

Plenumsregning – Uke 14

Geir-Arne Fuglstad, IMF

Vår 2020 – Oppgave 6

For n tilfeldig valgte personer modellerer vi sammenhengen mellom fettprosent og BMI (body mass index) ved

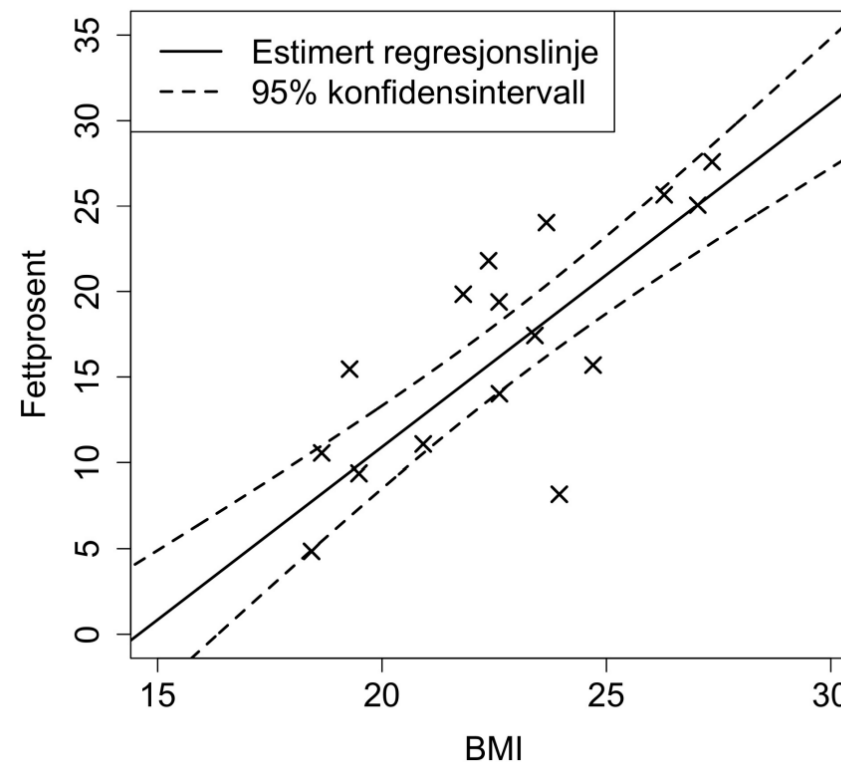
$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \dots, n,$$

der Y_i og x_i er henholdsvis fettprosent og BMI for person i , β_0 og β_1 er regresjonskoeffisienter, og residualene $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ er identiske og uavhengige normalfordelte med forventningsverdi 0 og varians σ^2 .

Vi måler 20 personer der kryssene viser målingene, den heltrukne linjen viser regresjonslinjen, og de stiplede linjene viser nedre og øvre grenser for punktvis 95% konfidensintervaller for regresjonslinjen

a)

Diskuter om antagelsene i den lineære regresjonsmodellen virker å være tilfredstilt for målingene vist i figuren.

