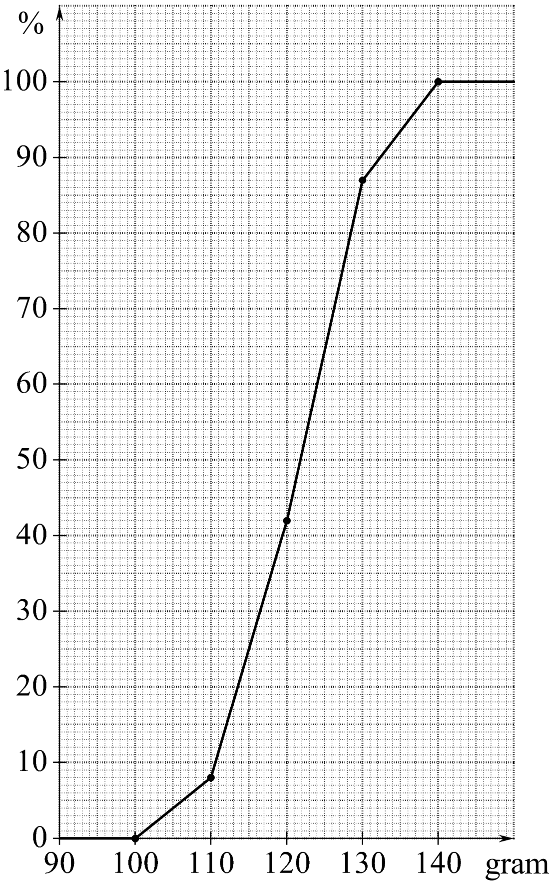
0 % er mindre end 100, så ved $x = 100$ afsætter vi et punkt ud for 0 % på y-aksen.

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.
(Se evt. begrundelsen for dette i afsnit 5 på side 9).



3.32 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 nedenfor.

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

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

Tallene i 3. række udregner vi sådan: $34 + 58 \ = \ 92$, $34 + 58 + 91 \ = \ 183$, osv.

Tallene i 4. række udregner vi sådan: $\dfrac{34}{282} \ = \ 0{,}120567$, $\dfrac{92}{282} \ = \ 0{,}326241$, osv.

Længde i meter	0,5-2	2-3	3-4	4-5	5-8
Hyppighed	34	58	91	72	27
Kumuleret hyppighed	34	92	183	255	282
Kumuleret frekvens	12,1 %	32,6 %	64,9 %	90,4 %	100,0 %

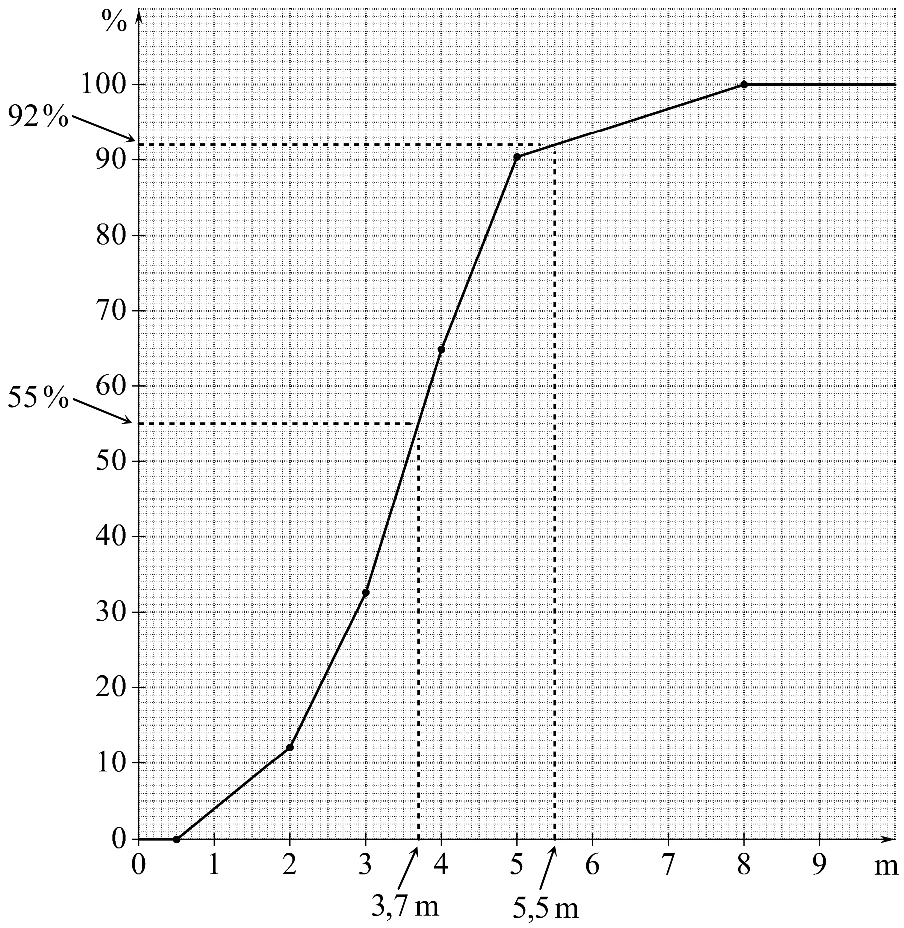
3.33 Hvis start-tal og/eller slut-tal mangler.

