NTNU
Kunnskap for ei betre verd

Lineær regresjon

Torstein Fjeldstad
Institutt for matematiske fag, NTNU
15.11.2018

I dag

- Informasjon
- Repetisjon
- Eksempel OL

Eksamen

- Onsdag 28.11 kl. 09:00 - 13:00
- Alle deloppgåver 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 eigne handskrivne formlar og notat (du kan skrive på begge sider av arket)

Korleis førebu seg til eksamen?

— Kræsjkurs: onsdag 21.11 klokka 16:15 i F1

— Eksamenslab:

- Måndag 26.11 klokka 11:00 - 14:00 i S3
- Tysdag 27.11 klokka 11:00 - 14:00 i S3

— Gamle eksamensoppgåver er god trening

Naturlege neste kurs i statistikk

1. TMA4255 Anvendt statistikk (V2019)
   - statistisk inferens
   - multippel lineær regresjon, forsøksplanlegging
2. TMA4267 Lineære statistiske modeller (V2019)
   - Som TMA4255, men noko meir matematisk
3. TMA4268 Statistisk læring (V2019)
   - Regresjon, ikkje-linearitet, klassifikasjon
   - Fokus på programmering i R
4. TM4265 Stokastisk modellering (H2019)
   - sannsynsrekning
   - rekne på prosesser som utviklar seg i tid

Repetisjon

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

Eigenskapar til estimatorane

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

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

Merk

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

me nyttar derfor

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

$$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$$

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

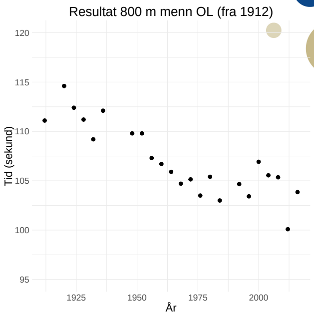
for $i = 1, 2, \ldots, 23$

Eksempel

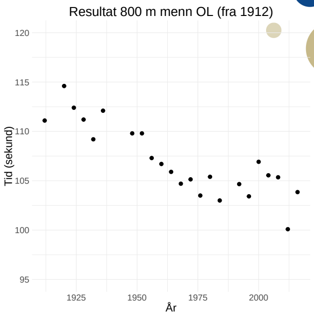
Vinnertid på 800 m løping for menn i OL (siden 1912).

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

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



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 $i = 1, 2, \ldots, 23$

Anta følgjande lineære samanheng

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

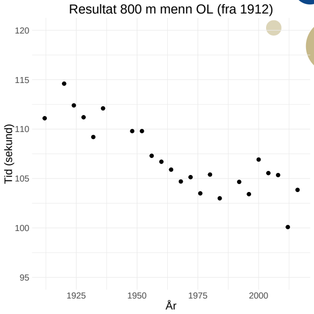
der $\varepsilon_i \sim n(\epsilon; 0, \sigma)$ og uavhengige.

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 $i = 1, 2, \ldots, 23$



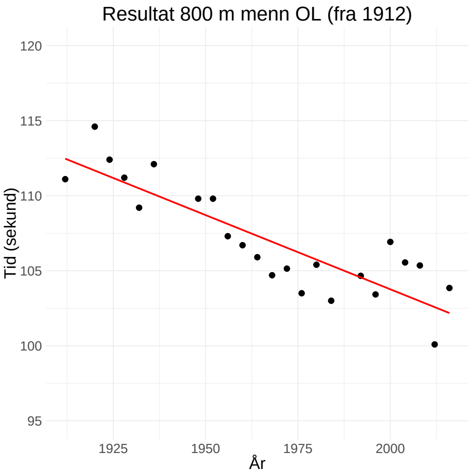
Anta følgjande lineære samanheng

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

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

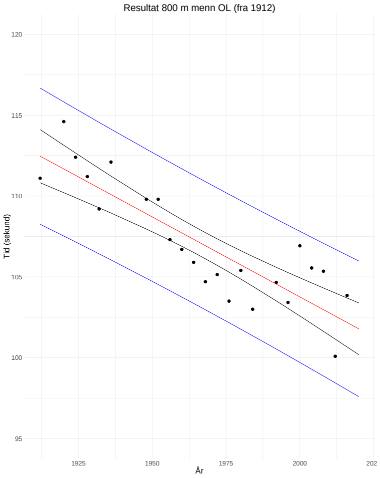
Utlei eit $(1 - \alpha) \cdot 100$ % konfidensintervall for β

Resultat eksempel



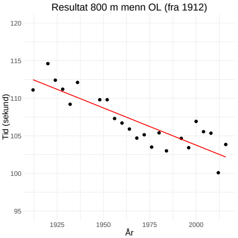
To typer intervall

- Konfidensintervall for $\mu_{Y|x_0}$ (regresjonslinja)
- Prediksjonsintervall for ein ny observasjon Y_0


Resultat 800 m menn OL (fra 1912)
Tid (sekund)
120
115
110
105
100
95
1925
1950
1975
2000
202
År

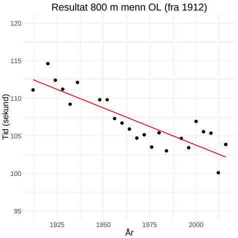
Undersøke modellantakingar

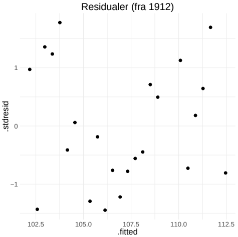
Frå 1912



Undersøke modellantakingar

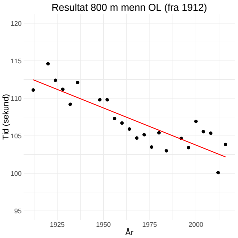
Frå 1912

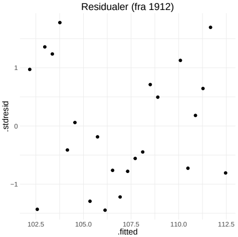




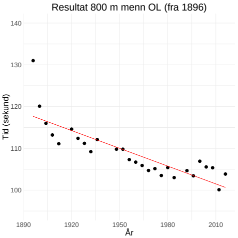
Undersøke modellantakingar

Frå 1912





Frå 1896



Neste veke

- Repetisjon