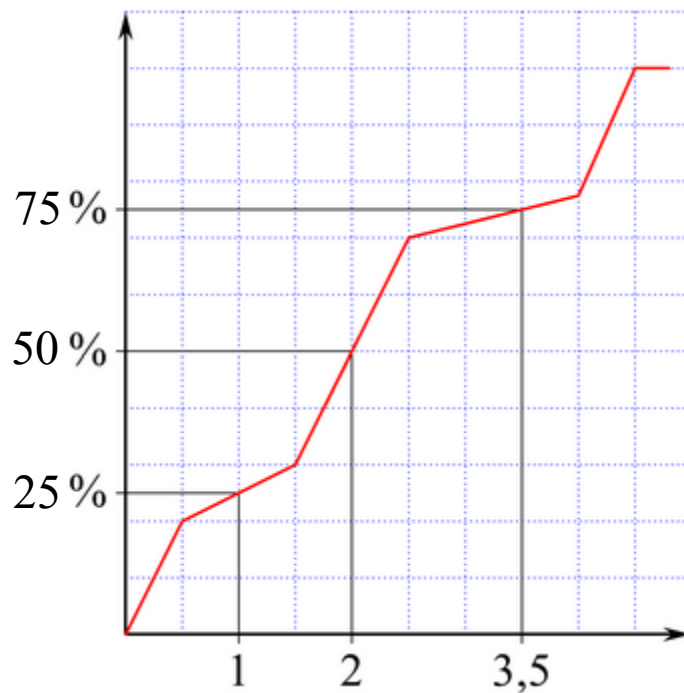


Statistik

for
gymnasiet og hf



2017 Karsten Juul

Nyeste version af dette hæfte kan downloades fra <http://mat1.dk/noter.htm>

Hæftet må benyttes i undervisningen hvis læreren med det samme sender en e-mail til kj@mat1.dk som oplyser at dette hæfte benyttes (angiv fulde titel og årstal), og oplyser hold, niveau, lærer og skole.

Tidligere udgaver af dette hæfte har skiftet adresser til

http://mat1.dk/statistik_for_gymnasiet_og_hf_2011.pdf

http://mat1.dk/statistik_for_gymnasiet_og_hf_2013.pdf

http://mat1.dk/statistik_for_gymnasiet_og_hf_2014.pdf

http://mat1.dk/statistik_for_gymnasiet_og_hf_2015.pdf

http://mat1.dk/statistik_for_gymnasiet_og_hf_2016.pdf

DESKRIPTIV STATISTIK

Hvad er 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 (middelværdi) for ugrupperede data?	1
2.11	Hvordan udregner vi middeltallet når der er få data?	1
2.12	Hvordan udregner vi middeltallet når der er mange data?	1
2.2	Hvordan finder vi medianen for ugrupperede data?	2
2.21	Hvordan finder vi medianen når der er få data?	2
2.22	Hvordan finder vi medianen når der er mange data?	2
2.3	Hvordan finder vi kvartilsættet for ugrupperede data?	3
2.31	Hvis der er et midterste tal	3
2.32	Hvis der ikke er et midterste tal	3
2.4	Hvordan tegner vi boksplot?	3
2.5	Hvordan sammenligner vi boksplot?	4
2.51	opgave	4
2.52	opgave	4
2.53	opgave	4

Grupperede data

3.1	Hvordan tegner vi et histogram?	5
3.2	Et grupperet datasæt er en model af virkeligheden der er meget forenklet	5
3.3	Hvordan tegner vi en sumkurve?	6
3.31	Kumuleret frekvens og sumkurve	6
3.32	Hvis der er oplyst procent for hvert interval	6
3.33	Hvis der er oplyst antal for hvert interval	6
3.34	Hvis start-tal og/eller slut-tal mangler	6
3.4	Hvordan aflæser vi på en sumkurve?	7
3.41	Hvor mange procent af rørene er UNDER 3,7 meter?	7
3.42	Hvor mange procent af rørene er OVER 5,5 meter?	7
3.43	Hvor mange procent af rørene er MELLEM 3,7 og 5,5 meter?	7
3.44	Hvor mange procent af rørene er LIG 3,7 meter ELLER DERUNDER ?	7

3.5	Hvordan finder vi medianen for grupperede data?	8
3.6	Hvordan finder vi kvartilsættet for grupperede data?	8
3.61	Nedre kvartil.	8
3.62	Øvre kvartil.	8
3.63	Kvartilsæt.	8
3.7	Middeltal for grupperede data når antal (hyppighed) er oplyst	9
3.8	Middeltal for grupperede data når procent (frekvens) er oplyst	9
4.1	Hvordan grupperer vi data?	9
4.2	Hvor brede skal vi gøre intervallerne når vi grupperer data?	10
5	Sumkurve og lineær sammenhæng.	10

TEST

6	Stikprøver.	11
6.1	Hvad er populationen?	11
6.2	Hvad er stikprøven?	11
6.3	Systematiske fejl ved valg af stikprøven	11
6.4	Tilfældige fejl ved valg af stikprøven.	11
6.5	Er der skjulte variable?	12
7	Hvad er sandsynlighed?	12
7.1	Eksempel.	12
7.2	Eksempel.	12
8	Test af hypotese..	12
8.1	Signifikansniveau.	12
8.2	Hvornår har vi vist noget med en test?	12
8.3	Tankegangen bag det at forkaste eller ikke forkaste en hypotese	12
9	Test for uafhængighed i 2×2 tabel.	13
9.1	Sådan udregner vi FORVENTEDE TAL	13
9.2	Sådan udregner vi χ^2 .	14
9.3	Sådan udregner vi p .	14
9.4	Sådan skriver vi konklusionen.	14
9.5	Misforstå ikke procenterne	14
10	Hvordan udregner vi antal FRIHEDSGRADER i test for uafhængighed?	15
10.1	Frihedsgrader for 2 gange 2 tabel.	15
10.2	Frihedsgrader for 2 gange 3 tabel.	15
10.3	Frihedsgrader for m gange n tabel.	15
11	Eksempel med to frihedsgrader i test for uafhængighed.	16
12	Nulhypotese.	17
13	Test for fordeling når stikprøven er angivet som antal.	17
13.1	Sådan udregner vi forventede tal.	17
13.2	Sådan udregner vi χ^2 .	17
13.3	Sådan udregner vi antal frihedsgrader.	18
13.4	Sådan udregner vi p .	18
13.5	Sådan skriver vi konklusionen.	18
13.6	Misforstå ikke procenttallene.	18
14	Test for fordeling når stikprøven er angivet med procenter.	19
15	Kritisk værdi.	19

FORDELINGER

Fordeling af teststørrelse

16 Fordeling af teststørrelse.....	20
------------------------------------	----

Normalfordeling

17 Tal der er normalfordelt.....	21
18 Graf for tal der er normalfordelt.....	21
19 Middelværdi for tal der er normalfordelt.....	21
20 Spredning for tal der er normalfordelt.....	21
21 Forskrift for funktion der viser fordelingen af tal der er normalfordelt.....	22
22 En anvendelse af normalfordeling.....	22

χ^2 -fordeling

23 χ^2 -fordeling.....	23
24 Forskrift for g.....	23
25 χ^2 -fordeling når antal frihedsgrader ikke er 1.....	24
26 χ^2 -fordeling og test.....	24

NSPIRE-BESVARELSER

27 Nspire: Bestem middeltal og kvartilsæt ud fra liste.....	25
28 Nspire: Bestem middeltal og kvartilsæt ud fra tabel.....	25
29 Nspire: Tegn boksplot ud fra liste.....	26
30 Nspire: Tegn boksplot ud fra tabel.....	26
31 Nspire: Tegn boksplot ud fra kvartilsæt.....	27
32 Nspire: Tegn flere boksplot sammen.....	27
33 Nspire: Tegn histogram ud fra procent.....	28
34 Nspire: Tegn sumkurve uden brug af regneark.....	28
35 Nspire: Tegn sumkurve ud fra antal.....	29
36 Nspire: Tegn sumkurve ud fra procenter.....	29
37 Nspire: Aflæs sumkurve.....	30
38 Nspire: Test for uafhængighed.....	32
39 Nspire: Test for uafhængighed, mm.....	32
40 Nspire: Test for fordeling, antal i stikprøven.....	33
41 Nspire: Test for fordeling, procent i stikprøven.....	34

DESKRIPTIV STATISTIK

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

1.2.1 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

1.2.2 Eksempel på grupperede data.

Vi har vejret 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 (middelværdi) for ugrupperede data?

For grupperede data skal vi gøre noget andet. Se afsnit 3.7-3.8 på side 9.

Middeltallet for nogle tal er det vi plejer at kalde gennemsnittet.

Vi kan udregne middeltallet (middelværdien) ved at lægge tallene sammen og dividere resultatet med antallet af tal.

2.1.1 Hvordan udregner vi middeltallet når der er få data? (Nspire: side 25 afsnit 27)

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

Sådan udregner vi middeltallet:

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

Middeltallet for elevens pointtal er 7,9

2.1.2 Hvordan udregner vi middeltallet når der er mange data? (Nspire: side 25 afsnit 28)

De nye elever på en skole har været til en prøve:

Point	1	2	3	4	5	6
Antal elever	5	22	58	49	62	18

I tabellen ser vi at 5 elever har fået 1 point, 22 elever har fået 2 point, osv.

Antallet af pointtal er altså

$$5 + 22 + 58 + 49 + 62 + 18 = 214$$

Vi behøver ikke lægge de 58 tretaller sammen.

Vi får det samme ved at udregne $3 \cdot 58$.

Middeltallet kan vi altså udregne sådan:

$$\frac{1 \cdot 5 + 2 \cdot 22 + 3 \cdot 58 + 4 \cdot 49 + 5 \cdot 62 + 6 \cdot 18}{214} = 3,91121$$

Middeltallet for elevernes pointtal er altså 3,9

Har tabellen brøker eller procenttal i stedet for antal, så tast blot dette i stedet i udregningerne, men tast ikke %-tegnet i udregningerne.

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.

2.21 Hvordan finder vi medianen når der er få data? (Nspire: side 25 afsnit 27)

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:

$\overbrace{15\ 20\ 23\ 27\ 27\ 36\ 39\ 47}^{10\text{ tal}}\ 48\ \overbrace{48\ 49\ 52\ 52\ 62\ 69\ 70\ 71}^{8\text{ tal}}$

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:

$\overbrace{3\ 3\ 4\ 5}^{4\text{ tal}}\ \overbrace{6\ 6\ 8\ 9}^{4\text{ tal}}$

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.

2.22 Hvordan finder vi medianen når der er mange data? (Nspire: side 25 afsnit 28)

De nye elever på en skole har været til en prøve:

Point	1	2	3	4	5	6
Antal elever	5	22	58	49	62	18

I tabellen ser vi at 5 elever har fået 1 point, 22 elever har fået 2 point, osv.

Antallet af pointtal er altså

$$5 + 22 + 58 + 49 + 62 + 18 = 214$$

Da $214 : 2 = 107$, ser det sådan ud:

$\overbrace{1\ 1\ \dots\ ?}^{107}\ \overbrace{?\ \dots\ 6\ 6}^{107}$

Tal nr. 6 i denne række er første total da der ifølge tabellen er 5 ettaller.

Tal nr. 28 er første tretal da $5 + 22 = 27$.

Tal nr. 86 er første firtal da $27 + 58 = 85$.

Tal nr. 135 er første femtal da $85 + 49 = 134$.

De to midterste tal, dvs. nr. 107 og 108, er altså begge firtaller.

Da der ikke er noget midterste tal, er medianen gennemsnittet af de to midterste tal.

Medianen for elevernes pointtal er altså 4,0

I tabellen ovenfor ændrer vi antallet 22 til 23. Så ser det sådan ud.

$\overbrace{1\ 1\ \dots\ ?}^{107}\ \overbrace{?\ \dots\ 6\ 6}^{107}$

Vi kan se at nu er det tal nr. 108 der er medianen, dvs. medianen er 4.

2.3 Hvordan finder vi kvartilsættet for ugrupperede data? (Nspire: side 25 afsnit 27 og 28)

(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 boksplot?

(Nspire: Tegn med linjestykker på grafskærm eller gør som beskrevet side 26-27 afsnit 29-32)

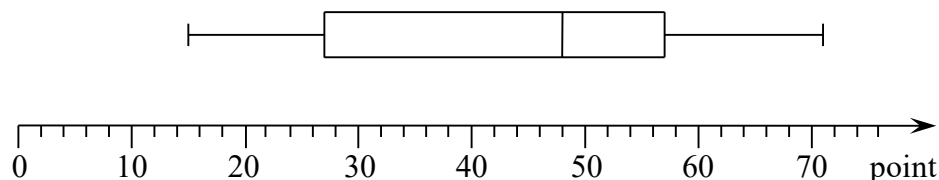
Ved at undersøge datasættet

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.

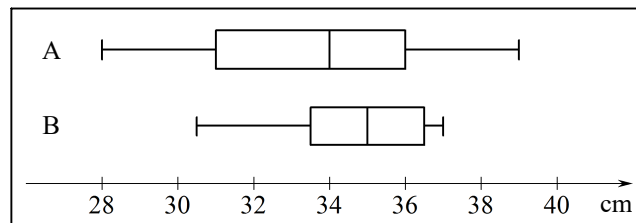
Den vandrette streg til venstre anskueliggør at den fjerdedel af tallene der er mindst, ligger i intervallet fra 15 til 27.

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

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

←..... begrundelse

←..... resultat

Sammenlign spredning

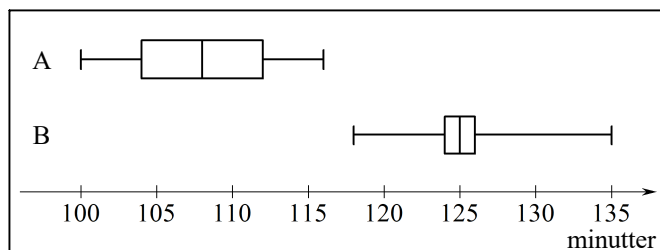
Både hele diagrammet og kassen er bredere på A's diagram end på B's, så højderne fra A er mere spredt end højderne fra B.

←..... begrundelse

←..... 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, så B's mindste tid er større end A's største tid.