Længde (m)	-2	2-3	3-4	4-5	5-
Hyppighed	34	58	91	72	27

I denne tabel mangler start-tal og slut-tal. Så skal vi kun tegne kurven i de andre intervaller. Kurven består altså af tre linjestykker. Spørgsmålene i opgaven vil kunne besvares ved hjælp af den del af kurven vi har tegnet.

3.4 Hvordan aflæser vi på en sumkurve?

Figuren viser sumkurven for rørene fra tabellen på foregående side.



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

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

3.42 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 8% af rørene over 5,5 meter.

3.43 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 37% af rørene mellem 3,7 og 5,5 meter.

3.44 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 (se 3.2).

3.5 Hvordan finder vi medianen for grupperede data?

For ugrupperede data skal vi gøre noget helt andet. Se afsnit 2.2 på side 1.

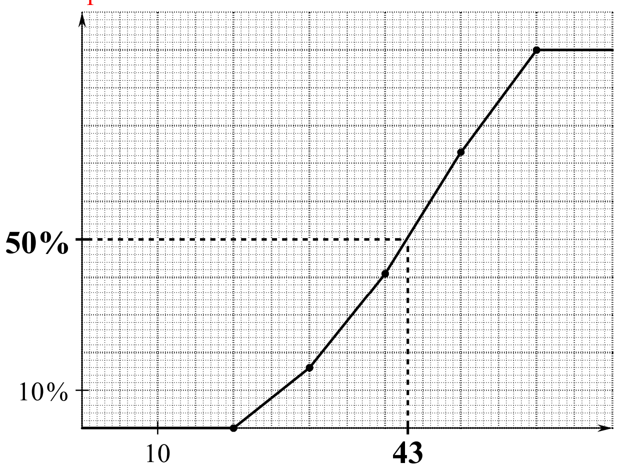
For at finde medianen skal vi bruge sumkurven når det er grupperede data.

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.

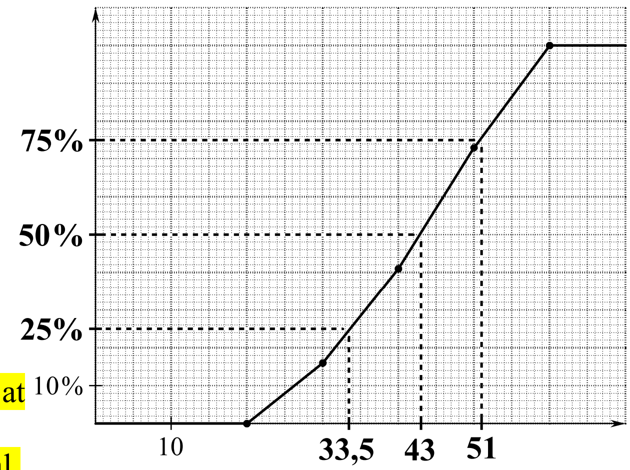


Dette har du brug for at vide når du har fundet medianen og skal svare på hvad dette tal fortæller. I dit svar skal du i stedet for "data" skrive det ord der står i opgaven, f.eks. "længde", og i stedet for "dette tal" skal du skrive det tal du har fundet, f.eks. 43.

3.6 Hvordan finder vi kvartilsættet for grupperede data?

For ugrupperede data skal vi gøre noget helt andet. Se afsnit 2.3 på side 2.

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



3.61 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.

3.62 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 .

3.63 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 .

3.7 Middeltal for grupperede data når antal (hyppighed) er oplyst

Vi vil udregne middeltallet for følgende grupperede datasæt:

Længde i meter	0,5-2	2-3	3-4	4-5	5-8
Antal rør	34	58	91	72	27

For at udregne middeltallet forestiller vi os at
de 34 tal i første interval alle er lig tallet i midten af dette interval,
de 58 tal i andet interval alle er lig tallet i midten af dette interval, osv.
Dette ændrer ikke middeltallet da tallene er jævnt fordelt i hvert interval.

Tallet i midten af intervallet udregner vi sådan: $\frac{0{,}5+2}{2}=1{,}25$, $\frac{2+3}{2}=2{,}5$, osv.

Tal i midten af intervallet	1,25	2,5	3,5	4,5	6,5
Hyppighed	34	58	91	72	27

Antal data er $34+58+91+72+27 = 282$. Middeltallet udregnes sådan:

$$\frac{1{,}25\cdot 34+2{,}5\cdot 58+3{,}5\cdot 91+4{,}5\cdot 72+6{,}5\cdot 27}{282} = 3{,}56560$$

Middeltal for rørs længde er $\underline{\underline{3{,}57\,\text{cm}}}$.

