



MATEMATIKBANKENS STATISTIKKOMPENDIUM

Statistiske begreber	Enkelte observationer	
Grupperede data	Diagrammer	Boksplot
Vurdering af grafisk statistik	Manipulation	

Helle Fjord
Morten Graae
Kim Lorentzen
Kristine Møller-Nielsen

MATEMATISKE KOMPETENCER:**Modellering**

Eleven kan vurdere matematiske modeller.

Eleven har viden om kriterier til vurdering af matematiske modeller.

STOFOMRÅDER:**Statistik & Sandsynlighed****Statistik:**

Eleven kan undersøge sammenhænge i omverdenen med datasæt.

Eleven har viden om metoder til undersøgelse af sammenhænge mellem datasæt, herunder med digitale værktøjer.

Eleven kan kritisk vurdere statistiske undersøgelser og præsentationer af data.

Eleven har viden om stikprøveundersøgelser og virkemidler i præsentation af data.

Eleven kan kritisk vurdere mediers anvendelse af statistik.

Eleven har viden om statistiks muligheder og begrænsninger som beskrivelsesmiddel og beslutningsgrundlag.


LÆRINGSMÅL:

1. Kunne anvende og forstå de statistiske deskriptorer der passer til observationstypen
 - a. Hyppighed, summeret hyppighed, frekvens, summeret frekvens, gennemsnit, median, typetal, kvartilsæt, variationsbredde, største og mindste værdi, interval og intervalmidtpunkt
2. Have forståelse for åbne og lukkede intervaller.
3. Kunne arbejde med enkeltobservationer, grupperede observationer og rådata
 - a. Skal være dynamisk
4. Kunne arbejde med diagrammer
 - a. Cirkeldiagram, stolpediagram, sumkurve, observationstabel og boksplot.
5. Kende til betydning af outliers og kunne tilpasse datasæt i forhold til det, som man undersøger. (Outliers betyder undtagelser)
6. Kunne bruge ovenstående læringsmål til sammenligning/vurdering af en statistisk undersøgelse


LEKTIONSPLAN 12 LEKTIONER FORDELT PÅ 6 GANGE 2 LEKTIONER**1. Dobbeltlektion**

- a. Excelkursus. Hent det her: <http://matematikbanken.dk/L/152/>
- b. Lav opgave 1.1 til 1.5 i Matematikbankens Excelkompendium

2. Dobbeltlektion

- a.  Praktisk arbejde med boksplot i hallen (45 min)
(Se Aktiv matematik - Øvelse #22 <http://matematikbanken.dk/L/148/>)
Link til skabelon til indsamling af elevers data fra jetonkast:
<http://matematikbanken.dk/L/153/> (Lav en kopi og giv linket til eleverne.)
- b. Grundig gennemgang af observationstabel med deskriptorer (Side 6)
(Det er vigtigt, at observationstabellen bliver dynamisk.)
Se formelsamling under statistik, for uddybning af begreber
- c. Eleverne laver selv observationstabel for "pigerne" (opgave 2-5)
- d. De elever som bliver hurtigt færdige, kan selv gå i gang med redegørelsen.


3. Dobbeltlektion

- a. Arbejde med observationstabeller for grupperede observationer
- b. Læreren gennemgår grupperede observationer for "drengene" og introducerer intervalmidtpunkt samt hjælp til gennemsnit
- c. Lav opgave 6-8
- d. Praktisk øvelse
 (Aktiv matematik - øvelse #23 <http://matematikbanken.dk/L/148/>)
- e. Opgave 9 (Obs: Vi bruger skostørrelser, da man kan aflæse kvartilsæt direkte ud fra $F(x)$, når vi bruger enkelobservationer.)

4. Dobbeltlektion

- a. Læreren gennemgår stolpediagrammer, cirkeldiagrammer, sumkurver, boksplot ud fra siderne i kompendiet. Eleverne laver opgave 10.
- b. Statistik ud fra rådata
- c. Lave boksplot ud fra praktisk boksplot (lektion 2) Sammenlign indbydes blandt de grupper eleverne var sammen med. (Boksplot skal være i samme diagram)
- d. Nå til og med opgave 12

5. Dobbeltlektion

- a. Lav opgave 13 og 14. ca. 35 min
- b. Spil "Det store statistikspil"
 (Aktiv matematik - øvelse #24 <http://matematikbanken.dk/L/148/>)
- c. Snak om citater. Hvad menes der?
- d. Gennemgå guide til at se manipuleret diagrammer. (Snak med elever om de 2 forskellige diagrammer i hvert afsnit. (Maks. 10 min)
- e. Lav opgave 15-17

6. Dobbeltlektion

- a. Opsamling på ovenstående
- b. Påbegynde redegørelse (Eleverne skal bruge fornuftig tid på egen undersøgelse)

INDHOLDSFORTEGNELSE

MATEMATISKE KOMPETENCER:.....	2
Modellering.....	2
STOFOMRÅDER:	2
Statistik & Sandsynlighed.....	2
Læringsmål:.....	2
Lektionsplan 12 lektioner fordelt på 6 gange 2 lektioner	3
Indholdsfortegnelse	4
Praktisk boksplot i hallen	5
Introduktionsopgave.....	6
Drenge:.....	6
Piger:	6
Gennemgang af opgave	7
”Den statistiske værktøjskasse” – Statistiske deskriptorer	8
Opgave: 1: Snak med sidemanden.....	8
Opgave: 2:	9
Opgave 3: Snak med sidemanden.....	10
Opgave 4: Snak med sidemanden.....	10
Opgave 5: Snak med sidemanden Aflæs/forklar for både drenge og piger	11
Grupperede og ikke-grupperede observationer	11
Grupperede observationer	11
Opgave 6: Snak med sidemanden.....	12
Opgave 7:	13
Opgave 8: Snak med sidemanden Find både for drenge og piger.....	13
Opgave 9: Snak med sidemanden.....	14
Diagrammer til ikke-grupperede observationer	15
Pindediagram	15
Diagrammer til grupperede observationer.....	17
Søjlediagrammer.....	17
Cirkeldiagram	17
Opgave 10	18
Statistik ud fra rå data i GeoGebra.	19
Opgave 11: Hvad kan du aflæse ud fra boksplottene?.....	21

Opgave 12 Boksplot ud fra jetonkast data.....	23
Aflæsning af boksplot.....	24
Opgave 13: Læsning af boksplot - Fartbump	24
Opgave 14: Er eleverne blevet mindre	24
2. aksens er brudt eller har en stor maksimumværdi.....	26
Sammenligning af 2 ting der ikke har noget med hinanden at gøre	27
Husk at sammenligningsgrundlag kan være forskellig.	28
Perspektiv for at det skal se pænt ud kan ødelægge diagrammet	28
Opgave 15: Lav et manipuleret diagram.....	30
Opgave 16: Vurdering af grafik	30
Opgave 17: Samme statistik:.....	31

PRAKTISK BOKSPLOT I HALLEN

a) Følg vejledningen du får af din lærer.

b) Når I er færdig, skal I indtast jeres data i et google docs. Link får du af din lærer.

a. INDSÆT LINK her:

INTRODUKTION TIL STATISTIK

Statistik er analyse af indsamlet data. Det vil sige, at man bearbejder et datamateriale, som i matematik næsten altid er tal. Derved får man et samlet overblik over talmaterialet. Rundt omkring i samfundet bliver statistik meget ofte brugt som baggrund for forskellige beslutninger. Derfor er statistik også en vigtig del af matematik i skolen.

INTRODUKTIONSOPGAVE

Eleverne på skolen skal på skitur, og i forhold til bestillingen, skal skisportsstedet vide, hvor mange skistøvler af forskellige størrelser, de skal have på lager.

