

Utvalgsfordelingar

- Utvalgsfordeling for gjennomsnitt (med kjent varians) ($\bar{X}$)
- Sentralgrenseteoremet (SGT)
- Utvalgsfordeling for varians (normalfordeling)
- Utvalgsfordeling for gjennomsnitt (normalfordeling og ukjent varians)

Databeskrivelse:

- Mest i kapittel 1 (anbefalte oppgaver 1)
- Kapittel 8: Normal-plott (er dataene frå ei normalfordeling?)

Frå eit datasett finn vi (observerer vi) typisk:

Empirisk gjennomsnitt: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

Empirisk varians: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

Empirisk median: $\tilde{x} = x_{(\frac{n+1}{2})}$ (dersom n odde)

Frå eit datasett finn vi (observerer vi) typisk:

Empirisk gjennomsnitt: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

Empirisk varians: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

Empirisk median: $\tilde{x} = x_{(\frac{n+1}{2})}$ (dersom n odde)

Om vi ser på desse teoretisk (desse er observatorar):

Gjennomsnitt: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

Utvallsvarians: $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

Median: $\tilde{X} = X_{(\frac{n+1}{2})}$ (dersom n odde)

Frå eit datasett finn vi (observerer vi) typisk:

Empirisk gjennomsnitt: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

Empirisk varians: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

Empirisk median: $\tilde{x} = x_{(\frac{n+1}{2})}$ (dersom n odde)

Om vi ser på desse teoretisk (desse er observatorar):

Gjennomsnitt: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

Utvallsvarians: $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

Median: $\tilde{X} = X_{(\frac{n+1}{2})}$ (dersom n odde)

Def. 8.4 Observator

Ein funksjon av stokastiske variable som representerer eit tilfeldig utval blir kalla ein observator (engelsk: statistic) $\Rightarrow$ har sannsynsfordeling.

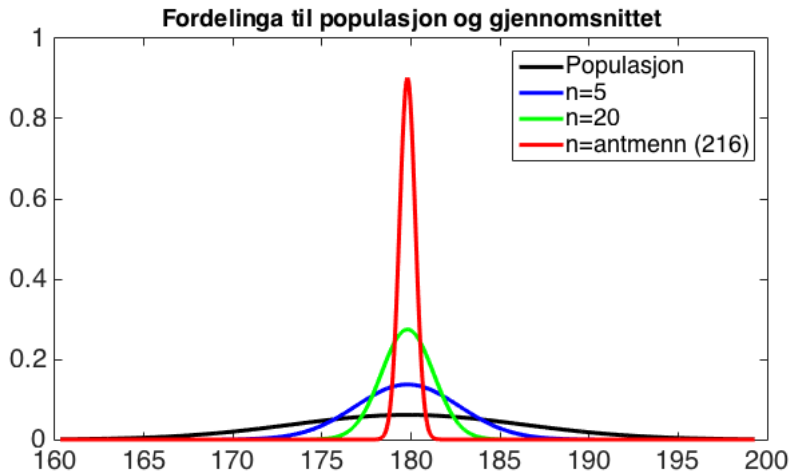
Ein observator er ein stokastisk variabel, då det er ein funksjon av stokastiske variable.

Dersom X_i er normalfordelt, og kjent varians

$X_i \sim N(\mu, \sigma^2)$, for $i = 1, 2, \dots, n$. Dvs eit utvalg på n .

Då er

$$\bar{X} \sim N(\mu, \sigma^2/n)$$



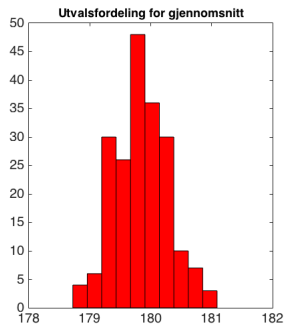
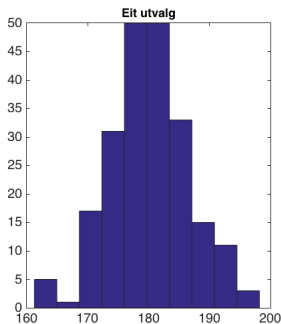
Utvalgsfordelinga til $\bar{X}$

Algoritme

For $m = 1 : M$

- Trekk $n=216$ datapunkt frå $N(179.8, 6.5^2)$ $m = 1, \dots, M$ gongar $\Rightarrow x_{m1}, x_{m2}, \dots, x_{mn}$.
- Finn gjennomsnittet $\bar{x}_m = 1/n \sum_{i=1}^n x_{mi}$

Plott histogram for av $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_M)$



Sentralgrenseteoremet. Fordelinga til gjennomsnittet, $\bar{X}$.

Sentralgrenseteoremet, teorem 8.2

La X_i , $i = 1, 2, \dots, n$ vere uavhengige identisk fordelte (u.i.f.) stokastiske variable med $E(X_i) = \mu$ og $\text{Var}(X_i) = \sigma^2 < \infty$ (endeleg varians). La $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ og $Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$. Då

$$Z \rightarrow N(0, 1)$$

når $n \rightarrow \infty$.