Det er nemmere at gange med 27 end 27 gange at skrive 6,5.

3.8 Middeltal for grupperede data når procent (frekvens) er oplyst

Vi vil udregne middeltallet for følgende grupperede datasæt:

Længde i meter	0,5-2	2-3	3-4	4-5	5-8
Frekvens	12 %	18 %	35 %	25 %	10 %

For at udregne middeltallet forestiller vi os at
de 12 % i første interval alle er lig tallet i midten af dette interval,
de 18 % i andet interval alle er lig tallet i midten af dette interval, osv.
Dette ændrer ikke middeltallet da tallene er jævnt fordelt i hvert interval.

Tallet i midten af intervallet udregner vi sådan: $\frac{0{,}5+2}{2}=1{,}25$, $\frac{2+3}{2}=2{,}5$, osv.

Tal i midten af intervallet	1,25	2,5	3,5	4,5	6,5
Frekvens	12 %	18 %	35 %	25 %	10 %

Middeltallet udregnes sådan:

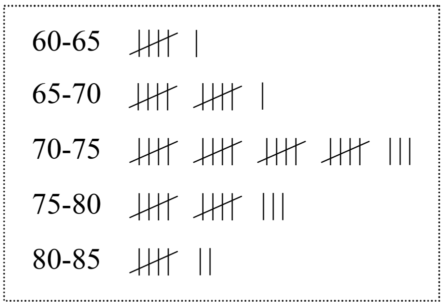
$$\frac{1{,}25\cdot 12+2{,}5\cdot 18+3{,}5\cdot 35+4{,}5\cdot 25+6{,}5\cdot 10}{100} = 3{,}6$$

Middeltal for rørs længde er $\underline{\underline{3{,}6\,\text{cm}}}$.

4.1 Hvordan grupperer vi data?

Vi har modtaget et datasæt på 60 tal (enhed er mm):

63	71	72	78	67	78	84	74	73	66
66	70	72	75	71	72	76	75	82	77
71	62	73	66	75	74	79	68	64	71
72	76	76	82	71	63	62	69	70	69
73	72	78	79	82	75	72	76	77	63
80	83	68	83	66	75	75	82	73	77



For at få overblik over disse tal vil vi gruppere dem i følgende intervaller:

60-65 65-70 70-75 75-80 80-85

I rammen har vi skrevet disse intervaller under hinanden. Første tal i datasættet er 63. Derfor sætter vi en streg ud for 60-65. Andet tal i datasættet er 71. Derfor sætter vi en streg ud for 70-75. Osv.

<u>Et tal i datasættet der er lig et af intervalendepunkterne, tæller vi med i intervallet til venstre for tallet.</u>

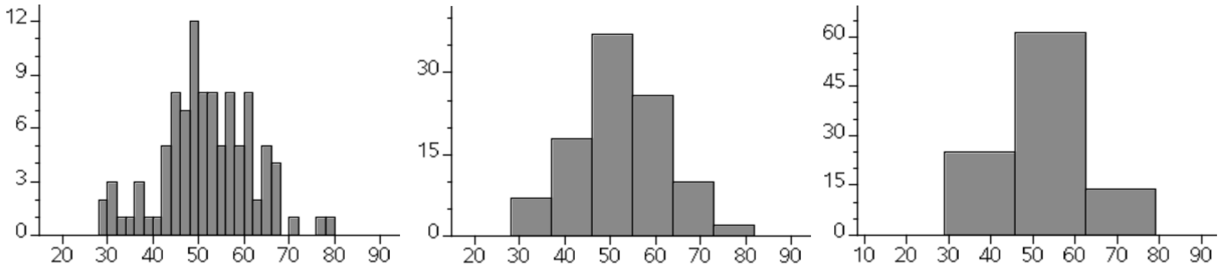
Dette er ikke eneste måde at gøre det på, men der er tradition for at bruge denne måde i det danske hf.

Efter at vi har foretaget denne optælling, kan vi opskrive det grupperede datasæt:

Længde i mm	60-65	65-70	70-75	75-80	80-85
Antal	6	11	23	13	7

4.2 Hvor brede skal vi gøre intervallerne når vi grupperer data?

På computer kan vi nemt ændre intervallernes bredde og se hvordan histogrammet ændres. Histogrammerne viser tre forskellige grupperinger af samme data. På y-aksen står antal.

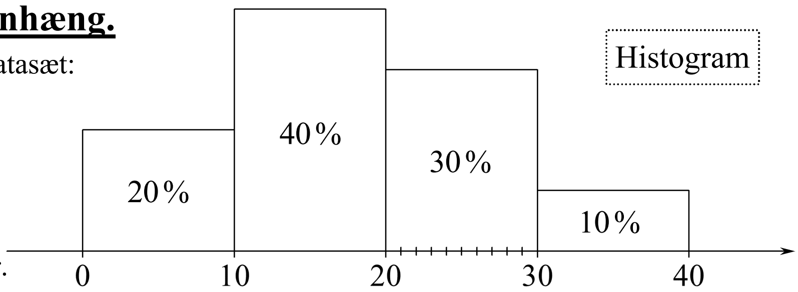


Venstre figur: For lille intervalbredde. Få data i hvert interval får højde til at svinge tilfældigt.

