

Deskriptiv statistik

Medier og statistik

I medierne præsenteres vi ofte for tal og diagrammer, der skal fortælle om verdens tilstand, effekten af en politik osv. Men kan vi tro på oplysningerne eller bliver vi manipuleret? Vi vil i det følgende se på nogle eksempler, der viser, at det er vigtigt, at man forholder sig kritisk til den måde, hvorpå tallene mv. præsenteres.

Eksempel 1: Ventetider i det danske sundhedsvæsen¹

Figur 1 herunder viser, hvordan udviklingen i ventetiden på at blive opereret var i perioden 2011-14, hvor Danmark havde en socialdemokratisk ledet regering.

- Hvad er der sket med ventetiden i den omtalte periode?
- Hvor mange procent er ventetiden faldet fra 2011 til 2014?
- Hvordan svarer resultatet i b) til dit umiddelbare indtryk, når du ser figuren?



Figur 1.

Figur 2 er en anden måde at præsentere de samme data på som er illustreret på figur 1.

- Hvori består forskellen mellem de to diagrammer?
- Diskutér forskellen på de to diagrammer ud fra et matematik-fagligt og moralsk perspektiv.

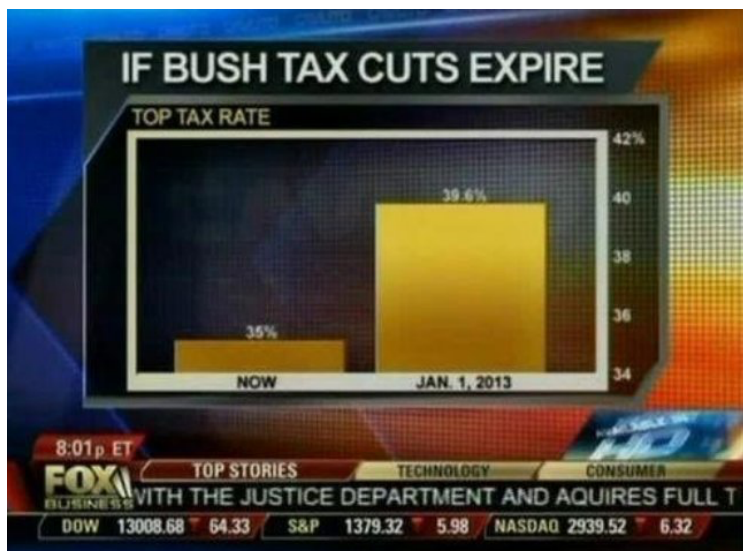


Figur 2.

¹ <http://www.karenmelchior.eu/ventetid-pa-operationer-manipulation-med-grafer/>

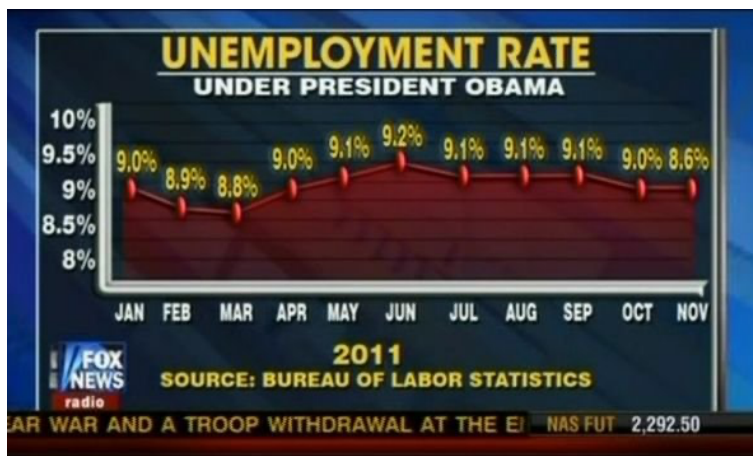
Eksempel 2: Fox News og amerikansk politik²

Fox News er en amerikansk nyhedsstation, og vi vil i det følgende se nærmere på nogle figurer, som Fox News har brugt til at informere den amerikanske befolkning.



Figur 3. Udviklingen i skattetrykket i USA ved fjernelse af de skattesænkninger, som den republikanske præsident Bush havde indført.

