

## Case 5: Tofaktormodeller med tilfældig(e) effekt(er)

Denne opgave går ud på at analysere datasættet fra kompendiets opgave 7.2 (*Virus in fish*). Det tilhørende pensum er afsnit 7.2, og det er bestemt tilladt at læse dette afsnittet undervejs.

Datasættet kan hentes under ugeplanen for uge 4 på Absalon.

På sidste side er der hjælp til nogle af spørgsmålene, men prøv først hvor langt du kan komme uden at benytte hjælpen!

### Indlæsning af data og overblik over faktorer

1. Læs om datasættet i kompendiets opgave 7.2 (til og med tabellen). Indlæs data og sørg for at R faktisk opfatter alle faktorer som faktorer.
2. Tegn et faktordiagram. Du behøver ikke overveje hvilke faktorer der skal være systematiske hhv. tilfældige endnu.

### Analyse 1: specielt udvalgte vira

I spørgsmål 3–9 skal vi antage at de tre vira er specielt udvalgt med det formål at sammenligne netop disse tre vira.

3. Foreslå en statistisk model og færdiggør faktordiagrammet (med firkantede parenteser om de tilfældige faktorer).
4. Fit modellen med `lme`.

Husk at en vekselvirkning ikke kan skrives ind i random med `:` men skal defineres separat. Husk også syntaksen ved nestede tilfældige faktorer:

```
random = 1 | grovfaktor / finfaktor
```

5. Undersøg om nogle af varianskomponenterne kan sættes til nul.
  - Undersøg først om varianskomponenten hørende til den tilfældige vekselvirkning kan sættes til nul. Her kan du direkte få  $R$  til at beregne en  $F$ -teststørrelsen ved at sammenligne relevante modeller fitted med `lm`.
  - Undersøg dernæst om varianskomponenten hørende til laboratorie kan sættes til nul. Her skal du selv konstruere den relevante  $F$ -teststørrelse ud fra  $R$ -outputtet (-se forelæsningslides!).

*Vink:* Fit den tilsvarende lineære model, hvor alle faktorer (også dem der egentlig er tilfældige) optræder som systematiske, og brug MS'erne fra denne model til at udføre de rette  $F$ -test.

6. Undersøg om der er forskel på de tre vira. Angiv estimater (også for variansparametre) i slutmodellen og konkluder.

## Repeterbarhed og reproducerbarhed

I forsøg af denne type er man ofte interesseret i *repeterbarhed* og *reproducerbarhed*.

7. Læs om repeterbarhed og reproducerbarhed i kompendiet (side 126–127).
8. Estimér repeterbarheden og forklar hvad det betyder.
9. Estimér reproducerbarheden og forklar det betyder.

## Analyse 2: tilfældigt udvalgte vira

I de følgende spørgsmål skal vi antage, at de tre vira er udtaget tilfældigt.

10. Opstil en statistisk model der svarer til denne situation, og fit modellen i R.  
*Vink (til R-delen):* De tilfældige faktorer er ikke længere nestede, så random-delen af modellen skal specificeres vha. `lmer` som i R-notatet til kapitel 5.2.2. Konstant middelværdi kan angives med  $y \sim 1$ .
11. Estimér varianskomponenterne. Beregn også et estimat for spredningen på observationerne,  $Y_i$  (beregnet først variansen).
12. Angiv et estimat og et konfidensinterval for den forventede “maximal titration with CPE”.

## Hjælp

Nedenfor følger vink til nogle af spørgsmålene. Prøv dog først at løse spørgsmålene uden af gøre brug af vinkene.

3. Husk at hvis en faktor er tilfældig, så skal alle finere faktorer også være tilfældige.
4. Definer fx. `labvirus = virus:lab` og skriv `random =~ 1 | lab / labvirus` i `lme`-kaldet.
5. Hvis den lineære model hedder `modelA.lm`, så brug `anova(modelA.lm)` til at få beregnet MS'erne. Brug faktordiagrammet til at se hvilken faktor den tilfældige faktor skal testes mod, og brug den rette MS'er i nævneren. Frihedsgraderne for  $F$ -testet kan enten aflæses fra `anova(modelA.lm)` eller fra faktordiagrammet.
6. Estimerer mm. skal beregnes i `lme`-modellen. Til konklusion kan fx. man beregne LSD-værdi eller konfidensintervaller.
11. Hvad er variansen på  $Y_i$  (som funktion af variansparametrene)? Indsæt estimerterne.
12. For `lmer` kan man ikke konstruere konfidensintervaller vha. `intervals`. Modellen har dog kun en systematisk parameter (intercept), der estimeres som gennemsnittet af alle observationerne. Brug spredningsestimatet fra R-outputtet til at konstruere et konfidens-interval for estimatet. Husk at responsen er log-transformeret.