Midterste figur: Intervallernes bredde er passende.

Nederste figur: For stor intervalbredde. Unødig forenklet beskrivelse af hvordan data er fordelt.

5 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 observationer. (x, y) er 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

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

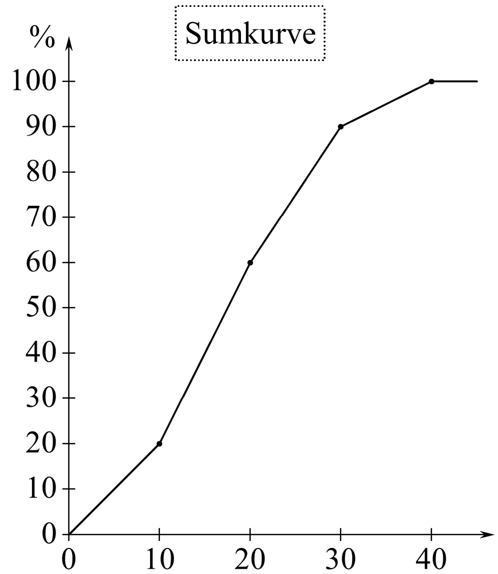
Vi udregner b :

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 .

6.1 Opgave.

Målt i cm har nogle planter følgende højder: 57, 59, 62, 63, 71, 73, 76, 80, 80, 81, 84, 89, 91.
Bestem median og middeltal for disse højder.

Besvarelse 1

Målt i cm har nogle planter følgende højder: 57, 59, 62, 63, 71, 73, 76, 80, 80, 81, 84, 89, 91.

Median

Tallene er sorteret efter størrelse.

Da der er 13 tal må det midterste være nr. syv, altså 76, så medianen er 76 cm.

Middeltal

Da $\frac{57+59+62+63+71+73+76+80+80+81+84+89+91}{13} = 74.3077$ er middeltallet 74 cm.

Besvarelse 2

Skal ikke kunnes, men er en god kontrol.

Målt i cm har nogle planter de højder som jeg har tastet i søjlen **højde** i regnearket.

Jeg vælger **Statistik med en variabel** så Nspire bl.a. udregner:

middeltal = 74,3077

median = 76

Medianen er 76 cm.

Middeltallet 74 cm.

højde	B	C
		=OneVar('
57	Titel	Statistik ...
59	x̄	74.3077
62	Σx	966.
63	Σx²	73308.
71	sx := sn-...	11.2797
73	σx := σn...	10.8371
76	n	13.
80	MinX	57.
80	Q₁X	62.5
81	MedianX...	76.
84	Q₃X	82.5
89	MaxX	91.
91	SSX := Σ...	1526.77

6.2 Opgave.

Målt i cm har nogle planter følgende højder: 91, 89, 59, 76, 73, 81, 57, 71, 62, 84, 80, 63, 80.
Bestem kvartilsættet for fordelingen af planternes højder.

Besvarelse 1

Målt i cm har nogle planter følgende højder: 91, 89, 59, 76, 73, 81, 57, 71, 62, 84, 80, 63, 80.

Vi sorterer tallene efter størrelse: 57, 59, **62**, **63**, 71, 73, **76**, 80, 80, **81**, **84**, 89, 91.

Du skal ikke forklare hvordan du har fået sorteret tallene. Du kan få Nspire til at sortere tallene: Markér søjlen nedefra og op, højreklik i markering, vælg Sortér efter, osv.

Det midterste tal er 76, så median er 76 cm.

Af tallene til venstre for medianen er 62 og 63 de to midterste.

$\frac{62+63}{2} = 62.5$ så nedre kvartil er 62,5 cm.

Af tallene til højre for medianen er 81 og 84 de to midterste.

$\frac{81+84}{2} = 82.5$ så øvre kvartil er 82,5 cm.

Besvarelse 2

Skal ikke kunnes, men er en god kontrol.

Målt i cm har nogle planter følgende højder:

91, 89, 59, 76, 73, 81, 57, 71, 62, 84, 80, 63, 80.

Vi taster disse tal i en søjle i regnearket og vælger **Statistik med én variabel**. Så udregner Nspire bl.a.:

Nedre kvartil er 62,5 cm

Median er 76 cm

Øvre kvartil er 82,5 cm

højde	B	C
		=OneVar('
91	Titel	Statistik ...
89	x̄	74.3077
59	Σx	966.
76	Σx²	73308.
73	sx := sn-...	11.2797
81	σx := σn...	10.8371
57	n	13.
71	MinX	57.
62	Q₁X	62.5
84	MedianX...	76.
80	Q₃X	82.5
63	MaxX	91.
80	SSX := Σ...	1526.77

6.3 Opgave.

Målt i cm har nogle planter følgende højder:

57, 59, 62, 63, 71, 73, 76, 80, 80, 81, 84, 89, 91.