←..... begrundelse

←..... resultat

Sammenlign spredning

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

←..... begrundelse

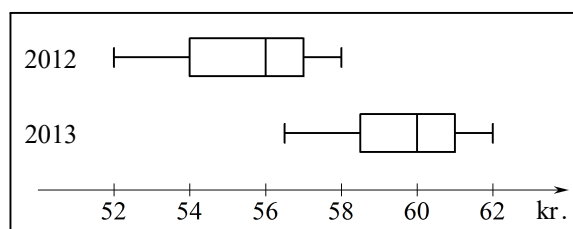
←..... resultat

←..... begrundelse

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

←..... begrundelse

←..... resultat

Hele 2012-diagrammet ligger til venstre for kassen i 2013-diagrammet, så alle 2012-priserne er lavere end de 75 % højeste 2013-priser.

←..... begrundelse

←..... 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, dvs. 53,50 kr. er i den nederste fjerdedel af 2012-priserne.

←..... begrundelse

←..... 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 56,00 kr. og 57,00 kr.

←..... begrundelse

←..... resultat

Grupperede data

3.1 Hvordan tegner vi et histogram? (Nspire: side 28 afsnit 33)

Tabellen viser fordelingen af nogle frugters vægt.

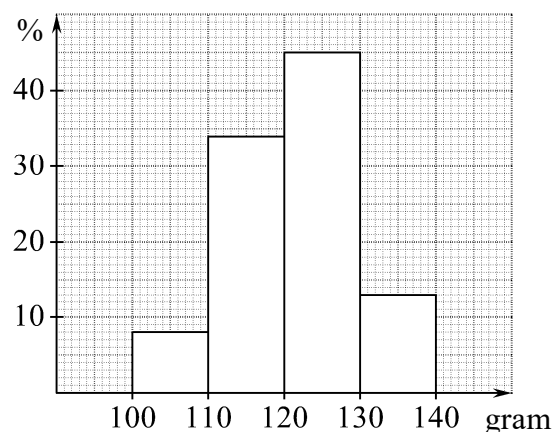
Vægt i gram	100-110	110-120	120-130	130-140
Procent	8	34	45	13

Histogrammet til højre viser oplysningerne i tabellen.

Rektanglet over intervallet 100-110 har højden 8 %.

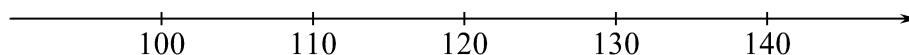
Dette viser at 8 % af frugterne vejer mellem 100 og 110 gram.

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.

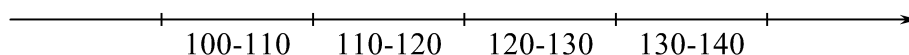


Advarsel: Den vandrette akse skal tegnes som en sædvanlig tallinje.

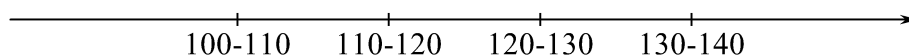
RIGTIGT:



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 opgave hvor du selv skal gruppere nogle data. Se afsnit 4.1 side 9.

3.3 Hvordan tegner vi en sumkurve? (Nspire: side 28-29 afsnit 34 og 36)

3.31 Kumuleret frekvens og sumkurve

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.

3.32 Hvis der er oplyst procent for hvert interval.

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

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

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

For at tegne sumkurven gør vi sådan:

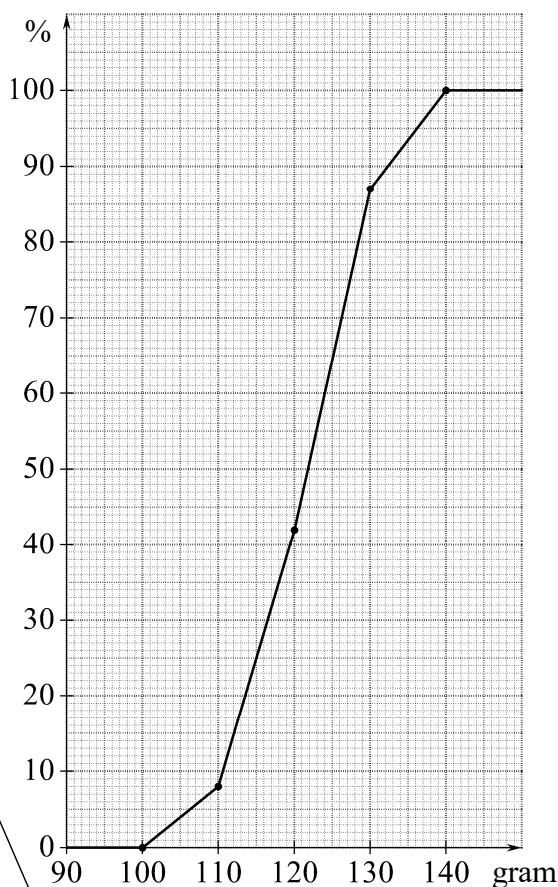
0 % er mindre end 100, så ved $x = 100$ afsætter vi et punkt ud for 0 % på y -aksen.

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. (Se evt. begrundelse for dette i afsnit 5 på side 10).



3.33 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

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:

$34 + 58 = 92$, $92 + 91 = 183$, osv.

Kumuleret frekvens udregner vi sådan:

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

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

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 %

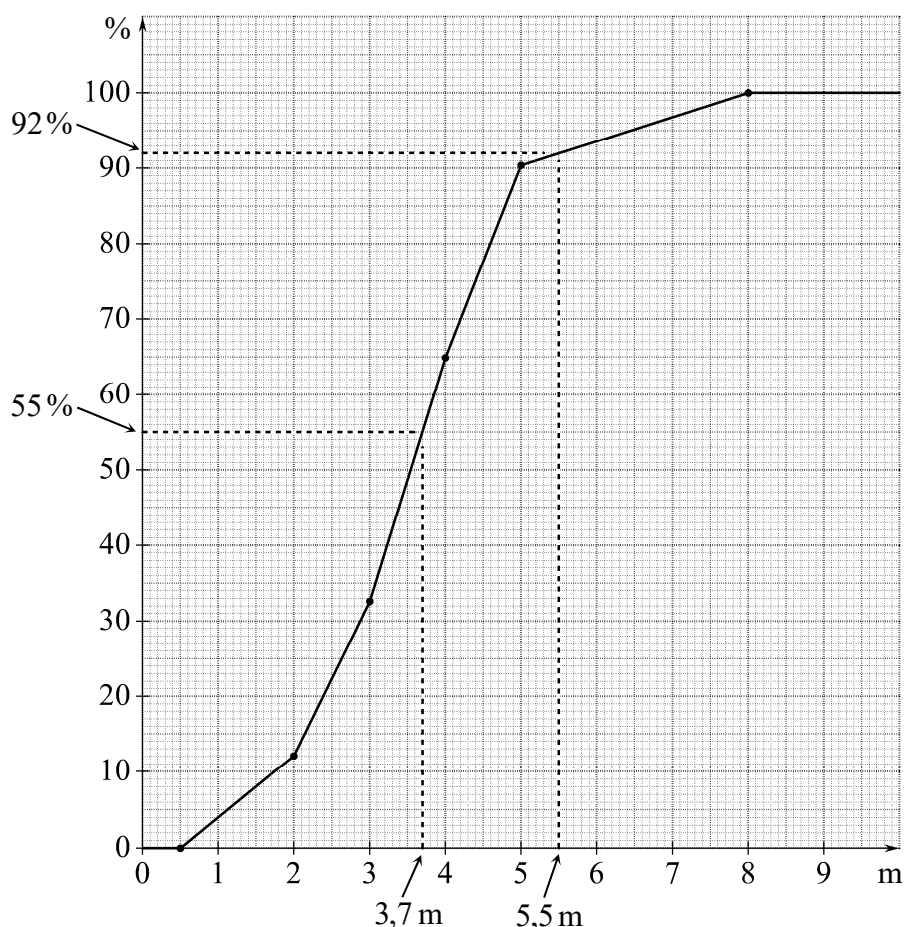
3.34 Hvis start-tal og/eller slut-tal mangler.

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? (Nspire: side 30 afsnit 37)

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... 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 3.2 på side 5).

3.5 Hvordan finder vi medianen for grupperede data? (Nspire: side 30 afsnit 37)

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.

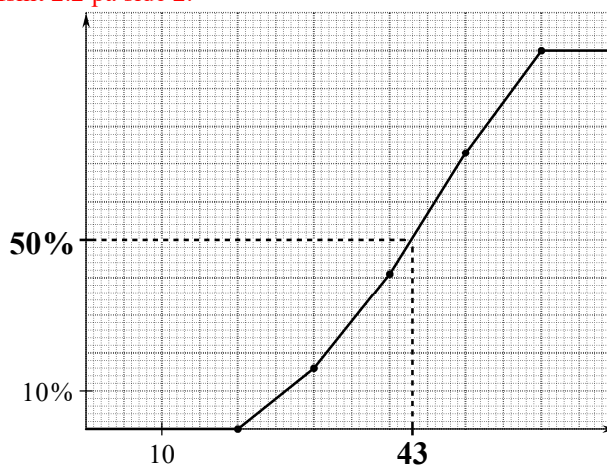
Vi starter i 50 % på y-aksen, går vandret hen til sumkurven, går lodret ned på x-aksen, og aflæser x-værdien.

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" 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? (Nspire: side 30 afsnit 36)

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.

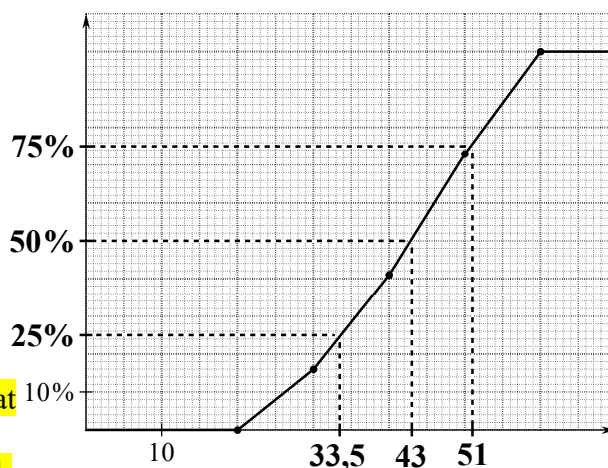
3.61 Nedre kvartil.

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.

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.

↑ 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.

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:

$$\frac{1,25 \cdot 34 + 2,5 \cdot 58 + 3,5 \cdot 91 + 4,5 \cdot 72 + 6,5 \cdot 27}{282} = 3,56560 \quad \text{Middeltal for rørs længde er } \underline{3,57 \text{ cm}} .$$

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:

$$\frac{1,25 \cdot 12 + 2,5 \cdot 18 + 3,5 \cdot 35 + 4,5 \cdot 25 + 6,5 \cdot 10}{100} = 3,6 \quad \text{Middeltal for rørs længde er } \underline{3,6 \text{ cm}} .$$

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

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

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.

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

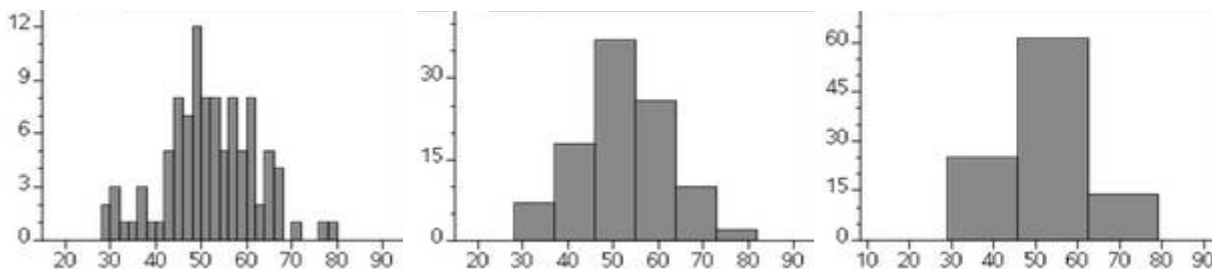
Det er ikke eneste måde at gøre det på, men der er tradition for at bruge denne måde i det danske gymnasium og hf.

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 intervallerne 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: Intervallerne bredde er passende.

Højre figur: For stor intervalbredde. Unødigt forenklet beskrivelse af hvordan data er fordelt.

5 Sumkurve og lineær sammenhæng.

Histogrammet viser et grupperet datasæt:

Intervallerne 20-30 deles 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

$$y = 0,03x.$$

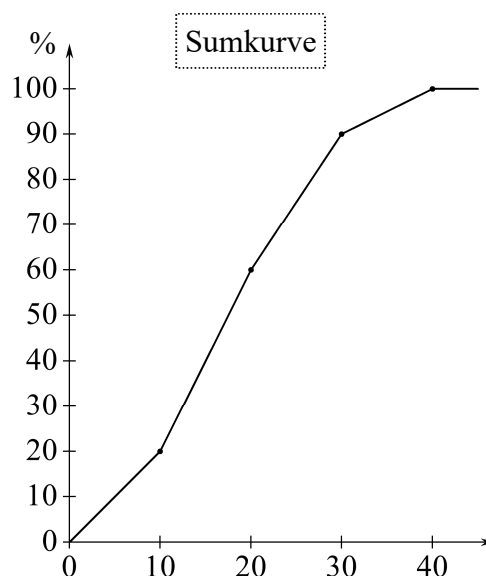
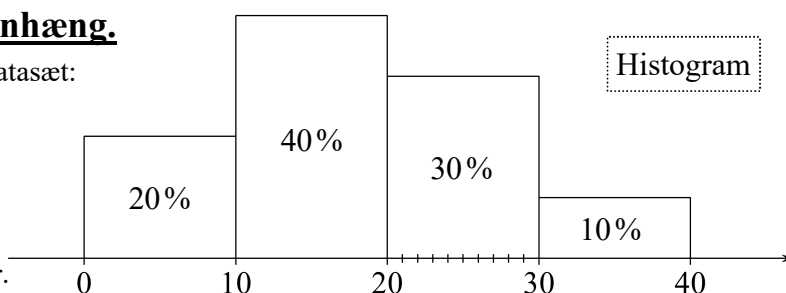
For de fire intervaller er ligningerne:

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.

TEST

6 Stikprøver.

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

6.1 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.

6.2 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.

6.3 Systematiske fejl ved valg af stikprøven.

Eksempel 1

Population: Eleverne på vores gymnasium.