- Hvad sker der med skatteprocenten, hvis Bush's skattesænkninger fjernes?
- Hvordan vurderer du den grafiske fremstilling i forhold til den forventede ændring i skatteprocenten?



Figur 4. Udviklingen i arbejdsløshed under den demokratiske præsident Obama.

- Hvordan har arbejdsløsheden i USA udviklet sig de første 11 måneder af 2011?
- Kig nøje på tallene, der viser arbejdsløshedsprocenten de enkelte måneder. Er grafen lavet korrekt?
- Diskutér ud fra figur 3 og 4 om Fox News helst vil have en republikansk eller en demokratisk præsident i USA.

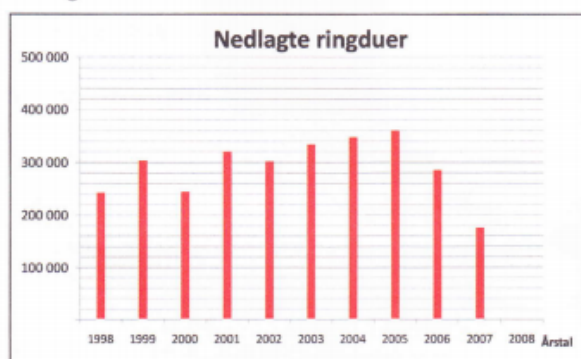
² <http://www.businessinsider.com/fox-news-charts-tricks-data-2012-11?IR=T>

Gensyn med statistik fra grundskolen

Løs følgende opgaver fra en af folkeskolens afgangsprøver:

Statistik og sandsynlighed

Diagrammet viser, hvor mange ringduer der blev nedlagt i Danmark i årene 1998 til 2007.



Kilde: Danmarks miljøundersøgelser

39. I år 2001 blev der nedlagt _____ ringduer.

40. Hvilket årstal blev der nedlagt flest ringduer? _____

I år 2008 blev der nedlagt 174 800 ringduer.

41. Tegn en pind i diagrammet, der viser antallet af nedlagte ringduer i 2008.

Statistik og sandsynlighed

Clara har kastet en tegnestift 1000 gange for at undersøge, hvordan den lander. Resultaterne ses i tabellen herunder.

Spidsen opad	Spidsen nedad
	
800	200

12. Ifølge undersøgelsen er sandsynligheden for, at tegnestiften lander med spidsen opad _____

13. Ifølge undersøgelsen er det i 500 kast mest sandsynligt, at tegnestiften lander med spidsen nedad (sæt et kryds)

50 gange 200 gange

100 gange 250 gange

150 gange

14. Ifølge undersøgelsen er sandsynligheden for, at tegnestiften lander med spidsen nedad to gange i træk _____

Statistik og sandsynlighed

Herunder er et observationssæt, der består af ti karakterer:

7, 4, 7, 7, 10, 4, 7, 7, 7, 10

23. Observationssættets middeltal er _____

24. Observationssættets variationsbredde er _____

Grundlæggende begreber i deskriptiv statistik

Når man skal præsentere et stort talmateriale bruges *deskriptiv statistik*. Ordet deskriptiv kommer af det latinske *descriptivus*, der betyder beskrivende, og det er samme ord som det engelske 'to describe'.

Når man i statistikken skal beskrive datamateriale skelner man mellem to kategorier:

- 1) Ikke-grupperede observationer
- 2) Grupperede observationer

Ved *ikke-grupperede observationer* er observationerne af en type, hvor der på forhånd er en naturlig opdeling i nogle velafgrænsede størrelser. Hvis man fx interesserer sig for dansk-karakterer, så er der kun syv mulige udfald (nemlig karaktererne fra -3 til 12). Hvis man spørger til skostørrelse, er der også kun forholdsvis få mulige svar; nemlig størrelserne 36, 37, 38 osv. Hvis man fx i en gymnasieklasse undersøger fordelingen af dansk-karakterer eller skostørrelser, er der altså tale om en ikke-grupperet observation.