Tegn et boksplot for fordelingen af planternes højder.

Besvarelse 1.

Målt i cm har nogle planter følgende højder:

57, 59, **62**, **63**, 71, 73, **76**, 80, 80, **81**, **84**, 89, 91.

- Det midterste tal er 76, så median er 76. Her er den lodrette streg inden i kassen.
- Af tallene til venstre for medianen er 62 og 63 de to midterste.

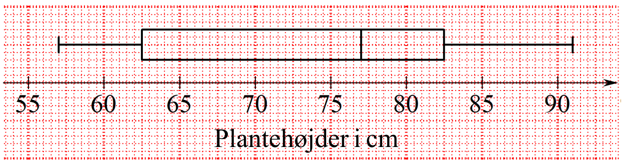
$\frac{62+63}{2} = 62.5$ så nedre kvartil er 62,5. Her er kassens venstre kant.

- Af tallene til højre for medianen er 81 og 84 de to midterste.

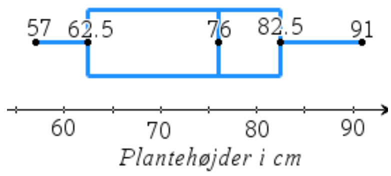
$\frac{81+84}{2} = 82.5$ så øvre kvartil er 82,5. Her er kassens højre kant.

- Mindste tal er 57. Her starter boksplot.
- Største tal er 91. Her slutter boksplot.

Boksplottet er tegnet på vedlagte millimeterpapir.



I stedet for at tegne på millimeterpapir, kan du tegne i et koordinatsystem på Nspire. Start med at afsætte punkter med *x*-koordinater 57 , 62,5 , 76 , 82,5 , 91 . Disse punkter skal have ens *y*-koordinater. Så kan du tegne resten som skitse.



Du kan også få Nspire til at tegne boksplottet (se nedenfor), men så træner du ikke den metode som du er nødt til at kende.

Besvarelse 2 Skal ikke kunnes.

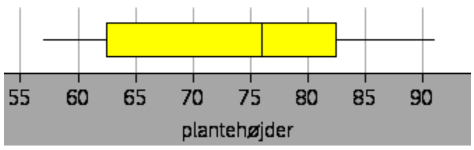
Nogle planters højder er målt i cm.

En søjle i regnearket kalder vi plantehøjder.

Her taster vi højderne.

I statistikvinduet vælger vi på den vandrette akse variablen plantehøjder.

Vi vælger diagramtypen Boksplot.



Hvis tallene i en opgave ikke er i størrelsesorden, så skal du blot taste dem i den orden de står i.

	A plantehøjder
=	
1	57
2	59
3	62
4	63
5	71
6	73
7	76
8	80
9	80
10	81
11	84
12	89
13	91

Opgave 6.4

Nogle planters højder målt i cm kan beskrives ved følgende:

	Mindste	Nedre kvartil	Median	Øvre kvartil	Største
Højde i cm:	57	62,5	76	82,5	91

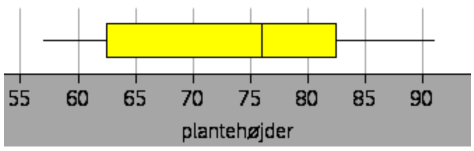
Tegn et boksplot for fordelingen af planternes højder.

Besvarelse Du må gerne selv tegne boksplottet.

Nogle planters højder er målt i cm.

Jeg tegner et boksplot sådan:

- Median er 76. Her er den lodrette streg inden i kassen.
- Nedre kvartil er 62,5. Her er kassens venstre kant.
- Øvre kvartil er 82,5. Her er kassens højre kant.
- Mindste tal er 57. Her starter boksplot.
- Største tal er 91. Her slutter boksplot.



For at få Nspire til at tegne som du har beskrevet, skal du gøre som i 6.3 besvarelse 2 bortset fra at du skal taste tallene 57 , 62,5 , 76 , 76 , 76 , 82,5 , 91 , hvor median er tastet tre gange.

Opgave 6.5

Figuren på bilaget viser et boksplot for fordelingen af nogle planters højder målt cm.
Bestem kvartilsættet for fordelingen af planternes højder.

Besvarelse

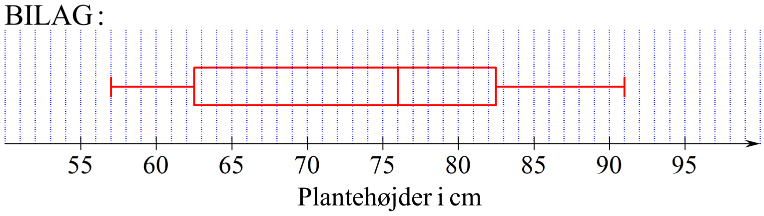
På bilaget ser vi at

kassens venstre kant er ved 62,5
lodret linje inde i kasse er ved 76
kassens højre kant er ved 82,5

så kvartilsættet er

nedre kvartil: 62,5 cm
median: 76 cm
øvre kvartil: 82,5 cm

BILAG:



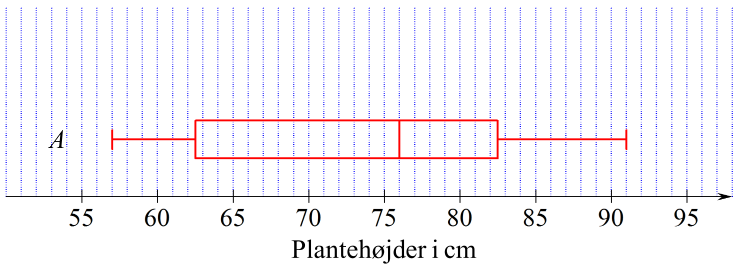
Opgave 6.6

Boksplottet på bilaget viser højder i cm for 13 planter af type A.
Målt i cm har 14 planter af type B følgende højder:

56, 57, 60, 62, 65, 69, 72, 72, 72, 73, 75, 75, 78, 81.

Tegn et boksplot for højderne af planterne af type B, og sammenlign ved hjælp af boksplottene fordelingen af højderne af planterne af type A med fordelingen af højderne af planterne af type B.

BILAG:



Besvarelse

Målt i cm har nogle planter af type B følgende højder:

56 57 60 **62** 65 69 **72** **72** 72 73 **75** 75 78 81

Da der er 14 tal, er det nr 7 og 8 der er de to midterste tal.

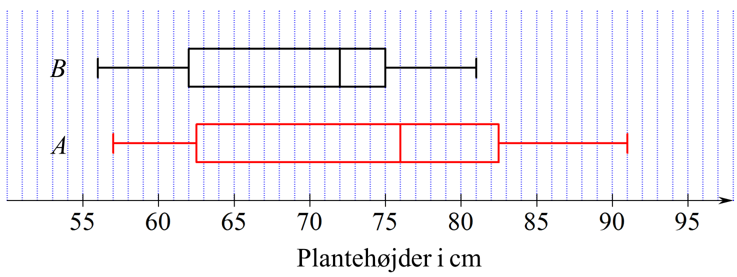
De er begge 72, så median er 72.

For de 7 tal til venstre er det midterste tal nr. 4, dvs. 62, så nedre kvartil er 62.

For de 7 tal til højre er det midterste 75, så øvre kvartil er 75.

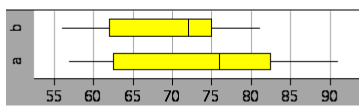
På bilag tegner jeg boksplot for B:

Boksplot starter ved mindste tal 56.
Kasses venstre kant er ved nedre kvartil 62.
Streg inde i kasse er ved median 72.
Kasses højre kant er ved øvre kvartil 75.
Boksplot slutter ved største tal 81.



Sammenligning af boksplot:

B–boksplot ligger længere til venstre end A–boksplot, så
B–højder er mindre end A–højder,
f.eks. er B–boksplots sluting til venstre for A–kasses højre kant, så
alle B–planter er lavere end højeste fjerdedel af A–planter,
og B–kasses højre kant er til venstre for linje inde i A–kasse, så
de 75 % laveste B–planter er mindre end de 50 % højeste A–planter.



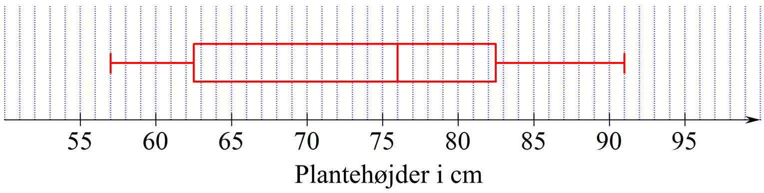
Bemærkning. (Skal ikke kunnes).
Du kan få Nspire til at tegne de to boksplot. Klik under x-akse, vælg b-søjlen, højreklik under x-akse og vælg Tilføj x-variabel, og vælg a-søjlen. Vælg i værktøjsmemuen Diagramtyper/Boksplot.

På side 3 er der flere eksempler på sammenligning af boksplot.

Opgave 6. 7

Figuren på bilaget viser et boksplot for fordelingen af nogle planters højder målt cm.
En plante kan sælges når dens højde er 80 cm eller mere.
Kommentér dette ud fra de oplysninger boksplottet giver.

BILAG:



Besvarelse

En plante kan sælges hvis dens højde er 80 cm eller mere. På boksplottet ser vi at 80 ligger mellem median (ved lodret streg inde i kasse) og øvre kvartil (ved højre kant af kasse), så det er mellem 50 % og 25 % der er 80 cm eller derover. Mellem 50 % og 25 % af planterne kan sælges.

Nederst på side 3 er der flere eksempler på kommentarer af denne type.

Opgave 6.8

Nedenstående tabel viser fordelingen af vægtene af nogle sten.

