

# ALT OM STATISTIK OG DIAGRAMMER

## TÆNK SELV!

Statistik er noget af det sværeste i stoffet, fordi man ud fra statistik og diagrammer skal sige noget selvstændigt. Selve de ting man skal kunne er dog ikke svære og det meste kan gøres med hjælp af regneark, Geogebra eller WordMat. Data kan enten være grupperet eller ugrupperet (se overskrift senere).

## SÅDAN ARBEJDER DU MED STATISTIK

- Start med at finde nogle **Statistiske Deskriptorer**,
- Find sammenhænge og sig noget om deskriptorerne
- Lav et passende diagram, find sammenhænge og sig noget om dem.

## STATISTISKE DESKRIPTORER

En samling data hedder et **datasæt** eller et **observationssæt**. Følgende tal (deskriptorer) kan være en god ide at finde for et observationssæt, men det kommer an på hvad det handler om. I det følgende betyder (...) at man skal markere de celler der skal behandles på dette sted. Der står først hvad du skal skrive på dansk Excel, der efter hvad du skal skrive på engelsk Excel. Der er vist eksempler på hvordan hver funktion *kan* bruges, ude til højre.

**Gennemsnit eller middelværdi:** Læg alle tal sammen og del med antallet, eller brug **=MIDDEL(...)** eller **=AVERAGE(...)**.

=MIDDEL(F2:G12)

**Mindsteværdi:** Den mindste værdi i sættet. **=MIN(...)** (eller **KVARTIL(...;0)** eller **QUARTILE(...;0)**)

=MIN(A:A)

**Størsteværdi:** Den største værdi i sættet. **=MAKS(...)** eller **=MAX(...)**(eller **KVARTIL(...;4)** eller **QUARTILE(...;4)**)

=MAKS(A1:A10000)

**Variationsbredde:** Forskellen på størsteværdi og mindsteværdi. **=MAKS(...)-MIN(...)**

=MAKS(A:A)-MIN(A:A)

**Nedre kvartil:** Den største af de 25 % mindste data (den midterste af den nedre halvdel). **=KVARTIL(...;1)**, eller **QUARTILE(...;1)**

=KVARTIL(A2:A55;1)

**Median:** Den største af de 50 % mindste data. (den midterste). **=MEDIAN(...)** eller **=KVARTIL(...;2)**, eller **=QUARTILE(...;2)**.

=MEDIAN(A:B)

**Øvre kvartil:** Den største af de 75 % mindste data (den midterste af den øvre halvdel). **=KVARTIL(...;3)**, eller **QUARTILE(...;3)** (Quartile-funktionen virker stadig, trods en advarsel)

=KVARTIL(A2:A129;3)

**Typetal:** Det tal der forekommer flest gange. Hvis flere tal forekommer flest gange, er de alle typetal. **=HYPPIGST.FLERE(...)** eller **=MODE.MULT(...)**. Vær opmærksom på at dette giver flere tal som så skal sorteres, f.eks. med **MIN(...)**. *TIP!* Det er ofte en fordel at lave en hyppighedstabel som beskrevet nedenfor.

=HYPPIGST.FLERE(A1:A300)

## HYPPIGHED OG FREKVENS -TABELLER

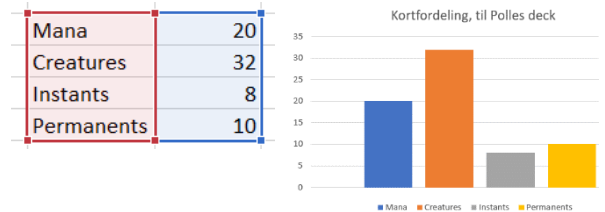
**Hypighed(-stabel):** Hvor mange gange et tal forekommer i datasættet. Tælles for hvert tal, gerne i en tabel, altså en hypighedstabel. Få f.eks. Excel til at sortere data først. Du kan bruge =TÆL.HVIS(...;tal) til at tælle hvert enkelt tal (hedder COUNTIF på engelsk). Husk også at tælle dataværdier der ikke er i et observationsæt (som for eksempel hvis der ikke har været 20 grader i en tabel over temperaturer i løbet af en måned).

=TÆL.HVIS(A:A;F2)

**Frekvens(tabel):** Hvor mange procent af gangene forekommer et tal. (hypighed/antal). Skal ganges med 100 hvis du laver det i hånden, ellers har Excel en tast der kan få data vist som procent...

## SÅDAN FÅR DU DATA IND I ET DIAGRAM

- Marker de data du vil have vist.
- Marker også gerne eventuelle labels du vil have vist.
- Tryk på indsæt -> diagram
- Vælg det der bedst viser det du vil vise
- Tilpas diagrammet så det du vil vises gøres endnu tydeligere.



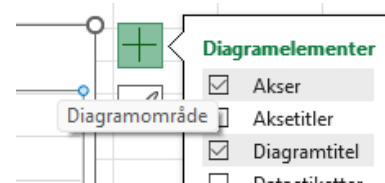
## ÆNDRING PÅ HVAD DER VISES

Så snart du indsætter et diagram, eller trykker på et diagram i dit Excel, viser der sig følgende menuer i toppen:

**Diagramdesign** **Formatér**

Via disse menuer er der adgang til måder at tilpasse dit diagram.

