

## Opgaver til Statistisk Dataanalyse 1

### Opgave HS.18 (Tandsten hos hunde)

Dette er en omskrivning af opgave 3.2 fra bogen, men kunne i øvrigt sagtens have været en eksamensopgave.

Mange hunde har problemer med tandsten, og et eksperiment med hunde blev udført for at undersøge effekten af to behandlinger. 26 hunde indgik i forsøget og blev tilfældigt fordelt i tre grupper svarende til de to behandlinger og en kontrolgruppe.

Efter fire ugers behandling blev hundene undersøgt, og man gav hver hund et index der angiver i hvor høj grad hunden har problemer med tandsten. Et højt index betyder at hunden er hårdt ramt af tandsten.

Data ligger som **tartar** i *isdals*-pakken. Grupperne er navngivet Control, P2O7, HMP.

1. Gør rede for datastrukturen passer med en ensidet ANOVA. Angiv den `lm`-kommando som du vil bruge til at fitte modellen, og kørs kommandoen.
2. Forklar hvilke antagelser der ligger bag modellen, og udfør modelkontrol. Du skal altså lave de relevante figurer og kommentere dem. Du må gerne nøjes med at lave skitser af figurerne hvis du laver din besvarelse i hånden, men de skal naturligvis ligne dem du har lavet i R.

Du skal fortsætte med den ensidede variansanalyse uanset hvad du konkluderede om modelkontrol i spørgsmål 2.

3. Angiv estimatet for den forventede værdi af indexet for en tilfældig hund der ikke er blevet behandlet (dvs. en hund fra kontrolgruppen).  
Angiv også estimatet for den forventede værdi af indexet for en tilfældig hund der er blevet behandlet med P2O7.
4. Undersøg med et enkelt hypotesetest om den forventede værdi er den samme i alle tre grupper.
5. Angiv et estimat og et 95% konfidensinterval for forskellen i forventet værdi mellem en kontrolhund og en hund behandlet med HMP. Er der evidens i data for at HMP-behandlingen virker?

### Opgave HS.19 (Model med proportionalitet)

I et eksperiment har man indsprøjet et stof i blodbanen på 150 mus med en dosis der varierer fra mus til mus. Efter 24 timer har man taget en blodprøve på hver mus og målt koncentrationen af stoffet. Data er tilgængelige i filerne `mus_prop.xlsx` og `mus_prop.txt`. Variablene hedder `dosis0` og `konc24`.

Man forventer at koncentrationen efter 24 timer er *proportional* med den indsprøjtede dosis, men kender ikke proportionalitetsfaktoren.

1. Indlæs data og lav et scatterplot (punktplot). Ovevej hvilken variabel der bør være på  $x$ -aksen hhv.  $y$ -aksen.
2. Antagelsen om proportionalitet er et specialtilfælde af modellen for lineær regression. Hvad er restriktionen i forhold til den sædvanlige lineære regressionsmodel?
3. Modellen med proportionalitet kan fittes med følgende kommando (hvor `mus_prop` er navnet på datasættet):

```
lm(konc24 ~ dosis0-1, data=mus_prop)
```

Kør koden, og angiv estimat, standard error og konfidensinterval for proportionalitetskonstanten.

4. Hvordan kunne du teste teorien om proportionalitet?

### Opgave HS.20 (Forrentningsprocenter)

*Dette er en omskrivning af Marts 2010, opgave 2.*

Over to år har man målt forrentningsprocenterne for 11 tilfældigt udvalgte landbrugsbedrifter på Vestfyn. Alle 11 bedrifter er således målt begge år. Summariske størrelser er givet i tabellen nedenfor.

	Gennemsnit	Stikprøvespredning
År 1	6.182	2.621
År 2	8.664	3.999
Forskel (år 2 minus år 1)	2.482	3.304

1. Hvilken statistisk model kan bruges til at undersøge om der er forskel på den gennemsnitlige forrentningsprocent de to år. Angiv estimater for samtlige parametre i modellen.
2. Bestem et 95% konfidensinterval for forskellen i gennemsnitlig forrentningsprocent mellem de to år.
3. Undersøg med et hypotesetest om der er forskel på den gennemsnitlige forrentningsprocent de to år.
4. Antag at man vil lave en ny dataindsamling fordi man ønsker et 95% konfidensinterval med en samlet bredde på 4 (dvs. svarende til 4 procentpoint). Hvor mange landbrugsbedrifter bør man indsamle data fra?