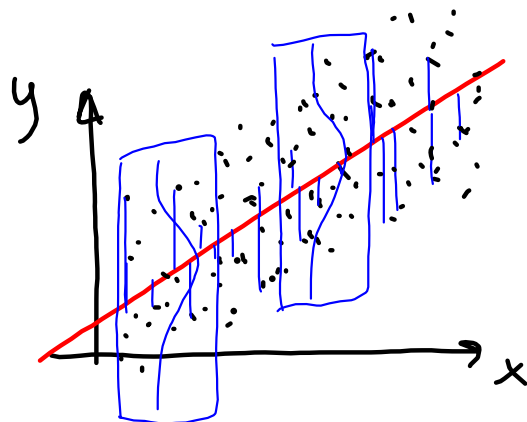


Regressjons ligning :  $y$  respons variabel  
 respons på endring i  $X$  : forklaringsvariabel  
 Kovariat  
 (prediktor)



Linear sammenheng :  $y = \beta_0 + \beta_1 x + \varepsilon$  ,  $\varepsilon \sim N(0, \sigma)$

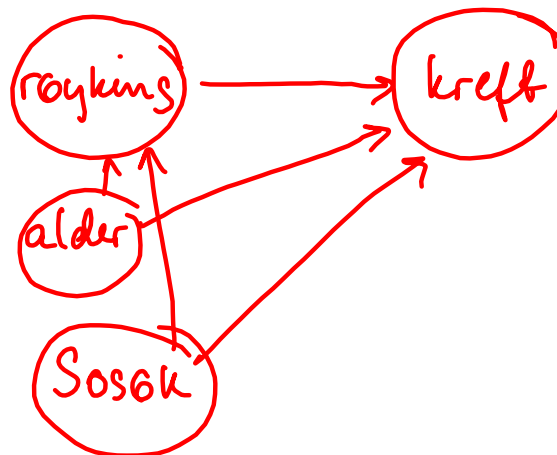
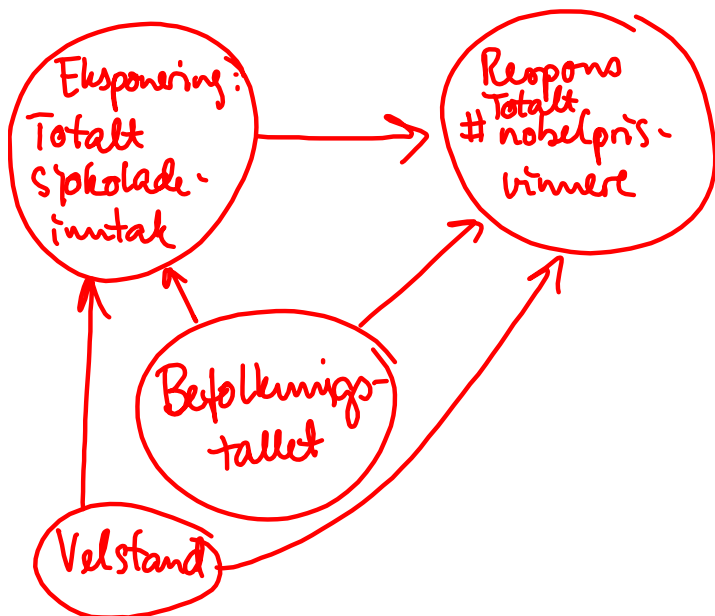
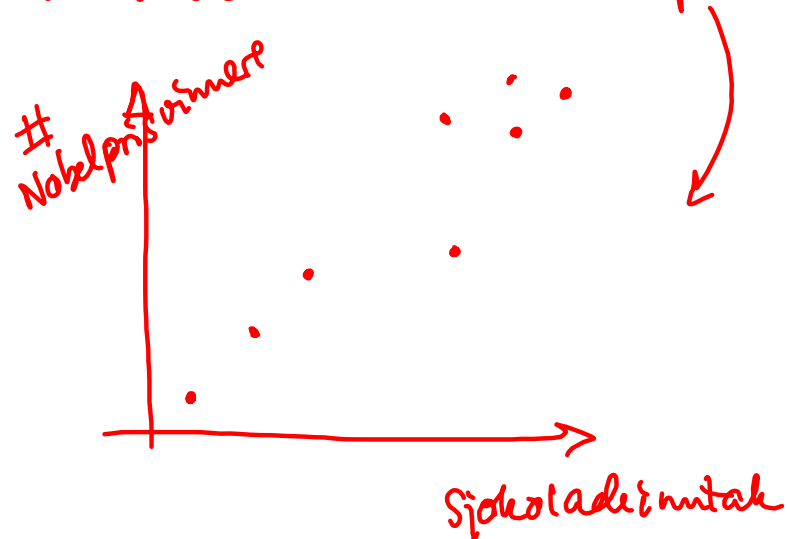
Sammenhengen mellom  $x$  og  $y$  : Hvor mye øker  $y$  (forventes  $y$  å øke) når  $x$  øker med en enhet? Effekt målet er  $\beta_1$