Drenge:

36	38	39	39	39	39	40	40	40	40
41	41	41	41	41	41	41	42	42	42
42	42	42	42	42	42	42	42	42	42
42	42	42	42	42	43	43	43	43	43
43	43	43	43	43	43	43	44	44	44
44	44	44	44	44	44	44	44	44	44
44	44	44	45	45	45	46	46	47	

Piger:

35	35	35	36	36	36	36	36	36	36
37	37	37	37	37	37	37	37	37	37
37	37	37	37	37	37	37	37	37	37
38	38	38	38	39	39	39	39	39	39
39	39	39	39	39	39	39	39	39	39
39	40	40	40	40	40	41	41	41	41
41	41	41	41	41	41	41	41	41	41
41	41	41	41	41	42	42	42	42	42
42	43	43	44	45	43				

Ovenstående data bruges til at give forståelse for de kommende værktøjer der bliver gennemgået på de næste sider.

GENNEMGANG AF OPGAVE

Tal, som er grundlag for en statistik, kalder man for *observationer*. Observationerne i tabellen ovenfor er skostørrelsen for eleverne. Alle observationer udgør tilsammen et *observationssæt*.

Statistik er ligesom en værktøjskasse. Der findes rigtig mange måder at bearbejde observationerne på. Så hvis man skal arbejde rigtigt med statistik, skal man både vide hvordan og hvornår, man skal bruge det forskellige værktøj. Det er jo heller ikke nok at have en værktøjskasse med super godt værktøj, hvis man tror, man skal bruge skruetrækkeren til at banke søm i med...vel?☺

Herunder gennemgås det vigtigste værktøj, som du næsten altid vil få brug for i folkeskolematematik. Hvornår værktøjet skal bruges, må du lære gennem træning.

Denne "statistiske værktøjskasse" som du gerne skulle have, når du er færdig med denne gennemgang af statistik, er i øvrigt grundlag for meget af den statistisk, som findes i matematik på højere niveau.

"DEN STATISTISKE VÆRKTØJSKASSE" – STATISTISKE DESKRIPTORER

Hyppighed - $h(x)$

Hyppigheden angiver, hvor ofte (hyppigt) de forskellige observationer forekommer. Det er altså antallet af gange, en observation forekommer. Normalt angiver man hyppigheden med " $h(x)$ "

Her er en observationstabel for drengenes skostørrelser.

Observationer Drengene:	$h(x)$	$H(x)$	$f(x)$	$F(x)$	Til gennemsnit
36	1	1	1%	1%	36
37	0	1	0%	1%	0
38	1	2	1%	3%	38
39	4	6	6%	9%	156
40	4	10	6%	14%	160
41	7	17	10%	25%	287
42	18	35	26%	51%	756
43	12	47	17%	68%	516
44	16	63	23%	91%	704
45	3	66	4%	96%	135
46	2	68	3%	99%	92
47	1	69	1%	100%	47
48	0	69	0%	100%	0
I alt	69				2927
Gennemsnit					42,42029



Video der viser hvordan man laver ovenstående observationstabel

<http://matematikbanken.dk/L/158/>

Her er formler for ovenstående beregninger:

	C	D	E	F	G	H
1	Observationer	$h(x)$	$H(x)$	$f(x)$	$F(x)$	til gennemsnit
2	36	1	=D2	=D2/\$D\$15	=F2	=D2*C2
3	37	0	=D3+E2	=D3/\$D\$15	=F3+G2	=D3*C3
4	38	1	=D4+E3	=D4/\$D\$15	=F4+G3	=D4*C4
5	39	4	=D5+E4	=D5/\$D\$15	=F5+G4	=D5*C5
6	40	4	=D6+E5	=D6/\$D\$15	=F6+G5	=D6*C6
7	41	7	=D7+E6	=D7/\$D\$15	=F7+G6	=D7*C7
8	42	18	=D8+E7	=D8/\$D\$15	=F8+G7	=D8*C8
9	43	12	=D9+E8	=D9/\$D\$15	=F9+G8	=D9*C9
10	44	16	=D10+E9	=D10/\$D\$15	=F10+G9	=D10*C10
11	45	3	=D11+E10	=D11/\$D\$15	=F11+G10	=D11*C11
12	46	2	=D12+E11	=D12/\$D\$15	=F12+G11	=D12*C12
13	47	1	=D13+E12	=D13/\$D\$15	=F13+G12	=D13*C13
14	48	0	=D14+E13	=D14/\$D\$15	=F14+G13	=D14*C14
15	I alt	=SUM(D2:D14)				=SUM(H2:H14)
16	Gennemsnit					=H15/D15
17						

Opgave: 1: Snak med sidemanden

- Hvad er typetallet
- Hvad er frekvensen for skostørrelse 41
- Hvad er variationsbredden?
- Hvad er gennemsnitsskostørrelsen for drengene
- Hvad kan man evt. bruge den summerede frekvens til?

Summeret hyppighed - $H(x)$

- Den summerede hyppighed er hyppighederne lagt sammen med de foregående hyppigheder.
- Dvs. at den summerede hyppighed for 8 er hyppigheden for 6, 7 og 8 lagt sammen.
- Den summerede hyppighed skrives " $H(x)$ "

Frekvens - $f(x)$

- Den hyppighed observationen kommer med i forhold til det samlede antal observationer.
- Det vil sige hyppighed divideret med antallet af observationer. Dette vil give et resultat i form af en brøk eller decimaltal.
- Vil man have resultatet i procent, skal man gange med 100. Frekvens kan enten være i procent, brøk eller decimaltal. Det bestemmer du selv! Det vil sige, at 10%, $\frac{1}{10}$ eller 0,10 er det samme resultat på forskellige måde.

Summeret frekvens - $F(x)$

- Er ligesom ved summeret hyppighed, men her er det bare frekvenserne, som skal lægges sammen.

Typetallet

Typetallet er det tal, som er "typisk" for observationssættet. Det vil sige den observation, som forekommer flest gange i observationssættet.

Gennemsnittet

Gennemsnittet eller middeltallet er det tal, som man får, hvis man lægger alle observationer sammen og dividerer dette tal med antallet af observationer.

Observationer	$h(x)$	$H(x)$	$f(x)$	$F(x)$	Til gennemsnit
35	3				
36	7				
37	20				
38	4				
39	17				
40	5				
41	19				
42	6				
43	3				
44	1				
45	1				
46	0				
47	0				
I alt	86				
Gennemsnit					

Opgave: 2:

Lav observationstabel for pigerne

Opgave 3: Snak med sidemanden

- a) *Hvad er typetallet*
- b) *Hvad er frekvensen for skostørrelse 41*
- c) *Hvad er gennemsnitsskostørrelsen for pigerne*

Medianen

Den observation, som står i midten, hvis man stiller observationerne op i rækkefølge med de mindste tal først. Hvis der er et lige antal observationer, så der ikke er **et** tal i midten, tager du gennemsnittet af de 2 tal.¹

Opgave 4: Snak med sidemanden.

- a) *Hvordan finder man medianen ud fra en observationstabel*

Median, typetal eller gennemsnit

Den ene deskriptor er ikke "bedre" at bruge end de andre. Det kommer an på observationssættet og det man vil undersøge. F.eks. er en af fordelene ved medianen i forhold til gennemsnittet, at medianen er mindre påvirket af ekstreme observationer. Er der stor forskel på median og gennemsnit, kan der måske være fejl i observationerne (f.eks. målefejl, tastefejl eller kommafejl) eller der kan bare være en stor spredning. De tre deskriptorer afhænger af fordelingen af observationerne.

Størsteværdi

Den største observation i observationssættet.

- NB. Det er **ikke** det største **antal gange** en observation forekommer!

Mindsteværdi

Den mindste observation i observationssættet.

- NB. Det er **ikke** det mindste **antal gange** en observation forekommer!

Variationsbredden

Variationsbredden er forskellen på den største og den mindste observation i sættet.

- Variationsbredden finder man ved at trække største værdien og mindsteværdien fra hinanden.

