

# Case 5: tofaktormodeller med tilfældige effekter

## Statistisk Dataanalyse 2

Anders Tolver

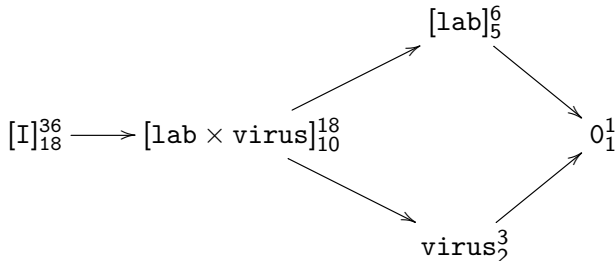
Institut for Matematiske Fag

Uge 4, torsdag d. 28/9-2017

## Case 5: faktordiagram og statistisk model (spm. 1-4)

Model:  $Y_i = \alpha(\text{virus}_i) + b(\text{lab}_i) + c(\text{lab} \times \text{virus}_i) + e_i.$

- ▶  $b(1), \dots, b(6)$  er uafhængige  $\sim N(0, \sigma_B^2)$
- ▶  $c(1, 1), \dots, c(6, 3)$  er uafhængige  $\sim N(0, \sigma_C^2)$
- ▶  $e_1, \dots, e_{36}$  er uafhængige  $\sim N(0, \sigma^2)$



Modellen fittes i R (-husk første library(nlme))

```
> LV=L:V
```

```
> lme(cpe~V, random=~1|L/LV)
```

## Case 5: test af varianskomponenter (spm. 5)

Test for effekt af **lab**  $\times$  **virus** svarer til hypotesen  $H_0 : \sigma_C^2 = 0$ .

**lab**  $\times$  **virus** skal testes mod den identiske faktor I, så vi kan udføre testet i R vha. `lm` og `anova`.

```
> mod0=lm(cpe~LV)
> mod1=lm(cpe~L+V)
> mod2a=lm(cpe~V)
```

```
> anova(mod1,mod0)
Analysis of Variance Table
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	28	11.815				
2	18	1.670	10	10.145	10.935	9.355e-06 ***

Vi konkluderer, at der er en signifikant effekt af **lab**  $\times$  **virus**.

## Case 5: test af varianskomponenter (spm. 5)

Test for effekt af **lab** svarer til hypotesen  $H_0 : \sigma_B^2 = 0$ .

Faktoren **lab** skal testes imod **lab**×**virus**.

Beregning af  $MS_e^{\text{lab} \times \text{virus}}$  (-se også forrige slide)

```
> SSe_LV=deviance(mod0)
> SSe_LpV=deviance(mod1)
> DF_LV=df.residual(mod1)-df.residual(mod0)
> (SSe_LpV-SSe_LV)/DF_LV
[1] 1.0145
```

Beregning af  $MS_e^{\text{lab}}$

```
> SSe_V=deviance(mod2a)
> DF_L=df.residual(mod2a)-df.residual(mod1)
> (SSe_V-SSe_LpV)/DF_L
[1] 4.898667
```

F-teststørrelse

$$F_{\text{meat}} = \frac{MS_e^{\text{lab}}}{MS_e^{\text{lab} \times \text{virus}}} = \frac{4.899}{1.0145} = 4.829 \sim F(5, 10) \Rightarrow p = 0.017$$

## Case 5: test for effekt af virus (spm. 6)

Likelihoodratio test med simuleret p-værdi

```
> modelA=lme(cpe~V,random=~1|L/LV,method="ML")
> modelB=lme(cpe~1,random=~1|L/LV,method="ML")
> anova(modelB,modelA)
```

	Model	df		AIC	BIC	logLik	Test	L.Ratio	p-value
modelB	1	4	82.35415	88.68822	-37.17707				
modelA	2	6	77.79245	87.29356	-32.89622	1 vs 2	8.561697	0.0138	

```
> sim<-simulate.lme(modelB,m2=modelA,nsim=1000)
> lr.sim<-2*(sim$alt$ML-sim$null$ML)
> psim<-sum(lr.sim>8.562)/1000
> psim
[1] 0.024
```

Alternativt kan man lave F-test som giver:  $F = 5.205$ ;  $p = 0.028$ . Der er således signifikant forskel på de forskellige vira.

Hvilken faktor skal virus testes imod?

## Case 5: slutmodel og estimerer (spm. 6)

Slutmodel:  $Y_i = \alpha(\text{virus}_i) + b(\text{lab}_i) + c(\text{lab} \times \text{virus}_i) + e_i$

- ▶  $b(1), \dots, b(6)$  er uafhængige  $\sim N(0, \sigma_B^2)$
- ▶  $c(1, 1), \dots, c(6, 3)$  er uafhængige  $\sim N(0, \sigma_C^2)$
- ▶  $e_1, \dots, e_{36}$  er uafhængige  $\sim N(0, \sigma^2)$

```
> model=lme(cpe~V-1,random=~1|L/LV,method="REML")
```

```
> summary(model)
```

```
Random effects:
```

```
Formula: ~1 | L
```

```
(Intercept)
```

```
StdDev: 0.8045887
```

```
Formula: ~1 | LV %in% L
```

```
(Intercept) Residual
```

```
StdDev: 0.6788688 0.3045940
```

```
> intervals(model)
```

	lower	est.	upper
V1	2.472571	3.450000	4.427429
V2	3.764238	4.741667	5.719096
V3	3.380904	4.358333	5.335762

Hvad bliver paramterestimerterne og hvor mange er der?

## Case 5: repeterbarhed og reproducerbarhed

### Repeterbarhed (spm. 8)

Beskriver 95%–konfidensinterval for forskellen ml. CPE målinger på samme virus i samme laboratorium

$$r = 2.83 \cdot \hat{\sigma} = 2.83 \cdot 0.305 = 0.86$$

### Reproducerbarhed (spm. 9)

Beskriver 95%–konfidensinterval for forskellen ml. CPE målinger på samme virus i forskellige laboratorier

$$r = 2.83 \cdot \sqrt{\hat{\sigma}_B^2 + \hat{\sigma}_C^2 + \hat{\sigma}^2} = 2.83 \cdot \sqrt{0.805^2 + 0.679^2 + 0.305^2} = 3.10$$

## Case 5: tre tilfældige effekter (spm. 10-12)

```
> library(lme4)
> randmod=lmer(cpe~1+(1|L)+(1|V)+(1|LV))
> summary(randmod)
Formula: cpe ~ 1 + (1 | L) + (1 | V) + (1 | LV)
```

Random effects:

Groups	Name	Variance	Std.Dev.
LV	(Intercept)	0.460861	0.67887
L	(Intercept)	0.647361	0.80459
V	(Intercept)	0.355528	0.59626
Residual		0.092778	0.30459

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	4.18333	0.50456	8.291

Estimat for forventede “maximal titration with CPE”

$$4.1833 \pm qt(0.975, 35) \cdot 0.505 = [3.16, 5.21]$$