Stikprøve: Eleverne i en sproglig klasse.

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.

6.4 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.

Det er muligheden for tilfældige fejl vi beskæftiger os med når vi udregner tallet p . (se afsnit 9.3 og 13.4).

Afsnit 6 fortsætter på næste side.

6.5 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.

7 Hvad er sandsynlighed?

7.1 Eksempel. At sandsynligheden for at vinde = 25 %
betyder at vi vinder 25 % af gangene.

7.2 Eksempel. At sandsynligheden for at en pose har 3 eller flere defekte = 4 %
betyder at 4 % af poserne har 3 eller flere defekte.

8 Test af hypotese.

8.1 Signifikansniveau.

Hypotese: *Halvdelen af brikkerne er gule.*

Vi tager en stikprøve for at teste hypotesen.

Vi får en stikprøve der ligger langt fra hypotesen.

Vi udregner at hvis hypotesen er rigtig,

så er sandsynligheden kun 3 %

for at få en stikprøve der ligger så langt eller længere fra hypotesen,

Hvis 3 % er mindre end det valgte signifikansniveau, så forkaster vi hypotesen.

Hvis vi har valgt signifikansniveau = 5 % , så forkaster vi da $3\% < 5\%$.

Hvis vi har valgt signifikansniveau = 1 % , så forkaster vi ikke da $3\% \geq 1\%$.

8.2 Hvornår har vi vist noget med en test?

Kun når vi forkaster en hypotese, har vi vist noget.

Når vi ikke forkaster hypotesen, har vi IKKE vist at hypotesen sandsynligvis er rigtig.

Mange (også lærere og lærebøger) tror at når vi ikke kan forkaste hypotesen, så er det sandsynligt at hypotesen er rigtig. Dette er en katastrofal misforståelse. Det har ingen forbindelse med virkeligheden. I nogle eksamensopgaver og vejledende opgaver spørges om der er belæg for uafhængighed. At spørge sådan er en grov fejl da testen aldrig kan give belæg for uafhængighed.

8.3 Tankegangen bag det at forkaste eller ikke forkaste en hypotese.

Antag at vi har fundet ud af følgende:

Hvis det hypotese A siger, er rigtigt, så er det usandsynligt at få en stikprøve der ser ud som den vi har fået.

Så tror vi ikke at det hypotese A siger, er rigtigt. Vi forkaster hypotese A.

Antag at vi har fundet ud af følgende:

Hvis det hypotese A siger, er rigtigt, så er det IKKE usandsynligt at få en stikprøve der ser ud som den vi har fået.

Så forkaster vi ikke A.

Hvis vi ikke forkaster A, er det så sandsynligt at A er rigtig? NEJ, for der er altid flere hypoteser B, C, D ... hvorom der gælder:

Hvis det hypotesen siger, er rigtigt, så er det IKKE usandsynligt at få en stikprøve der ser ud som den vi har fået.

9. Test for uafhængighed i 2x2 tabel.

Vi vil teste om piger og drenge har samme holdning til et spørgsmål.

Populationen er alle danskere hvis alder er 16 til 18 år.

Hypotese: *Andelen der siger ja, er ens for piger og drenge.*

Hypotesen er altså at svaret er uafhængigt af om det er en pige eller en dreng.

I den slags test vi laver her, gælder altid: Hypotesen er at noget er ens.

Vi vælger: **signifikansniveau = 5%**

Nogle tilfældigt udvalgte piger og drenge får stillet samme spørgsmål.

Stikprøven er de piger og drenge som svarede.

De svarede sådan:

Faktiske tal	ja	nej
piger	87	46
drenge	71	21

9.1 Sådan udregner vi FORVENTEDE TAL.

For at teste hypotesen, udregner vi først noget vi kalder de forventede tal, dvs. hvordan svarene skulle være fordelt mellem de fire felter hvis ja-andelen i tabellen skulle være ens for piger og drenge.

For at kunne udregne de forventede tal udregner vi først følgende tal:

Antal piger:	$87 + 46 = 133$
Antal drenge:	$71 + 21 = 92$
Antal ja-sigere:	$87 + 71 = 158$
Antal nej-sigere:	$46 + 21 = 67$
Antal piger og drenge:	$133 + 92 = 225$

Disse tal skriver vi i et skema:

	ja	nej	i alt
piger			133
drenge			92
i alt	158	67	225

I tabellen med forventede tal skal andelen af ja-sigere være den samme for piger og drenge, så da 158 af de 225 elever svarede ja, skal vi i denne tabel skrive at $\frac{158}{225}$ af de

133 piger svarede ja:

forventet antal ja-svar fra piger:	$133 \cdot \frac{158}{225} = 93,40$
forventet antal nej-svar fra piger:	$133 - 93,40 = 39,60$
forventet antal ja-svar fra drenge:	$158 - 93,40 = 64,60$
forventet antal nej-svar fra drenge:	$92 - 64,60 = 27,40$

Her er de fire forventede tal skrevet ind i skemaet:

Forventet	ja	nej	i alt
piger	93,40	39,60	133
drenge	64,60	27,40	92
i alt	158	67	225

9.2 Sådan udregner vi χ^2 .

Symbolet χ^2 læses sådan: *ki i anden*

For hvert af de fire felter udregner vi

$$\frac{(\text{faktisk} - \text{forventet})^2}{\text{forventet}}$$

Ved at lægge disse fire tal sammen får vi tallet χ^2 :

$$\begin{aligned}\chi^2 &= \frac{(87 - 93,40)^2}{93,40} + \frac{(46 - 39,60)^2}{39,60} + \frac{(71 - 64,60)^2}{64,60} + \frac{(21 - 27,40)^2}{27,40} \\ \chi^2 &= 0.438544 + 1.03434 + 0.634056 + 1.49489 \\ \chi^2 &= 3,60\end{aligned}$$

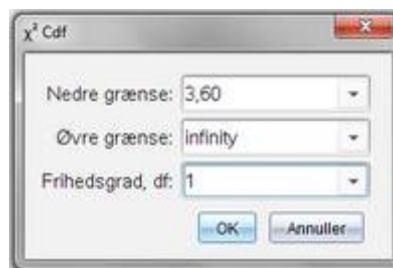
Teststørrelsen 3,60 er afstanden mellem de faktiske tal og de forventede tal.

Vi ser at det er drenge/nej der giver det **største bidrag til teststørrelsen**. Dette kan du blive spurgt om!

9.3 Sådan udregner vi p .

Ovenfor udregnede vi at afstanden mellem faktiske og forventede tal er $\chi^2 = 3,60$.

I beregningsmenuen vælger vi Statistik / Fordelinger / χ^2 Cdf... og udfylder sådan:



Tallet 1 er antal frihedsgrader. Se afsnit 10.

Vi kan skrive:

Nspire: $\chi^2\text{Cdf}(3.6, \infty, 1) = 0.05778$

Dvs. $p = 5,8\%$

hvor p er sandsynligheden hvis hypotesen er rigtig, for at afstanden χ^2 er 3,60 eller større når antal frihedsgrader er 1.

9.4 Sådan skriver vi konklusionen.

Da sigifikansniveauet er 5%, og p ikke er mindre end 5%, kan vi ikke forkaste hypotesen.

Stikprøven giver ikke belæg for at hævde at andelen der siger ja, er forskellig for piger og drenge.

9.5 Misforstå ikke procenterne.

5,8% er IKKE sandsynligheden for at hypotesen er rigtig.

5,8% er IKKE sandsynligheden for at hypotesen er forkert.

94,2% er IKKE sandsynligheden for at hypotesen er rigtig.

94,2% er IKKE sandsynligheden for at hypotesen er forkert.

De 5,8% er udregnet under den forudsætning at hypotesen er rigtig og er sandsynligheden for at få en stikprøve hvis afvigelse fra hypotesen er så stor som eller større end afvigelsen i den stikprøve vi fik.

10 Hvordan udregner vi antal FRIHEDSGRADER i test for uafhængighed?

10.1 Frihedsgrader for 2 gange 2 tabel.

Tabellen stammer fra en undersøgelse af om der er forskel på drenge og pigers holdning til det stillede spørgsmål.

I tabellen er der

2 rækker (piger og drenge)

og

2 søjler (ja og nej).

Hvis vi skriver et tal i ét af de fire felter, så er det fastlagt hvad der skal stå i de tre andre felter. Det er fordi både antal piger, antal drenge, antal ja og antal nej er kendt.

Fordi man kun kan vælge 1 tal, er antallet af frihedsgrader 1.

	ja	nej	i alt
piger	87		133
drenge			92
i alt	158	67	225

10.2 Frihedsgrader for 2 gange 3 tabel.

Tabellen stammer fra en undersøgelse af om der er forskel på drenge og pigers holdning til det stillede spørgsmål.

I tabellen er der

2 rækker (piger og drenge)

og

3 søjler (ja, nej og ?).

Hvis vi skriver tal i to af de seks felter, så er det fastlagt hvad der skal stå i de fire andre felter. Det er fordi både antal piger, antal drenge, antal ?, antal ja og antal nej er kendt.

Fordi man kun kan vælge 2 tal, er antallet af frihedsgrader 2.

	ja	nej	?	i alt
piger	35	12		61
drenge				45
i alt	58	23	25	106

10.3 Frihedsgrader for m gange n tabel.

Hvis der i test for uafhængighed er

m rækker og n søjler,

så er

$$\text{antal frihedsgrader} = (m-1) \cdot (n-1)$$

Nogle gange skal vi udregne antal frihedsgrader på en anden måde. Se afsnit 13.3 .

Hvis $m = 2$ og $n = 2$, så får vi af denne formel at

$$\text{antal frihedsgrader} = (2-1) \cdot (2-1) = 1 \cdot 1 = 1$$

ADVARSEL: Nogle gange skal vi udregne antal frihedsgrader på en anden måde. Se afsnit 13.3 .

11 Eksempel med to frihedsgrader i test for uafhængighed.

Populationen er alle danske elever i 1.g som har matematik på A- eller B-niveau. Vi vil teste følgende hypotese på 5%-signifikansniveau:

Hypotese: *Eleverne på A-niveau og B-niveau er lige gode til brøkgregning.*

Stikprøven er nogle elever som fik en prøve i brøkgregning. Hver elev fik
under middel, *middel* eller *over middel*.

Resultatet står i tabellen.

Faktiske tal:

	under	middel	over	i alt
mat A	51	65	78	194
mat B	44	72	46	162
i alt	95	137	124	356

Vi udregner de forventede værdier:

95 af de 356 elever fik *under middel*, så hvis hypotesen er rigtig, må vi forvente at også $\frac{95}{356}$ af mat A-eleverne fik *under middel*.

Det forventede antal er altså $\frac{95}{356} \cdot 194 = 51,77$.

Det forventede antal mat A-elever med karakteren *middel* er $\frac{137}{356} \cdot 194 = 74,66$.

Vi kan fortsætte sådan for at udregne resten af de forventede tal, men vi kan også udregne resten af de forventede værdier ved at bruge at vi kender summen af hver række og søjle.

Det forventede antal mat B-elever med karakteren *under middel* kan vi altså udregne sådan: $95 - 51,77 = 43,23$.

Forventede tal:

	under	middel	over	i alt
mat A	51,77	74,66	67,57	194
mat B	43,23	62,34	56,43	162
i alt	95	137	124	356

Vi udregner teststørrelsen χ^2 som er afstanden mellem de faktiske tal og de forventede tal:

$$\chi^2 = \frac{(51-51,77)^2}{51,77} + \frac{(65-74,66)^2}{74,66} + \frac{(78-67,57)^2}{67,57} + \frac{(44-43,23)^2}{43,23} + \frac{(72-62,34)^2}{62,34} + \frac{(46-56,43)^2}{56,43}$$

Vi får $\chi^2 = 6,31$.

Antal frihedsgrader = (antal rækker - 1) · (antal søjler - 1) = (2 - 1) · (3 - 1) = 2

Nogle gange skal vi udregne antal frihedsgrader på en anden måde. Se afsnit 13.3.

Vi udregner p som er sandsynligheden hvis hypotesen er rigtig for at χ^2 er i hvert fald 6,31:

Nspire: $\chi^2\text{Cdf}(6.31, \infty, 2) = 0.042638$

Dvs. $p = 4,3\%$

hvor p er sandsynligheden hvis hypotesen er rigtig, for at afstanden χ^2 er 6,31 eller større når antal frihedsgrader er 2.

Da p er mindre end 5%, forkaster vi hypotesen, så på 5%-signifikansniveau har vi vist:

Eleverne på A-niveau og B-niveau er ikke lige dygtige til brøkgregning.

12 Nulhypotese.

Når vi udfører en test, så undersøger vi om en bestemt hypotese kan forkastes. Denne hypotese kaldes nulhypotesen. Nulhypotesen skal påstå at noget er ens.

Dette kan udtrykkes på flere måder. Følgende fire sætninger er samme hypotese:

Der er ikke forskel på pigers og drenges holdning til spørgsmålet.

Pigers og drenges holdning til spørgsmålet er ens.

Elevers holdning til spørgsmålet er uafhængig af køn.

Elevers køn har ikke betydning for deres holdning til spørgsmålet.

Der er ikke sammenhæng mellem køn og holdning til spørgsmålet.

Den alternative hypotese er den hypotese som vi har belæg for at hævde hvis testen forkaster nulhypotesen.

Hvis vi vil undersøge om der er forskel på pigers og drenges holdning til et spørgsmål, så skal vi skrive følgende nulhypotese:

Nulhypotese: *Pigers og drenges holdning er ens.*

Alternativ hypotese: *Pigers og drenges holdning er forskellig.*

13 Test for fordeling når stikprøven er angivet som antal.

I et land er det muligt at stemme på på fire partier A, B, C og D.

Hypotese: *Tilslutningen til partierne er som vist i tabellen.*

Hypotese:

A	B	C	D
33 %	8%	39%	20%

Vi spørger folk der er tilfældigt valgt blandt dem der må stemme, og får følgende 150 svar:

Faktiske tal:

A	B	C	D
46	22	55	27

Populationen er dem der må stemme. Stikprøven er dem der svarede. Vi vil teste hypotesen på signifikansniveau 5 %, samt angive det parti der giver det største bidrag til teststørrelse.