¹ OBS: Medianen kan bestemmes på flere forskellige måder, men denne måde bruger GeoGebra

Opgave 5: Snak med sidemanden**Aflæs/forklar for både drenge og piger**

- Aflæs: Hvad er frekvens for skostørrelse 38?*
- Aflæs: Hvor stor en procentdel har skostørrelse 42?*
- Aflæs: Hvad er hyppigheden for skostørrelse 42?*
- Forklar for skuldermakker hvorfor man bruger frekvens frem for hyppighed.*
- Forklar hvorfor gennemsnittet og medianen ikke er ens.*
- Hvad fortæller medianen*
- Hvad fortæller gennemsnittet*

GRUPPEREDE OG IKKE-GRUPPEREDE OBSERVATIONER

I nogle tilfælde kan det være en fordel at dele observationerne ind i grupper. F.eks. hvis man skulle lave en statistik over en skoleklasse med 25 elever, som springer længdespring i en idrætstime. Højest sandsynlig vil man få 25 forskellige resultater med en hyppighed på 1. Det giver os ikke et så meget bedre overblik over tallene. Derfor vil man ofte se, at tallene bliver inddelt i grupper. F.eks. 0-1 meter, 1 til 2 meter osv. Disse grupper kalder man i statistik for *intervaller*.

Ovenfor er der lavet statistik på baggrund af ikke grupperede observationer, hvor observationerne altså ikke er inddelt i intervaller.

Nedenfor vil vi bruge de samme observationer som ovenfor, men nu vil vi lave intervaller, som vi samler observationerne i.

GRUPPEREDE OBSERVATIONER**Intervaller**

Hvis man har mange uens observationer, kan man inddele oplysningerne i grupper, som også kaldes *intervaller*.

Ved grupperede observationer vil man normalt ikke kunne finde hverken typetal, største værdi, mindsteværdi og variationsbredde, fordi man ofte ikke kender de enkelte observationer, men kun har observationerne samlet i en observationstabel. I nogle sammenhænge kan man dog snakke om et *typeinterval*, som er det interval, hvor der er flest observationer. Man kan også finde et gennemsnit, median og kvartilerne, men man gør det normalt på en lidt anden måde ved grupperede observationer.

Ofte ser man, at der er "firkantede parenteser" omkring intervallerne "[" og "] " Disse parenteser angiver, om tallet er med eller ej. Hvis parenteser vender ind mod tallet, er tallet med. Vender parenteser væk fra tallet, betyder det, at tallet ikke er med, men tallene op til tallet er med.

- Eks. I intervallet $[2;4[$ er tallet 2 med og så er tallene op til 4 også med, men tallet 4 er ikke med. Det vil sige 3,99999999999999999999 osv. er med. Så man kan sige fra og med 2 til og ikke med 4.

Gennemsnit i forhold til intervalmidtpunkt

Hvis man skal finde gennemsnittet af observationer, som er inddelt i intervaller, hvor man ikke kan finde tilbage til de oprindelige observationer, skal man i første omgang finde intervalmidtpunktet. Det vil sige, man finder den midterste værdi i intervallet. Eks. hvis intervallet går fra 0 til 10, så er midtpunktet 5. Man finder intervalmidtpunktet, fordi man ikke ved hvordan observationerne fordeler sig i intervallet. Derfor går man ud fra, at observationerne fordeler sig jævnt omkring midten af intervallet.

Hvis man havde kendt observationerne, ville man lægge dem sammen og så til sidst dividere med det samlede antal. Faktisk gør man lidt det samme, når man har observationerne i intervaller. Dog er det lettere at gange intervalmidtpunkterne.

- Eks. hvis intervalmidtpunktet er 5 og hyppigheden af intervallet er 3, så svarer det til, at man har observationerne 5, 5 og 5. Derfor er det lettere at sige 5 gange 3 end $5+5+5$.
- De tal, som man får ud for de enkelte intervaller, lægger man sammen og dividerer med antallet af observationer (ikke antallet af intervaller).

Opgave 6: Snak med sidemanden

a) Kan gennemsnittet ligge uden for variationsbredden? Begrund hvorfor / hvorfor ikke

Når eleverne skal på ski, skal skisportsstedet også bruge deres højder i forhold til valg af længde af ski. Eleverne skal angive om de er mellem $[160;165[$, $[165;170[$ osv. Det præcist nok i forhold til længde af ski hvorimod skistøvlerne skal passe helt præcist.

Observationstabel for højde for drenge

Observationer	$h(x)$	$H(x)$	$f(x)$	$F(x)$	Interval midtpunkt	Til gennemsnit
[155-160[1	1	1%	1%	157,5	157,50
[160-165[1	2	1%	3%	162,5	162,50
[165-170[2	4	3%	6%	167,5	335,00
[170-175[10	14	14%	20%	172,5	1.725,00
[175-180[22	36	32%	52%	177,5	3.905,00
[180-185[16	52	23%	75%	182,5	2.920,00
[185-190[12	64	17%	93%	187,5	2.250,00
[190-195[5	69	7%	100%	192,5	962,50
I alt	69				Højde i alt	12.417,50
					Gennemsnit	179,96



Video der viser hvordan man laver ovenstående observationstabel

<http://matematikbanken.dk/L/159/>

Observationstabel for højde for piger

Observationer	$h(x)$	$H(x)$	$f(x)$	$F(x)$	Interval midtpunkt	Til gennemsnit
[155-160[8					
[160-165[17					
[165-170[19					
[170-175[33					
[175-180[4					
[180-185[3					
[185-190[2					
[190-195[0					
I alt	86				Højde i alt	
					Gennemsnit	168,95

Opgave 7:

a) Færdiggør observationstabellen for pigernes højde

Opgave 8: Snak med sidemanden

Find både for drenge og piger

- Aflæs: Frekvensen for $[175;180[$
- Aflæs: Intervalmidtpunktet $[175;180[$
- Forklar: Hvad er intervalmidtpunktet og hvad bruges det til?
- Aflæs: Hvad er gennemsnitshøjden?
- Aflæs: Største- og mindsteværdi i forhold til observations Tabellen.
- Hvad er fordelene ved grupperede observationer
- Hvad er ulemperne ved grupperede observationer
- Er det rigtigt, at gennemsnittet altid vil ligge imellem største og mindsteværdi? Hvorfor/hvorfor ikke?

Kvartiler

1. kvartil, 2. kvartil og 3. kvartil er de observationer, som forekommer efter henholdsvis 25%, 50% og 75% af observationerne, når observationerne er stillet i rækkefølge med det mindste først.

Bemærk:

- At 1. kvartilen også kaldes 0,25-kvartilen eller nedre kvartil.
- At 2. kvartilen også kaldes medianen eller 0,50-kvartil.
- At 3. kvartilen også kaldes øvre kvartil eller 0,75-kvartilen
- Kvartiler kan umiddelbart findes i enkeltobservationer ud fra den summeret frekvens
- Kvartiler kan også findes ved grupperet observationer, men det kræver en sumkurve først

Opgave 9: Snak med sidemanden

Tag hensyn til både piger og drenge

- a) Hvad er mindsteværdien for skostørrelse?*
- b) Hvad er 1. kvartilen for skostørrelse?*
- c) Hvad er median (2. kvartilen) for skostørrelse?*
- d) Hvad er 3. kvartil for skostørrelse?*
- e) Hvad er størsteværdien for skostørrelse?*

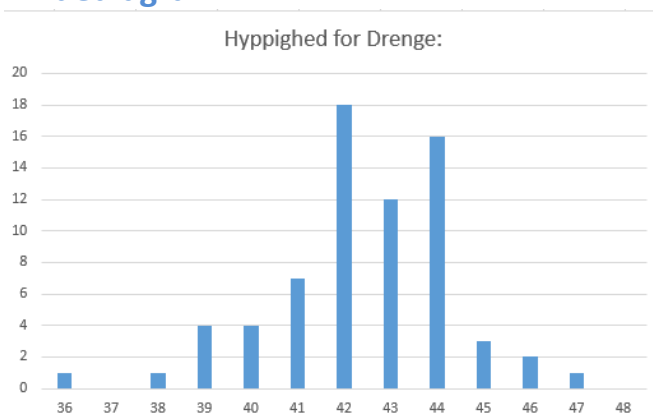
DIAGRAMMER

Det er ikke alle diagramtyper, som bruges ved både grupperede og ikke-grupperede observationer. Nedenfor kan du se, hvornår de forskellige diagramtyper bruges.