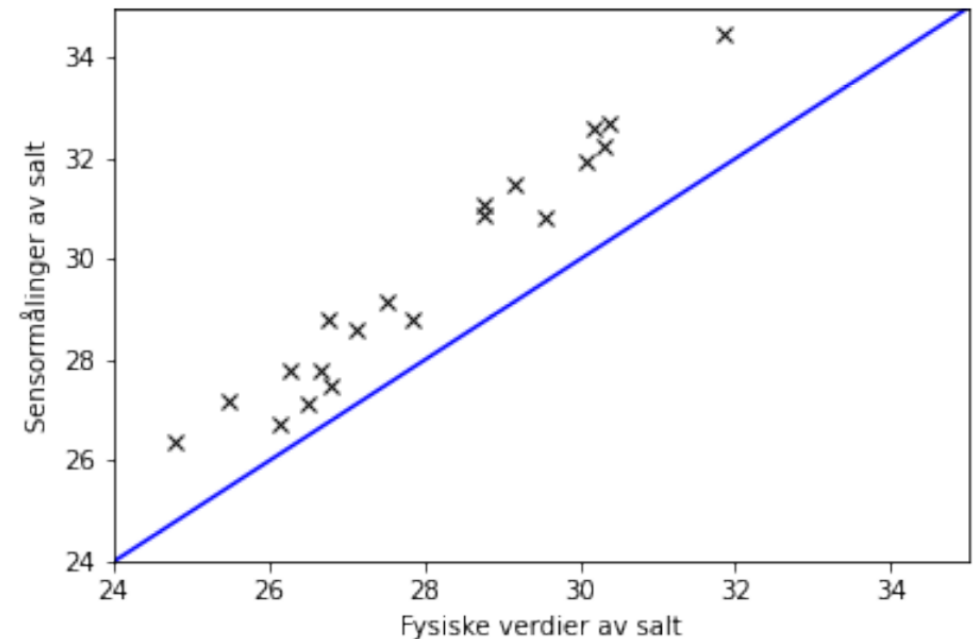


Høst 2024 – Oppgave 4 (forkortet tekst og oppgave)

Vi har sensormåling (y_i) og resultat fra fysisk modell (x_i) for saltinnhold ved lokasjoner $i = 1, \dots, n$ i Trondheimsfjorden. Regresjonsmodellen $Y_i = \alpha + \beta x_i + \epsilon_i$, $i = 1, \dots, n$, er foreslått der vi antar $\epsilon_1, \dots, \epsilon_n \sim_{\text{iid}} N(0, \sigma^2)$. De $n = 20$ målingene i figuren gir $\bar{x} = 28.0$, $\bar{y} = 29.6$, $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = 64.6$ og $S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 77.2$.

a)

- (Forrige uke: Regn ut estimatene for α og β)
- Utfra residualene estimeres σ^2 med $s^2 = 0.48^2$.
Regn ut et 90 % konfidensintervall for β .

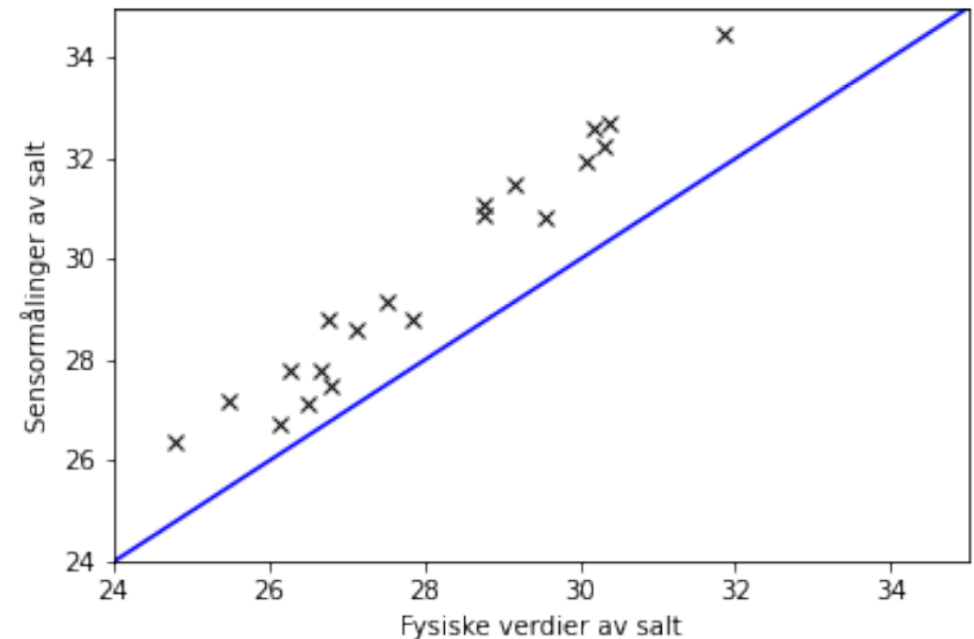


Høst 2024 – Oppgave 4 (forkortet tekst og oppgave)

