Hypotese test.

Sammenligning af 2 populationer

Case

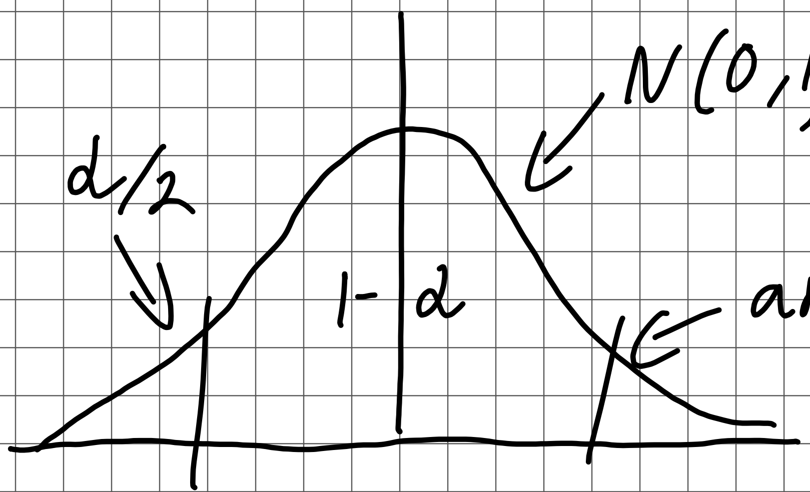
Model

kendt

pop 1 $X_{1i} \sim N(\mu_1, \sigma^2)$

Test

enten vælg α — 1%, 5%, 10%



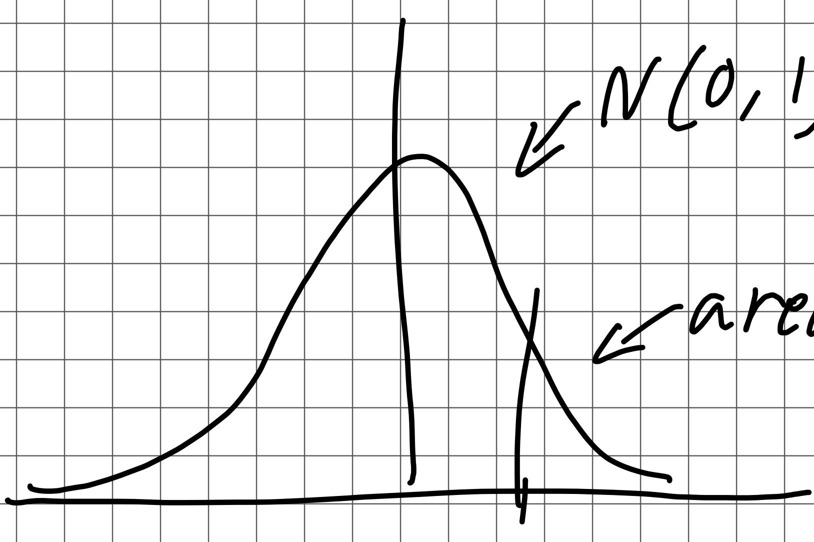
N(0,1) pdf

areal $\alpha/2$

$-z_{\alpha/2}$ $\qquad z_{\alpha/2} = \mathrm{norminv}(1-\frac{\alpha}{2})$

$|z_0| > z_{\alpha/2}$ -> forkast H_0

Eller



N(0,1) pdf

areal $\frac{p\text{-verdi}}{2}$

z_0

$p\text{-verdi} = 2 \cdot P(Z > |z_0|)$

$$= 2(1 - \underbrace{P(Z \leq |z_0|)}_{\substack{\text{normcdf}(|z_0|) \\ \Phi(|z_0|)}})$$

p-verdi = Lille

$\downarrow$

forkast H_0

Case A   ukendte $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Test statistik $$Z_0 = \frac{(\bar{X}_1 - \bar{X}_2) \cdot \Delta_0}{\sqrt{\sigma^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

antal samples

σ^2 ukendt

$$X_1 \to S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2$$

$$X_2 \to S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n} (x_{2i} - \bar{x}_2)^2$$

hvilken skal vi valge?