13.1 Sådan udregner vi forventede tal.

Hvis hypotesen er rigtig, må vi forvente følgende tal:

A: $150 \cdot 0,33 = 49,5$ B: $150 \cdot 0,08 = 12$ C: $150 \cdot 0,39 = 58,5$ D: $150 \cdot 0,20 = 30$

Forventede tal:

A	B	C	D
49,5	12	58,5	30

13.2 Sådan udregner vi χ^2 .

Symbolet χ^2 læses *ki i anden*

For hvert af de fire felter udregner vi

$$\frac{(\text{faktisk} - \text{forventet})^2}{\text{forventet}}$$

Ved at lægge disse fire tal sammen får vi teststørrelsen χ^2 som er afstanden mellem de faktiske tal og de forventede tal:

$$\chi^2 = \frac{(46 - 49,5)^2}{49,5} + \frac{(22 - 12)^2}{12} + \frac{(55 - 58,5)^2}{58,5} + \frac{(27 - 30)^2}{30}$$

$$\chi^2 = 0,247475 + 8,33333 + 0,209402 + 0,3$$

$$\chi^2 = 9,09$$

Langt det største bidrag til teststørrelsen er 8,33333 som stammer fra B.

Dette kan du blive spurgt om!

13.3 Sådan udregner vi antal frihedsgrader.

Ovenfor skrev vi tal i følgende skema:

A	B	C	D	I alt
				150

Når vi har skrevet tal i tre af felterne, så er det fastlagt hvad der skal stå i det sidste felt da summen skal være 150.

Antal frihedsgrader er 3 da vi kan vælge tre af tallene.

I test for fordeling gælder:

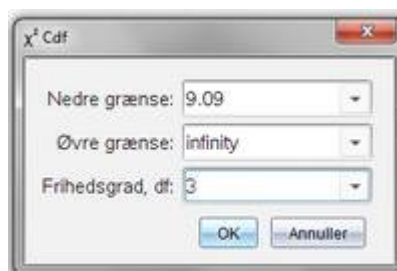
$$\text{antal frihedsgrader} = \text{antal felter} - 1.$$

Nogle gange skal vi udregne antal frihedsgrader på en anden måde. Se afsnit 10.3 .

13.4 Sådan udregner vi p .

Ovenfor udregnede vi at afstanden mellem faktiske og forventede tal er $\chi^2 = 9,09$.

I beregningsmenuen vælger vi Statistik / Fordelinger / χ^2 Cdf... og udfylder sådan:



Vi kan skrive:

Nspire: $\chi^2\text{Cdf}(9.09, \infty, 3) = 0.028118$

Dvs. $p = 2,8\%$

hvor p er sandsynligheden hvis hypotesen er rigtig, for at afstanden χ^2 er 9,09 eller større når antal frihedsgrader er 3.

13.5 Sådan skriver vi konklusionen.

Da p er mindre end 5 %, forkaster vi hypotesen. På 5 %-signifikansniveau har vi belæg for at hævde at:

Fordelingen er ikke som vist i første tabel.

13.6 Misforstå ikke procenttallene.

2,5 % er IKKE sandsynligheden for at hypotesen er rigtig.

2,5 % er IKKE sandsynligheden for at hypotesen er forkert.

97,5 % er IKKE sandsynligheden for at hypotesen er rigtig.

97,5 % er IKKE sandsynligheden for at hypotesen er forkert.

De 2,5 % er udregnet under den forudsætning at hypotesen er rigtig og er sandsynligheden for at få en stikprøve hvis afvigelse fra hypotesen er så stor som eller større end afvigelsen i den stikprøve vi fik.

14 Test for fordeling når stikprøven er angivet med procenter.

I et land er det muligt at stemme på på fire partier A, B, C og D.

Hypotese: *Tilslutningen til partierne er som vist i tabellen.*

Hypotese:

A	B	C	D
33 %	8%	39 %	20 %

Nogen har spurgt folk der er tilfældigt valgt blandt dem der må stemme, og har fået svar fra 150. De har angivet resultatet i procenter sådan:

Vores data:

A	B	C	D
30,7 %	14,7 %	36,7 %	18,0 %

For at kunne teste hypotesen må vi udregne de faktiske tal som disse procenter er udregnet ud fra.

I tabellen med faktiske tal SKAL alle tal være hele tal.

(I tabellen med forventede tal må der gerne stå kommatal).

$$A: 150 \cdot 0,307 = 46,05 \quad B: 150 \cdot 0,147 = 22,05 \quad C: 150 \cdot 0,367 = 55,05 \quad D: 150 \cdot 0,18 = 27$$

Faktiske tal:

A	B	C	D
46	22	55	27

Da vi nu har de faktiske tal, kan vi udføre testen på samme måde som i afsnit 13.

15 Kritisk værdi.

Hver dag undersøger vi nogle varer med en χ^2 -test med 2 frihedsgrader og signifikansniveau 5 %.

En dag er $\chi^2 = 6,88$. Så er $p = 3,2\%$, dvs. vi forkaster.

En dag er $\chi^2 = 4,68$. Så er $p = 9,6\%$, dvs. vi forkaster ikke.

Når $\chi^2 = 6,88$ forkaster vi, og når $\chi^2 = 4,68$ forkaster vi ikke.

Der må være en χ^2 -værdi α mellem 6,88 og 4,68 som skiller mellem at forkaste og ikke forkaste.

Vi vil finde den værdi der skiller, dvs. den χ^2 -værdi hvor $p = 5\%$:

Nspire

løser ligningen $\chi^2\text{Cdf}(\alpha, \infty, 2) = 0,05$ mht. α

og får $\alpha = 5,99$.

Tallet 5,99 kaldes den kritiske værdi.

De dage hvor χ^2 er større end 5,99, forkaster vi hypotesen.

(I andre test vil de kritiske værdier normalt være andre tal).

FORDELINGER

Fordeling af teststørrelse

16 Fordeling af teststørrelse

I skemaet er nogle elever inddelt på to måder:

måde 1: pige eller dreng

måde 2: svarer ja eller nej

	ja	nej
pige		
dreng		

De to inddelinger er uafhængige af hinanden

hvis H_0 : sandsynlighed for ja er ens for pige og drenge

dvs. hvis H_0 : brøkdelen af piger der siger ja = brøkdelen af drenge der siger ja.

Man vælger tilfældigt en stikprøve blandt eleverne og udfylder skemaet ovenfor.

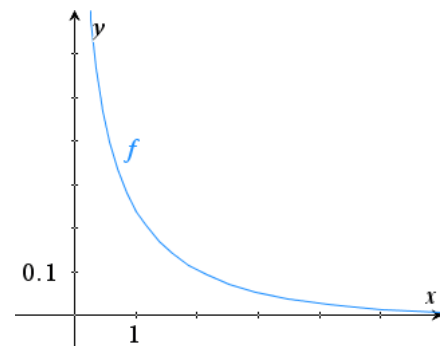
Selv om populationen opfylder brøkdelen af piger der siger ja = brøkdelen af drenge der siger ja, kan vi ikke forvente at dette gælder stikprøven (skemaet). Hvis skemaet er meget langt fra at opfylde dette, så gætter vi på at H_0 er forert. Vi siger at vi forkaster nulhypotesen.

Ud fra stikprøven (skemaet) udregner vi et tal (teststørrelsen χ^2) som er et udtryk for stikprøvens afstand fra hypotese. (Tallet udregnes ved at bruge formelen i afsnit 9.2).

Hvis hypotesen er rigtig og vi mange gange tager en stikprøve og udregner tallet χ^2 , så får vi tal der er fordelt som kurven viser.

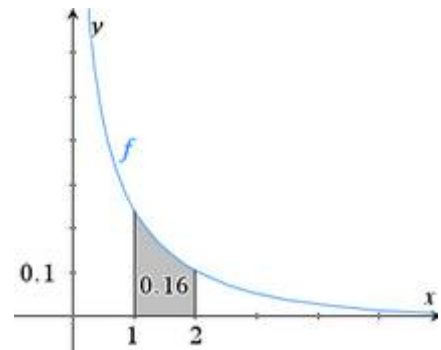
$$f(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} \cdot e^{-\frac{x}{2}}, \quad x > 0.$$

Forskriften for f skal du IKKE huske!



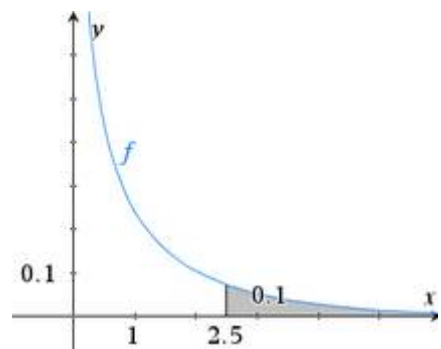
Eksempel: Areal mellem graf og x-akse i intervallet $1 \leq x \leq 2$ er 0,16. Det betyder at 16% af tallene ligger i intervallet. Hele arealet mellem graf og x-akse må være 1 (100%).

$$\int_1^2 f(x) dx = 0,16.$$



Hvis vi får en stikprøve med $\chi^2 = 2,5$, så er p lig sandsynligheden hvis hypotesen er rigtig, for at få en stikprøve hvor χ^2 er 2,5 eller større. På sidste figur ser vi at p er 0,1, altså 10%, så vi forkaster ikke hypotesen hvis signifikansniveau er 5%.

$$\int_{2,5}^{\infty} f(x) dx = 0,1.$$



Normalfordeling

17 Tal der er normalfordelt.

Vi måler højden af alle drenge i 3.g .

De målte tal afsætter vi som prikker på en tallinje.

Tallene ligger ikke jævnt fordelt. De er fordelt sådan:

Der er et sted på tallinjen hvor tallene ligger tæt.

Jo længere man kommer fra dette sted, jo mindre tæt ligger tallene.

Når vi måler ting af samme slags, vil tallene ofte være fordelt på en bestemt måde som vi kalder **normalfordelt**. F.eks. er højderne af drengene i 3.g normalfordelt.

18 Graf for tal der er normalfordelt.

Grafen til venstre viser hvordan nogle målte tal er fordelt.

Hvis vi afsætter tallene som prikker på x -aksen, vil prikkerne ligge tættest der hvor y er størst, dvs. nær 13 på x -aksen ligger prikkerne tættest.

Nær 11 på x -aksen er y mindre, så nær 11 vil prikkerne ligge mindre tæt.

Nær 6 vil prikkerne ligge meget mindre tæt da y er tæt på 0.

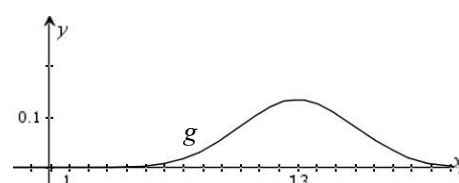
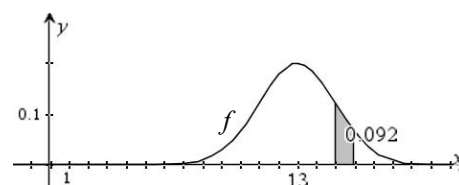
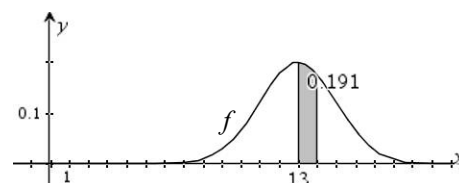
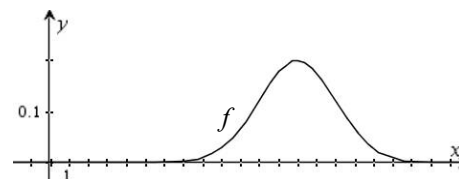
Arealet under grafen er $1 = 100\%$.

I intervallet $13 \leq x \leq 14$ er arealet under grafen 0,191 , dvs. i dette interval ligger 19,1 % af de målte tal.

I intervallet $15 \leq x \leq 16$ er arealet under grafen 0,092 , dvs. i dette interval ligger 9,2 % af de målte tal.

Grafen for g er bredere end grafen for f . Til gengæld er den lavere da arealet under grafen skal være 1.

Når grafen er bredere, ligger tallene mere spredt.



19 Middelværdi for tal der er normalfordelt.

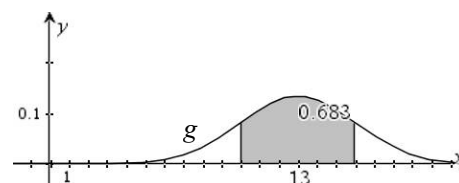
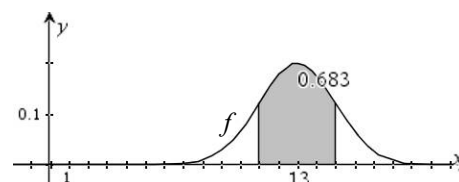
Middelværdien af de målte tal er 13, da grafen er symmetrisk om $x=13$. Dette gælder både for f og g .

20 Spredning for tal der er normalfordelt.

På f -grafene ser vi at hvis vi fra middelværdien går 2 ud til begge sider, så får vi 68,3 % af tallene med.

Vi siger at tallenes spredning er 2 fordi vi skal gå 2 ud fra middelværdien for at få 68,3 % af tallene med.

På g -grafene ser vi at spredningen er 3.



21 Forskrift for funktionen der viser fordelingen af tal der er normalfordelt.

Når tal er normalfordelt med middelværdi m og spredning s , så vil funktionen f der viser fordelingen, have følgende forskrift:

$$f(x) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2}$$

Denne forskrift skal du IKKE huske!

22 En anvendelse af normalfordeling.

Ved fremstillingen af en vare er der på grund af tilfældigheder stor forskel på vægtene af varerne. Vægtene er normalfordelt med spredning 1,5 gram. Varer der vejer under 32 gram skal kasseres. (Jo tungere varen er, jo dyrere er det at fremstille den).

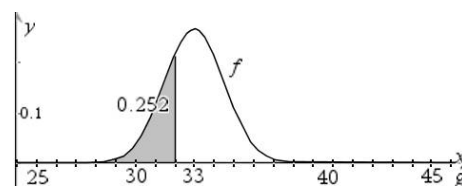
Opgave: Hvor stor en del af varerne skal kasseres hvis maskinen indstilles til at fremstille varer med middelværdi 33 gram ?

Svar: Lad f være funktionen der viser fordelingen af tal der er normalfordelt med middelværdi $m = 33$ og spredning $s = 1,5$.

Arealet under f -grafnen i intervallet $-\infty < x < 32$ er

$$\int_{-\infty}^{32} f(x) dx = 0,252 \quad \text{Udregnet af Nspire.}$$

så 25,2 % af varerne skal kasseres.



