

Residualer og residualplott

TMA4240/TMA4245 Statistikk

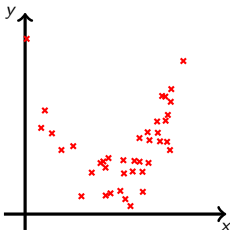
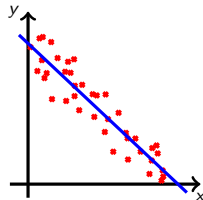
Håkon Tjelmeland

Institutt for matematiske fag

Norges teknisk-naturvitenskapelige universitet

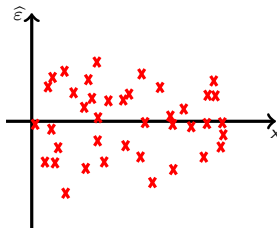
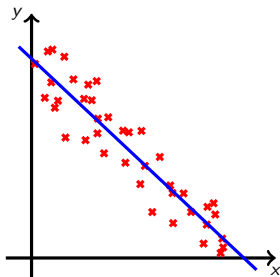
Enkel lineær regresjon

- ★ Situasjon: Har observasjonspaar $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$
- ★ Tilpasser en rett linje til de observerte parene
- ★ Sannsynlighetsmodell: $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$,
 $Y_1, Y_2, \dots, Y_n$ uavhengige
- ★ Husk:
 - α og β er ukjente parametre (tall)
 - betrakter Y_i -ene som stokastiske
 - betrakter x_i som tall (altså ikke stokastiske variabler)
- ★ Har funnet estimatorer for α og β (og for σ^2)
- ★ Fokus i denne videoen:
 - er sannsynlighetsmodellen vi har satt opp fornuftig når vi ser på dataene?
- ★ Eksempel hvor sannsynlighetsmodellen vi har satt opp åpenbart ikke er fornuftig



Residualer og residualplott

- ★ Sannsynlighetsmodell: $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$, $Y_1, Y_2, \dots, Y_n$ uavhengige
- ★ Alternativ formulering: $Y_i = \alpha + \beta x_i + \varepsilon_i$, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ uavhengige og $\varepsilon_i \sim N(0, \sigma^2)$
- ★ Residualer: $\varepsilon_i = y_i - (\alpha + \beta x_i)$, $i = 1, 2, \dots, n$
- ★ (Estimerte) residualer: $\hat{\varepsilon}_i = y_i - (\hat{\alpha} + \hat{\beta} x_i)$, $i = 1, 2, \dots, n$
- ★ Kan vurdere sannsynlighetsmodellen ved å «se på» de (estimerte) residualene
 - ★ ser de ut til å være fra en normalfordeling med forventning lik null?
 - ★ er variansen den samme for alle residualene? Eller varierer variansen med x_i ?

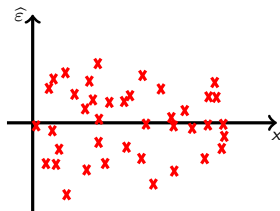


- ★ Er residualene normalfordelt? Sjekk ved å lage normalfordelingsplott for $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n$
- ★ Hvordan lære hvordan residualplott ser ut når modellen ikke passer?
 - simuler fra en modell som avviker fra den vi har antatt, estimer og lag residualplott

Oppsummering

- ★ I en enkel lineær regresjonsmodell har vi definert

- residualer
- residualplott



- ★ Har diskutert

- hvordan vi kan bruke residualplott til å vurdere om antatt modell passer til dataene
- kan bruke normalfordelingsplott til å vurdere om residualene er normalfordelt
- kan bruke stokastisk simulering til å lære hvordan det ser ut når modellen ikke passer