

# Enkel lineær regresjon

- ★ Situasjon:

- har observert  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$
- ønsker å tilpasse en rett linje til disse dataene

- ★ Stokastisk modell:

- betrakter  $y_1, y_2, \dots, y_n$  som realisasjoner av stokastiske variabler  $Y_1, Y_2, \dots, Y_n$
- betrakter  $x_1, x_2, \dots, x_n$  som (kjente) tall
- antar  $Y_1, Y_2, \dots, Y_n$  uavhengige og

$$Y_i \sim n(y_i; \alpha + \beta x_i, \sigma)$$

- ukjente parametre:  $\alpha$ ,  $\beta$  og  $\sigma^2$

- ★ Metoder for å estimere parametrene

- sannsynlighetsmaksimeringsprinsippet
- minste kvadraters metode

## Enkel lineær regresjon — SME

- ★ Modell:  $Y_1, Y_2, \dots, Y_n$  er uavhengige og  $Y_i \sim n(y_i; \alpha + \beta x_i, \sigma)$
- ★ Log-rimelighetsfunksjon:

$$l(\alpha, \beta, \sigma^2) = n \ln \left( \frac{1}{\sqrt{2\pi}} \right) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$$

- ★ Vi fant maksimum ved å løse ligningssystemet

$$\frac{\partial l}{\partial \alpha} = 0 \quad , \quad \frac{\partial l}{\partial \beta} = 0 \quad , \quad \frac{\partial l}{\partial \sigma^2} = 0$$

med hensyn på  $\alpha$ ,  $\beta$  og  $\sigma^2$

- ★ SME

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad , \quad \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{x} \quad , \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$

## Egenskaper til estimatorene

- ★ Vi fant at  $\hat{\beta}$  er normalfordelt og

$$E[\hat{\beta}] = \beta$$

og

$$\begin{aligned} \text{Var}[\hat{\beta}] &= \text{Var}\left[\frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right] \\ &= \left(\frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)^2 \text{Var}\left[\sum_{i=1}^n (x_i - \bar{x}) Y_i\right] \\ &= \left(\frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)^2 \sum_{i=1}^n \text{Var}[(x_i - \bar{x}) Y_i] \\ &= \dots \end{aligned}$$

# Egenskaper til estimatorene

- ★ Videre skal vi å finne egenskaper til  $\hat{\alpha}$  og  $\hat{\sigma}^2$ :
  - forventingsrett?
  - varians
  - fordeling
- ★ Konfidensintervall og hypotesetest for  $\alpha$  og  $\beta$ 
  - når  $\sigma^2$  er kjent
  - når  $\sigma^2$  er ukjent
- ★ Prediksjonsintervall for ny observasjon  $Y_0$  med  $x = x_0$
- ★ Vurdere om antatt modell er rimelig — residualplott