NTNU

Kunnskap for ei betre verd

Lineær regresjon

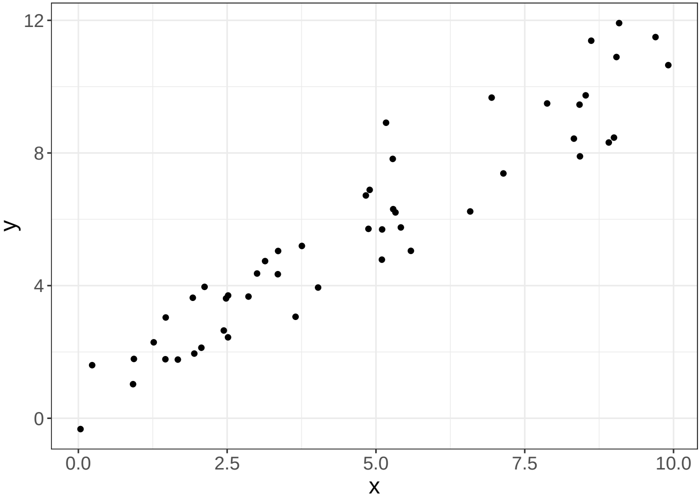
Torstein Fjeldstad

Institutt for matematiske fag, NTNU

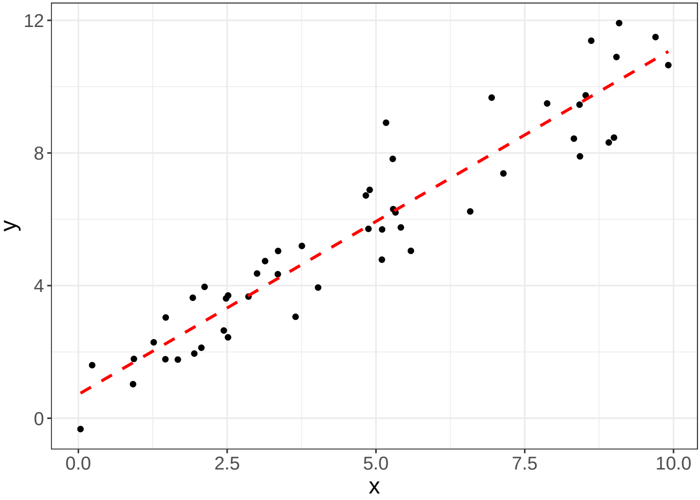
28.03.2019

I dag

- Eksempel
- Eigenskapar til estimatorane (lineær regresjon)



12
8
4
0
y
0.0
2.5
5.0
7.5
10.0
x



12
8
4
0
y
0.0
2.5
5.0
7.5
10.0
x

Enkel lineær regresjon

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

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

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

Enkel lineær regresjon

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

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

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

Merk:

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

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

Enkel lineær regresjon

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

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

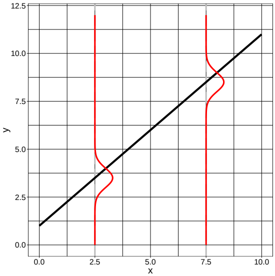
$$Y_i | x_i \sim n(y_i; \alpha + \beta x_i, \sigma)$$

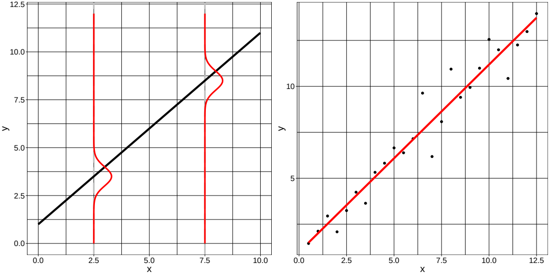
Merk:

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

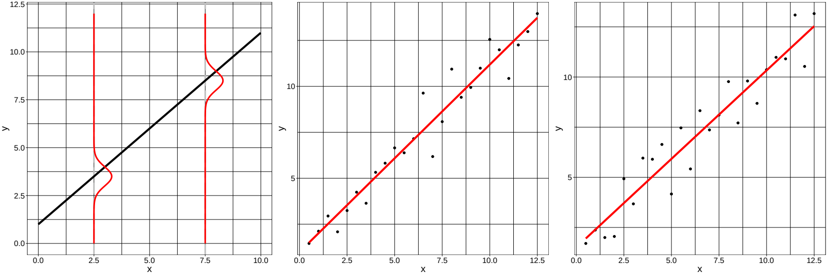
$$Var(Y_i | x_i) = \sigma^2$$

Mål: estimere α, β og σ^2


12.5
10.0
7.5
5.0
2.5
0.0
y
0.0
2.5
5.0
7.5
10.0
x



y
x
0.0
2.5
5.0
7.5
10.0
12.5
5
10



12.5
10.0
7.5
5.0
2.5
0.0
y
0.0
2.5
5.0
7.5
10.0
x
10
5
y
0.0
2.5
5.0
7.5
10.0
12.5
x
10
5
y
0.0
2.5
5.0
7.5
10.0
12.5
x

SME enkel lineær regresjon I

Maksimer log-rimelighetsfunksjon

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

Får følgjande likningssystem (deriver mhp α, β og σ^2):

$$n\alpha + \beta\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{1}$$

$$\alpha\sum_{i=1}^{n} x_i + \beta\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \tag{2}$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2 = n \tag{3}$$

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

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

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

Eigenskapar til estimatorane

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

Torsdag

- Prediksjonsintervall
- Diskusjon av modellantakingar