vi tager et vægtet gennemsnit

pooling

vægt gennemsnit

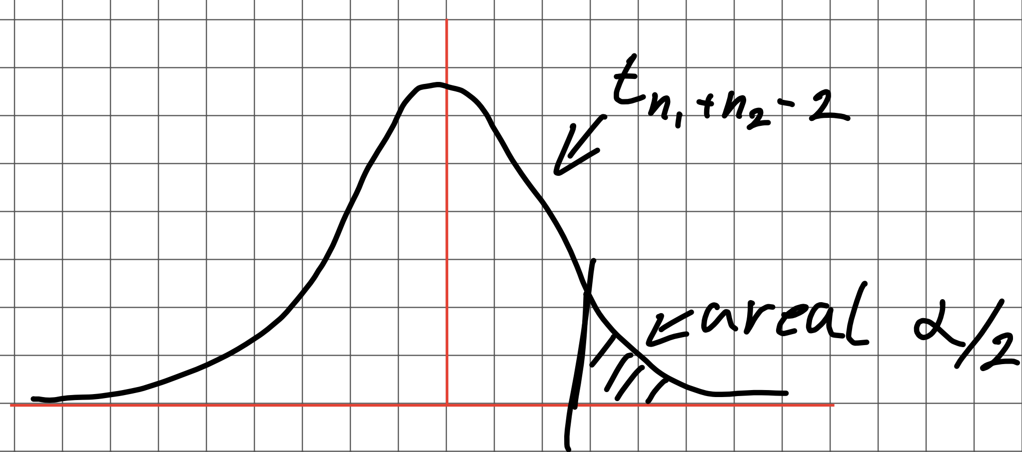
$$S_p^2 = \frac{(n_1 - 1) S_1 + (n_2 - 1) S_2}{n_1 + n_2 - 2} \sim \frac{\sigma^2}{n_1 + n_2 - 2}$$

på den måde vægter den med flest samples højst

Test Statistic

$$T_0 = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$$

Data, H_0: $T_0 = t_0$ (talværdi)



$t_{n_1+n_2-2,\,\alpha/2}$

$\uparrow$

$\text{tinv}\left(1-\frac{\alpha}{2}, n_1+n_2-2\right)$

forkast $H_0 \iff |t_0| > t_{n_1+n_2-2,\,\alpha/2}$

$$\text{p-værdi} = 2\left(1-\underbrace{P(t_{n_1+n_2-2} > |t_0|)}_{\downarrow}\right)$$

$\text{tcdf}(|t_0|, n_1+n_2-2)$

Case B ukendt $\sigma_1^2 \neq \sigma_2^2$

Large sample approximate test

$n_1, n_2 \gg 1$

$$Z_0^* = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{\text{approx}}{\sim} N(0,1)$$

Data $Z_0^* = z_0^*$ → hypoteret
→ CI

Paired test

hvert test objekt skal kandeles i 2 og gives hver sin "treatment"

Model

paneede data $(X_{1i}, X_{2i})\ i=1 \ldots n$

$X_{1i} \sim N(\mu_1, \sigma^2)$

$X_{2i} \sim N(\mu_2, \sigma^2)$

Sks en mark med 2 forskellige typer gødning.