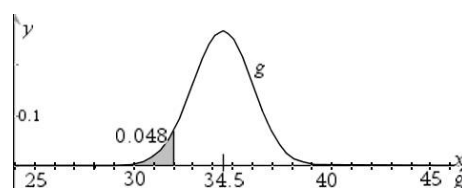
Opgave: Hvor stor en del af varerne skal kasseres hvis maskinen indstilles til at fremstille varer med middelværdi 34,5 gram ?

Svar: Lad g være funktionen der viser fordelingen af tal der er normalfordelt med middelværdi $m = 34,5$ og spredning $s = 1,5$.

Arealet under g -grafnen i intervallet $-\infty < x < 32$ er

$$\int_{-\infty}^{32} g(x) dx = 0,048 \quad \text{Udregnet af Nspire.}$$

så 4,8 % af varerne skal kasseres.



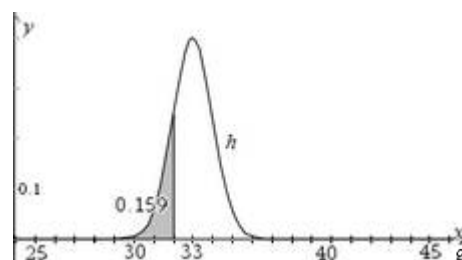
Opgave: En ny maskine laver varer der er normalfordelt med spredning 1 gram. Hvor stor en del af varerne skal kasseres hvis maskinen indstilles til at fremstille varer med middelværdi 33 gram ?

Svar: Lad h være funktionen der viser fordelingen af tal der er normalfordelt med middelværdi $m = 33$ og spredning $s = 1$.

Arealet under h -grafnen i intervallet $-\infty < x < 32$ er

$$\int_{-\infty}^{32} h(x) dx = 0,159 \quad \text{Udregnet af Nspire.}$$

så 15,9 % af varerne skal kasseres.



23 χ^2 -fordeling.

Nogle tal er normalfordelt med middelværdi 0 og spredning 1.

På figur 9 har vi tegnet nogle få af disse tal som prikker.

Disse få af tallene har vi valgt sådan at de giver et indtryk af hvordan alle tallene er fordelt.

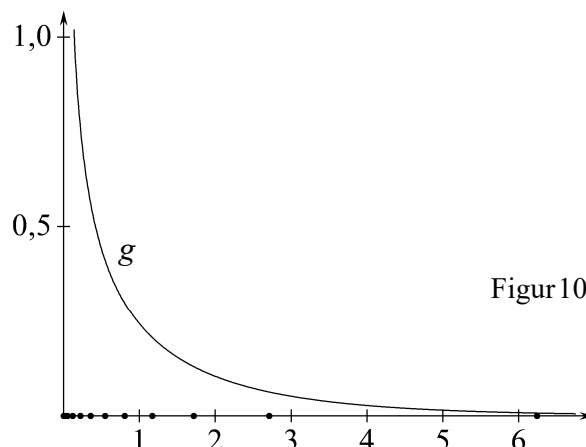
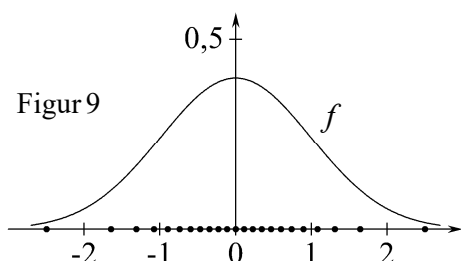
Hvert af tallene opløfter vi til anden:

$$(-0,6)^2 = 0,36$$

$$2,5^2 = 6,25$$

osv.

De tal vi får på denne måde, er fordelt som antydnet på figur 10. Vi vil finde forskriften for funktionen g der viser hvordan disse tal er fordelt. (Kun få af tallene er vist som prikker).



g -grafen viser hvordan tallene χ^2 fra 9.3 er fordelt, dvs.

**hvis vi mange gange tager en stikprøve og hver gang udregner χ^2 ,
så vil vi få nogle tal der er fordelt som g -grafen viser.**

Da vi i afsnit 9 tog en stikprøve hvor $\chi^2 = 3,60$, kunne vi altså have udregnet p sådan:

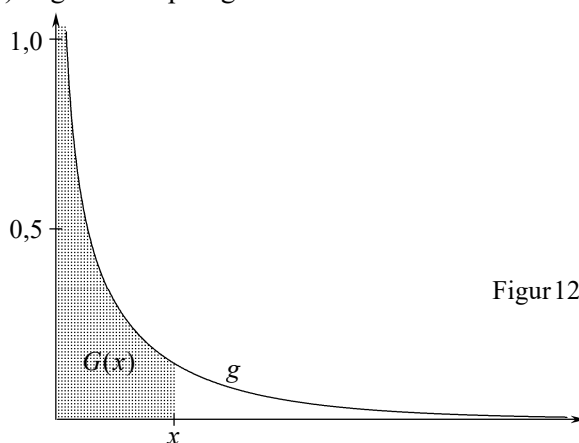
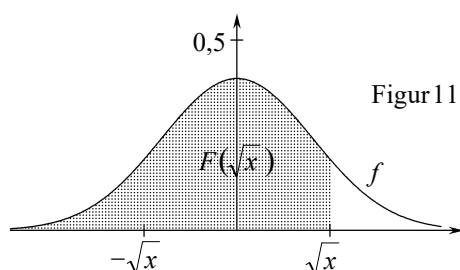
$$p = \int_{3,60}^{\infty} g(x) dx = 0,058.$$

Her har vi brugt forskriften for g som står i linjen (9) i afsnit 25.

24 Forskrift for g .

f og g er funktionerne fra afsnit 23. Arealfunktionerne for f og g kalder vi F og G , dvs.

$F(x)$ = gråt areal på figur 11 og $G(x)$ = gråt areal på figur 12.



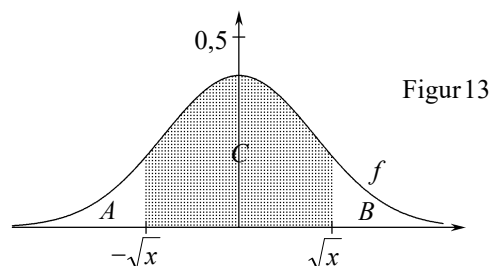
Tallene i intervallet $0 \leq t \leq x$ stammer fra tallene i intervallet $-\sqrt{x} \leq t \leq \sqrt{x}$, så

gråt areal på figur 12 = gråt areal på figur 13

dvs.

$$(1) \quad G(x) = C$$

Af figur 11 og 13 ser vi at $C = F(\sqrt{x}) - A$.



Vi indsætter dette i (1):

$$(2) \quad G(x) = F(\sqrt{x}) - A$$

Da f -grafen er symmetrisk, er $A = B$, så i (2) kan vi erstatte A med B :

$$(3) \quad G(x) = F(\sqrt{x}) - B$$

Hele arealet under f -grafen er 1, så af figur 13 og 11 ser vi at $B = 1 - F(\sqrt{x})$.

Dette indsætter vi i (3):

$$(4) \quad G(x) = F(\sqrt{x}) - (1 - F(\sqrt{x}))$$

Vi reducerer (4) og får:

$$(5) \quad G(x) = 2F(\sqrt{x}) - 1$$

Da G er arealfunktion for g , er

$$(6) \quad g(x) = G'(x)$$

Af (5) og (6) får vi

$$(7) \quad g(x) = (2F(\sqrt{x}) - 1)'$$

Nspire udregner højresiden og får

$$(8) \quad g(x) = \frac{\sqrt{2} \cdot e^{-\frac{x}{2}}}{2 \cdot \sqrt{\pi \cdot x}} \quad \text{Denne formel skal du IKKE huske!}$$

I rammen ses hvordan vi taster for at få (8).

Du skal IKKE kunne denne indtastning.

Define $f(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x}{2}}$ • Udført
fs står for F
Define $fs(x) = \int_{-\infty}^x f(t) dt$ • Udført
Define $g(x) = \frac{d}{dx} (2 \cdot fs(\sqrt{x}) - 1)$ • Udført
 $g(x) = \frac{\sqrt{2} \cdot e^{-\frac{x}{2}}}{2 \cdot \sqrt{\pi \cdot x}}$ ⚠
 $p = \int_{3.6}^{\infty} g(x) dx = 0.05778$

25 χ^2 -fordeling når antal frihedsgrader ikke er 1.

Forskriften (8) omskriver vi til

$$(9) \quad g(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} \cdot e^{-\frac{x}{2}}, \quad x > 0.$$

funktion viser fordelingen af tal der er χ^2 -fordelt med 1 frihedsgrad. I (9) har x eksponenten $-\frac{1}{2}$.

Hver gang vi lægger $\frac{1}{2}$ til eksponenten, bliver antallet af

frihedsgrader 1 større. Samtidig må vi erstatte konstanten $\frac{1}{\sqrt{2\pi}}$ med en anden konstant så

arealet under grafen stadig er 1. Hvis antallet af frihedsgrader er 4, så er tæthedsfunktionen

altså af typen

$$g(x) = k \cdot x \cdot e^{-\frac{x}{2}}$$

Nspire løser ligningen

$$\int_0^{\infty} k \cdot x \cdot e^{-\frac{x}{2}} dx = 1$$

mht. k og får $k = \frac{1}{4}$, så

$$(10) \quad g(x) = \frac{1}{4} \cdot x \cdot e^{-\frac{x}{2}}, \quad x > 0$$

er funktionen der viser fordelingen af tal der er χ^2 -fordelt med 4 frihedsgrader.

26 χ^2 -fordeling og test.

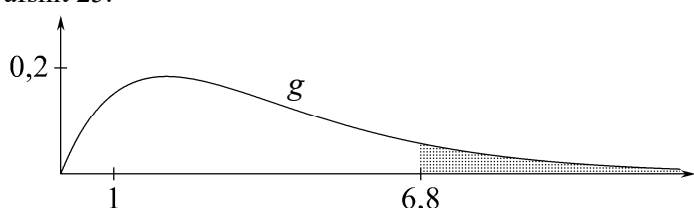
Figuren viser grafen for funktionen (10) fra afsnit 25.

Hvis vi i en χ^2 -test med 4 frihedsgrader får at χ^2 er 6,8, så er p lig det grå areal:

$$p = \int_{6,8}^{\infty} \frac{1}{4} \cdot x \cdot e^{-\frac{x}{2}} dx = 0,146842$$

Normalt regner vi dette tal ud sådan:

$$p = \chi^2 \text{Cdf}(6,8, \infty, 4) = 0,146842$$



Nspire-besvarelser

27 Nspire: Bestem middeltal og kvartilsæt ud fra liste

OPGAVE Vægten i gram af kødstykkerne i en pakke er

10 8 7 8 16 9 10 23 8 23 11 17 19 16 21 7 23 25

Bestem middeltallet og kvartilsættet.

Ordene **middeltal** og **middelværdi** bruges i samme betydning.

Median er IKKE det samme som **middeltal**. Medianen er et af de tre tal i kvartilsættet.

BESVARELSE er vist i de røde rammer.

Gør sådan:

- Vælg **Indsæt / Opgave** og vælg **Indsæt / Side**.
- Klik i det første af vinduerne og vælg **Tilføj Noter**.
- Klik i det andet af vinduerne og vælg **Tilføj Liter og regneark**.
- I det øverste af de to grå felter: Skriv **vægt**.
- I de hvide felter i samme søjle: Skriv tallene 10, 8 7, osv.
- Klik på  (Værktøjsmenuen) og vælg **Statistik / Statistiske beregninger / Statistik med én variabel / Antal lister = 1**. Der fremkommer et vindue hvor du som **X1-liste** skal vælge (IKKE taste) søjlen **vægt**. I feltet **Frekvensliste** skal der stå **1**.
- De fire tal der er spurgt om, står til højre for $\bar{x}$, Q_1 , **MedianX**, Q_3 .
- I det første vindue (Noter): Tast i et matematikfelt **vægt** og tryk på **enter**. Så fremkommer de tastede tal.
- Tilføj tekst og tal som vist i den store røde ramme ovenfor.

Vægten i gram af kødstykkerne i en pakke er tastet i søjlen **vægt**

vægt = { 10,8,7,8,16,9,10,23,8,23,11,17,19,16,21,7,23,25 }

Nspire beregner middeltallet:

middeltal er **14.5 gram** ($\bar{x}$)

Nspire beregner kvartilsættet:

nedre kvartil er **8 gram** (Q_1 , X)

median er **13.5 g** (MedianX)

Øvre kvartil er **21 g** (Q_3 , X)

A vægt	B	C
		=OneVar(
10	Titel	Statistik ...
8	$\bar{x}$	14.5
7	Q_1	261

28 Nspire: Bestem middeltal og kvartilsæt ud fra tabel

OPGAVE I tabellen ser vi at 5 elever fik 1 point, 22 fik 2, osv.

Point: 1 2 3 4 5 6

Antal: 5 22 58 49 62 18 I stedet for **Antal** kan man skrive **Hyppighed**.

Bestem middeltal og kvartilsæt.

Står der brøker eller procenttal i stedet for antal, så tast blot dette i stedet, men tast ikke %-tegnet.

Ordene **middeltal** og **middelværdi** bruges i samme betydning.

Median er IKKE det samme som **middeltal**. Medianen er et af de tre tal i kvartilsættet.

BESVARELSE er vist i den røde ramme.

Gør sådan:

- Vælg **Indsæt / Opgave**.
- Klik i vinduet og vælg **Tilføj Noter**.
- Klik på  og del siden op i et venstre og et højre vindue: .
- Klik i det højre vindue og vælg **Tilføj Liter og regneark**.
- I de øverste grå felter: Skriv **point** og **antal**.
- I de hvide felter i samme søjler: Skriv tallene fra den oplyste tabel.
- Tast evt. brøk eller procenttal (uden %) i stedet for antal.
- Vælg i Værktøjsmenuen: **Statistik / Statistiske beregninger / Statistik med én variabel / Antal lister=1**. I **X1-liste** skal du vælge (IKKE taste) **point**. I **Frekvensliste** skal du vælge (IKKE taste) **antal**.
- De fire tal der er spurgt om, står til højre for $\bar{x}$, Q_1 , **MedianX**, Q_3 .
- Tilføj tekst og tal som vist i den røde ramme.

