

Kommentarer til spørgsmål vedr. 2. forelæsning/øvelse

Du skal undersøge en evt. forskel i blodtryk hos rygere og ikke-rygere, og du har et sample bestående af 230 rygere og 290 ikke-rygere, hver med et målt systolisk blodtryk.

1. Du finder P-værdien for test af ens middelværdi for de to grupper ved at udføre et
 - (a) Parret T-test
Forkert: Der er ingen parring mellem en ryger og en ikke-ryger
 - (b) Levenes test
Forkert: Levenes test tester identitet af varianser (eller spredninger) i to grupper
 - (c) Uparret T-test
Korrekt: Da vi har ret store grupper (og da blodtrykket sædvanligvis er rimeligt symmetrisk fordelt), er det fornuftigt at benytte et parametrisk test baseret på en normalfordeling, da man så i tilgift kan få et konfidensinterval for forskellen.
 - (d) Mann-Whitney test
Korrekt: Men ikke så informativt, da man ikke får et konfidensinterval for forskellen på de to middelværdier.
2. For at få den fulde information om forskellen i middelværdi, skal du udregne et 95% konfidensinterval. Hvad vil det egentlig sige?
 - (a) Et interval, der indeholder middelværdien for begge grupper
Forkert: Det er forskellen i middelværdier, der interesserer os, og desuden kan man aldrig lave intervaller, der *med sikkerhed* omslutter nogetsomhelst.
 - (b) Et interval, der med 95% sandsynlighed indeholder middelværdien for begge grupper
Forkert: Det er forskellen i middelværdier, der interesserer os.
 - (c) Et interval, hvori den sande forskel på middelværdierne ligger med 95% sandsynlighed
Korrekt: Og så alligevel ikke helt, fordi den sande forskel er et fast tal, som vi bare ikke kender. Enten ligger det der, eller også gør det ikke, se valgmuligheden nedenfor.

- (d) Et interval, som med 95% sandsynlighed indeholder den sande forskel på middelværdierne

Korrekt: Det er intervallet, som vi konstruerer ud fra vores observationer, som man med rimelighed kan komme med sandsynlighedsudsagn om. Men det gør ikke så meget, hvis man tænker på det som muligheden ovenfor.

3. Er det kritisk, om blodtryksværdierne ser ud til at være normalfordelte? Og hvordan vil du checke denne antagelse?

- (a) Kritisk?

Ja eller nej

Nej, det er ikke kritisk, fordi vi har så mange observationer, og fordi fordelingen af blodtryk sædvanligvis er pænt symmetrisk.

- (b) Check?

- i. Ved at tegne histogram/fraktildiagram over samtlige 520 værdier

Forkert: Hvis der er forskel på middelværdierne, vil vi se en blanding af to fordelinger, måske med to toppe

- ii. Ved at tegne histogram/fraktildiagram for rygere og ikke-rygere hver for sig

Korrekt: men ikke praktisk med mindre grupperne er store, så kan det bedre betale sig at gøre som nedenfor:

- iii. Ved at tegne histogram/fraktildiagram over residualerne fra T-test analysen

Korrekt: Det er helt generelt *residualerne*, man checker fordelingen af, efter at man har trukket middelværdistrukturen ud (her bare trukket det relevante gruppegennemsnit fra observationerne).

Du skal nu opdele ovennævnte data i mænd og kvinder.

1. Du vil gerne vide, om rygning har en forskellig effekt på mænd og kvinder. Hvad vil du gøre?
 - (a) Lave et uparret T-test for hvert køn for sig
Forkert: Ved denne fremgangsmåde kan du se, om rygning har en effekt blandt mænd, og om rygning har en effekt blandt kvinder, men ikke om disse to effekter er forskellige.
 - (b) Udføre en tosidet variansanalyse, med såvel køn som rygning som kovariater/faktorer
Forkert: Når man kun har de to faktorer med som marginale effekter (dvs. uden interaktionsled), svarer det til at *antage*, at effekten af rygning er den samme for mænd og kvinder.
 - (c) Som ovenfor, men yderligere inkludere et interaktionsled
Korrekt: Interaktionsleddet estimerer forskellen i effekt af rygning for de to køn, og tester om denne er forskellig fra 0.
 - (d) Undersøge om lige mange mænd og kvinder ryger
Forkert: Frekvensen af rygning har ikke noget med effekten af rygning at gøre.
2. Hvilke tal ville du inkludere i artiklen som en del af argumentationen for, om der var en sådan forskel på kønnene (altså forskel i effekt af rygning) eller ej?
 - (a) Det gennemsnitlige blodtryk for mænd hhv. kvinder, med konfidensinterval
Forkert: Dette siger intet om effekten af rygning
 - (b) Differensen på blodtrykket blandt rygere og ikke-rygere, med konfidensinterval
Forkert: Her skelnes ikke mellem mænd og kvinder
 - (c) Frekvenserne af rygning for hvert køn
Forkert: Frekvensen af rygning har ikke noget med effekten af
 - (d) Forskellen på de to differenser (effekt af rygning) for de to køn, med konfidensinterval
Korrekt: og dette er estimatet for interaktionsleddet
 - (e) Differensen på blodtrykket blandt rygere og ikke-rygere, med konfidensinterval, opdelt efter køn
Korrekt: men ikke fyldestgørende, fordi det ikke indeholder en direkte sammenligning af kønnene

- (f) Differensen på blodtrykket hos mænd og kvinder, med konfidens-interval

Forkert: Dette siger intet om effekten af rygning