I dag

Bruker ein enkel liner regresjonsmodell til å svare på:

- Er det ein samanheng mellom røyking og hjarteinfart? (Kap 11.5)
- Kor høg kjem Torgeir, som har foreldre med gjennomsnitthøgde 182.5, til å bli? (Kap 11.6)

I dag

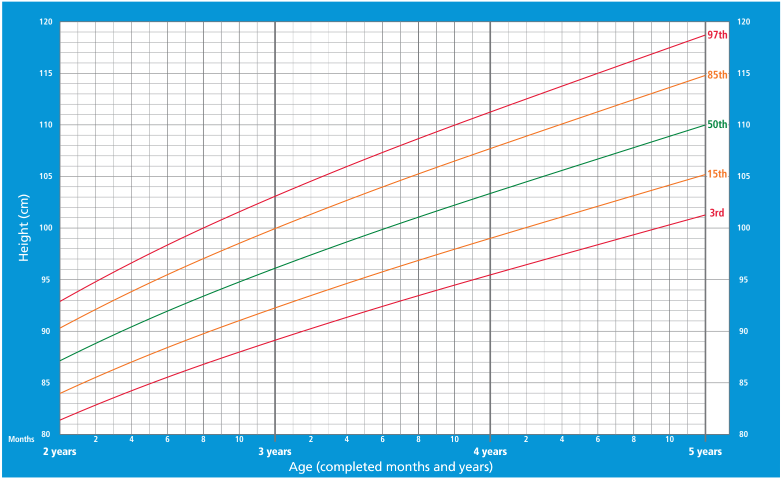
Bruker ein enkel liner regresjonsmodell til å svare på:

- Er det ein samanheng mellom røyking og hjarteinfart? (Kap 11.5) Hypotesetest
- Kor høg kjem Torgeir, som har foreldre med gjennomsnitthøgde 182.5, til å bli? (Kap 11.6) Prediksjon

Height-for-age BOYS

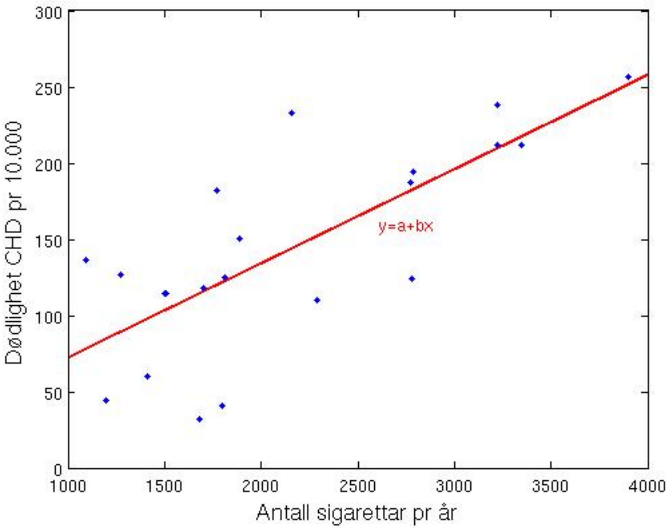
2 to 5 years (percentiles)

World Health Organization



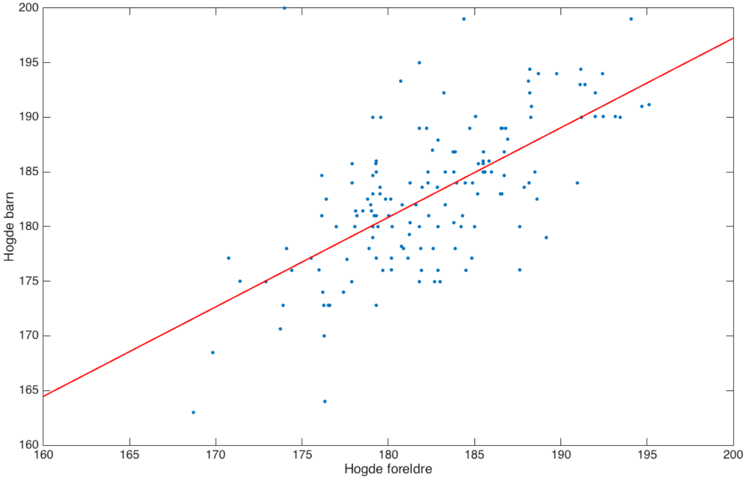
WHO Child Growth Standards

Estimat: $a = 11.41$ og $b = 0.0616$



Høgde foreldre-barn, våre data

Estimat: $a = 33.4$, $b = 0.82$



Enkel lineær regresjon

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

- Y: Respons (stok. var)
- x_i: Forklaringsvariabel (kjent, tal)
- α og β: Regresjonsparameter (param, tal, ukjent)
- ϵ_i: Tilfeldig støy ('feilen', stok.var)
  - $E(\epsilon_i) = 0$, $Var(\epsilon) = \sigma^2_\epsilon$
  - σ^2_ϵ (param., tal, ukjent)

