Ett $100(1-\alpha)\%$ konfidensintervall är ett slumpintervall för en parameter θ sådant att

$$P[L_1 \leq \theta \leq L_2] = 1-\alpha$$

$\bar{X}$, stickprovsmedlet, är en slumpvariabel med storlek n.

Antag normalfördelning: då är

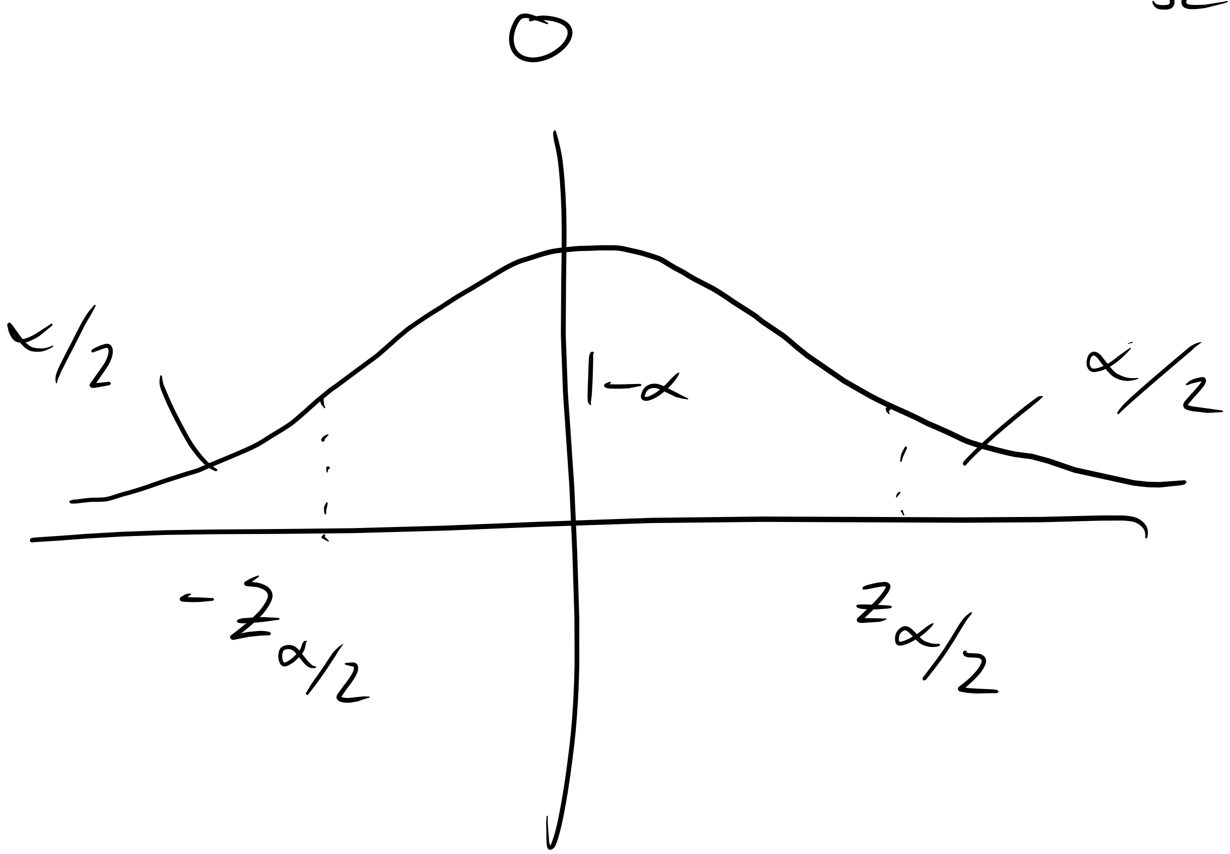
$$\bar{X} \sim N(\mu, \sigma^2/n)$$

Konfidens interval på medlet när μ och σ^2 är kända.

$$\bar{x} \pm z_{\alpha/2}\,\sigma/\sqrt{n}$$

hur hitta $z_{\alpha/2}$?

scip. stats. ppf ($\alpha/2$)



0.5 ± 0.02

$[0.418 ,\ 0.502]$

Konfidensintervall på stickprovsmedlet när μ och σ^2 är okända.

$$\bar{X} \pm t_{\alpha/2} S\sqrt{n}$$

T-fördelningen har en viktig parameter: n, stickprovets storlek. Om jag har ett stickprov $n = 123$ då är det T_{122}-fördelningen jag skall använda!
Vi säger att T-fördelningen har $n-1$ frihetsgrader (degrees of freedom).

Konfidensintervall på variansen (ovanligt mått) :

$$(n-1)\,S^2/\sigma^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/\sigma^2$$

denna följer en χ^2-fördelning.

$$L_1 = (n-1)\,S^2/\chi^2_{\alpha/2} \qquad L_2 = (n-1)\,S^2/\chi^2_{1-\alpha/2}$$

Centrala gränsvärdessatsen.

