

Statistikk

Jo Eidsvik

Matematiske fag, NTNU

Regresjon

Data kommer i form av kjente kovariater (eller forklaringsvariable) og målinger eller responsvariable.

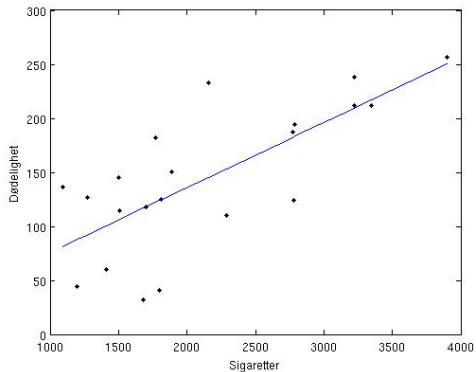
- ▶ Forklaringsvariable: x_1, \dots, x_n
- ▶ Responsvariable: Y_1, \dots, Y_n .

Modell for linear regresjon

For $i = 1, \dots, n$:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \text{ uavhengige} \quad (1)$$

Dødelighet av hjerte og karsykdommer (Y) mot sigaretter (x)



Mål er parameterestimering og prediksjon

1. **Parameterestimering:** Finn $\hat{\beta}_0$ og $\hat{\beta}_1$ utfra data: responser og kovariater.
2. **Prediksjon:** Finn $E(Y_0)$ og $\text{Var}(Y_0)$ der Y_0 er en ny måling med kovariat x_0 .

Lag konfidensintervall for β_0 og β_1 , lag prediksjonsintervall for Y_0 , og gjennomfør hypotesetester.

Parameterestimering

Minste kvadratsums metode finner linja ($\hat{Y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$) som minimerer kvadratiske (vertikale) avvik til data.

$$SSE(\beta_0, \beta_1) = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2 \quad (2)$$

SSE sigarett eksempel

$\hat{\beta}_0 = 15.8$, $\hat{\beta}_1 = 0.06$. (Blå linje.)