Estimeres med  $\hat{\beta}_1$

To variabler er assosierte (med hverandre) hvis de har en tallmessig sammenheng (Tallmessig sammenheng kan måles oha korrelasjonsanalyse, t-test, WRS-test,  $\chi^2$ -test, Einers anova, Kruskal-Wallis test)

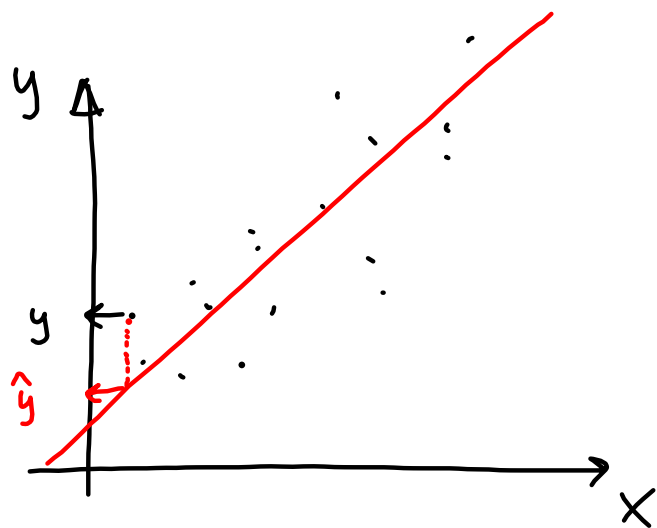
Assosiasjon  $\neq$  sammenheng.

Vi skiller mellom assosiasjoner og årsakssammenhenger.



Prediksjon :

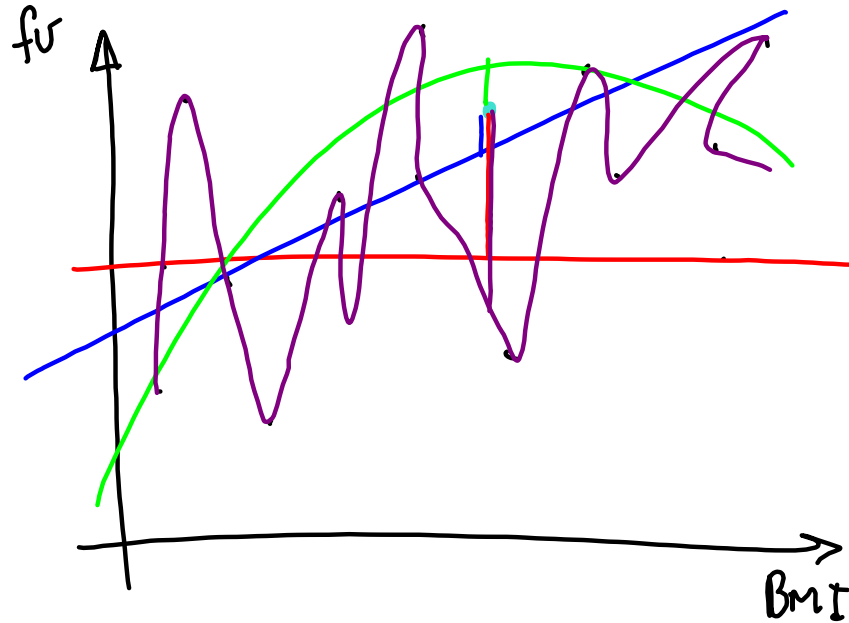
Respons  $y$  : predikeres av prediktoren  $x$



Linear sammenheng :  $y = \beta_0 + \beta_1 \cdot x + \varepsilon$  ,  $\varepsilon \sim N(0, \sigma)$

Prediksjonsfeil kan for eksempel være  $\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 \cdot \frac{1}{n}}$

"Root mean squared error of prediction" : RMSEP



$$y = \beta_0 + \epsilon$$

$$\hat{\beta}_0 = \bar{y}$$

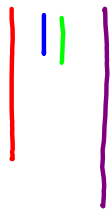
$$y = \beta_0 + \beta_1 x + \epsilon$$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$$

$$y = \text{veldig kompleks funksjon av } x + \epsilon$$

Nytt datapunkt

Hva er prediksjonstikken ift disse fire modellene?



For enkel modell

Bedre

Også ok

Over tilpasset