

Opgave 1: Vital Capacity

I filen '`cadmium.txt`' ligger observationer fra et eksempel omhandlende lungefunktionen hos arbejdere i cadmium industrien (hentet fra P. Armitage & G. Berry: Statistical methods in medical research. 2nd ed. Blackwell, 1987).

Datsættet består af sammenhørende værdier af alder og vital capacity (liter) for 3 grupper af personer, fordelt således:

- Gruppe 1: Eksponeret for cadmium i mere end 10 år ($n = 12$)
- Gruppe 2: Eksponeret for cadmium i mindre end 10 år ($n = 28$)
- Gruppe 3: Ikke eksponeret for cadmium ($n = 44$)

Den første linie i data indeholder variabelnavnene `group`, `age` og `vitcap`.

1. Konstruer en faktor (klassevariabel) med beskrivende navne til de 3 grupper, f.eks. `expo>10`, `expo<10` og `no-expo`.
2. Beskriv fordelingen af vital capacity i de 3 grupper ved hjælp af passende valgte summary statistics. Lav også passende plots. Ser der umiddelbart ud til at være forskel på grupperne?
3. Ignorer i første omgang `age`-variablen. Er der forskel på vital capacity i de 3 grupper? Giv et estimat for forskellen i vital capacity på de langtidseksponerede og de ueksponerede, naturligvis med konfidensgrænser.
4. Udregn korrelationen mellem alder og vital capacity for hver gruppe for sig, samt for datamaterialet som helhed. Hvad kan vi slutte af dette?
5. Foretag for hver af grupperne en lineær regressionsanalyse af vital capacity mod alder. Hvor stærk er sammenhængen i de tre grupper?
6. Giv et estimat for forskellen i vitalkapacitet mellem 40-årige langtidseksponerede og 40-årige ueksponerede.
 - Sammenlign med det i spørgsmål 3 fundne estimat.
 - Hvordan kan man anskueliggøre modellen til grund for dette estimat?
 - Kan sammenhængen mellem alder og vital capacity påvises at være forskellig for de tre grupper?
7. Hvor meget hurtigere taber man 1 liter i vitalkapacitet, når man er langtidseksponeret i forhold til, hvis man var ueksponeret?

Opgave 2: Væksthormon

Filen |juul2.txt| indeholder en variant af Anders Juuls datamateriale om IGF-I (Insulin-like Growth Factor) hos normale mennesker.

Filen indeholder (i den nævnte rækkefølge):

- **age**: – alder i år
- **height**: – højde i cm
- **sex**: – Køn (1/2=M/F)
- **igf1**: – Serum IGF-I
- **tanner**: – Tanner's pubertetsklassifikation (1–5)
- **weight**: – vægt (kg)

1. Lav en lineær regressionsanalyse for præpubertale individer (Tanner stadium 1), for hvert køn for sig, og med logaritmetransformeret **igf1** som outcome, og alderen som forklarende variabel.
Giv en forståelig fortolkning af hældningen i form af den procentuelle øgning af **igf1** på 5 år.
2. Undersøg om regressionslinjerne er ens for de to køn, og om der samlet set er en effekt af alder.
Giv også et estimat for forskellen på drenge og piger i 7-års alderen.
3. Forklar hvorfor en lineær regression af $\log(\mathbf{igf1})$ overfor alder ville være misvisende, hvis man analyserede hele materialet på en gang.
4. Fokuser nu på de postpubertale (alder > 25 år) og estimer det årlige fald i **igf1**. Er der forskel på kønnene?
5. Udvid modellen fra forrige spørgsmål med en ekstra kovariat, idet $BMI = \text{vægt}/\text{højde}^2$ inddrages.
Estimer igen det årlige fald i **igf1**, og kommenter på forskellen til spørgsmål 4.