Dette tilsvarer

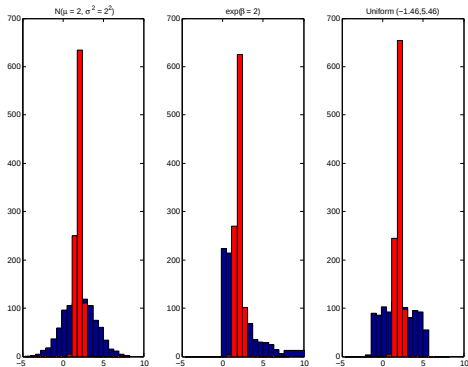
$$\bar{X} \rightarrow N\left(\mu, \frac{\sigma^2}{n}\right)$$

når $n \rightarrow \infty$.

PS: Gjeld uansett fordeling for X_i .

PSS: $Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$ er og ein observator

Histogram populasjon og gj.snitt av 30



Ønsker å finne ut om det er kvalitetsforskjell på ventilane A og B .

- Har testa $n = 30$ av kvar, og fått $\bar{x}_A = 2100$ og $\bar{x}_B = 2157$.
- Antar kjente variansar $\sigma_A^2 = 100^2$ og $\sigma_B^2 = 200^2$
- Forskjell mellom dei, eller tilfeldig variasjon?

PS: Seier ingenting om fordelinga til kvaliteten!

Skal sjå på:

- $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
- $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$

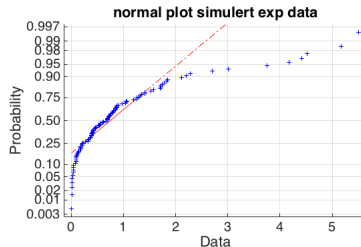
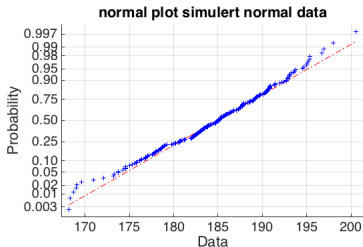
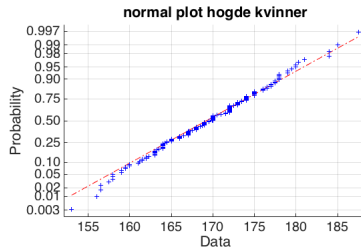
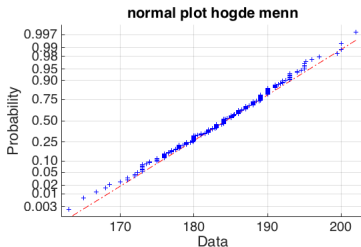
Gjeld kun for $X_1, \dots, X_n \sim N(\mu, \sigma^2)$

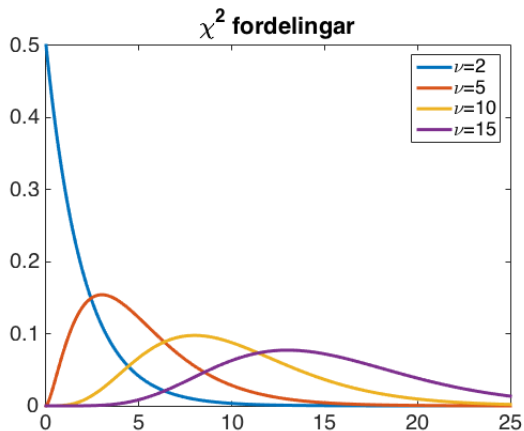
- Korleis finn vi ut om dataene våre er normalfordelte?

Kan bruke **normal-plott** (i Matlab *normplot*)

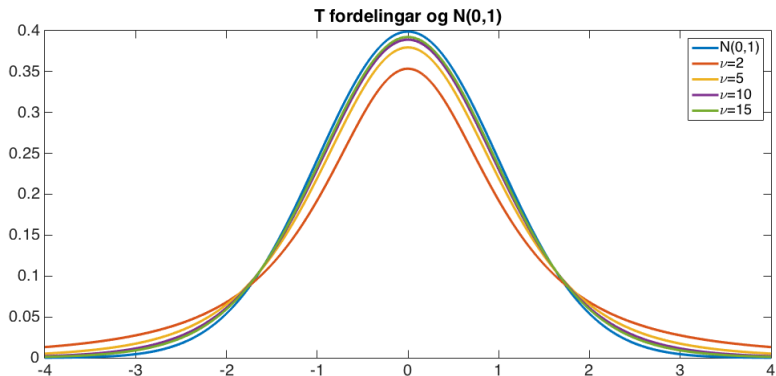
Normal-plot: Dersom normalfordelt, ca på rett linje.

Normal-plott





Student t-fordelingar



- Utvalsfordelingar
- Utvalsfordeling for gjennomsnitt (med kjent varians)
 - Sentralgrenseteoremet (SGT) uansett fordeling
 - Utvalsfordeling for varians normalfordeling, χ^2_{n-1}
 - Utvalsfordeling for gjennomsnitt normalfordeling og ukjent varians, T_{n-1}

Databeskrivelse: boksplott og qq-plot