Point-tal står i søjlen **point**.

Antal elever der fik point-tallet står i søjlen **antal**.

Nspire beregner middeltallet:

middeltal er 3.91121 point $\approx$

3.9 point ($\bar{x}$)

Nspire beregner kvartilsættet:

nedre kvartil er **3 point** (Q_1 , X)

median er **4 point** (MedianX)

Øvre kvartil er **5 point** (Q_3 , X)

	A point	B antal
=		
1	1	5
2	2	22
3	3	58
4	4	49
5	5	62
6	6	18
7		

29 Nspire: Tegn boksplot ud fra liste

OPGAVE Vægten i gram af kødstykkerne i en pakke er

10 8 7 8 16 9 10 23 8 23 11 17 19 16 21 7 23 25

Tegn et boksplot.

BESVARELSE er vist i den røde ramme.

Gør sådan:

- Vælg **Indsæt / Opgave** og **Indsæt / Side**.
- Klik i det første af vinduerne og vælg **Tilføj Noter**.
- Klik i det andet af vinduerne og vælg **Tilføj Liter og regneark**.
- I det øverste af de to grå felter: Skriv **vægt**.
- I de hvide felter i samme søjle: Skriv tallene 10, 8, 7, osv.
- Skift til det første vindue (Noter).
- Klik på  og del siden op i et øvre og et nedre vindue: .
- Klik i det nedre vindue og vælg **Tilføj Diagrammer og statistik**.
- Klik under x-aksen og vælg **vægt**.
- Klik på  (Værktøjsmenuen) og vælg **Diagramtyper / Boxplot**.
- I det øvre vindue (Noter): Tast i et matematikfelt **vægt** og tryk på **enter**. Så fremkommer de tastede tal.
- Tilføj tekst og tal som vist i den røde ramme ovenfor.
- Hvis dit boksplot har prikker som vist nedenfor, så højreklik og vælg **Udvid boksplotgrænser**.

A vægt
10
8
7

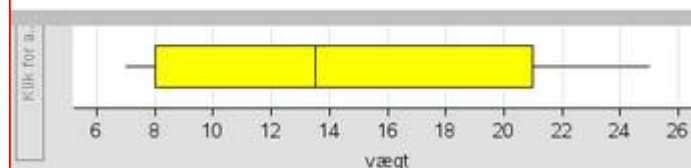
Vægten i gram af kødstykkerne i en pakke testes i søjlen **vægt**.

vægt = { 7,7,8,8,8,9,10,10,11,16,16,17,19,21,23,23,23,25 }

Nspire tegner **boksplot** nedenfor.

Ved elektronisk aflæsning på boksplot får vi:

Mindste værdi: 7
Nedre kvartil: 8
Median: 13.5
Øvre kvartil: 21
Største værdi: 25



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

30 Nspire: Tegn boksplot ud fra tabel

OPGAVE I tabellen ser vi at 5 elever fik 1 point hver, 22 elever fik 2 point hver, osv.

Point: 1 2 3 4 5 6

Antal: 5 22 58 49 62 18

Tegn et boksplot.

BESVARELSE er vist i den røde ramme.

Gør sådan:

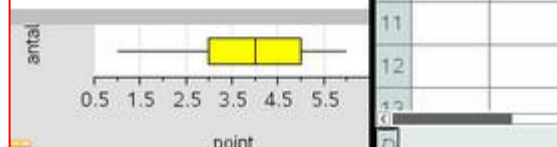
- Vælg **Indsæt / Opgave**.
- Klik på  og del siden op i tre vinduer: .
- Klik i de tre vinduer og tilføj som vist i den røde ramme: **Noter**, **Lister og Regneark**, **Diagrammer og statistik**.
- Tast tabel som vist.
- Klik under x-akse og vælg **point** (navn på første søjle).
- Højreklik til venstre for y-akse vælg **Tilføj Y-værdiliste** og vælg **antal** (navn på anden søjle).
- Klik på  (Værktøjsmenuen) og vælg **Diagramtyper / Boxplot**.
- Tilføj tekst og tal som vist i den røde ramme.
- Hvis dit boksplot har prikker som vist nedenfor, så højreklik og vælg **Udvid boksplotgrænser**.

I søjlen **point** er tastet mulige point. For hver af disse er i søjlen **antal** tastet antal med dette pointtal.

Nspire tegner **boksplot** nedenfor.

Elektronisk aflæsning på boksplot:

Mindste værdi: 1
Nedre kvartil: 3
Median: 4
Øvre kvartil: 5
Største værdi: 6



Du skal IKKE tilføje variabel, læs hvad der står.

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

31 Nspire: Tegn boksplot ud fra kvartilsæt

OPGAVE Nogle planters højder målt i cm kan beskrives ved følgende:

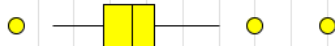
Mindste: 57
Nedre kvartil: 62,5
Median: 76
Øvre kvartil: 82,5
Største: 91

Tegn et boksplot.

BESVARELSE er vist i den røde ramme.

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 de givne tal som vist i højre delvindue.
Bemærk at medianen skal skrives to gange!
- Klik under x-akse og vælg **højde** (navnet på søjlen i regnearket).
- Klik på  (Værktøjsmenuen) og vælg **Diagramtyper / Boxplot**.
- Tilføj tekst og tal som vist i den røde ramme.
- Hvis dit boksplot har prikker som vist nedenfor, så højreklik og vælg **Udvid boxplotgrænser**.



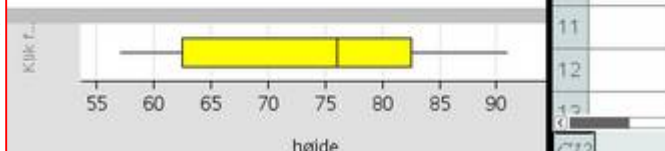
- Hvis du ændrer på tallene, så tilpas boksplot-vinduet ved at højreklikke et tomt sted og vælge **Zoom / Zoom-Data**.

Nogle planters højder målt i cm

kan beskrives ved følgende:

Mindste: 57 boksplots startpunkt
Nedre kvartil: 62,5 kasses venstre kant
Median: 76 lodret streg i kasse
Øvre kvartil: 82,5 kasses højre kant
Største: 91 boksplots slutpunkt

Se **boksplot** nedenfor.



32 Nspire: Tegn flere boksplot sammen

- Hvis du skal tegne mere end et boksplot, og et datasæt er givet ved tabel, så tast dette datasæt i kun én søjle i regnearket ved at gøre følgende:

Bestem kvartilsæt med metod fra afsnit 28, tast derefter en søjle som vist i afsnit 31 hvor medianen er tastet to gange.

OPGAVE Nogle planter af typerne A og B har følgende højder i cm.

A: 57 59 62 63 71 73 76 80 80 81 84 89 91
B: 56 57 60 62 65 69 72 72 72 73 75 75 78 81
Lav på samme figur boksplot for de to datasæt.

BESVARELSE er vist i den røde ramme.

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 de givne tal som vist i højre delvindue.
- Klik under x-akse og vælg søjlen **a**. Højreklik under x-akse og vælg **"Tilføj X-variabel"** og vælg søjlen **b**.
- Klik på  (Værktøjsmenuen) og vælg **Diagramtyper / Boxplot**.
- Tilføj tekst og tal som vist i den røde ramme.
- Hvis dit boksplot har prikker som vist nedenfor, så højreklik og vælg **Udvid boxplotgrænser**.



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

Højder (cm) af nogle planter af typer A og B er tastet i søjler a og b.

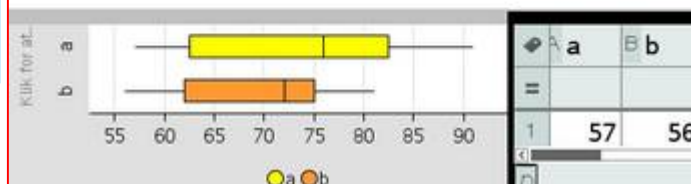
$a = \{57, 59, 62, 63, 71, 73, 76, 80, 80, 81, 84, 89, 91\}$

$b = \{56, 57, 60, 62, 65, 69, 72, 72, 72, 73, 75, 75, 78, 81\}$

Nspire tegner **boksplot** nedenfor.

Elektronisk aflæsning på boksplot:

	mindst	nedre kvartil	median	øvre kvartil	størst
A	57	62.5	76	82.5	91
B	56	62	72	75	81



33 Nspire: Tegn histogram ud fra procent

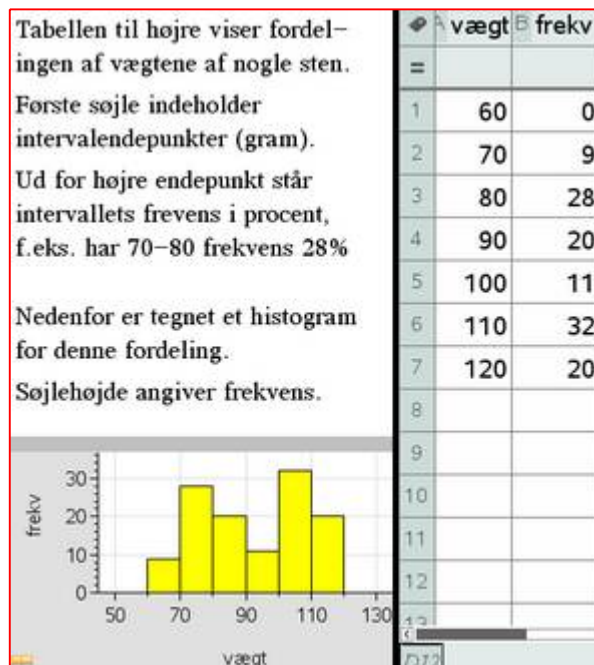
OPGAVE Vægt af nogle sten er fordelt sådan:

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. 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 intervallets frekvens i procent, f.eks. har 70–80 frekvens 28%.
- Nedenfor er tegnet et histogram for denne fordeling. Søjlehøjde angiver frekvens.
- Klik under x-aksen og vælg **vægt** fordi det er det navn som vi har givet første søjle.
- **Højreklik** 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**.

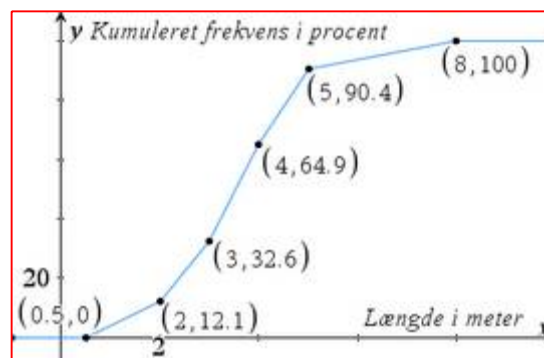


34 Nspire: Tegn sumkurve uden brug af regneark

I afsnit 3.32 og 3.33 er de kumulerede frekvenser udregnet uden brug af regneark. Ud fra disse procenttal kan vi tegne sumkurven på skærmen.

Gør sådan (vi bruger de kumulerede frekvenser fra 3.33):

- 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.



35 Nspire: Tegn sumkurve ud fra antal med brug af regneark

OPGAVE Tabellen viser fordelingen af vægtene af nogle frugter.

Vægt (g): 80-90 90-100 100-110 110-120 120-130

Antal: 6 11 15 32 27

Bestem de kumulerede frekvenser og tegn sumkurven.

BESVARELSE er vist i den røde ramme på denne side og den røde ramme på næste side.

En mulighed er at du selv udregner procenterne og derefter bruger metoden fra afsnit 34. Ellers gør sådan:

- Vælg **Indsæt / Opgave**. Klik på  og del siden op i to vinduer: .
- Klik i venstre vindue og vælg **Tilføj Noter**, og klik i det højre vindue og vælg **Tilføj Lister og regneark**.
- I de øverste grå felter: Skriv de viste fire navne på søjlerne:
- I de hvide felter i to første søjler: tast den oplyste tabel som vist, dvs. antallet af frugter i et interval skal skrives ud for intervallets højre endepunkt.
- Klik i det grå felt under **kumant** og skriv $\text{cumulative sum}(\text{ant})$ og tryk på **enter**.
- Klik i det grå felt under **kumfrk** og skriv $\text{round}(\frac{\text{kumant} \cdot 100}{\text{sum}(\text{ant})}, 1)$ og tryk på **enter**.
- I venstre vindue: Tast den viste forklaring.

Betydning af tal i højre vindue:

endpkt: intervalendepunkter (vægt i gram)

ant: antal i interval står ved højre endepunkt

kumant: kumuleret antal for tallet i

endpkt. Udregnes ved at tallet i **ant** i denne række og over lægges sammen.

kumfrk: kumuleret frekvens for tallet i

endpkt. Er udregnet som

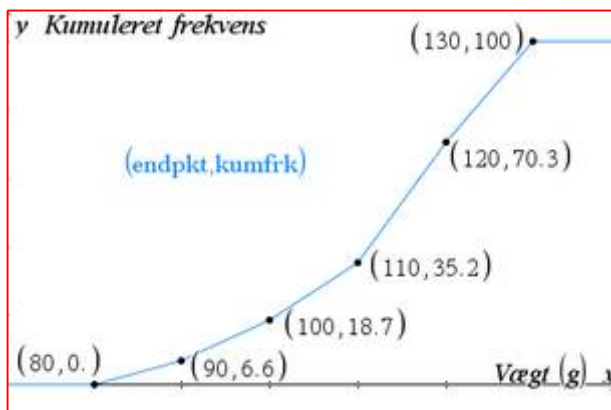
$$\frac{\text{kumant}}{(\text{sum af tal i ant})} \cdot 100$$

og afrundet til én decimal.

Sumkurve er tegnet i næste vindue.

	A endpkt	B ant	C kumant	D kumfrk
=			=cumulative sum	=round(kumant/sum(ant)*100,1)
1	70	0	0	0.
2	80	0	0	0.
3	90	6	6	6.6
4	100	11	17	18.7
5	110	15	32	35.2
6	120	32	64	70.3
7	130	27	91	100.
8	140	0	91	100.