Der er også en hurtigere mulighed oppe i højre hjørne af diagrammet:



## FORSKELLIGE DIAGRAMMER TIL VISNING AF DATASÆT

Ofte bliver data mere forståelige når de sættes ind i et diagram. Men hvor skal du bruge hvad? BRUG PINDEDIAGRAM HVIS DU ER I TVIVL, MEN TÆNK DIG OM!

**Histogram** viser en fordeling af de enkelte datas antal. Laves ud fra en hypigheds- eller frekvenstabel. Se hvordan nedenfor.

**Boksplot** bruges til at vise sammenligninger mellem datasæt. *Se hvordan nedenfor.*

**Pindediagram** eller **søjlediagram** til at vise niveauer og niveauforskelle.

**Kurvediagram** til at vise hvordan noget udvikler sig.

**Cirkeldiagram** til at vise hvordan noget er fordelt ud af et hele. Fordeling. Det er MEGET sjældent cirkeldiagram du skal bruge.

Lad være med at få noget til at se fancy ud! Det forvirrer ofte mere end godt er.

Brug fanen Anbefalede diagrammer. Den virker mega godt. Er der noget galt med akserne eller labels, må du forsøge dig med forskellige ting med at højreklikke på diagrammet og trykke "vælg data" eller lignende.



## DIAGRAMMER TIL SAMMENLIGNING AF DATASÆT

Hvis du skal sammenligne datasæt, skal de behandles på en eller anden måde. **Boksplot** er skabt til dette, men en anden mulighed er at fremstille et histogram, en frekvens- eller hyppighedstabel.

## STATISTISK FREMSKRIVNING AF DATA

Nogle gange vil man bruge målinger til at forudsige fremtiden. Her gælder det om at finde en form for vækst som man kan bruge til at *modellere* hvad der vil ske. Til at finde en vækstmodel bruges regression (se ALT om *Find den rette funktion eller regression*).

## SUMKURVER OG TRAPPEDIAGRAMMER.. ETC.

Jeg anbefaler generelt ikke at I bruger sumkurver, men det er der andre der gør. Hvis I skal lave en sumkurve eller trappediagram, er det nemmest at paste data ind i WordMats Statistikdel. Så tegner den automatisk en sumkurve. Vær opmærksom på at WordMat åbner et nyt excelark som du nu kan paste data ind i. HUSK at trykke på:



Kopier til øvrige ark

, og vælg en fane hvor den tegner det du vil have vist.

## GRUPPEREDE DATA OG UGRUPPEREDE DATA

Ugrupperede data er rene tal, og altså enkelte observationer.

Grupperede data er optællinger af grupper af data, som bliver inddelt i et interval. Man laver ofte gruppering af data når der er rigtig meget data. Når man skal *regne* med grupperede data, regner man normalt ud fra intervalmidtpunktet.

*Eksempel:*

```
=TÆL.HVIS(D2:D77;"<6")-TÆL.HVIS(D2:D77;"<1")
```

I denne celle vises hvor mange af tallene i celle D2 til D77 er mellem 1 og 5.

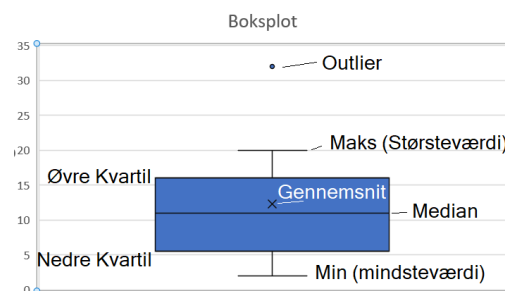
Intervalmidtpunktet ville være 3.

## SÅDAN LAVER DU ET BOKSPLOT

Du kan lave boksplot i både, Excel, Geogebra og WordMat.

Herunder beskriver jeg hvordan det gøres i Excel da det PT er nemmest at bruge.

1. Marker data
2. Tryk på Indsæt
3. Vælg (som altid) anbefalede diagrammer
4. Vælg alle diagrammer
5. Vælg "Kasse med Hale"



Hvis du markerer flere kolonner, laver den flere boksplot i samme diagram (husk at navngive det eller dem). Derfor skal du naturligvis samle dine data i en enkelt kolonne, hvis du skal lave et enkelt boksplot.

## OUTLIERS M.M.

Outliers er data som ligger så langt fra det forventede at de markeres som prikker udenfor, for at gøre opmærksom på at de netop virker som noget der ikke passer med datasættet. Om du skal tage højde for dem, afhænger af, hvordan datasættet er fremkommet, så altså på din egen vurdering.

Hvis du vil ændre på hvordan diagrammet fremstår, f.eks. fjerne outliers, skal du højreklikke på det specifikke diagram, vælge *formater dataserie*, og tilvælge/fravælge i menuen.

## SÅDAN LAVER DU ET HISTOGRAM I EXCEL (OG GEOGEBRA)

1. Start med at lave en hyppighedstabel, altså hvor meget af datasættet der ligger på en bestemt værdi, eller interval af værdier, hvis du grupperer dine data.
2. Husk at medtage alle tal i det interval du vil vise, også hvor der ikke er nogle data med den pågældende værdi.
3. Gør det samme for hvert datasæt, hvis du vil sammenligne.
4. Giv hver kolonne en fornuftig titel.
5. Marker dine data og titler og indsæt et pindediagram.

Hvis hellere vil gøre det i geogebra, skal du paste data ind i geogebra's regneark, markere dem og vælge enkeltvariabelanalyse. Her er der en mulighed for at lave et histogram.