

Statistikk-quiz Digital Samling 1.

Vi tar et tilfeldig utvalg av 6 studenter i dette kurset. Vi måler studentenes motivasjon (skala 1-10) for å lære statistikk, og teller antall samlinger de kommer på, og noterer eksamensscore. Vi noterer også om studenten har hatt minst to statistikkurs før dette kurset (MinstTo)

ID	motivasjon	oppmøte	eksamensscore	MinstTo
1	9	12	15.00	ja
2	4	6	7.00	nei
3	7	9	12.50	ja
4	1	4	3.00	nei
5	2	5	4.50	nei
6	7	8	11.00	ja

1. Ser det ut til å være en sammenheng mellom motivasjon og oppmøte? Er den negativ eller positiv?

2. Lag et spredningsplott med motivasjon på x akse og oppmøte på y akse.

```
motivasjon <- c(9, 4, 7, 1, 2, 7)
oppmote <- c(12, 6, 9, 4, 5, 8)
plot(motivasjon, oppmote)
```

3. Hva er standardavviket til oppmøte?

4. Du utfører en enkel regresjon med *oppmøte* som den avhengige variabelen og *motivasjon* som den uavhengige variabelen. Hva blir intercept og hva blir slope?

```
mymodel <- lm(oppmote ~ motivasjon)
```

5. Tolk slopen β_1 i forrige modell.

6. Hva blir residualet for studenten med ID=1?

```
summary(mymodel)
```

7. Hvor mye av variasjonen i oppmøte kan forklares ved varierende motivasjon?

8. Dummy-koding. Hva er effekten av *MinstTo* på motivasjon?

```
lm(motivasjon ~ MinstTo)
```

9. Enkel regresjon: Hva er effekt av oppmøte på eksamensscore?
10. Bruk modellen i forrige oppgave til å forutse hva en elev med motivasjon=5 vil få på eksamen

```
predict(model, newdata=data.frame(motivasjon=5))
```
11. Hva betyr det latinske uttrykket *ceteris paribus*?
12. Multippel regresjon. Kjør en regresjon der eksamensscore forklares av oppmøte OG motivasjon.
13. Gi en tolkning av koeffisienten for oppmøte i modellen i den forrige oppgaven.
14. Gi en tolkning av koeffisienten for motivasjon i modellen i den forrige oppgaven.
15. Hvor mye av variasjon i eksamensscore forklares av varierende motivasjon og oppmøte?
16. Nevn to ting som må være oppfylt dersom vi kan stole på regresjonsmodellen?
17. Du trekker et tilfeldig utvalg med $n = 16$ studenter fra samme klasse. Tror du p-verdiene til regresjonskoeffisienten da blir større eller mindre enn de er i vårt $n = 6$ utvalg? Eller like store omtrent?