$H_0: \mu_1 - \mu_2 = \Delta_0$

Difference model

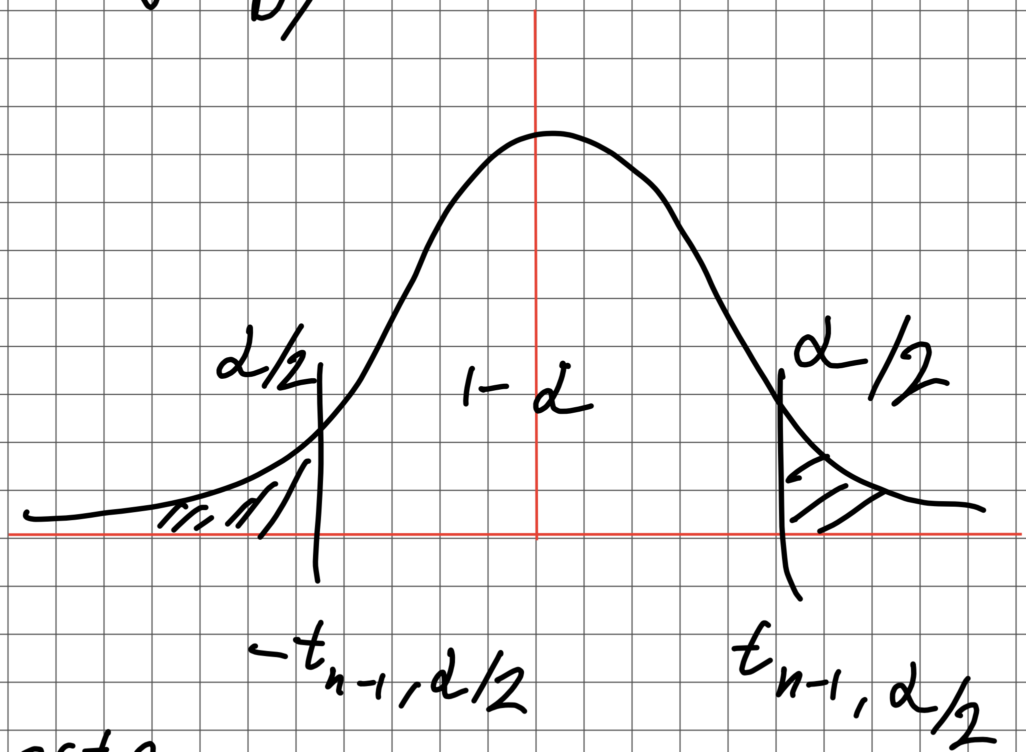
$$D_i = X_{1i} - X_{2i} \sim N(\mu_1 - \mu_2, \sigma_D^2)$$

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i \sim N\left(\mu_1 - \mu_2, \frac{\sigma_D^2}{n}\right)$$

$$\tilde{\sigma}_D^2 = S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(D_i - \overline{D}\right)^2$$

Test Statistic

$$T_{D,0} = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{S_D^2/n}} \sim t_{n-1}$$



Vartest2 i matlab kan sammenligne 2 varianser.

Varianshomogenitet

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

$$S_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n} (X_{1,i} - \bar{X}_1)^2 \sim \frac{\sigma_1^2}{n_1-1} \chi^2_{n_1-1}$$

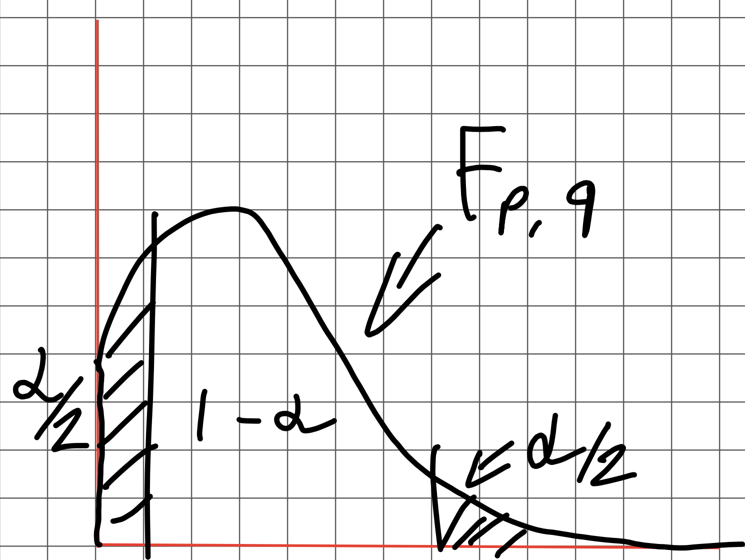
$$S_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n} (X_{1,i} - \bar{X}_1)^2 \sim \frac{\sigma_2^2}{n_2-1} \chi^2_{n_2-1}$$

Fisher fordeling

$SX \sim \chi^2_p$

$Y \sim \chi^2_0$

$\frac{X/p}{Y/q} \overset{def}{\sim} F_{p,q}$



$F_{n_1-1,n_2-1,\alpha/2}$

finv$(1-\alpha/2, n_1-1, n_2-1)$

alle 4: brande til KBH: 396

2 ~~brand~~ brande til Kbh : 198 + 120 - pladser

2' odense til Kbh : 134

total: $\underline{\underline{452}}$

hjem

2 Kbh til brande: 198 + 120 - plads

2 Kbh til odens : 178

total: 496

total-total: 948