DIAGRAMMER TIL IKKE-GRUPPEREDE OBSERVATIONER

Hvis det er observationer, som ikke er inddelt i intervaller, vil man normalt bruge følgende diagrammer:

Pindediagram



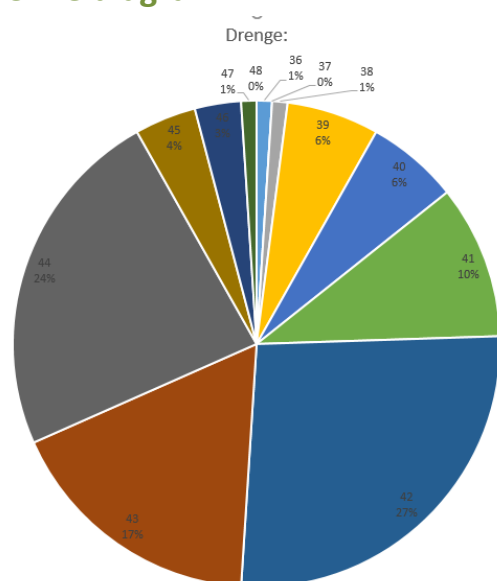
Video der viser hvordan man laver et søjlediagram/pindediagram

<https://matematikbanken.dk/L/174/>

Ofte bliver der brugt mange navne om et diagram, som dette. Nogen kalder det stolpediagram, andre søjlediagram. Der findes ikke en entydig definition på, hvad der er pinde, stolpe- og søjlediagrammer. Man skal dog lægge mærke til, at under hver "pind" er der kun **et** tal.

- Til pindediagrammet bruger man kolonnen $h(x)$ som serieværdi og skostørrelse som kategori

Cirkeldiagram



- Til cirkeldiagrammet bruges søjlen $f(x)$ som serienavn, og skostørrelsen som kategori.

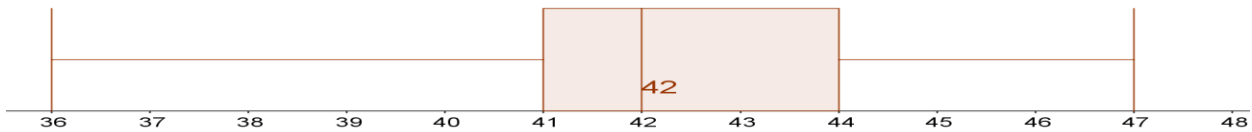


Video der viser hvordan man laver et cirkeldiagram

<https://matematikbanken.dk/L/175/>

Boksplot

Viser drengenes skostørrelse fordeling



Et boksplot viser, mindsteværdi, 1. kvartil, median, 3. kvartil, største værdi og variationsbredden.

Ud fra boksplottet kan man se:

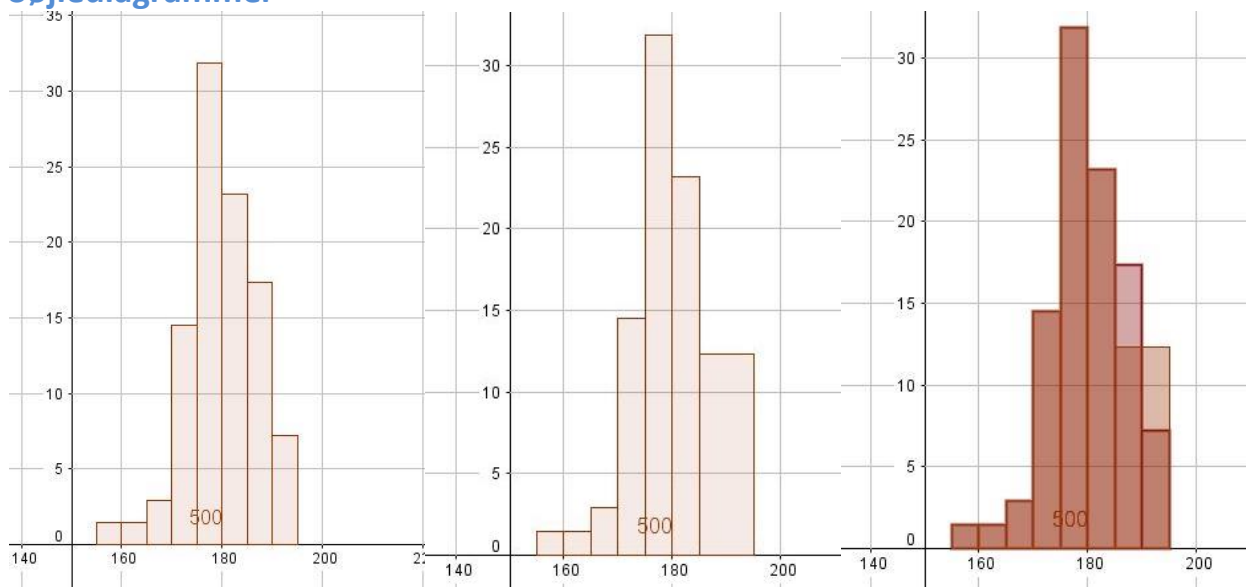
- at de midterste 50 % har en skostørrelse mellem 41 og 44 (begge inkl.)
- at de første 25 % har en skostørrelse mellem 36 og 41 (begge inkl.)
- at de sidste 25 % har en skostørrelse mellem 44 og 47 (begge inkl.)
- at 25 % har en skostørrelse på max 41
- at 75 % har en skostørrelse på mindst 41
- at 75 % har en skostørrelse på max 44
- at de sidste 25 % har en skostørrelse på 44 eller derover.
- at forskellen i skostørrelsen i blandt de første 25% af observationerne er større end sidste 25%

Bemærk: Hvis man har grupperet observationer, så er man nødt til at lave en sumkurve for at finde kvartilsættet, før man kan lave et boksplot

DIAGRAMMER TIL GRUPPEREDE OBSERVATIONER

I forbindelse med oplysninger, som er sat i intervaller, vil man normalt bruge følgende diagrammer.

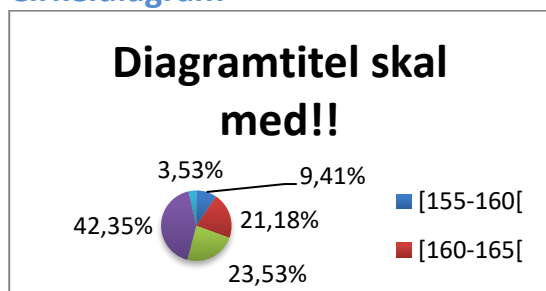
Søjlediagrammer



Ligesom beskrevet i forbindelse med pindediagrammet, er der ikke nogen fast regel for, hvad der er søjle- og stolpediagrammer. Dog kalder man det kun enten søjle- eller stolpediagram og ikke pindediagram. Det skyldes at det er vigtigt at søjlerne hænger sammen og ikke står som pinde med luft i mellem.

Bemærk at søjlen går **mellem to tal**. F.eks. fra 1 til 2. Det vil sige at i intervallet fra 1 til 2 er der to observationer, hvis man aflæser søjlediagrammet ovenfor.

