

## Opgaver til Statistisk Dataanalyse 1

### Opgave HS.33 (Kropstemperatur)

Det antages sædvanligvis at kropstemperaturen hos raske mennesker er normalfordelt med middelværdi  $37^{\circ}\text{C}$  og spredning  $0.4^{\circ}\text{C}$ .

1. Bestem et interval som indeholder kropstemperaturen for 95% af populationen af raske mennesker (under ovennævnte antagelse).
2. Beregn sandsynligheden for at et tilfældigt rask menneske har en kropstemperatur på  $37.5^{\circ}\text{C}$  eller derover.
3. Beregn sandsynligheden for at der blandt 10 tilfældige raske personer er præcis to med en kropstemperatur på  $37.5^{\circ}\text{C}$  eller derover.

I et medicinsk studie blev kropstemperaturen målt for 130 raske personer. Blandt disse havde 4 personer en kropstemperatur på  $37.5^{\circ}\text{C}$  eller derover.

4. Beregn et estimat og et konfidensinterval for sandsynligheden for at en tilfældig person (fra populationen) har en kropstemperatur på  $37.5^{\circ}\text{C}$  eller derover.  
Hvordan passer data med antagelsen om at gennemsnitstemperaturen er  $37^{\circ}\text{C}$ ?

Bemærk at første del af spørgsmål 4 kan besvares uafhængigt af spørgsmål 1–3.

### Opgave HS.34 (Fuldmåne og henvendelser til psykiatrisk skadestue)

*Dette er en omskrivning af Februar 2016, opgave 3.*

I folketroen har fuldmånen en ond og mørk kraft der kan styre vores adfærd. Men er der noget om snakken? Datasættet til denne opgave består af det gennemsnitlige antal henvendelser per dag i den psykiatriske skadestue på *Virginia mental health clinic* i dagene før, under og efter de 12 fuldmåner fra august 1971 til juli 1972.

Data er tilgængelig i filerne `patienter.xlsx` og `patienter.txt`. Der er følgende variable: *maaned* (kategorisk variabel med 12 niveauer: *71.Aug, ..., 72.Jul*); *fuldmaane* (kategorisk variabel med 3 niveauer: *Ja, Nej.Efter, Nej.Foer*), og *antal* (antal henvendelser per dag).

- (a) Forklar hvorfor det er naturligt at bruge en tosidet ANOVA med *antal* som respons og *maaned* og *fuldmaane* som forklarende variable, og hvorfor man ikke kan inddrage vekselvirkning i modellen.

Nedenstående spørgsmål skal baseres på den additive (uden vekselvirkning) tosidede ANOVA med *antal* som respons og med *maaned* og *fuldmaane* som forklarende variable.

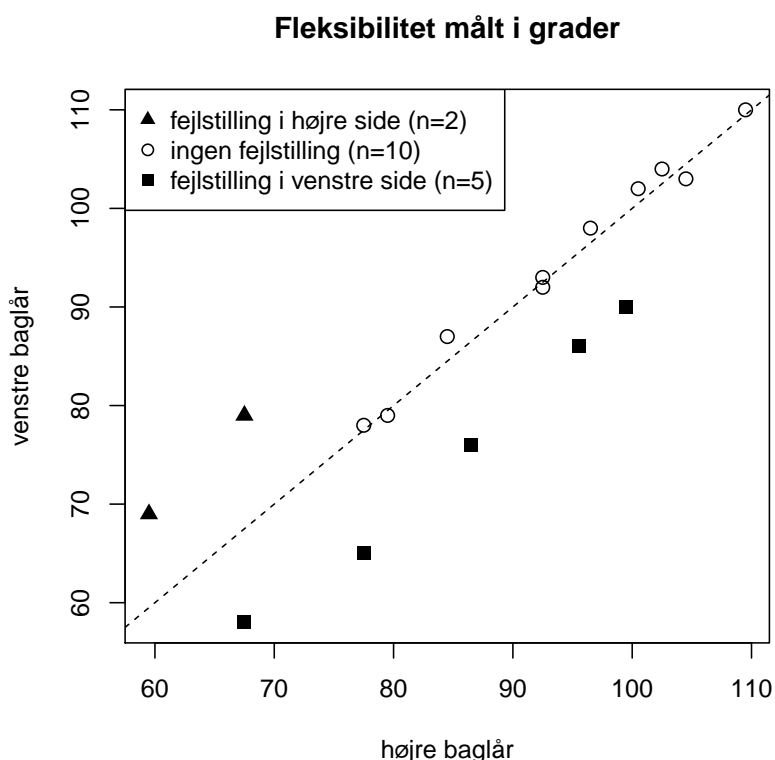
- (b) Fit modellen og udfør modelkontrol. Lav herunder en skitse af de relevante plot og kommentér figureerne.

