

Kommentarer til spørgsmål vedr. 4. forelæsning/øvelse

Du skal konstruere nogle grænser, der skal bruges til at diagnosticere personer med for højt eller for lavt kalium. Kalium refererer her til “total body potassium”, justeret for kropsvægt, og enheden er mmol.

1. Hvad har du brug for?

(a) Et stort sample, med lige mange mænd og kvinder.

Delvist korrekt: Et stort sample er jo altid rart, og hvis man interesserer sig for begge køn, er det naturligvis nødvendigt at inkludere begge køn i materialet. Men det er ikke vigtigt, at der er lige mange mænd og kvinder.

(b) Et repræsentativt sample af raske personer

Korrekt: Det er vigtigt, at man samler repræsentativt, og f.eks. ikke blot medtager patienter til udredning af kalium el.lign.

(c) Et sample, hvor kalium-værdierne er normalfordelt

Forkert: Man kan ikke på forhånd bestemme fordelingen af det, man undersøger. Desuden er det i efterfølgende analyser kun *residualerne*, der antages normalfordelt.

2. Det viser sig, at kalium-koncentrationen ændrer sig lineært med alderen. Hvad vil du inkludere i din artikel omkring denne problemstilling?

(a) Korrelationskoefficienten mellem alder og kalium, for det samlede materiale.

Forkert: Korrelationskoefficienten er afhængig af, hvordan aldersfordelingen tilfældigvis blev i materialet - og desuden er den vanskelig at tillægge nogen god fortolkning. Ydermere har man her glemt at opdele i mænd og kvinder.

(b) Korrelationskoefficienten mellem alder og kalium, for kvinder og mænd hver for sig

Forkert: Korrelationskoefficienten er afhængig af, hvordan aldersfordelingen tilfældigvis blev i materialet - og desuden er den vanskelig at tillægge nogen god fortolkning.

(c) Hælningskoefficienten for relationen mellem kalium og alder, for det samlede materiale.

Forkert: I hvert fald kan man først tillade sig dette, hvis man har godtgjort, at der næppe er nogen forskel på kønnene.

- (d) Hædningskoefficienten for relationen mellem kalium og alder, for kvinder og mænd hver for sig.

Korrekt: Disse to størrelser vil naturligt indgå som en del af dokumentationen. Husk konfidensinterval, og enheder (her mmol/år).

- (e) Forskellen mellem hædningskoefficienterne for mænd og kvinder.

Korrekt: Det vil formentlig være af interesse at se, om aldersafhængigheden ser ens ud for de to køn. Næste gang lærer I at sammenligne sådanne to hældninger.

- (f) Forskellen mellem korrelationskoefficienterne for mænd og kvinder.

Forkert: Korrelationskoefficienterne er næppe særligt forståelige, og dermed heller ikke forskellen imellem dem.

3. Hvilke af følgende udsagn er korrekte?

- (a) Prediktionsgrænsernes bredde er stærkt afhængigt af sample size.

Forkert: Det er kun konfidensgrænserne, der ændrer sig på systematisk vis med sample size (divideres ned med $\sqrt{N}$). Prediktionsgrænserne ændrer sig ikke nævneværdigt.

- (b) Prediktionsgrænserne er meget bredere end konfidensgrænserne.

Korrekt: Prediktionsgrænser svarer nemlig til normalområder (SD), hvorimod konfidensgrænser svarer til konfidensintervaller (SEM).

- (c) Korrelationskoefficienten er det samme som regressionskoefficienten.

Forkert: Men de er proportionale, så hvis den ene er 0, er den anden det også. Og de har samme fortegn.

- (d) Test af korrelationskoefficienten=0 giver det samme som test af regressionskoefficient=0.

Korrekt: Hvis man *bare* skal have en P-værdi, er det således ligegyldigt, hvilken af de to, man vælger at fokusere på.

- (e) Regressionskoefficienten er det samme som hældningen.

Korrekt: Ja, men når vi når til flere forklarende variable (multipel regression), kalder vi det ikke hældninger mere, fordi man så ikke mere kan tegne det som linier.

- (f) Intercept og hældning er uafhængige af hinanden.

Forkert: De vil som oftest være negativt korrelerede: Lav hældning vil svare til højt intercept, og omvendt.

4. Vi har nu fundet estimatet for interceptet for mændene til 4667, og det tilsvarende hældningsestimat til -18.33.
- (a) Hvordan skal man fortolke interceptet på 4667?
Det er en (urimelig) ekstrapolation, der angiver det forventede kalium niveau for en nyfødt (alder=0) dreng.
 - (b) Hvordan skal man fortolke hældningen på -18.33?
For hvert år, man bliver ældre, forventes ens kalium niveau at falde med 18.33 mmol.
 - (c) Hvad er den predikterede værdi for en mand på 70 år?
Vi udregner efter liniens ligning: $4667 - 70 \times 18.33 = 3384$ mmol
5. Kvinderne i materialet var i alderen 20-65 år, bortset fra en enkelt kvinde på 82 år. Hvad skal vi gøre med hende?
- (a) Udelade hende.
Korrekt: Da hun ligger så langt væk fra resten af materialet, bør hun udgå, og vi skal så tilsvarende begrænse vores konklusion til kun at gælde i aldersintervallet 20-65 år.
 - (b) Checke om der er tale om en fejl, og hvis der ikke er det, så skal hun med i analysen.
Forkert: Selvfølgelig skal vi checke for fejl, men som argumenteret for ovenfor, skal hun udelades, hvis alderen er korrekt registreret.
 - (c) Checke om hun har stor indflydelse på estimationen, og udelade hende, hvis hun har.
Forkert: Hvis man konsekvent udelader observationer med stor indflydelse på resultatet, får man til sidst noget meget usikkert. Men man kan lave sensitivitetsbetragtninger, dvs. analyser både med og uden en person, for at se, om konklusionerne holder. Lige i dette tilfælde skal personen dog udgå, se ovenfor.