Låt $X_1, X_2, \ldots, X_n$ vara ett stickprov med storlek n från en fördelning med medel μ och varians σ^2.

För stora n så är $\bar{X} \sim N(\mu, \sigma^2/n)$ och

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

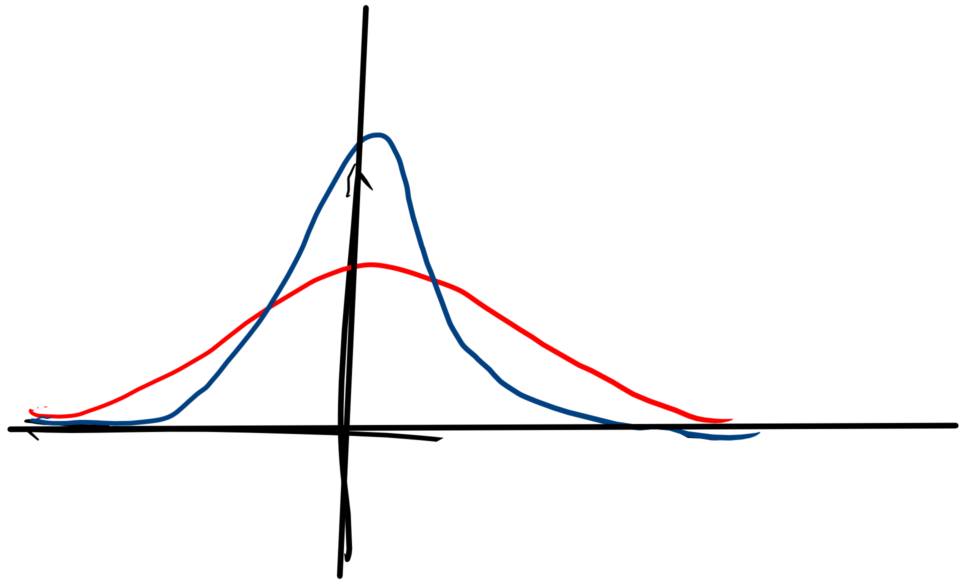
T-fördelning (student-t)

$$T = \frac{Z}{\sqrt{\chi_r^2 / r}}$$

där χ_r är en χ^2-slumpvariabel och r en parameter.

har en parametern $n-1$ (som blir r ovan), stickprovets storlek med Bessels korrigering $(n-1)$.

Egenskaper: Det finns oändligt många T-fördelningar (en för varje n)



Exempel konfidensintervall:

En dödlig sjukdom leder till att patienten avlider i medel 13 månader efter diagnos med standard avvikelse 3 månader. En ny behandling undersökes. Låt $X_1, X_2, \dots, X_n$ vara ett stickprov som beskriver hur länge patienten överlever.

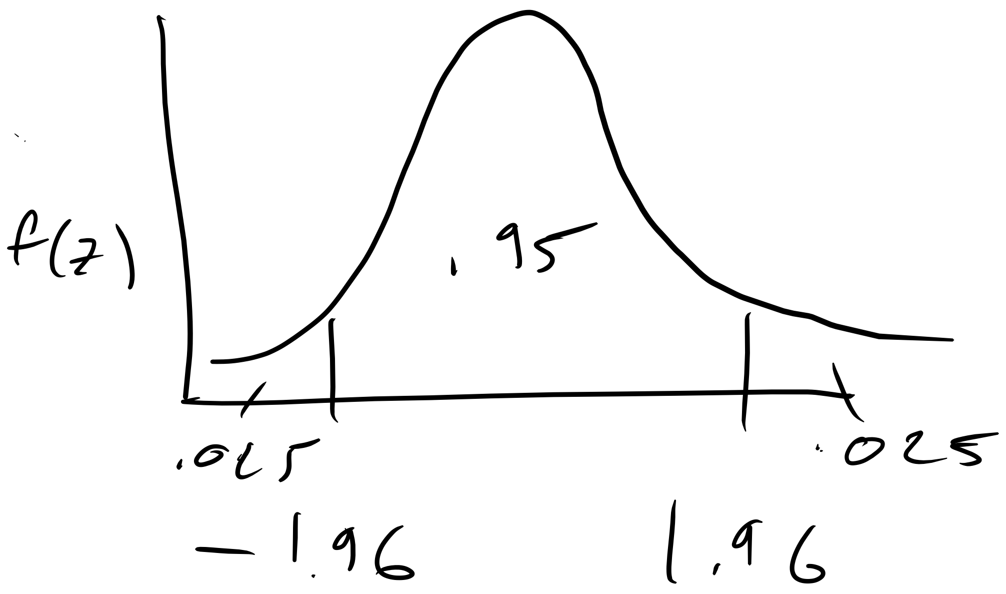
En dödlig sjukdom leder till att patienten avlider i medel 13 månader efter diagnos med standard avvikelse 3 månader. En ny behandling undersöks.

