



# Lineær regresjon

Torstein Fjeldstad

Institutt for matematiske fag, NTNU

01.04.2019

# I dag



- Repetisjon
- Egenskapar til estimatorane (fortsetjing)
- Eksempel

# Eksamen



- Fredag 07.06 kl. 09:00 - 13:00
- 10 deloppgåver der alle tel like mykje (men dei er ikkje like vanskelege)
- Ta med naturlege mellomrekningar. Merk det er skilnad på "skriv opp" og "utlei"

## Tillate hjelpemidler på eksamen



- Tabeller og formler i statistikk
- Bestemd, enkel kalkulator
- Stempla gult A5-ark med egne handskrivne formlar og notat (du kan skrive på begge sider av arket)



# Repetisjon

# Enkel lineær regresjon



Situasjon: har observert par  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ .

- $x_1, x_2, \dots, x_n$  er kjende tal
- $y_1, y_2, \dots, y_n$  er realisasjoner frå uavhengige stokastiske variablar  $Y_1, Y_2, \dots, Y_n$  med

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

# Enkel lineær regresjon



Situasjon: har observert par  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ .

- $x_1, x_2, \dots, x_n$  er kjende tal
- $y_1, y_2, \dots, y_n$  er realisasjoner frå uavhengige stokastiske variablar  $Y_1, Y_2, \dots, Y_n$  med

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

Merk:

$$E(Y_i|x_i) = \alpha + \beta x_i = \mu_i$$

$$\text{Var}(Y_i|x_i) = \sigma^2$$

# Enkel lineær regresjon



Situasjon: har observert par  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ .

- $x_1, x_2, \dots, x_n$  er kjende tal
- $y_1, y_2, \dots, y_n$  er realisasjoner frå uavhengige stokastiske variablar  $Y_1, Y_2, \dots, Y_n$  med

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

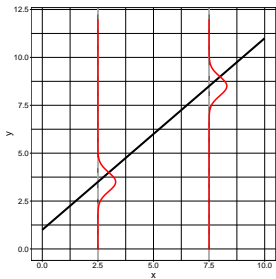
Merk:

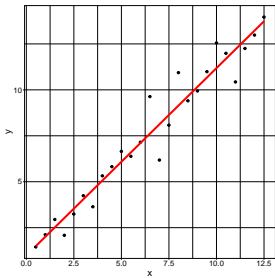
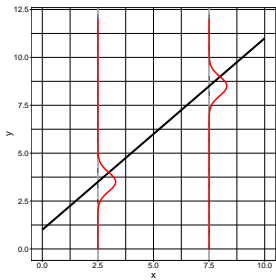
$$E(Y_i|x_i) = \alpha + \beta x_i = \mu_i$$

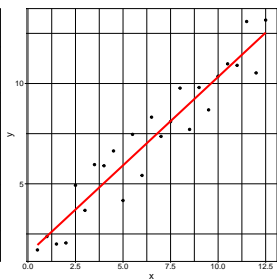
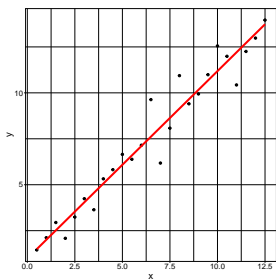
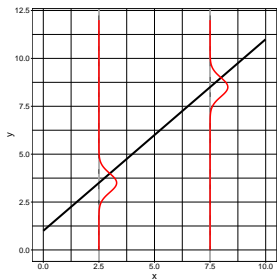
$$\text{Var}(Y_i|x_i) = \sigma^2$$

Mål: estimere  $\alpha, \beta$  og  $\sigma^2$









# SME enkel lineær regresjon I

Maksimer log-likelihoodfunksjon

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



## SME enkel lineær regresjon II

Definer:

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x})y_i$$

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$$

SME

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



Kva er fordelinga til  $\hat{\beta}$ ?