- Vælg **Indsæt / Side** (IKKE **Opgave**).
- Vælg **Tilføj Grafer**.
- Klik på  (Værktøjsmenuen) og vælg **Grafindtastning/Rediger / Punktplo** og tast de søjler der indeholder x og y:
 $x \leftarrow \text{endpkt}$
 $y \leftarrow \text{kumfrk}$
- Klik på  (Værktøjsmenuen) og vælg **Vindue/Zoom / Zoom-Data**
- Forbind punkter med linjestykker: Vælg i Værktøjsmenu **Geometri / Punkter og linjer / Linjestykke**. Så fremkommer linjestykke-ikonen. Før markøren hen til det ene punkt så prikken bliver større, og klik. Før markøren hen til det andet punkt så prikken bliver større, og klik. **HUSK** at trykke på **Esc** for at fjerne ikonen når du er færdig med at tegne linjestykker.
- Fremkald hvert punkts koordinatsæt ved at højreklikke på punktet og vælge **Koordinater og ligninger**. **Dette virker først efter at du har tegnet linjestykker.**
- Skriv tekst på akser.



36 Nspire: Tegn sumkurve ud fra procenter

OPGAVE Den tid det tager en robot at rydde op, varierer som vist i tabellen.

Tid (sekunder): 6.8-7.2 7.2-7.6 7.6-8.0 8.0-8.4 8.4-8.8 8.8-9.2

Procent: 8 12 19 27 20 14

Bestem de kumulerede frekvenser, og tegn sumkurven.

BESVARELSE er vist i de to røde rammer nedenfor. Gør sådan:

- Vælg **Indsæt / Opgave**. Klik på  og del siden op i to vinduer: .
- Klik i venstre vindue og vælg **Tilføj Noter**, og klik i det højre vindue og vælg **Tilføj Lister og regneark**.

- I de øverste grå felter: Skriv de tre viste navne på søjlerne:
- I de hvide felter i to første søjler: tast den oplyste tabel som vist, dvs. antallet af procent i et interval skal skrives ud for intervallets højre endepunkt.
- Klik i det grå felt under **kumfrk** og skriv **cumulativesum(pct)** og tryk på **enter**.
- I venstre vindue: Tast den viste forklaring.
- Vælg **Indsæt / Side** (IKKE **Opgave**).
- Klik i vinduet og vælg **Tilføj Grafer**.
- Klik på  (Værktøjsmenuen) og vælg **Grafindtastning/Rediger / Punktplot** og tast de søjler der indeholder x og y :
 $x \leftarrow \text{endpkt}$
 $y \leftarrow \text{kumfrk}$
- Klik på  (Værktøjsmenuen) og vælg **Vindue/Zoom / Zoom-Data**
- Forbind punkter med linjestykker. Du kan forbinde to punkter sådan: Vælg i Værktøjsmenu **Geometri / Punkter og linjer / Linjestykke**. Så fremkommer en ikon i vinduets øverste venstre hjørne. Før markøren hen til det ene punkt så prikken bliver større, og klik. Før markøren hen til det andet punkt så prikken bliver større, og klik. **HUSK** at trykke på **Esc** for at fjerne ikonen når du er færdig med at tegne linjestykker (ellers kan du ikke gøre andet).
- Fremkald hvert punkts koordinatsæt ved at højreklikke på punktet og vælg **Koordinater og ligninger**.
- Skriv tekst på akser.

I søjlen **endpkt** står intervalendepunkter for tidsintervaller (sekund).

I søjlen **pct** står ud for et intervals højre endepunkt den procent af gangene hvor tiden ligger i intervallet.

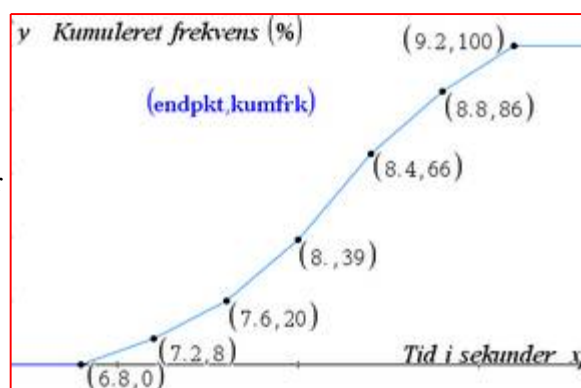
I søjlen **kumfrk** står kumuleret frekvens (%) af tallet i søjlen **endpkt**.

Dette tal udregnes ved at tallene i søjlen **pct** i denne og foregående rækker lægges sammen.

På næste side er tegnet **sumkurven**.

I hvert punkt (x,y) på sumkurven er y lig den kumulerede frekvens af x .

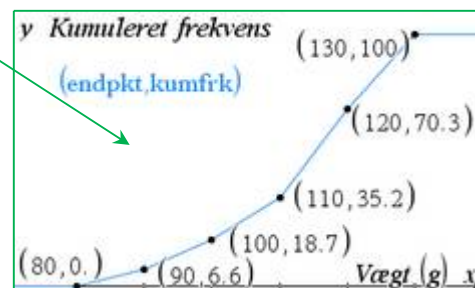
	A endpkt	B pct	C kumfrk
=			=cumula
1	6.4	0	0
2	6.8	0	0
3	7.2	8	8
4	7.6	12	20
5	8.	19	39
6	8.4	27	66
7	8.8	20	86
8	9.2	14	100
9	9.6	0	100
10			



37 Nspire: Aflæs sumkurve

OPGAVE Vi har tegnet sumkurven for vægtene af nogle frugter.

- Hvor mange procent af vægtene er under 103 g ?
- Hvor mange procent af vægtene er over 96 g ?
- Hvor mange procent af vægtene er mellem 115 g og 128 g ?
- Bestem den nedre kvartil for vægtene.
- Bestem medianen for vægtene.
- Bestem den øvre kvartil for vægtene.



BESVARELSE er vist i de to røde rammer på denne og næste side. Gør sådan:

- I spørgsmålet står x -værdien 103, så vi skal bruge linjestykket hvis endepunkter har x lig 100 og 110 da 103 ligger mellem 100 og 110. Klik på  (værktøjsmenuen), vælg **Geometri / Punkter og linjer / Punkt på**, og klik på det nævnte linjestykke, og klik igen på linjestykket. Nu er der afsat et punkt på linjestykket. Tryk på **Esc** for at fjerne Punkt-på-ikonen.

- Højreklik på punktet, og vælg **Koordinater og ligninger**. Så fremkommer punktets koordinatsæt. Dobbeltklik på punktets x -koordinat (tallet før kommaet), ret den til 103, og tryk på **enter**.

- Vi ser at punktets y -koordinat ændres til 23,7. Det betyder at 23,7 % af vægtene er under 103 g.

Du kan ændre antal decimaler på tal i grafvinduet: Højreklik på tallet og vælg attributter.

Klik på de sorte pile for at ændre antal decimaler.

- 

- Med metoden fra spørgsmål a) finder vi ud af at 13,9 % af vægtene er under 96. Fra 100 % trækker vi de 13,9 % for at finde svaret.

- Med metoden fra spørgsmål a) finder vi ud af at 52,8 % af vægtene er under 115, og at 94,1 % af vægtene er under 115. Fra de 94,1 % trækker vi de 52,8 % for at finde svaret.

a) På sumkurven afsætter vi et punkt og taster at x er 103. Vi ser at y er 23,7.

23,7% af vægtene er under 103 g .

b) På sumkurven afsætter vi et punkt og taster at x er 96. Vi ser at y er 13,9.

13,9% af vægtene er under 96 g.

$$100 - 13.9 = 86.1$$

86,1% af vægtene er over 96 g .

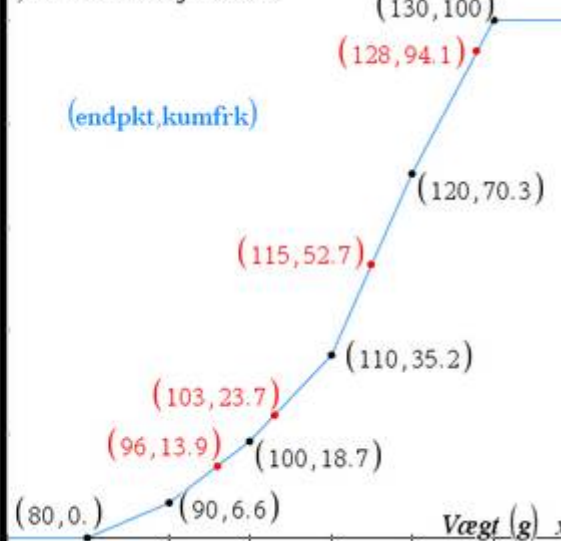
c) På sumkurven afsætter vi to punkter og taster at x er 115 og 128. Vi ser at y er 52,8 og 94,1 .

52,8% er under 115 g og 94,1% er under 120.

$$94.1 - 52.8 = 41.3$$

41,3% af vægtene er mellem 115 og 128 .

y Kumuleret frekvens



d)

- For at finde nedre kvartil skal vi bruge y-værdien 25, så vi skal bruge linjestykket hvis endepunkter har y lig 18,7 og 35,2 da 25 ligger mellem 18,7 og 35,2. Klik på (værktøjsmenuen), vælg **Geometri / Punkter og linjer / Punkt på**, og klik på det nævnte linjestykke, og klik igen på linjestykket. Nu er der afsat et punkt på linjestykket. Tryk på **Esc** for at fjerne Punkt-på-ikonen.
- Højreklik på punktet, og vælg **Koordinater og ligninger**. Så fremkommer punktets koordinatsæt. Dobbeltklik på punktets y-koordinat (tallet efter kommaet), ret den til 25, og tryk på **enter**.
- Vi ser at punktets x-koordinat ændres til 104, dvs. nedre kvartil er 104.

e)

- Vi finder median på næsten samme måde som vi fandt nedre kvartil på i spørgsmål d). Den eneste forskel er at vi skal bruge y-værdien 50 i stedet for y-værdien 25.

f)

- Vi finder øvre kvartil på næsten samme måde som vi fandt nedre kvartil på i spørgsmål d). Den eneste forskel er at vi skal bruge y-værdien 75 i stedet for y-værdien 25.

d) På sumkurven afsætter vi et punkt og taster at y er 25. Vi ser at x er 104.

Nedre kvartil er 104 g .

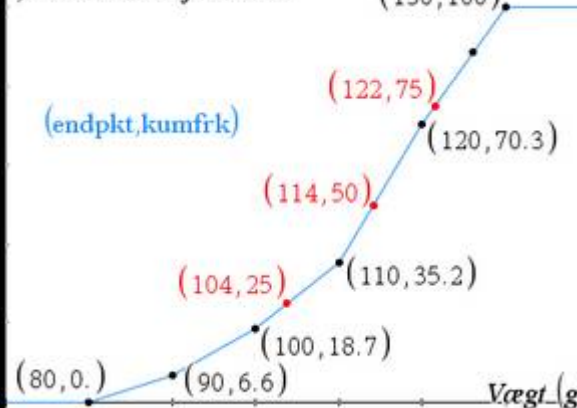
e) På sumkurven afsætter vi et punkt og taster at y er 50. Vi ser at x er 114.

Median er 114 g .

f) På sumkurven afsætter vi et punkt og taster at y er 75. Vi ser at x er 122.

Øvre kvartil er 122 g .

y Kumuleret frekvens



38 Nspire: Test for uafhængighed

OPGAVE Tilfældigt valgte elever er spurgt om de synes bedst om film A eller B. Se resultatet i tabellen.

	A	B
Pige	55	48
Dreng	32	52

Brug en statistisk test på signifikansniveau 5% til at undersøge om meningen om A og B afhænger af køn.

BESVARELSE er vist i den røde ramme.

Gør sådan:

- Vælg **Indsæt / Opgave**, og vælg **Tilføj Noter**.
- Skriv de oplysninger fra opgaven som du skal bruge. (Se linje 1-4 i rød ramme).
- Skriv nulhypotesen. Nulhypotesen er altid at der ikke er forskel. (Se linje 5 i rød ramme).
- Skriv teksten i linje 6. Lav det viste matematikfelt hvor t indføres som betegnelse for tabellen med de givne tal. For at skrive tabellen skal du bruge skabelonen $\begin{bmatrix} \square & \square \\ \square & \square \end{bmatrix}$. Hvis tabellen har flere tal, så bug $\begin{bmatrix} \square & \square & \square \\ \square & \square & \square \end{bmatrix}$ hvor man bliver spurgt om størrelsen. Tegnet $:=$ bevirker at når du trykker på enter, så kommer t til at betegne tabellen.
- Flyt markør til næste linje. Klik på  (værktøjsmenuen) og vælg **Beregninger / Statistik / Statistiske tests / χ^2 -uafhængighedstest**. Der fremkommer den viste linje 7 i den røde ramme ovenfor.
- p -værdien er det tal der står til højre for "PVal". Tast de resterende 5 linjer fra den røde ramme.

Hvis $p \geq 5\%$, skal de to sidste linjer erstattes med:

Da $p \geq 5\%$, forkaster vi ikke nulhypotesen.

På signifikansniveau 5%: **Vi kan ikke slutte at mening om A og B afhænger af køn.**

Hvis der kun spørges om hypotesen forkastes, så skal vi ikke medtage sidste linje i den røde ramme.

Synes du bedst om A eller B? Svar:

	A	B
Pige	55	48
Dreng	32	52

Nulhypotese: Mening om A og B afhænger ikke af køn.

Nspire tester for uafhængighed på $t := \begin{bmatrix} 55 & 48 \\ 32 & 52 \end{bmatrix} = \begin{bmatrix} 55 & 48 \\ 32 & 52 \end{bmatrix}$:

	"Titel"	" χ^2 -uafhængighedstest"
χ^2 2way t: stat.results =	" χ^2 "	4.35494
	"PVal"	0.036902
	"df"	1.
	"ExpMatrix"	"[...]"
	"CompMatrix"	"[...]"

Sandsynligheden hvis hypotesen er rigtig, for at få en tabel der er mindst lige så langt fra hypotesen som t , er

$$p = 0.036902 \approx 3.7\%$$

Da $p < 5\%$, forkaster vi nulhypotese.

På signifikansniveau 5%: **mening om A og B afhænger af køn.**

39 Nspire: Test for uafhængighed, m.m.

OPGAVE Nogle elever vil sammenligne to skoler og tester nogle elever som er tilfældigt valgt blandt de to skolars 3.g-elever. Resultatet er vist i tabellen.

	dårlig	middel	god
A-skole	5	14	11
B-skole	12	7	6

- Hvad er **populationen**, og hvad er **stikprøven**? Opstil en **nulhypotese**, og lav på grundlag heraf en tabel over de **forventede værdier**.
- Bestem **teststørrelsen** og den tilsvarende **p -værdi**, og test hypotesen på signifikansniveau 5%. Inddrag de statistiske begreber "**systematisk fejl**" og "**skjult variabel**" i din kommentar til resultatet.