Hvis man derimod spørger til elevernes vægt, så vil der være mange mulige udfald. Én elev vejer 62,1 kg, en anden vejer 67,4 kg, en tredje vejer 57,6 kg osv. Hvis man spørger alle elever i en klasse, så er der faktisk en god mulighed for, at alle vejer noget forskelligt. Dermed kan det være svært at danne sig et overblik over datamaterialet. Derfor kan man med fordel undersøge, hvor mange elever der vejer 50-55 kg, 55-60 kg, 60-65 kg osv. Det betyder, at man grupperer de mulige udfald i nogle intervaller, og man kalder det en *grupperet observation*.

Opgave

Forestil dig, at man stiller alle elever på Silkeborg Gymnasium disse fire spørgsmål:

- a) Hvor høj er du?
- b) Hvor mange søskende har du?
- c) Hvor meget tjener din far på et år?
- d) Hvor mange gange har du været i Italien?

Overvej i hvert af de fire tilfælde, om det indsamlede talmateriale bedst behandles som en grupperet eller en ikke-grupperet observation.

Ikke-grupperede observationer

Vi vil repetere (eller måske introducere) en række grundlæggende begreber i den deskriptive statistik ved at se på et eksempel fra gymnasiet. Skemaet herunder viser karakterfordelingen i 2015 ved studentereksamen i AT for de 405 3.g-elever på Silkeborg Gymnasium. Vi kalder et sådan skema et *observationssæt*.

Karakter	-3	00	02	4	7	10	12
Antal elever	0	2	23	78	110	103	89

Karaktererne i øverste række kaldes *observationer*, og de beskriver altså de mulige svar, når man spørger en elev: Hvilken karakter fik du til studentereksamen i AT?

I den nederste række står antallet af elever, der har fået de forskellige karakterer. Vi kalder disse tal for de observerede *hyppigheder*.

Besvar nu nedenstående spørgsmål. I kan muligvis besvares med jeres viden om statistik fra grundskolen og ellers SKAL i læse på denne side i -bogen: <https://plusbstx.systeme.dk/?id=1412>

Når I arbejder med spørgsmålene, så vær opmærksom på at bruge i-bogen, når det er nødvendigt at læse om de statistiske begreber for at kunne besvare spørgsmålene. Vær også opmærksomme på om I kan omsætte de statistiske begreber til en beskrivelse og forståelse af 'den virkelige verden'.

- Hvad er hyppighederne for karaktererne 02 og 10?
- Hvad er *observationssættets størrelse*?
Hvordan finder man generelt observationssættets størrelse?
- Hvad er *variationsbredden* for observationssættet?
Hvad fortæller variationsbredden?
Hvordan finder man generelt variationsbredden for et observationssæt?
- Hvad er *typetallet* i observationssættet?
Hvad fortæller typetallet?
Hvordan finder man generelt typetallet for et observationssæt?
- Hvad er *middelværdien* (som også kaldes *gennemsnittet*)?
Hvordan beregnes middelværdien?
- Lav på et stykke papir et *prikdiagram* eller et *stolpediagram*, der illustrerer karakterfordelingen. Tag et billede og indsæt det her.
Hvad er forskellen på et prik- og et stolpediagram?
- Tilføj en tredje række til skemaet, hvor du i den nye række angiver *frekvensen* i procent for de enkelte observationer.
Hvad er summen af alle frekvenserne i den nye række? Er resultatet overraskende?

- h) Tilføj en fjerde række til skemaet, hvor du i den nye række angiver *den kumulerede frekvens* i procent. I grundskolen bliver kumuleret frekvens nogle gange kaldt summeret frekvens.
- i) Hvis du har regnet rigtigt, står der under observationen 7 i fjerde række i skemaet 52,6 %. Hvad fortæller dette tal om de karakterer, som eleverne opnåede i AT?
- j) Hvor stor en procentdel af eleverne dumpede til studentereksamen i AT?
- k) Lav på et stykke papir et *trappediagram*, der illustrerer udviklingen i den kumulerede frekvens. Indsæt et billede af figuren her
- l) Aflæs på trappediagrammet fra spørgsmål k) den *nedre kvartil*, *medianen* og den *øvre kvartil*. Hvad fortæller de tre tal om karaktererne, som blev opnået til AT-eksamen?
- m) Tegn værdierne ind i et boksplot på et stykke papir. Indsæt et billede af figuren her.

