

Opgave om 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 (12)
- Gruppe 2: Eksponeret for cadmium i mindre end 10 år (28)
- Gruppe 3: Ikke eksponeret for cadmium (44)

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

1. Indlæs data og konstruer en faktor (klassevariabel) med beskrivende navne til de 3 grupper, f.eks. ved (**inden run;**) at skrive

```
if group=1 then fgroup='expo>10';  
if group=2 then fgroup='expo<10';  
if group=3 then fgroup='no-expo';
```

2. Beskriv fordelingen af vital capacity i de 3 grupper ved hjælp af `proc means`. Lav også passende plots.
3. Er der forskel på vital capacity i de 3 grupper?
Giv estimater for forskellene i vital capacity, og suppler med konfidensgrænser for disse forskelle.
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. Kan sammenhængen mellem alder og vital capacity påvises at være forskellig for de tre grupper? Kvantificer forskellene!

vend!

Opgave 2: Vækst

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

Filen indeholder følgende variable:

- `age` – alder i år
- `height` – højde i cm
- `menarche` – er menarchen indtrådt? (1/2=nej/ja)
- `sex` – Køn (1/2=M/F)
- `igf1` – Serum IGF-I
- `tanner` – Tanner's pubertetsklassifikation (1–5)
- `testvol` – testisvolumen
- `weight` – vægt (kg)

Formatet er sådan at der er en “header”-linje med variabelnavne og tallene er adskilt med blanke. Uoplyste er markeret med et punktum.

1. Indlæs data.
2. Lav regressionsanalyser for præpubertale individer (Tanner stadium 1), for hvert køn for sig, med logaritmetransformeret `igf1` som outcome, og alder som forklarende variabel.
3. Undersøg om regressionslinjerne er ens for de to køn, og om der samlet set er en effekt af alder.
4. Gentag spørgsmål 2 og 3 for postpubertale (alder > 25 år).
5. Forklar hvorfor en lineær regression på alder ville være misvisende, hvis man analyserede hele materialet på en gang.
6. Udvid analysen i spørgsmål 4 til en multipel regressionsanalyse, idet $BMI = \text{vægt}/\text{højde}^2$ inddrages.