

Opgaver til Statistisk Dataanalyse 1

Opgave HS.21 (Ekstra hypotese i antibio-eksemplet)

Husk datasættet vedr. antibiotikabehandling og nedbrydning af organisk materiale i gødning fra bogen (side 53 og mange andre steder). Data ligger i datasættet **antibio** i *isdals*-pakken.

1. Udfør F -testet for sammenligning af de seks grupper. Testet er rapporteret på side 166, så du kan checke om du får det samme.

Antag at man allerede før eksperimentet blev udført havde en hypotese om at Fenbendazole, Ivermectin og Spiramycin har samme effekt. Dette svarer til hypotesen

$$H_0 : \alpha_{\text{fenbendazole}} = \alpha_{\text{ivermectin}} = \alpha_{\text{spiramycin}}$$

og svarer til at vi kan „slå de tre grupper sammen“.

2. Forklar hvorfor hypotesen kan testes på følgende måde:

- Først laves en ny variabel der inddeler de 34 observationer i fire grupper: Control, α -Cypermethrin, Enrofloxacin, samt en *fælles gruppe* for observationer hørende til Fenbendazole, Ivermectin og Spiramycin. Lad os kalde den nye variabel `type2`.
- Så fittes to modeller med `lm`: En model hvor `type` bruges som forklarende variabel og en model hvor `type2` bruges som forklarende variabel.
- Til sidst sammenlignes de to modeller med funktionen `anova`.

3. Udfør testet, dvs. udfør de tre skridt fra spørgsmål 2. Til første skridt kan følgende kode være nyttig (husk at forstå hvad der foregår):

```
antibio$type2 <- antibio$type
levels(antibio$type2)
levels(antibio$type2)[4:6] <- "Faelles"
levels(antibio$type2)
antibio$type2
```

En alternativ løsning til at ændre gruppenavne eller slå grupper sammen kunne være følgende:

```
library(tidyverse)
antibio$type2 <- recode(antibio$type, "Fenbenda" = "Faelles"
                        , "Ivermect" = "Faelles"
                        , "Spiramyc" = "Faelles")
antibio$type2
```

Hvad er konklusionen på testet?

Opgave HS.22 (Effekt af fenol på relativ bakterieaktivitet)

Dette er en omskrivning af Januar 2014, opgave 2.

Data til denne opgave indeholder 12 målinger af den relative bakterieaktivitet i jordprøver ved forskellige koncentrationer af miljøgiften fenol. Den relative bakterieaktivitet er aktiviteten i forhold til bakterieaktiviteten uden tilsætning af fenol. Der er tre målinger for hver af koncentrationerne 5 mM, 10 mM, 15 mM og 20 mM.

Data er tilgængelig i filerne `fenoldata.txt` og `fenoldata.xlsx`, med variabelnavne `fenol` og `rakt`. Du skal starte med at indlæse data i R.

I den første del af opgaven skal du bruge variabelen `fenol` som en *kvantitativ, kontinuert* variabel.

1. Angiv en statistisk model hvor du bruger `fenol` som en kvantitativ, kontinuert forklarende variabel, og opskriv kort (gerne i punktform) antagelserne bag modellen.
2. Fit modellen i R, og angiv estimater for samtlige parametre i modellen (incl. residualspredningen).
3. Angiv estimatet for det forventede fald i relativ bakteriaktivitet hvis fenolkoncentrationen øges med 1 mM. Bestem også det tilhørende 95% konfidensinterval.
4. Brug modellen til at bestemme et estimat for det forventede fald i relativ bakteriaktivitet hvis fenolkoncentrationen øges fra 5 mM til 10 mM. Bestem også det tilhørende 95% konfidensinterval.

I næste del af opgaven skal du bruge variabelen `fenol` som en *kategorisk* (og nominal) variabel.

5. Angiv en statistisk model hvor du bruger `fenol` som en kategorisk forklarende variabel, og forklar hvordan modellen adskiller sig fra modellen i spørgsmål 1.
6. Fit modellen i R. Angiv estimater for den forventede relative bakterieaktivitet ved 5 mM fenol og ved 20 mM fenol.
7. Brug modellen til at bestemme et estimat og et 95% konfidensinterval for det forventede fald i relativ bakteriaktivitet ved en fenolkoncentration på 10 mM fenol sammenlignet med 5 mM.
8. I den lineære regression var der præcis ét fornuftigt estimat for det forventede fald i relativ bakteriaktivitet når fenolkoncentrationen øges med 1 mM. Forklar hvorfor dette ikke er tilfældet for den nye model.

Til sidst skal du bruge begge modeller:

9. Forklar hvorfor de to modeller er "nestede", dvs. at den ene er et specialtilfælde af den anden. Brug denne egenskab til at lave et hypotesetest der undersøger om sammenhængen mellem `fenol` og relativ bakterieaktivitet er lineær.

Opgave HS.23 (To stikprøver med og uden ens spredning)

Dette er en omskrivning af Case 3, part II fra bogen.

I et forsøg på tidligere Landbohøjskole på Frederiksberg blev 11 lam udvalgt fra en population og delt i to grupper: Fem lam fik standardfoder, mens seks lam fik foder der var tilsat fisk. Efter

slagting og modning blev kødet fra hvert lam kemisk undersøgt for et stof der er relateret til fiskesmag. Man vil gerne vide om fisken fra foderet bliver overført til fiskesmag i kødet, hvilket man selv sagt ikke ønsker.

Data er angivet i tabellen:

Feed			Fish flavor			
Standard	3.81	3.00	3.85	3.30	3.78	
With fish	9.42	3.95	7.23	6.86	6.09	3.99

1. Forklar hvorfor der er tale om to afhængige (eller uparrede) stikprøver i stedet for parrede stikprøver.
2. Indlæs data i R, fx med følgende kommandoer:

```
standard <- c(3.81, 3.00, 3.85, 3.30, 3.78)
fish <- c(9.42, 3.95, 7.23, 6.86, 6.09, 3.99)
```

3. Beregn stikprøvespredningen for hver af stikprøverne.

I en ensidet ANOVA antager vi at spredningerne er ens i grupperne. På side 61 i bogen er der nævnt en tommelfingerregel om at der ikke bør være mere end en faktor 2 til forskel mellem stikprøvespredningerne. Er tommelfingerregelen opfyldt her?

4. Prøv kommandoerne

```
t.test(standard, fish, var.equal=FALSE)
t.test(standard, fish, var.equal=TRUE)
```

og kommenter forskellene. Bemærk specielt antallet af frihedsgrader. Se gerne side 129–131 og 139 i bogen.

Hvad er konklusionen mht. om fiskesmagens afsættes i kødet?

5. Vi kan som bekendt tænke på situationen med to uafhængige stikprøver som en ensidet ANOVA med to grupper og bruge `lm`. Hvordan skal data organiseres for at vi kan gøre dette, og hvilken af de to ovenstående analyser ville dette svare til?
6. Lav to nye variable med følgende kodelinier:

```
y <- c(standard, fish)
y
grp <- factor(c(rep("std", 5), rep("fish", 6)))
grp
```

Fit derefter den ensidede ANOVA med `y` som respons og `grp` som forklarende variabel, og undersøg om der er forskel på de to grupper. Hvilken af ovenstående `t.test`-kommandoer svarer det til?

7. Lav et residualplot for modellen og forklar sammenhængen med det du så i spørgsmål 3.