Opgaver vedrørende ikke-grupperede observationer

Besvar følgende opgaver:

Opgave 1

En klasse fik karaktererne anført i nedenstående tabel

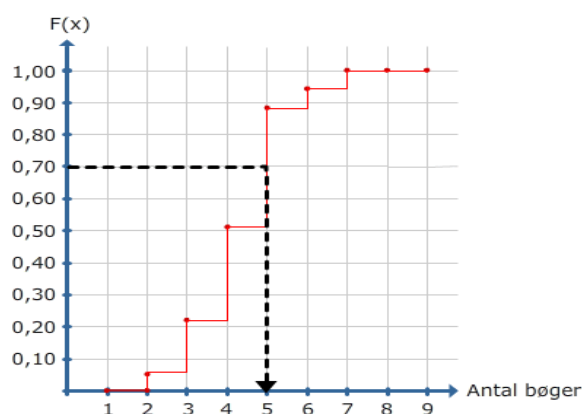
Obs.	00	02	4	7	10	12
Hyp.	3	0	5	7	5	2

- Angiv observationssættets størrelse, typetal og middelværdi. Tegn et stolpediagram over hyppigheden
- Beregn frekvensen og den kumulerede frekvens

Opgave 2

Trappediagrammet til højre viser fordelingen af antal læste bøger i en 1.g klasse 1 år.

- Hvor mange procent af eleverne har læst 4 bøger eller derunder?
- Hvor mange procent har læst 7 bøger eller derover?
- Hvor mange procent af eleverne har læst mellem 4 og 7 bøger, dvs. 4, 5, 6 eller 7 bøger?



Opgave 3

På Solby Gymnasium har man undersøgt, hvor lang tid eleverne bruger på deres lektier. Resultatet blev:

Timer pr. uge	0	1	2	3	4	5	6	7
Antal elever	20	58	62	105	85	23	12	5

- Hvad er observationssættets størrelse?
- Hvor lang tid bruger eleverne i gennemsnit pr. uge til lektier?
- Hvor mange procent af eleverne bruger 3 timer eller derunder?

Opgave 4

Nedenstående statistik over straffelovsovertrædelser kan man finde i tiårsoversigt for 1983

Antal afgørelser for straffelovsovertrædelser i 1983										
Alder	15	16	17	18	19	20	21	22	23	24
Mænd	1801	3867	4388	4705	5037	4390	3998	3327	3107	2749
Kvinder	260	465	542	610	593	563	475	448	453	379

- Tegn et stolpediagram der viser kriminalitet begået af mænd i alderen mellem 15 og 24 i 1983
- Hvor gamle er mændene i undersøgelsen, når de er mest kriminelle?
- Besvar de tilsvarende spørgsmål for kvinderne. Diskuter forskelle og ligheder

Status og selvevaluering vedrørende ikke-grupperede observationer

Kan du forklare betydningen af nedenstående statistiske begreber og kan du omsætte dem til praktisk forståelse af virkeligheden?	Det kan jeg	Det kan jeg næsten	Det kan jeg ikke	Hvad kan jeg gøre for at komme helt i mål?
Hyppighed				
Variationsbredde				
Typetal				
Middelværdi				
Pindediagram				
Frekvens				
Kumuleret frekvens				
Trappediagram				
Nedre kvartil				
Median				
Øvre kvartil				
Boksplot				
Jeg kan bruge statistik skabelonen fra Excel til at løse opgaver med ovenstående begreber				

Grupperede observationer

Vi er nu klar til at se på observationsæt, hvor man med fordel kan samle observationerne i nogle grupper af svar. Dvs. vi skal arbejde med grupperede observationer. Du vil i den forbindelse opdage, at der er mange statistiske begreber til beskrivelse af grupperede observationer, som ligner dem, der bruges til ikke-grupperede observationer, men at begreberne er tilpasset en smule.