Vi har sensormåling (y_i) og resultat fra fysisk modell (x_i) for saltinnhold ved lokasjoner $i = 1, \dots, n$ i Trondheimsfjorden. Regresjonsmodellen $Y_i = \alpha + \beta x_i + \epsilon_i$, $i = 1, \dots, n$, er foreslått der vi antar $\epsilon_1, \dots, \epsilon_n \sim_{\text{iid}} N(0, \sigma^2)$. De $n = 20$ målingene i figuren gir $\bar{x} = 28.0$, $\bar{y} = 29.6$, $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = 64.6$ og $S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 77.2$.

b)

- Et sted gir den fysiske modellen $x_0 = 30$. Finn predikert sensormåling der.
- For hvilket resultat i den fysiske modellen vil usikkerheten i predikert sensormåling være minst?
- Uten å regne, hva kan du si om vi i stedet skulle predikert en sensormåling for $x_0 = 20$?



Vår 2022 – Oppgave 6

En eiendomsmegler ønsker å undersøke hvordan boligpris Y_i (målt i kr) til ulike solgte eiendommer avhanger av antall kvadratmeter boligareal x_i (målt i m^2). Han bruker en regresjonsmodell hvor han antar

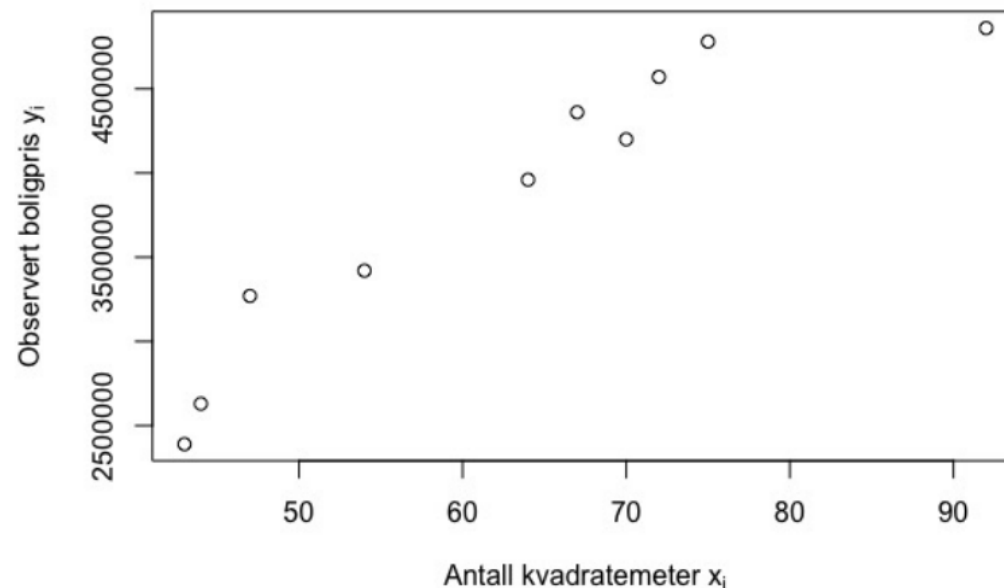
$$Y_i = \beta x_i + \epsilon_i, \quad i = 1, \dots, n,$$

der $\epsilon_1, \dots, \epsilon_n \sim_{\text{iid}} N(0, \sigma^2)$. Stigningstallet β kan da tolkes som det boligkjøpere må forvente å betale per kvadratmeter. Eiendomsmegleren har observasjoner på $n = 10$ solgte boliger i Trondheim. Vi skal anta kjent $\sigma = 300000$ (målt i kr). Det oppgis videre at $\sum_{i=1}^n x_i y_i = 2.53 \cdot 10^9$ og $\sum_{i=1}^n x_i^2 = 41668$.

Anta kjent at SME er $\hat{B} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$.

c)

Eiendomsmegleren ønsker å undersøke om dataene gir grunnlag for å påstå at kvadratmeterprisen i det aktuelle området er større enn 60000 kr. Formuler dette som en hypotesetest og foreslå en passende testobservator. Bestem testen forkastningsområde. Hva blir testens konklusjon for signifikansnivå $\alpha = 0.05$. Finn også p-verdi for testen.