Vægt (g)	60-70	70-80	80-90	90-100	100-110
Frekvens	9 %	28 %	20 %	11 %	32 %

Hvis der er oplyst hyppighed (dvs. antal) i stedet for procent, så skal du udregne de kumulerede frekvenser med den metode der står i afsnit 3.32.

a) Tegn en sumkurve for fordelingen.
De sten der vejer mere end 94 g skal males.
b) Hvor mage procent af stenene skal males?

A frekv	B
	=cumulativesum(frekv)
9	9
28	37
20	57
11	68
32	100

Vi kan få Nspire til at udregne de kumulerede frekvenser: Klik i tom søjle, vælg Data/Liste operationer/Kumuleret sum og skriv navnet på søjlen der indeholder frekvenserne.

Besvarelse

Vægt af nogle sten er fordelt sådan:

Vægt i g: 60−70 70−80 80−90 90−100 100−110

Frekvens: 9 % 28 % 20 % 11 % 32 %

a) Sådan kan vi udregne kumuleret frekvens:

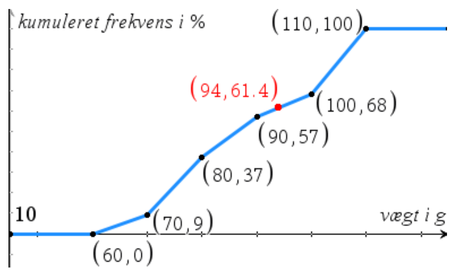
Kumuleret frekvens af 80 =

procentdel der er under 80 =

9 % + 28 % = 37 %

Vi udregner kumulerede frekvenser for at kunne tegne sumkurve:

Vægt i g:	60	70	80	90	100	110
Kum. frkv:	0 %	9 %	37 %	57 %	68 %	100 %



Punkterne bestemt ved denne tabel forbinder vi med rette linjestykker.
Figuren viser sumkurven.

b) Sten der vejer mere end 94 g skal males.

På kurven afsætter vi punkt med x−koordinat 94. y−koordinat bliver 61,4 .

Dvs. 61,4 % vejer under 94 g, så 100 % − 61,4 % = 38,6 % vejer over 94 g.

38,6 % af stenene skal males.

I stedet for at tegne sumkurven med Nspire, må du gerne tegne den på millimeterpapir og skrive aflæsninger herpå som vist side 5-7.
Ofte skal du ikke selv tegne sumkurven. Så skal du på det udleverede bilag skrive aflæsninger som vist side 6-7.

I stedet for at tegne sumkurvens knækpunkter ét ad gangen, kan man i værktøjsmenuen vælge Grafindtastning/Punktplot og angive to søjler hvor man har tastet x-koordinater og y-koordinater.

Hvis der spørges om hvor mange procent der vejer mellem 75 g og 85 g, så skal du på kurven afsætte punkter med *x*-koordinater 75 og 85, og trække mindste *y*-koordinat fra største.

Hvis der spørges om hvor tunge de 20 % tungeste sten er, så skal du på kurven afsætte et punkt med *y*-koordinat 80. *x*-koordinaten er den laveste vægt blandt de 20 % tungeste.

Hvis der spørges om kvartilsættet, så skal du på kurven afsætte punkter med *y*-koordinater 25, 50 og 75. *x*-koordinaterne er kvartilsættet. Se side 7.

Opgave 6.9

Nedenstående tabel viser fordelingen af vægtene af nogle sten.

Vægt (g)	60-70	70-80	80-90	90-100	100-110
Frekvens	9%	28%	20%	11%	32%

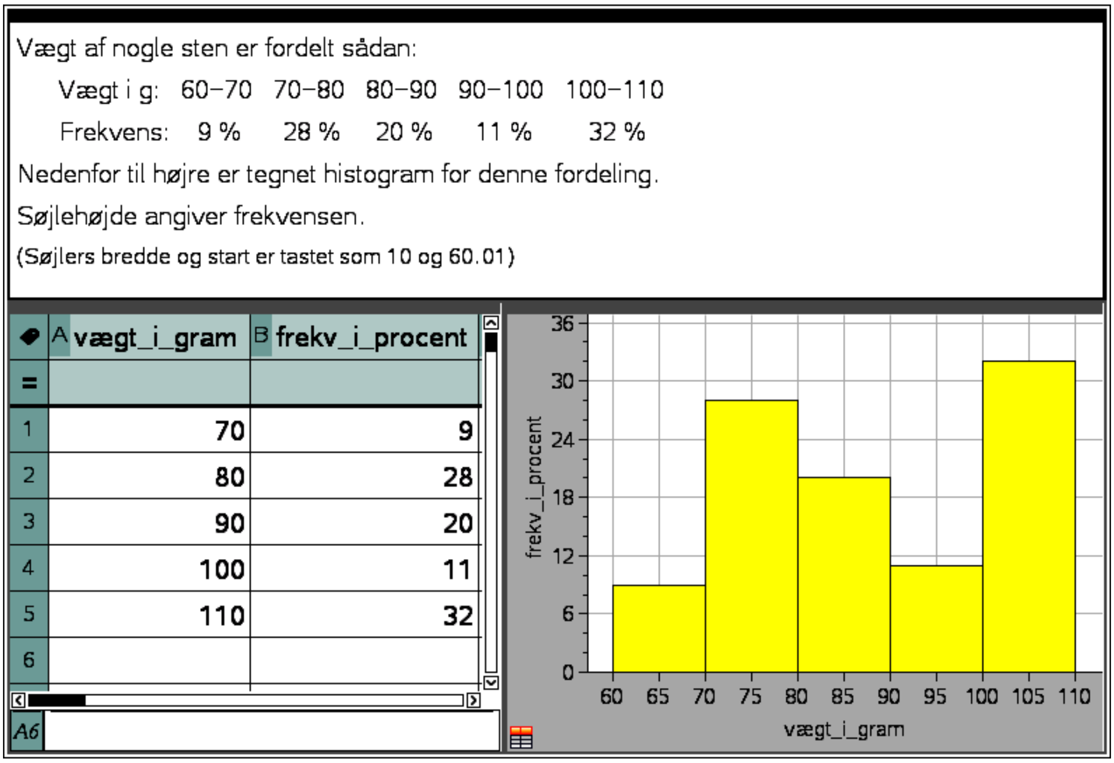
Her kan stå hyppighed (antal) i stedet for frekvens, men så skal du blot kalde søjlen dette i stedet for frekvens_i_procent.