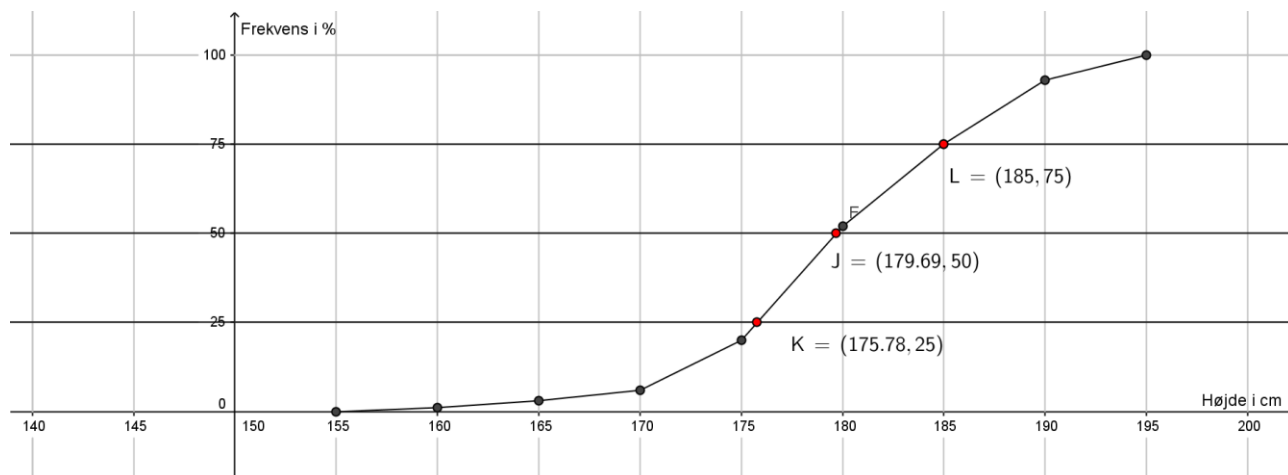
Cirkeldiagram



Husk diagramtitel og procentangivelse

Sumkurver

Hvis man vil lave en sumkurve er det normalt lettest at bruge den summerede frekvens som udgangspunkt.



Læg mærke til at kvartilerne er indtegnet. Ved de ikke-grupperede observationer kunne vi finde medianer og kvartiler ved at kigge på observationssættet eller observationstabellen. Det er ikke så let ved de grupperede observationer. Her er man nødt til at aflæse på grafen. På grafen ovenfor er 50% kvartilen som også er medianen ca. 180



Video der viser hvordan man laver en sumkurve

<https://matematikbanken.dk/L/176/>

Opgave 10

- Tegn stolpediagram for hyppigheden af pigernes højde.
- Tegn cirkeldiagram for fordelingen af frekvensen på de forskellige højder. (Et for drenge og et for piger)
- Tegn en sumkurve for både drengenes og pigernes højde.
 - Aflæs kvartiler
- Tegn i samme diagram boksplot af drenge og pigernes højde
- Hvad kan man sige om drengenes og pigernes højde i forhold til deres kostørrelser?

STATISTIK UD FRA RÅ DATA I GEOGEBRA.

Du har en række af data.

En skoleklasse vil undersøge hvor lang tid eleverne sidder foran en elektronisk skærm, på en normal hverdag. Fra de står op til de går i seng.

Oversigt over 9.a timer, der ikke er undervisningsrelevant, foran en skærm. (PC, Tablet, Fjernsyn, smartphone m.v.)

Rå data

1	2	0	4	6	2	5	3	2	1
5	5	3	2	3	4	7	9	3	7

9. b laver samme undersøgelse, men har sorteret deres data i en observationstabel

Obs	h(x)
0	0
1	1
2	3
3	4
4	6
5	5
6	3
7	0
8	0
9	1

For at man nemmest kan bruge data i GeoGebra, så skal man lave observationstabellen til rådata. Så ser det således ud.

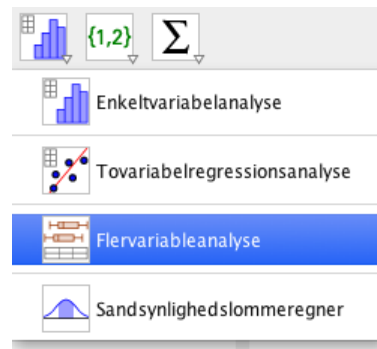
9. b. rå data

1	2	2	2	3	3	3	3	4	4
4	4	4	4	5	5	5	5	5	6
6	6	9							

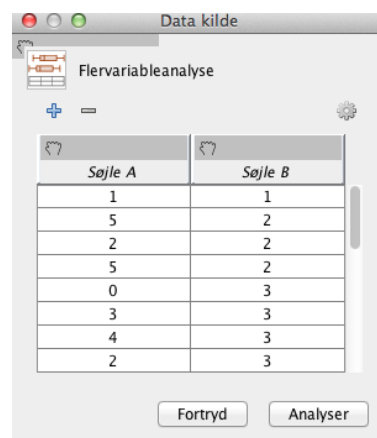
1. Start GeoGebra → Tryk "Vis" → Tryk "Regneark"
2. Kopiere 9. A data ind i kolonne A, og 9. B's data i kolonne B
 - a. Så ser det sådan her ud

	A	B
1	1	1
2	5	2
3	2	2
4	5	2
5	0	3
6	3	3
7	4	3
8	2	3
9	6	4
10	3	4
11	2	4
12	4	4
13	5	4
14	7	4
15	3	5
16	9	5
17	2	5
18	3	5
19	1	5
20	7	6
21		6
22		6
23		9
24		

Marker nu kolonne A og kolonne B




Tryk på flervariabelanalyse

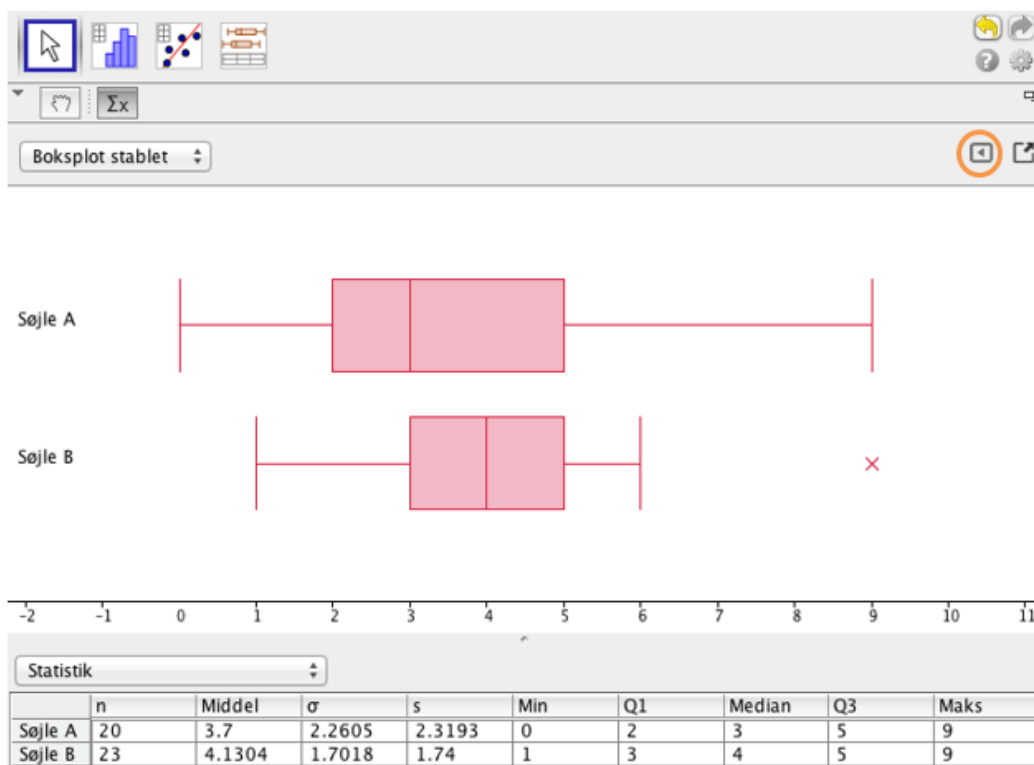


Tryk analyser



Tryk på  for at få udvidet oplysninger.

Nu skulle det gerne se således ud

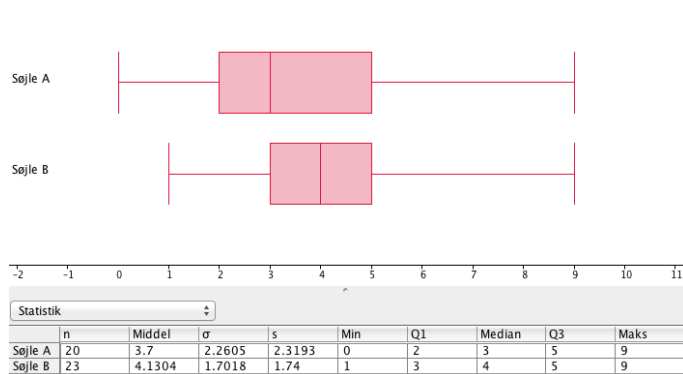


Læg mærke til krydset - det er en outliers (Undtagelse) - den kan du slå fra ved at trykke på.



fjern nu fluebenet i vis outliers

☐ Vis Outliers



Video der viser hvordan man laver et boksplot ud fra rådata
<https://matematikbanken.dk/L/178/>

Opgave 11: Hvad kan du aflæse ud fra boksplottene?

a) Snak med sidemanden, hvad kan du se ud fra boksplottet og statistikvinduet?