- (c) Er der en signifikant sammenhæng mellem antallet af henvendelser på den psykiatriske skadestue og om det er fuldmåne? Husk at forklare hvilken hypotese der testes.
- (d) Bestem et 95% konfidensinterval for forskellen mellem antal henvendelser per dag på den psykiatriske skadestue under og før fuldmåne.
- (e) Er der signifikant forskel på antallet af henvendelser på den psykiatriske skadestue før og efter fuldmåne? Husk at forklare hvilken hypotese der testes.

### Opgave HS.35 (Fleksibilitet i baglår hos fodboldspillere)

*Dette er en omskrivning af November 2014, opgave 1.*

Skader af spillere har stor sportslig og økonomisk betydning i moderne topfodbold. For at identificere en mulig årsag til hyppige skader hos nogle fodboldspillere målte en fysioterapeut fleksibiliteten af baglåret på begge ben hos 7 spillere med en såkaldt fejlstilling i bækkenet og hos 10 spillere uden fejlstilling i bækkenet. Fejlstillingen kan være i enten venstre eller højre side af bækkenet, og baglåret siden med fejlstilling formodes at være mindre fleksibel end baglåret i den anden side. Figuren nedenfor viser målingerne af fleksibiliteten i venstre og højre baglår hos de 7 spillere med fejlstilling (heraf havde 5 spillere fejlstilling i venstre side) og hos de 10 spillere uden fejlstilling.



Responsvariablen i denne opgave er *forskellen* mellem fleksibiliteten af venstre og højre baglår.

Data er tilgængelige i filerne `flex.txt` og `flex.xlsx`. Filerne indeholder 17 datalinier (en linie per person) og to variable. Variablen `forskel` er den netop omtalte forskel i fleksibilitet, mens variablen `side` har niveauerne `venstre`, `højre` og `neutral` afhængig af hvilken side personen har fejlstilling.

Det kan være nyttigt at lave vektorer med forskellene fra hver af de tre grupper (men opgaven kan også løses uden disse vektorer). Hvis det indlæste datasæt hedder `flex`, så kan vektorerne laves med kommandoerne nedenfor:

```
forskel.venstre <- subset(flex1, side=="venstre")$forskel  
forskel.neutral <- subset(flex1, side=="neutral")$forskel  
forskel.højre <- subset(flex1, side=="højre")$forskel
```

I hele opgaven kan det antages at forskellen er normalfordelt i alle 3 fejlstillingsgrupper og at variansen er den samme i de tre grupper.

I de første to spørgsmål skal du kun bruge data fra personer uden fejlstilling.

- (a) Lav et 95% konfidensinterval for forskellen mellem fleksibiliteten af venstre og højre baglår i populationen af spillere uden fejlstilling i bækkenet. Kan man antage at venstre og højre baglår er lige fleksible i denne population?

I næste spørgsmål skal du bruge data fra personer uden fejlstilling og personer med fejlstilling på venstre side.

- (b) Er forskellen i fleksibiliteten mellem venstre og højre baglår forskellig i populationen af spillere med fejlstilling i venstre side af bækkenet og i populationen af spillere uden fejlstilling i bækkenet?

I det sidste spørgsmål skal du kun benytte data fra personer med fejlstilling i venstre eller højre side.

- (c) Betragt nu de 7 spillere med fejlstilling i bækkenet og hypotesen om symmetri, dvs. at den forventede forskel i fleksibilitet mellem siden med fejlstilling og siden uden fejlstilling er den samme i de to grupper. Formulér hypotesen i termer af middelværdiparametrene for de tilsvarende populationer, og lav et statistisk test af denne hypotese.