## Egenskapar til estimatorane

$$\hat{\beta} \sim n \left( z; \beta, \sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right)$$



## Egenskapar til estimatorane

$$\hat{\beta} \sim n \left( z; \beta, \sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right)$$
$$\hat{\alpha} \sim n \left( z; \alpha, \sqrt{\frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}} \right)$$



## Egenskaper til estimatorane

$$\hat{\beta} \sim n \left( z; \beta, \sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right)$$
$$\hat{\alpha} \sim n \left( z; \alpha, \sqrt{\frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}} \right)$$

Merk

$$E(\hat{\sigma}^2) = \frac{n-2}{n} \sigma^2,$$

me nyttar derfor

$$S^2 = \frac{n}{n-2} \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$

## Egenskaper til estimatorane

$$\hat{\beta} \sim n \left( z; \beta, \sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right)$$
$$\hat{\alpha} \sim n \left( z; \alpha, \sqrt{\frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}} \right)$$

Merk

$$E(\hat{\sigma}^2) = \frac{n-2}{n} \sigma^2,$$

me nyttar derfor

$$S^2 = \frac{n}{n-2} \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$
$$\frac{(n-2)S^2}{\sigma^2} \sim \chi_{n-2}^2$$

## Konfidenzintervall $\beta$



$$\left[ \hat{\beta} - t_{\alpha/2, n-2} \sqrt{\frac{S^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}, \hat{\beta} + t_{\alpha/2, n-2} \sqrt{\frac{S^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right]$$



# Eksempel (eksamen desember 2012)

## Eksempel

Vinnertid på 800 m løping for menn i OL (siden 1912).

- $Y_i$  er vinnertid i OL nummer  $i$
- $x_i$  er årstal for OL nummer  $i$

for  $i = 1, 2, \dots, 23$



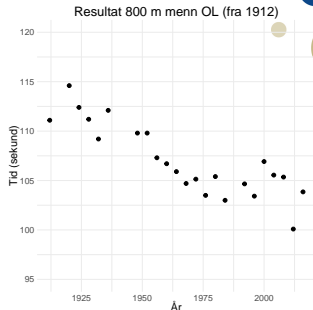
# Eksempel

Vinnertid på 800 m løping for menn i OL (siden 1912).

—  $Y_i$  er vinnertid i OL nummer  $i$

—  $x_i$  er årstal for OL nummer  $i$

for  $i = 1, 2, \dots, 23$



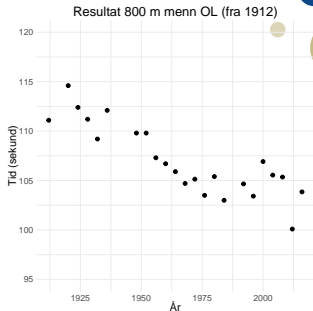
# Eksempel

Vinnertid på 800 m løping for menn i OL (siden 1912).

—  $Y_i$  er vinnertid i OL nummer  $i$

—  $x_i$  er årstal for OL nummer  $i$

for  $i = 1, 2, \dots, 23$



Anta følgende lineære sammenheng

$$Y_i = \alpha + \beta x_i + \varepsilon_i$$

der  $\varepsilon_i \sim n(\varepsilon; 0, \sigma)$  og uavhengige.

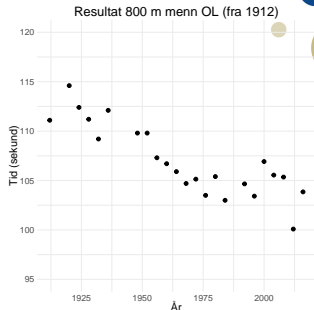
# Eksempel

Vinnertid på 800 m løping for menn i OL (siden 1912).

—  $Y_i$  er vinnertid i OL nummer  $i$

—  $x_i$  er årstal for OL nummer  $i$

for  $i = 1, 2, \dots, 23$



Anta følgende lineære sammenheng

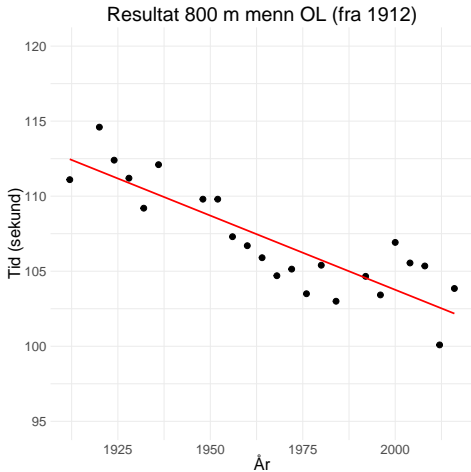
$$Y_i = \alpha + \beta x_i + \varepsilon_i$$

der  $\varepsilon_i \sim n(\varepsilon; 0, \sigma)$  og uavhengige.

Utlei eit  $(1 - \alpha) \cdot 100$  % konfidensintervall for  $\beta$



# Resultat eksempel



# Torsdag



- Prediksjonsintervall
- Diskusjon av modellantakingar