Jeg kan nu aflæse for 9.a

- 20 elever har et middeltal på 3,7 timer og en variationsbredde på $9-0 = 9$ timer
- 25 % af eleverne er 2 timer eller mindre foran en skærm
- 75% af eleverne er 2 timer eller mere foran en skærm
- Den midterste observation er 3, hvilket betyder, at 50% af eleverne bruger 3 timer eller mindre foran en skærm.
- 50% af eleverne bruger mindst 3 timer foran en skærm.
- 25% bruger mere end 5 timer foran en skærm
- De midterste 50% bruger mellem 2 og 5 timer foran en skærm

Jeg kan aflæse for 9.b.

- 23 elever, har et middeltal på 4,1 timer og en variationsbredde på $9-1 = 8$ timer
- 25% af eleverne er 3 timer eller mindre foran en skærm
- 75% af eleverne er 3 timer eller mere foran en skærm
- Medianen er 4 timer
- 50% af eleverne bruger 4 timere eller mindre foran en skærm
- 50% af eleverne bruger 4 timer eller mere foran en skærm
- 25% bruger mere end 5 timer foran en skærm.
- De midterste 50% bruger mellem 3 og 5 timer for an skærm.

For begge klasser kan man evt. sige.

- I 9. a ligger gennemsnittet længere væk fra medianen end gennemsnittet ligger fra medianen i 9.b. (Det kan tolkes således, at 9.a har mere usikre data end 9.b, da spredningen i data er større i 9.a end 9.b)
- 9. B har en mindre variationsbredde end 9.a
- 9. B har en mindre spredning i "normalområdet" (De midterste 50%)
- 9.B bruger mere tid foran en skærm end 9.a - da "normalområdet" er mere til højre end 9.A (Det fortæller både median og gennemsnit også)
- 9.A bruger mindre tid foran en skærm end 9.b. det kan man se på 2 medianen på 9.a. ligger samme sted som 9.b's første kvartil.
- 50% af eleverne i 9.a bruger 3 timer eller mindre, hvor 9.B er der det kun 25% der brugere 3 timere eller mindre.

Konklusionen må være at 9.b brugere mere tid foran en skærm end 9.a.

OBS:

Tag et skærmbillede af boksplot med udvidet oplysninger, og sæt det ind der, hvor du skal bruge det.

Opgave 12 Boksplot ud fra jetonkast data

Du skal nu lave boksplot ud fra jetonkast - linket får du af din lærer.



Video der viser hvordan man laver et boksplot udfra kvartilsæt

<https://matematikbanken.dk/L/177/>

a) Lav boksplot ud fra værdierne fra de 3 grupper.

Skriv i GeoGebra: (udskift, mindsteværdi osv. med de rigtige værdier)

Gruppe 1

Boksplot[1,1, Mindsteværd, 1. kvartil, Median, 3. kvartil, Største værdi]

Gruppe 2 (Obs: læg mærke til de ikke er helt ens)

Boksplot[4,1, Mindsteværdi, 1. kvartil, Median, 3. kvartil, Største værdi]

Gruppe 3 (Obs: læg mærke til de ikke er helt ens)

Boksplot[7,1, Mindsteværdi, 1. kvartil, Median, 3. kvartil, Største værdi]

b) Sammenlign de 3 grupper.

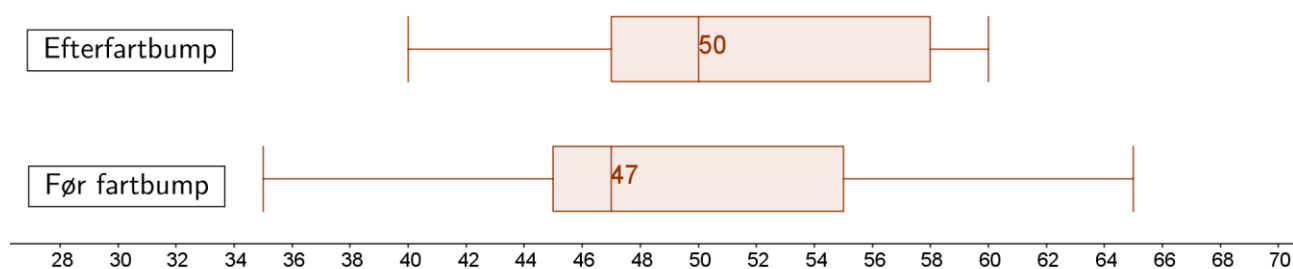
c) Hvilken gruppe skal efter din mening med til OL i jetonkast?

d) Ville gennemsnittet være en bedre model for udtagelsen - eller hvad tænker du?

AFLÆSNING AF BOKSPLOT

Opgave 13: Læsning af boksplot - Fartbump

På Lyngvej er der lavet fartbump for at sænke farten, men virker de?



- Den ene sidemakker skal argumentere for at farten er steget, den anden sidemakker skal argumentere for at farten er faldet efter der er lavet fartbump. Hvem kan overbevise den anden bedst?
- Hvad kan man eller sige om boksplottet ovenfor? Du skal komme med mindst 6 forskellige udsagn.

Opgave 14: Er eleverne blevet mindre

Bjergsnæs Efterskole er en skole med meget springgymnastik. I år 2006 fik skolen et nyt springcenter. En af skolens lærere vil undersøge om eleverne er blevet mindre siden 2006, fordi skolen nu tiltrækker flere øvede springgymnaster.

Data til boksplot kan hentes på: <http://goo.gl/hXvCBR>

- Undersøg: Er eleverne (drenge) blevet mindre siden skolen fik et springcenter i 2006?
- Sammenlign evt. med dette års drengehøjde, som du har arbejdet med i dette kompendium

CITATER

"Statistik er ligesom en bikini - den viser meget, men skjuler det væsentlige..." (Uffe Ellemand Jensen)

"Der findes tre typer af løgne: Løgn, forbandet løgn og statistik." (Mark Twain)

"At sige det med statistik, er at sige, at du aldrig er sikker." (Erich Segal)

"Der er ingen der er mere skeptisk overfor statistik, end statistikere."

"Statistik er kunsten at lyve ved hjælp af tal." (William Stegel)

"Det er bevist, at det er sundt at fejre fødselsdag! Statistikken viser, at de mennesker, der fejrer flest fødselsdage, bliver ældst." (Saul Below)

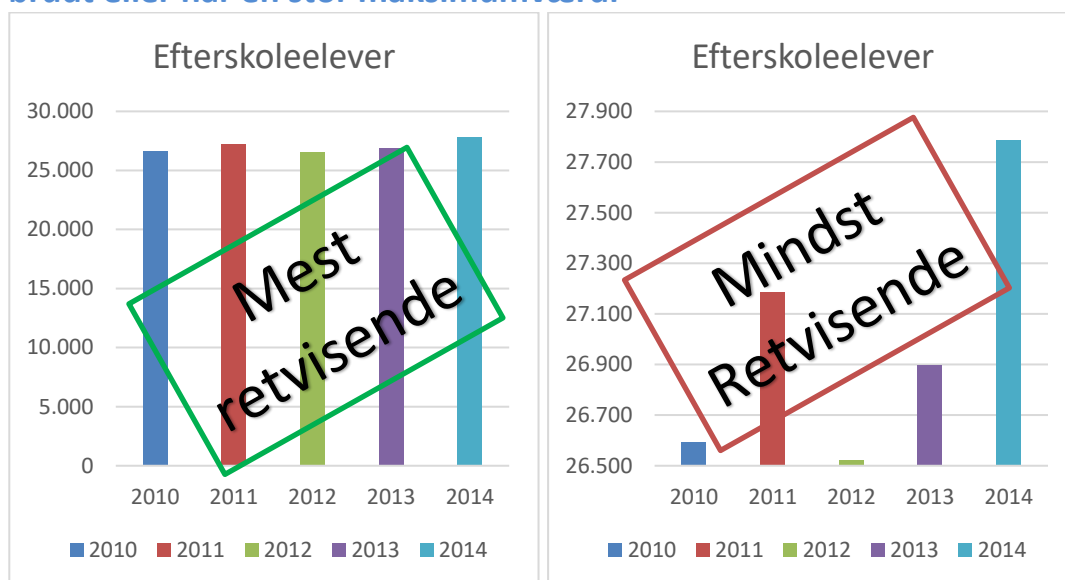
"Jeg kan bevise alt ved hjælp af statistik, bortset fra sandheden." (George Canning)