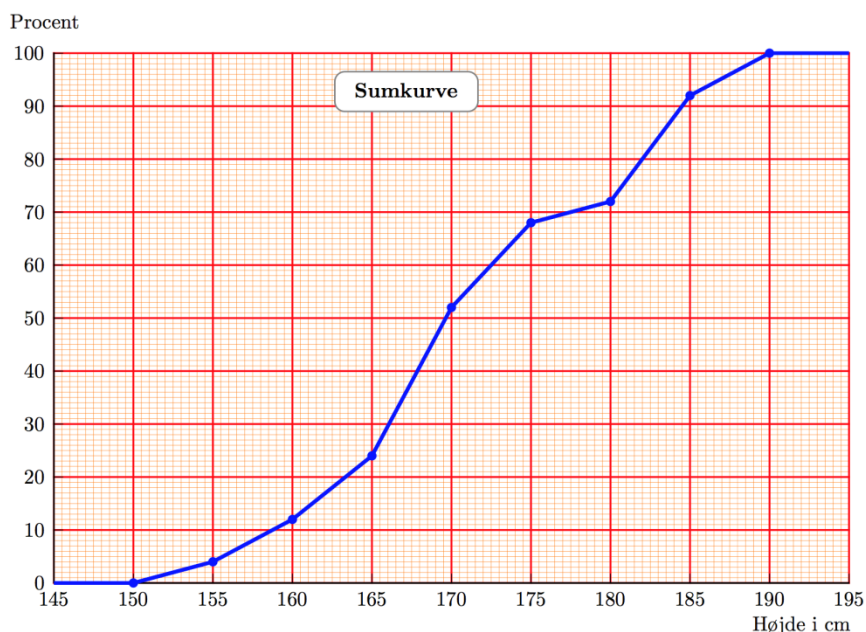
I kan finde alt det i skal vide i i-bogen på følgende side: <https://plusbstx.systime.dk/?id=1507>

Opgave 1: Start med at læse de to første afsnit i I-bogen og skim overskrifterne på resten af siden og find ud af, hvad en sumkurve er.

Løs nu følgende opgave, og brug bogen undervejs som hjælp.

Opgave 2:

Figuren viser sumkurven, der beskriver højden af eleverne på et gymnasium med 800 elever.



- Hvad kan du sige om højden af de højeste elever?
- Hvad kan du sige om højden af de laveste elever?
- I hvilke intervaller har man grupperet elevernes højde i forbindelse med undersøgelsen?
- Hvad er typeintervallet?
- Aflæs kvartilsættet, altså nedre kvartil, median og øvre kvartil. Hvad fortæller de tre tal om elevernes højde?
- Hvor stor en procentdel af eleverne har en højde i intervallet]170 cm; 175 cm]? Hvor mange elever er der tale om?

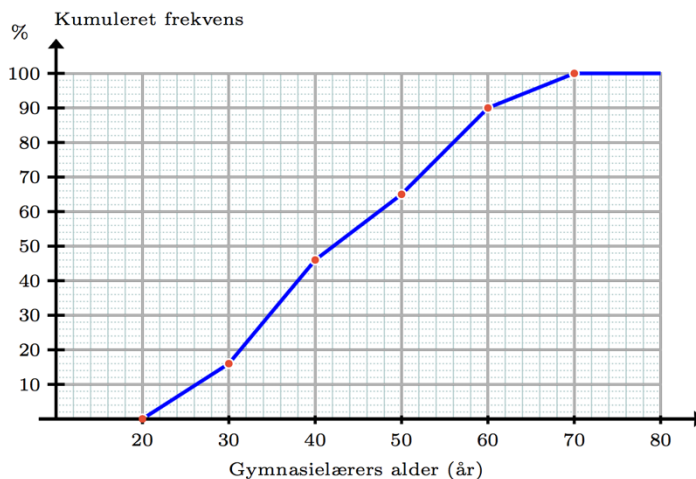
- g) De 20 % højeste elever kan blive udtaget til gymnasiets basket ball-hold. Hvor høj skal man være for at kunne komme på holdet?
- h) Hvad er gennemsnitshøjden for eleverne på gymnasiet?

Opgaver vedrørende grupperede observationer

Opgave 1

Aflæs på sumkurven til højre

- Hvor mange procent af lærerne er 50 år eller derunder?
- Hvor mange af lærerne er over 50 år?
- Hvor mange af lærerne er 42 år eller derunder?
- Hvor mange procent af lærerne er mellem 42 år og 50 år?