Vår 2022 – Oppgave 6

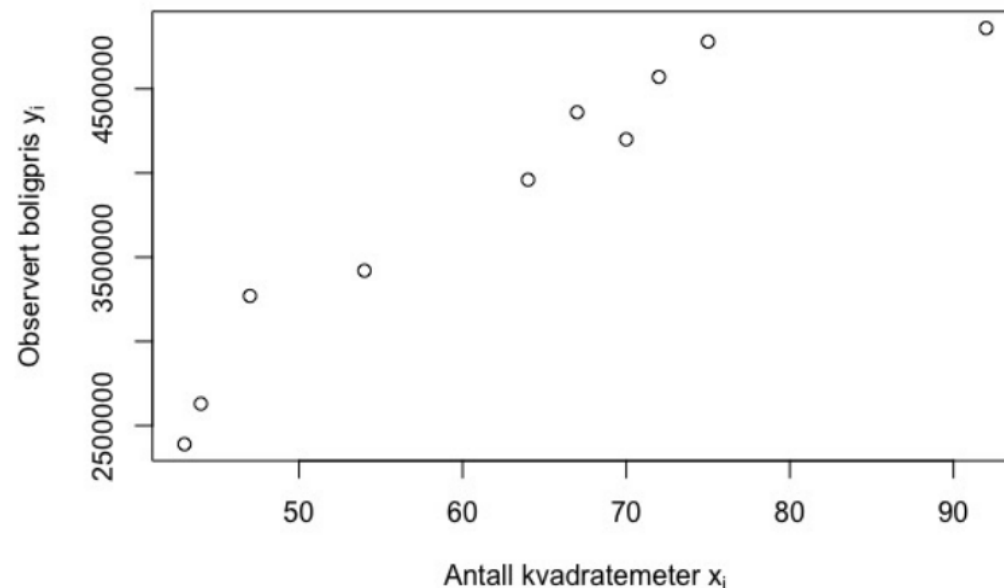
En eiendomsmegler ønsker å undersøke hvordan boligpris Y_i (målt i kr) til ulike solgte eiendommer avhenger av antall kvadratmeter boligareal x_i (målt i m^2). Han bruker en regresjonsmodell hvor han antar

$$Y_i = \beta x_i + \epsilon_i, \quad i = 1, \dots, n,$$

der $\epsilon_1, \dots, \epsilon_n \sim_{\text{iid}} N(0, \sigma^2)$. Stigningstallet β kan da tolkes som det boligkjøpere må forvente å betale per kvadratmeter. Eiendomsmegleren har observasjoner på $n = 10$ solgte boliger i Trondheim. Vi skal anta kjent $\sigma = 300000$ (målt i kr). Det oppgis videre at $\sum_{i=1}^n x_i y_i = 2.53 \cdot 10^9$ og $\sum_{i=1}^n x_i^2 = 41668$.

Anta kjent at SME er $\hat{B} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$.

d)
Finn teststyrken til hypotesetesten i punkt c) gitt at den faktiske kvadratmeterprisen er 65000 kr.



Diskusjon – hvis tid

Anta enkel lineær regresjon

$$Y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \dots, n,$$

med vanlige antagelser der n er antall observasjoner. Vi ønsker å gjøre en hypotesetest $H_0: \beta = 0$ mot $H_1: \beta \neq 0$. La $\hat{A}$ og $\hat{B}$ være standard estimatorer for α og β .

Hvorfor stemmer det ikke at

$$T_1 = \frac{\hat{B}}{S_1 / \sqrt{S_{xx}}} \sim t_{n-1},$$

der $S_1^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$ under H_0 ?

NB: Det vanlige er å bruke

$$T_2 = \frac{\hat{B}}{S_2 / \sqrt{S_{xx}}} \sim t_{n-2},$$

der $S_2^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{A} - \hat{B}x_i)^2$ under H_0 .