

## Lineær regresjon

ResponsvARIABLE  $y$

ForklarsvARIABLE  $x_1, \dots, x_p$

Populasjonsregresjonsligning

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i$$

Minste kvadraters metode  $\hat{\sigma}^2$

$\rightarrow b_0, b_1, \dots, b_p$

$N(0, \sigma)$

er Maximum Likelihood-estimatoren

## Statistisk inferens for lineær regresjon

Stigningstall  $\beta_j$  :

- konfidensintervall
- hypotesetester

Responsvariabelen  $y$ :  
gitt en  $x$ -verdi

- konfidensint. for  $\mu_y$
- prediksjonsint.  $\hat{y}$

# Maximum Likelihood -prinsippet

Parametrisk sannsynlighetsmodell

Eks :  $\text{Bin}(n, p)$

Eks :  $\mathcal{N}(\mu, \sigma)$

param

$X_1, \dots, X_n$

$\mathbb{P}(X_1=x_1, \dots, X_n=x_n) = f(\downarrow)$

## Logistisk regresjon

Responsvariabelen  $y_i \in \{0, 1\}$

EKS: syk eller ikke syk  
vant eller vant ikke

Vi modellerer suksess-sannsynligheten og hvordan den varierer med  $x$

$$0 \leq P(y = 1 | x_1, \dots, x_p) \leq 1$$

Den må være et tall mellom 0.0 og 1.0.

Tre ekvivalente def. av logistisk regresjonsmodell:

- i) definerer  $p_i$
- ii) definerer odds  $\frac{p_i}{1-p_i}$
- iii) definerer log-odds  $\log\left(\frac{p_i}{1-p_i}\right)$

Husk også å definere suksess-sannsynligheten (koblet til  $y_i$ )