

AARHUS UNIVERSITET

KVI

SEMESTER 3

NOTER

Studerende:

Mette HAMMER NIELSEN-KUDSK

Martin BANASIK

September 15, 2015



Biostatistik

Variationer når man begynder at måle. Systematisk variation

- Mænd er højere end kvinder
- Blodtrykket stiger med alderen
- Rygning øger risikoen for lunge cancer

Tilfældig variation

- Analytisk variation (målefejl)
- Biologisk variation
- Observatør variation (rapportering)

Hvorfor statistik?

- Behov for at kvantificere hvor meget af den observerede data der skyldes tilfældige variation og hvor meget der skyldes systematisk variation.
- Behov for at resumere mange enkelte observationer i nogle få tal.

Hvorfor stikprøver (samples)?

- Hurtigere, billigere, umligt at undersøge alle, kan være tilstrækkelig med en sample, statistiske metoder kan bruges til at vurdere usikkerhed.

Dataanalysen kan opdeles i

DESKRIPTIV STATISTIK Som er et data summary. Indeholder:

- gennemsnit / median / percentiler
- hyppigheder / relativ risiko / oddsratio
- Tegninger/figurer - Vigtig!

STATISTISK INFERENS

- Model / antagelser angående variationen i data
- Estimation af relevante parametre i modellen (sample fra populationen) (f.eks. middelværdi eller forskel mellem to grupper) ud fra stikprøven med tilgørende sikkerhedsintervaller
- Opstilling af statistiske hypoteser, statistiske test
- Statiske konklusioner
- Kvantificere usikkerheden på estimatet
- Faglige konklusioner

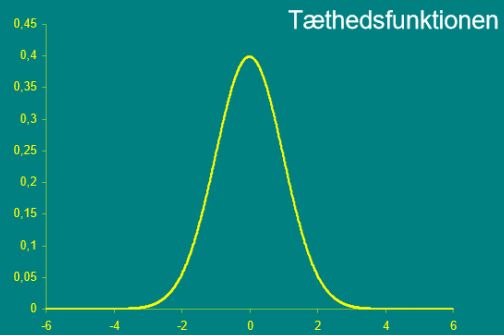
STATISTISK INFERENS

- Kvalitative data (kategoriske data)
 - F.eks. køn, blodtype (ABO)
- Kvantitative data (numeriske data)
 - F.eks fødselsvægt, blodtryk, cancer incidens
- Diskrete data
 - Når observationer kun kan antage visse numeriske værdier
- Kontinuære data
 - Som regel ved måling, f.eks FEV1, SVR

Normal fordelingen

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

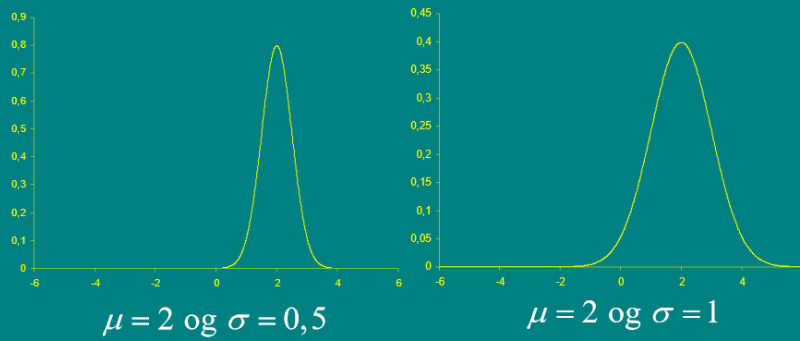
$\mu = 0$ og $\sigma = 1$



$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Normal fordelingen. Klokkeform. Spredningen er = 1 og middelværdi er = 0