BESVARELSE af a) er vist i den røde ramme til højre.

Se brugsanvisningen i afsnit 38.

Man får de forventede tal frem ved i et matematikfelt at skrive **stat.ExpMatrix**.

Til sidst i sidste linje i første røde ramme er brugt skabelonen $\begin{bmatrix} \square \\ \square \end{bmatrix}$.

Resultat af test af tilfældigt udvalgte elever fra 3.g på to skoler:

	dårlig	middel	god
A-skole	5	14	11
B-skole	12	7	6

a) Population er **3.g-elever på de to skoler**,

stikprøve er **de der blev testet**.

Nulhypotese: **Karakter afhænger ikke af skole.**

Nspire tester for uafhængighed på $t := \begin{bmatrix} 5 & 14 & 11 \\ 12 & 7 & 6 \end{bmatrix} = \begin{bmatrix} 5 & 14 & 11 \\ 12 & 7 & 6 \end{bmatrix}$:

	"Titel"	" χ^2 -uafhængighedstest"
χ^2 2way t: stat.results =	" χ^2 "	6.28366
	"PVal"	0.043204
	"df"	2.
	"ExpMatrix"	"[...]"
	"CompMatrix"	"[...]"

Nspire udregnede følgende forventede værdier:

	dårlig	middel	god
A-skole	9.27273	11.4545	9.27273
B-skole	7.72727	9.54545	7.72727

- Vælg **Indsæt / Side** (IKKE **Opgave**).

Der fremkommer et vindue hvor vi skriver besvarelse af b).

```
stat.results = [
  "Titel"      "χ²-ufafhængighedstest"
  "χ²"         6.28366
  "PVal"       0.043204
  "df"         2.
  "ExpMatrix"  "[...]"
  "CompMatrix" "[...]"
]
```

b) Nspire udregnede χ^2 -teststørrelsen til 6.28366. Dette tal er et udtryk for afstand mellem stikprøve og forventede værdier. Sandsynlighed hvis hypotese er rigtig, for at få denne afstand eller mere, udregnede Nspire til $p = 0.043204 \approx 4.3\%$

Da $p < 5\%$, forkaster vi nulhypotese.

På signifikansniveau 5%: **elevers dygtighed afhænger af skole.**

Måske har 3g på den ene skole de mest (eller mindst) effektive af skolens lærere dette år, så stikprøven er udtaget blandt elever der ikke er repræsentative for hele skolen (**systematisk fejl**).

En forskel på de to skolars resultater skyldes måske ikke en forskel på skolerne, men skyldes den **skjulte variabel** "forældrenes uddannelsesniveau" som måske er forskellig i de to skolars områder.

- I et matematikfelt skriver vi **stat.result** for at se nogle tal fra sidste vindue som vi skal bruge.
- Tallet til højre for " χ^2 " er teststørrelsen. Skiv besvarelse som vist i den røde ramme ovenfor.

40 Nspire: Test for fordeling, antal i stikprøven

OPGAVE For nogle bær er fordelingen af diametre i 2014 vist i nedenstående tabel.

Diameter i mm:	under 0,5	0,5-1,0	1,0-1,5	1,5-2,0	over 2,0
Antal i %:	6	23	32	28	11

Fordelingen af diametre i en stikprøve fra 2015 er vist i nedenstående tabel.

Diameter i mm:	under 0,5	0,5-1,0	1,0-1,5	1,5-2,0	over 2,0
Antal:	89	339	352	334	119

Man er interesseret i at vide om den procentvise fordeling af diametre i 2015 er den der er angivet for 2014.

- Opstil en nulhypotese, og beregn de forventede værdier med udgangspunkt i nulhypotesen.
- Afgør om nulhypotesen kan forkastes på et 5% signifikansniveau.
- Angiv teststørrelsen, og redegør for hvilket af diameter-intervallerne der giver størst bidrag til teststørrelsen.

BESVARELSE af a) er vist i første røde ramme.

Gør sådan:

- Vælg **Indsæt / Opgave**.
- Klik på  og del siden op i et øvre og et nedre vindue: .
- Klik i det første af vinduerne og vælg **Tilføj Noter**.
- Klik i det andet af vinduerne og vælg **Tilføj Lister og regneark**.
- Skriv de viste ord i øverste grå række.
- Tast " for skrive teksterne i de hvide felter i første søjle
- Skriv de givne tal i de hvide felter i de to næste søjler.
- Dobbeltklik i det tomme grå felt i søjlen **forv**. Der skal stå **forv**:= $1133 \cdot \frac{\text{ford}}{100}$. HUSK decimalpunktum.
- Skriv forklaringen i øverste vindue.

Fordeling i % af diametre fra 2014 står i to første søjler i nedenfor. Stikprøve fra 2015 står i søjlen **ant**.

a) Nulhypotese: Den procentvise fordeling af diametre i 2015 er den der er angivet for 2014.

Antal data fra 2015 er sum(**ant**) = 1133. For første interval er den forventede værdi 6 % af 1133, for næste er det 23 % osv. Forventede værdier står i søjlen **forv**.

	A diam	B ford	C ant	D forv
=				=1133*(ford/(100.))
1	under 0.5	6	89	67.98
2	0.5-1.0	23	239	260.59
3	1.0-1.5	32	352	362.56
4	1.5-2.0	28	334	317.24
5	over 2.0	11	119	124.63

BESVARELSE af b) og c) er vist i den røde ramme.

Gør sådan:

- Vælg **Indsæt / Side** (IKKE opgave) og **Tilføj Noter**.
- Klik på  (værktøjsmenuen) og vælg **Beregninger / Statistik / Statistiske tests / χ^2 -Goodness of Fit test** og vælg **ant** som observeret liste og **forv** som forventet liste, og tast 4 frihedsgrader.
- Tast i et matematikfelt **stat.CompList** og tryk på **enter**.
- Tilføj den forklaring der står i den røde ramme.

b) Nspire udfører χ^2 GOF-test med søjle **ant** (observeret) og **forv**, og 4 frihedsgrader (da 5 felter og $5-1=4$), og får at sandsynligheden hvis hypotese er rigtig, for at få en observeret søjle der er mindst lige så langt fra **forv** som **ant**, er $p = 0.045124 \approx 4.5\%$. Da $p < 5\%$, gælder:

På 5%-signifikansniveau: **Nulhypotesen kan forkastes**.

χ^2 GOF ant,forv,4: stat.results =

"Titel"	" χ^2 -Goodness of Fit test"
" χ^2 "	9.73565
"PVal"	0.045124
"df"	4.
"CompList"	"{...}"

c) Ved udførelse af testen fandt Nspire at teststørrelsen er $\chi^2=9.73565$ og at de fem intervaller bidrag er

-0.5 0.5-1 1-1.5 1.5-2 2-

stat.CompList = { 6.49956, 1.78874, 0.307573, 0.885442, 0.254328 }

Vi ser at 6.49956 er størst, så intervallet "under 0.5 mm" giver det største bidrag. (Når vi lægger bidragene sammen, får vi teststørrelsen som er et udtryk for afstanden mellem observerede og forventede tal).

41 Nspire: Test for fordeling, procent i stikprøven

OPGAVE For nogle bær er fordelingen af diametre i 2014 vist i nedenstående tabel.

Diameter i mm:	under 0,5	0,5-1,0	1,0-1,5	1,5-2,0	over 2,0
Antal i %:	6	23	32	28	11

Fordelingen af diametre i en stikprøve på 1133 bær fra 2015 er vist i nedenstående tabel.

Diameter i mm:	under 0,5	0,5-1,0	1,0-1,5	1,5-2,0	over 2,0
Antal i %:	7.9	21.1	31.1	29.5	10.5

Man er interesseret i at vide om den procentvise fordeling af diametre i 2015 er den der er angivet for 2014.

- Opstil en nulhypotese. Lav en tabel over antal bær fra stikprøven, og lav en tabel over forventet antal bær ifølge nulhypotesen.
- Afgør om nulhypotesen kan forkastes på et 5% signifikansniveau.

BESVARELSE af a) er vist i første røde ramme.

Fordeling i % af diametre fra 2014 er som vist i to første søjler nedenfor.

Stikprøve fra 2015 som procent står i søjlen **pct**. Stikprøvestørrelse er 1133.

a) Nulhypotese: **Den procentvise fordeling af diametre i 2015 er den der er angivet for 2014.**

Stikprøve som antal står i søjle **ant** og er for hvert interval udregnet ved afrunding til helt tal af 1133 gange 2015-procenten delt med 100. På grund af afrunding har den beregnede stikprøve i **ant** ikke præcis 1133 data. Det nøjagtige antal er sum(ant) som er sum af antallene i **ant**.

Forventede værdier står i søjle **forv** og er for hvert interval udregnet som 2014-procenten af sum(ant).

	A diam	B ford	C pct	D ant	E forv	F
=				=round(1133*(pct/100),0)	=sum(ant)*(ford/(100.))	
1	under 0.5	6	7.9	90.	68.04	
2	0.5-1.0	23	21.1	239.	260.82	
3	1.0-1.5	32	31.1	352.	362.88	
4	1.5-2.0	28	29.5	334.	317.52	
5	over 2.0	11	10.5	119.	124.74	

Gør sådan:

- Vælg **Indsæt / Opgave**.
- Klik på  og del siden op i et øvre og et nedre vindue: .
- Klik i det første af vinduerne og vælg **Tilføj Noter**.
- Klik i det andet af vinduerne og vælg **Tilføj Lister og regneark**.

- Skriv de viste ord i øverste grå række.
- Tast " for skrive teksterne i de hvide felter i første søjle
- Skriv de givne tal i de hvide felter i de to næste søjler.
- Dobbeltklik i det tomme grå felt i søjlen **ant**. Der skal stå $\text{ant} := \text{round}\left(1133 \cdot \frac{\text{pct}}{100}, 0\right)$.
- Dobbeltklik i det tomme grå felt i søjlen **forv**. Der skal stå $\text{forv} := \text{sum}(\text{ant}) \cdot \frac{\text{ford}}{100}$. ← HUSK punktum.
- Skriv forklaringen der står i note-vinduet.

BESVARELSE af b) er vist i den røde ramme nedenfor.

Gør sådan:

- Vælg **Indsæt / Side** (IKKE **opgave**) og **Tilføj Noter**.
- Klik på  (værktøjsmenuen) og vælg **Beregninger / Statistik / Statistiske tests / χ^2 -Goodness of Fit test** og vælg **ant** som observeret liste og **forv** som forventet liste, og tast 4 frihedsgrader.
- Tilføj den forklaring der står i den røde ramme.

b) Nspire udfører χ^2 GOF-test med søjler **ant** (observeret) og **forv**, og 4 frihedsgrader (da 5 felter og $5-1=4$), og får at sandsynligheden hvis hypotesen er rigtig, for at få en observeret søjle der er mindst lige så langt fra **forventet** som **antal**, er $p = 0.045124 \approx 4.5\%$.

Da $p < 5\%$, gælder: På 5%-signifikansniveau: **Nulhypotesen kan forkastes**.

"Titel"	" χ^2 -Goodness of Fit test"
" χ^2 "	10.3588
"PVal"	0.034799
"df"	4
"CompList"	"{...}"

χ^2 GOF ant,forv,4: stat.results =

Stikordsregister

#

χ^2 -fordeling.....23, 24

A

alternativ hypotese.....17

arealfunktion.....23

B

boksplot ud fra kvartilsæt.....27

boksplot, flere sammen.....27

boksplot, sammenligne.....4

boksplot, tegne.....3, 26

D

data.....1

deskriptiv statistik.....1

F

faktiske tal i test for fordeling.....17, 19

faktiske tal i test for uafhængighed.....13, 16

forkaste hypotese.....12, 14, 16, 18, 19

forventede tal i test for fordeling.....17

forventede tal i test for uafhængighed.....13, 16

forventede værdier.....32

frekvens.....6

frihedsgrader.....19, 24

frihedsgrader i test for fordeling.....18

frihedsgrader i test for uafhængighed.....15, 16

G

grupperede data.....1, 5

gruppering af data.....9, 10

H

histogram.....5, 10, 28

hypotese.....12, 13, 16, 17, 19

I

intervallers bredde.....10

intervallers endepunkter.....9

intervals frekvens.....6

K

kritisk værdi.....19

kumuleret frekvens.....6

kumuleret hyppighed.....6

kvartilsæt for grupperede.....30

kvartilsæt for grupperede data.....8

kvartilsæt for ugrupperede data.....3, 25

M

median for grupperede data.....8, 30

median for ugrupperede data.....2, 3

middeltal for grupperede data.....9

middeltal for ugrupperede data.....1, 25

middelværdi for grupperede data.....9

middelværdi for normalfordeling.....21

middelværdi for ugrupperede data.....1, 25

N

nedre kvartil for grupperede data.....8, 30

nedre kvartil for ugrupperede data.....3, 25

normalfordeling.....21

normalfordeling, forskrift.....21

nulhypotese.....17, 32

P

p11, 14, 16, 18, 19, 24, 32

population.....11, 13, 16, 17, 32

S

sandsynlighed.....12

signifikansniveau.....12, 16, 18, 19

skjult variabel.....12, 32, 33

spredning.....21

stikprøve.....11, 13, 16, 17, 32

sumkurve.....28, 29, 30

sumkurve og lineær sammenhæng.....10

sumkurve, aflæs.....7, 8, 30

sumkurve, tegn.....28

sumkurve, tegn når antal oplyst.....6, 29

sumkurve, tegn når procent oplyst.....6, 29

systematisk fejl.....11, 32, 33

T

test af hypotese.....12, 17, 19

test for fordeling ud fra antal.....17, 33, 34

test for fordeling ud fra procent.....19, 34

test for uafhængighed.....13, 16, 32

teststørrelse.....14, 16, 17, 32

teststørrelse, største bidrag.....17

tilfældig fejl.....11

U

uafhængig.....17

ugrupperede data.....1

Ø

øvre kvartil for grupperede data.....8, 30

øvre kvartil for ugrupperede data.....3, 25