"10 pct. af alle biltyve er kejthåndede. Alle isbjørne er kejthåndede. Så hvis din bil bliver stjålet, er der 10 pct. chance for, at den er stjålet af en isbjørn." (John Allen Paulos)

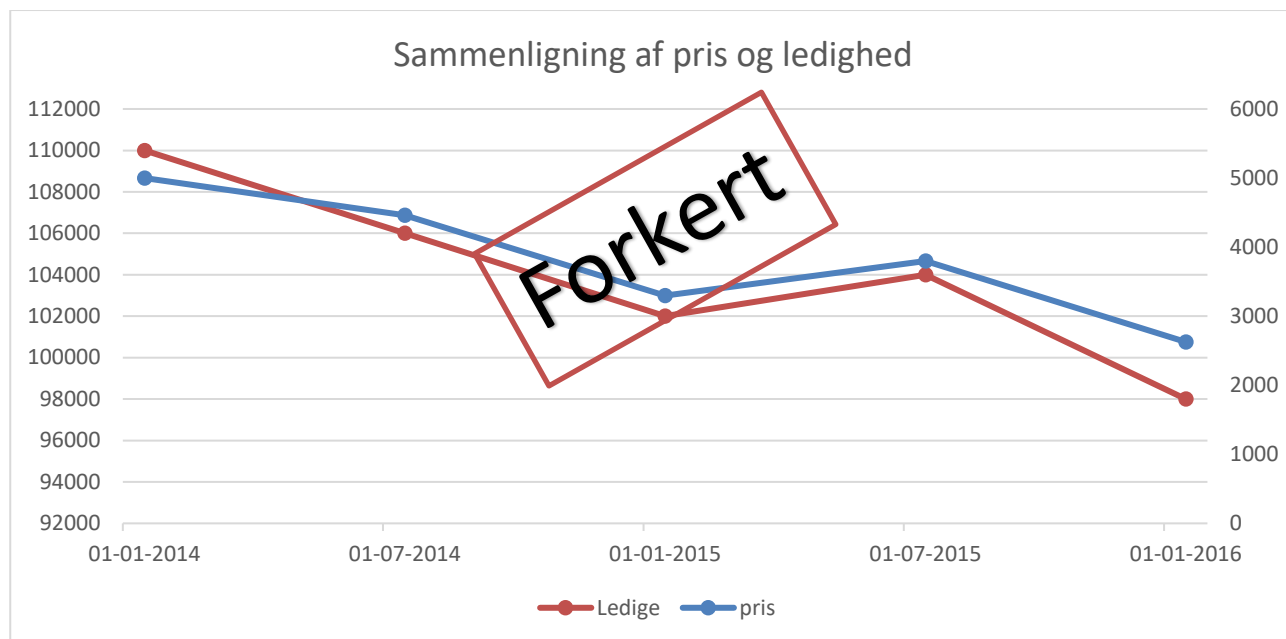
"Fordi noget kan tælles, behøver det nødvendigvis ikke at blive talt" (Harold L. Enarson)

GUIDE TIL AT SE MANIPULERET DIAGRAMMER

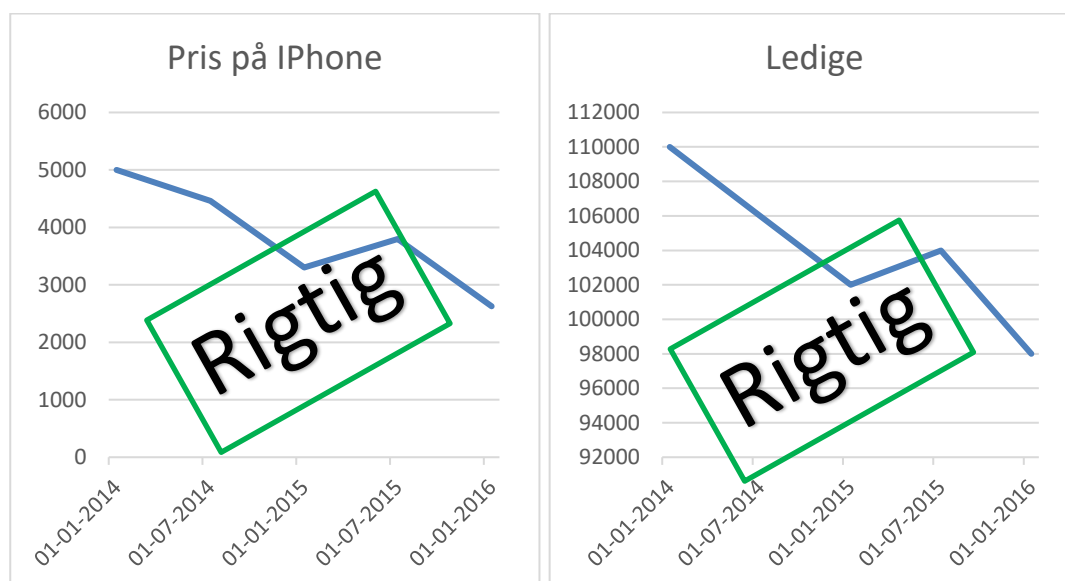
2. aksens er brudt eller har en stor maksimumværdi



Sammenligning af 2 ting der ikke har noget med hinanden at gøre



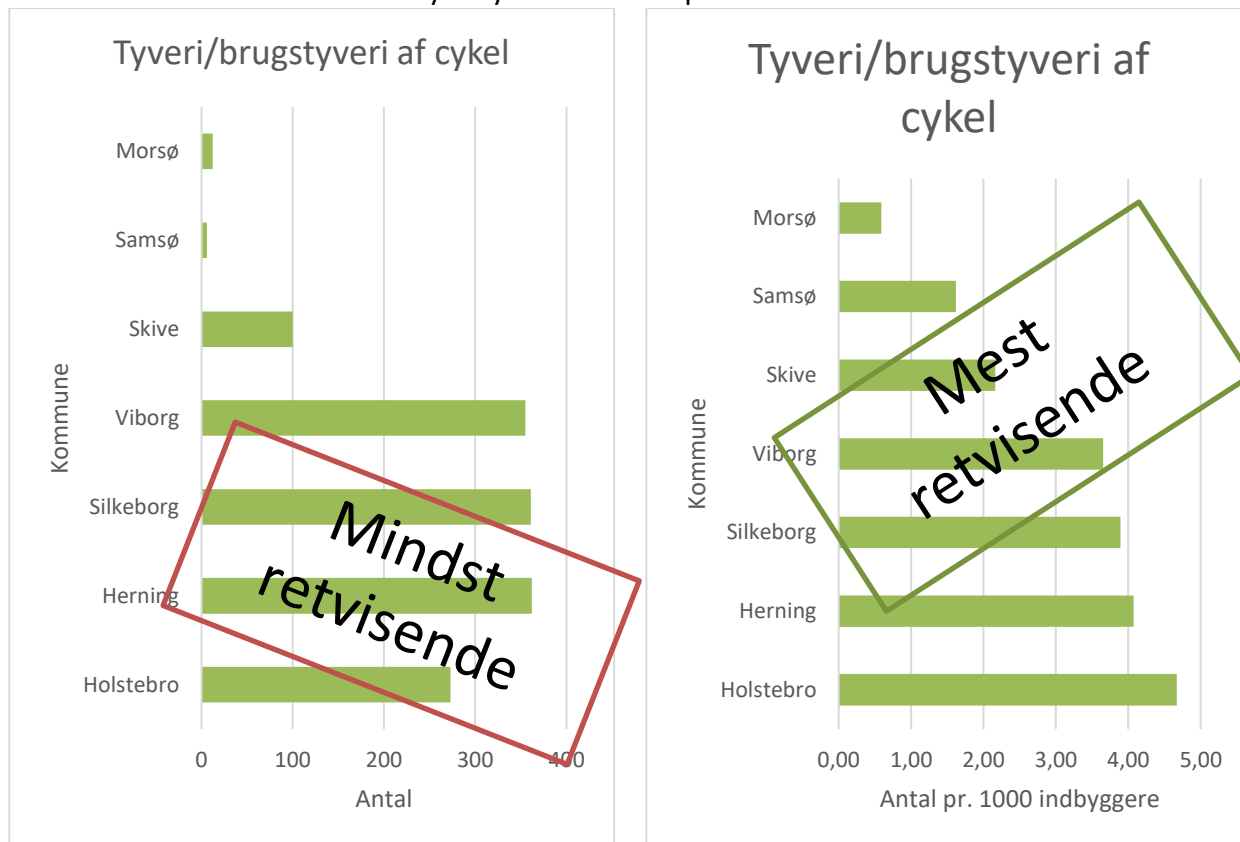
Her sammenlignes pris på iPhone 5s og antallet af ledige²