Opgave 2

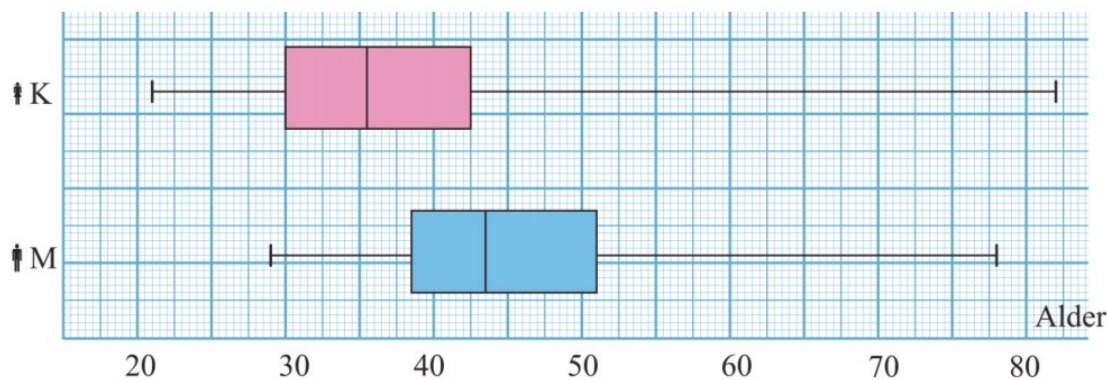
Aflæs på grafen fra opgave 1

- Aflæs 15 % fraktilen og forklar med egne ord, hvad den betyder
- Aflæs 90 % fraktilen
- Hvor gamle er de ældste 10 % af lærerne?
- Lærere over 62 år kan forventes at gå på pension inden for de næste 5 år. Hvor mange procent drejer det sig om?

Opgave 3

Læs om boksplot i i-bogen inden du begynder på denne opgave.

På figuren ses to boksplot over aldersfordelingen blandt Oscar-vindere i kategorien "Bedste skuespiller" fordelt på køn i perioden 1970-2012.



- Bestem kvartilsættet for de to køns aldersfordelinger, og benyt disse til at sammenligne aldersfordelingen for Oscar-vindere i kategorien "Bedste skuespiller" fordelt på køn.

Opgave 4



Kilde: Politiken.dk

Tabellen viser en statistik over antallet af solgte billetter til 667 danske biograffilm fra perioden 1976-2011.

Antal billetter (tusind)	0-50	50-100	100-200	200-400	400-600	600-
Antal film	323	81	104	100	39	20

- Bestem de kumulerede frekvenser.
- Tegn en sumkurve for fordelingen.
Hvad kan man sige om antallet af solgte billetter til de 20 % mest sete film?

Kilde: dfi.dk.

Opgave 5



DMI opgør hvert år antallet af solskinstimer i december.
Tabellen viser fordelingen af antallet af solskinstimer i december i en 50-årsperiode.

Antal solskinstimer	0-20	20-40	40-60	60-80
Hyppighed (antal år)	2	20	25	3

- Bestem middeltallet for antallet af solskinstimer i december i 50-årsperioden.
- Bestem de kumulerede frekvenser.
- Tegn en sumkurve.
Hvad gælder der om antallet af solskinstimer i en december måned, når den er blandt de 20 % mest solrige i 50-årsperioden?

Status vedrørende grupperede observationer

Kan du i forbindelse med grupperede observationer forklare betydningen af nedenstående statistiske begreber og kan du omsætte dem til praktisk forståelse af virkeligheden?	Det kan jeg	Det kan jeg næsten	Det kan jeg ikke	Hvad kan jeg gøre for at komme helt i mål?
Interval				
Intervalhyppighed				
Typeinterval				
Intervalfrekvens				
Kumuleret intervalefrekvens				
Histogram				
Sumkurve				
Nedre kvartil				
Median				
Øvre kvartil				
Boxplot (gælder både ikke-grupperede og grupperede observationer)				
Jeg kan bruge wordmat til at løse opgaver med ovenstående begreber				