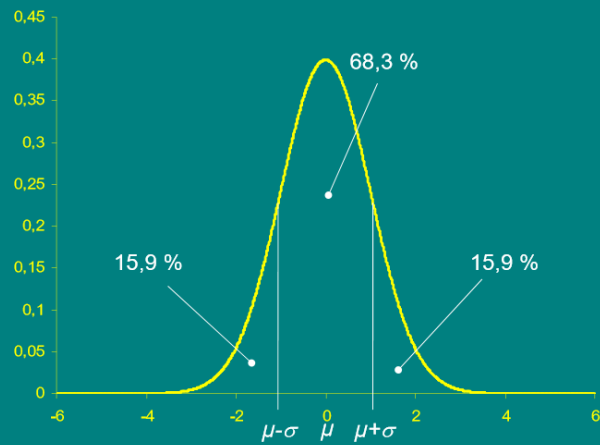
Normal fordelingen



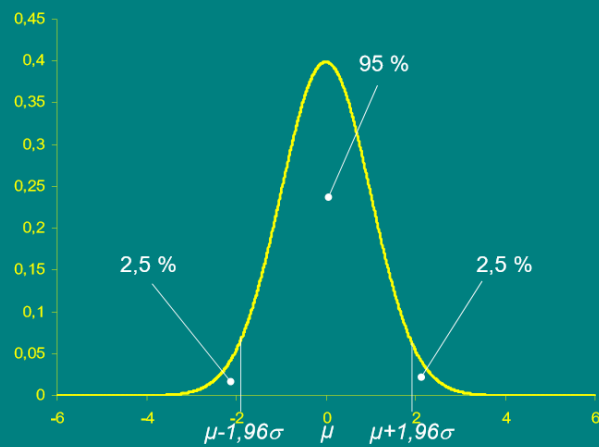
Til venstre viser den en middelværdi = 2 og en spredning = 0,5. til højre er middelværdi = 2 og en spredning = 1.

Middelværdien viser placeringen og spredningen viser bredden af klokkeformen.

Normal fordelingen



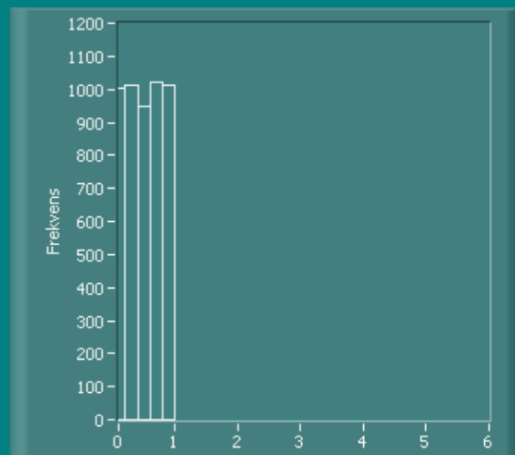
Normal fordelingen



Beskriver hvor meget vores data spreder sig. En lille spredning er lig færre stikprøver.

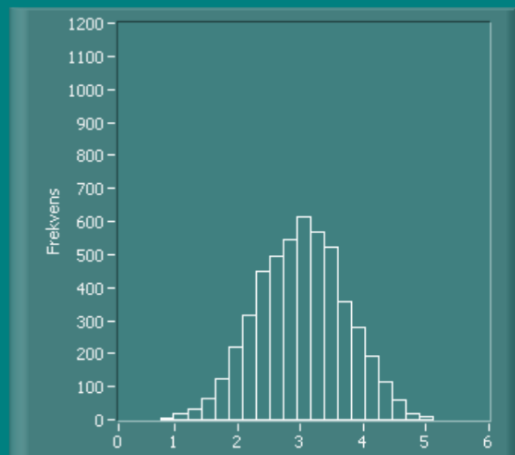
Den centrale grænseværdi sætning

5000 observationer af en jævnt fordelt variable



Den centrale grænseværdi sætning

5000 observationer af seks jævnt fordelte variable



Vil tilsidst blive til en normalfordeling jo mere de bliver summeret sammen (højere variable).