Tegn et histogram for denne fordeling.

Besvarelse

Du behøver ikke bruge Nspire til at tegne histogrammet.
I stedet kan du tegne histogrammet på millimeterpapir som vist i afsnit 3.1 .
Du må gerne tegne histogrammet i Nspire i et sædvanligt koordinatsystem.
Metoden nedenfor er hurtigere når man har lært den.

Når du har tastet de to søjler, så klik under x-akse og vælg variablen vægt_i_gram .
Højreklik til venstre for y-akse, vælg Tilføj y-værdiliste og vælg frekv_i_procent .
Vælg i værktøjsmenuen Diagramegenskaber/Egenskaber for histogram/Søjleindstillinger/Lige store intervaller og tast Bredde som 10 og Søjlestart som 60,01 .



Opgave 6.10

Nedenstående tabel viser fordelingen af vægtene af nogle sten.

Vægt (g)	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12
Antal	16	22	30	41	43	39	32	28	21	9	4

Bestem de kumulerede frekvenser.

Besvarelse

Du må gerne besvare denne opgave uden at bruge regneark. Se afsnit 3.32 .

Tabellen nedenfor viser fordelingen af vægtene af nogle sten.

Vægt (g):	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9	9–10	10–11	11–12
Antal:	16	22	30	41	43	39	32	28	21	9	4

Vi taster antallene i en søjle i regnearket.

Nspire bestemmer de kumulerede antal.

Vi ser at samlet antal sten er 285.

Nspire omskriver kumulerede antal til procent ved at dividere med 285 og gange med 100.

Vi taster punktum efter 100 for at få facitter som kommatal.

Vægt (g):	1	2	3	4	5	6	7	8	9	10	11	12
Kumuleret frekvens (%):	0	5,6	13	24	38	53	67	78	88	95	99	100

ADVARSEL: Ved venstre endepunkt af første interval skal der stå 0 % .

	A antal	B kum_ant	C kum_pct
=		=cumulativesum(antal)	=kum_ant/285*100.
1	16	16	5.61404
2	22	38	13.3333
3	30	68	23.8596
4	41	109	38.2456
5	43	152	53.3333
6	39	191	67.0175
7	32	223	78.2456
8	28	251	88.0702
9	21	272	95.4386
10	9	281	98.5965
11	4	285	100.
12			

Stikordsregister

B
boksplot, aflæs 2, 3, 12, 13
boksplot, sammenlign....................................3, 12
boksplot, tegn 2, 11, 12
D
data 1
deskriptiv statistik...1
F
frekvens 5
G
grupperede data 1, 4
gruppering af data..8
H
histogram 4, 8, 9, 14
hyppighed 5
I
intervallers bredde 8
intervallers endepunkter 8
intervals frekvens..5
K
kumuleret frekvens 5, 13, 15
kumuleret hyppighed 5
kvartilsæt for grupperede data 7, 13
kvartilsæt for ugrupperede data 2, 12
M
median for grupperede data 7
median for ugrupperede data 1, 10, 11, 12
middeltal for grupperede data 8
middeltal for ugrupperede data.......................1, 10
mindste tal.. 11
N
nedre kvartil for grupperede data......................... 7
nedre kvartil for ugrupperede data......2, 10, 11, 12
S
største tal.. 11
sumkurve og lineær sammenhæng...................... 9
sumkurve, aflæs 6, 7, 13
sumkurve, tegn når antal oplyst 5
sumkurve, tegn når procent oplyst.................5, 13
U
ugrupperede data.. 1
Ø
øvre kvartil for grupperede data.......................... 7
øvre kvartil for ugrupperede data........2, 10, 11, 12