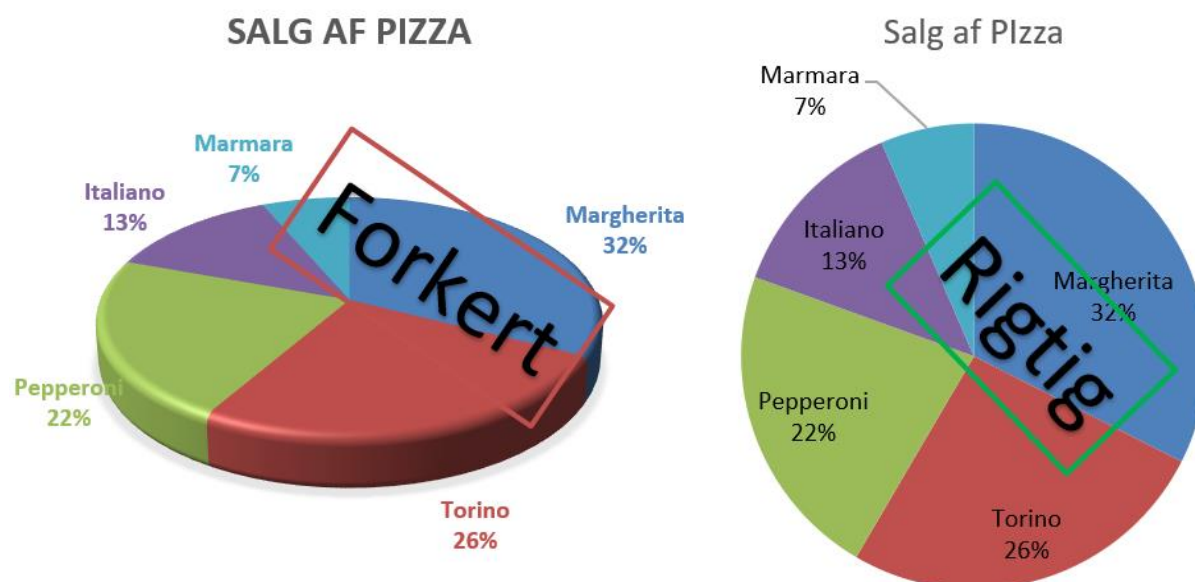
² Ledighedstallene er let manipuleret for at fremme forståelsen

Husk at sammenligningsgrundlag kan være forskellig.

Nedenstående viser antallet af cykeltyverier fordelt på kommuner.



Perspektiv for at det skal se pænt ud kan ødelægge diagrammet



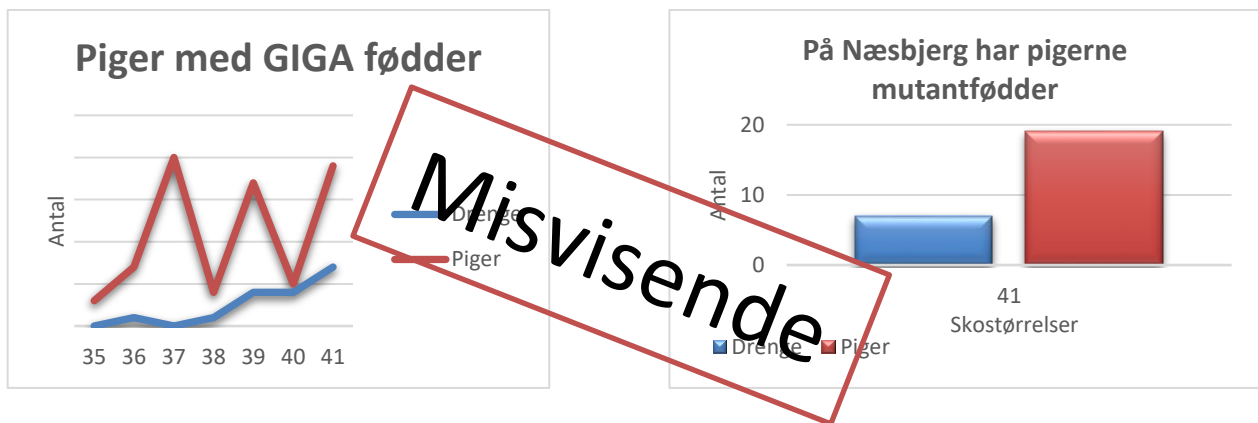
Torino ser ud til at være større end Margherita pga. perspektivet.

VURDERING AF GRAFISK STATISTIK

Eksempel

I vores tidligere opgave med skostørrelser kunne man fx se, at når vi sammenligner drengestørrelser med pigestørrelser, så får vi et lidt overraskende billede.

Disse diagrammer var for et par år siden lige ved at havne i Viborg Stifts Folkeblad, men blev stoppet inden avisen røg i trykken.



Et eksempel fra et dansk nyhedside, hvor de kun kigger på en lille del.



Opgave 15: Lav et manipuleret diagram

År	Besøgstal
2007	12000
2008	12500
2009	12750
2010	12745
2011	12700
2012	12750
2013	13000
2014	13100
2015	12400

Siden 2007 er besøgstallet til Palles Gavebod steget og steget grundet god kampagne fra deres reklamebureau. Men i 2015 går der noget galt.

Sidemakker A: Du er ejer af reklamebureauet og vil gerne sørge for, at blive ved med at lave reklame for "Palles Gavebod" - Lav et diagram og en tekst til diagrammet, der kan overtale Palle til at beholde jer til fortsat at lave reklame for Palles Gavebod.

Sidemakker B: Du er ejer af et andet reklamebureau, og du vil gerne overtage Palles Gavebod som kunde. Lav et diagram og en tekst, der gør at du får kontrakten for den fortsatte reklame.

- Slå sten, saks papir om hvem der er sidemakker A og hvem der er sidemakker B
- Lav et diagram og tekst i forhold til om du er A eller B
- Vis det til din lærer.

Opgave 16: Vurdering af grafik

- Find fejlen
- Begrund hvordan fejlen kan være fremkommet?
- Lav et pindediagram der er rigtig



Kilde: <http://politik.tv2.dk/2015-05-07-her-gaar-det-helt-galt-for-news-jeg-koeber-nye-briller>

Opgave 17: Samme statistik:

- d) *Begrund hvorfor de 2 grafer er ens*
e) *Hvilket diagram tror du, Socialdemokraterne har brugt i deres annonce, den gang de sad i regering og havde ansvaret for ventelisterne?*



Kilde: <http://www.karenmelchior.eu/ventetid-pa-operationer-manipulation-med-grafer/>