Låt $X_1, X_2, \ldots, X_n$ vara ett stickprov som beskriver hur länge patienten överlever.

Antag $\sigma^2 = 9$ och μ är okänd.

Vi vill konstruera ett 95% konfidensintervall för medlet:

$$P[L_1 \leq \mu \leq L_2] = .95$$



$L_1 = \bar{X} - 1.96\,\sigma/\sqrt{n}$

$L_2 = \bar{X} + 1.96\,\sigma/\sqrt{n}$

$\sigma = 3$

$n = ?$

Ett 95% konfidensintervall för medlet är

$\bar{X} \pm 1.96(.3119)$

$\bar{X} \pm z_{\alpha/2}\,\sigma/\sqrt{n}$

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p$$

För att se om regressioner (uppskattningar) är "bra" så testar vi <u>hypoteser</u> med statistik.

Exempel:

påstående: mer än 50% av alla bilar i trafik har skeva lampar.

noll-hypotesen $H_0 : p \leq .5$

$H_1 : p > .5$

1) Om vi avvisar H_0 när den var sann så har vi gjort ett Typ I-fel.

2) Om vi avvisar H_0 när H_1 är sann så har vi gjort rätt!

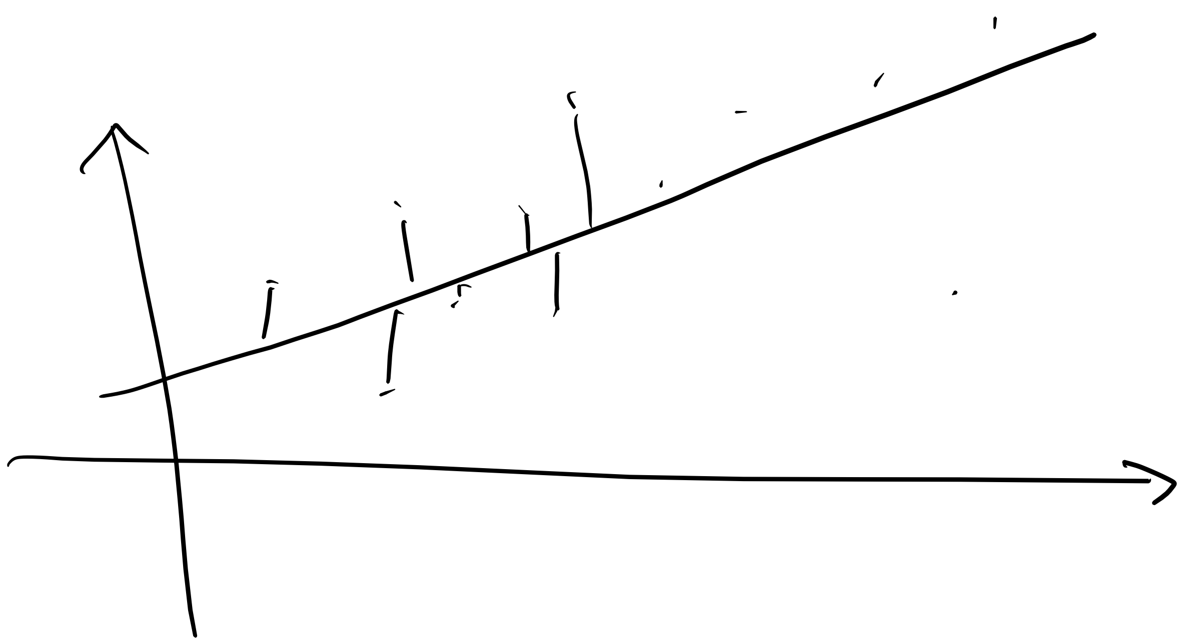
3) Om vi inte avvisar H_0 när H_1 är sann så har vi gjort ett Typ II-fel

4) Om vi inte avvisar H_0 när H_0 är sann har vi också gjort rätt.

Y - responsvariabel $\quad Y = X$

X - nu är X fasta värden (dvs händelser) $\quad X = x$

$\mu_{Y|x} = \beta_0 + \beta_1 x$



Minsta kvadratmetoden:

försöker minimera avståndet från regressionslinjen och alla punkter. (kvadraten på avståndet!)

Vårt optimeringsvillkor:

MSE (mean squared errors)
SSE (sum of —"—)

Hur räknar vi ut en MKM?

$Y = \beta_0 + \beta_1 X$

Det vi vill ha är β-värdena!

$Y = X\beta + E$

$\hat{\beta} = b = (X^T X)^{-1} X^T Y$

residualer

$y_1 = b_0 + b_1 x_1 + e_1$