Dersom $\epsilon \sim N(0, \sigma^2)$

$$Y_i \sim N(\alpha + \beta x_i, \sigma^2)$$

Typisk i lineær regresjon

Vi kjenner typisk ikkje parametra α, β og σ_ϵ^2.

1. Skaffar data
2. Estimerer α, β og σ_ϵ^2.

Vi gjer:

- Hypotesetest for α, β eller σ_ϵ^2.
- Prediksjon når α, β og σ_ϵ^2 er estimert.

Regresjonsparametra

Må estimere α, β, σ^2_ϵ frå data.
Data: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. Antar uavh.

Parametre: α, β, σ^2
Estimatorar: A, B, S^2_ϵ
Estimat: a, b, s^2_ϵ

Minste kvardraters metode:

Finn a og b slik at kvadratfeilen, SSE, blir minst mogeleg

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y - (a + bx_i))^2$$

Minste kvadraters estimatorar

- $A = \bar{Y} - B\bar{x}$
- $B = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$

Enkel lineær regresjon

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

Antar $\epsilon_i \overset{u.i.f.}{\sim} N(0, \sigma^2)$

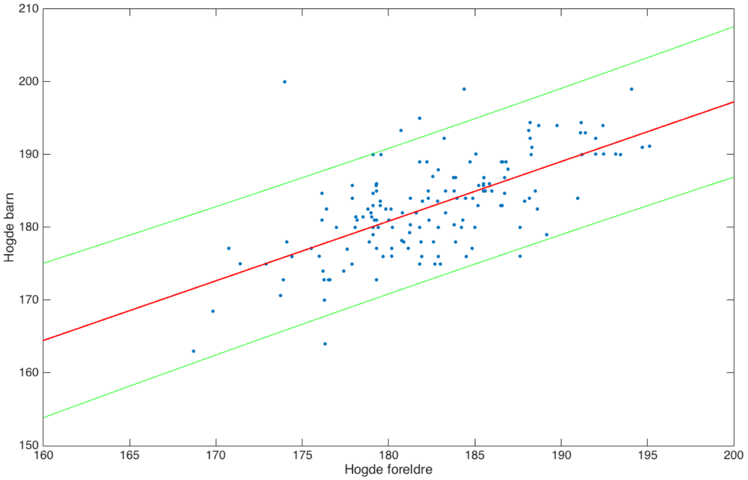
$$Y_i \sim N(\alpha + \beta x_i, \sigma^2)$$

Estimatorar

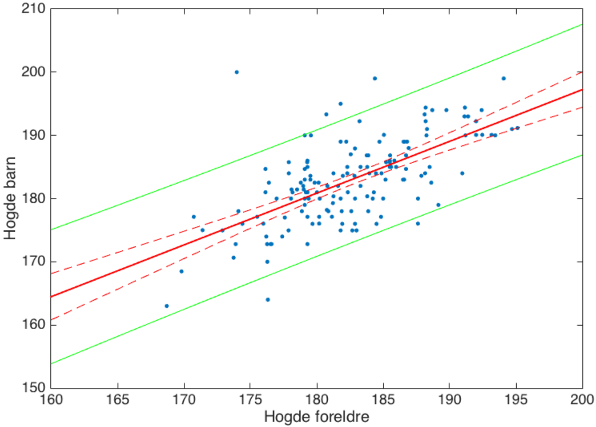
- $A \sim N(\alpha, \sigma_A^2)$
  - der $\sigma_A^2 = \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \sigma_\epsilon^2$
- $B \sim N(\beta, \sigma_B^2)$
  - der $\sigma_B^2 = Var(B) = \frac{\sigma_\epsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma_\epsilon^2}{S_{xx}}$
- $S_\epsilon^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2}$
  - der $\hat{Y}_i = A + Bx_i$

Høgde foreldre-barn, våre data, med 95%prediksjonsintervall

Estimat: $a = 33.4,\ b = 0.82,\ s_\epsilon^2 = 5.08^2$



Høgde foreldre-barn



Regresjonslinja med konfidensintervall (raud stripla) og prediksjonsintervall (grøn).