

Opgaver til kursusuge 4

Formålet med denne uges øvelser er at opnå forståelse for og fortrolighed med modeller med tilfældige effekter, herunder

- at opskrive modellerne (med papir og blyant)
- at fitte modellerne med `lme` og trække parameterestimer, konfidensintervaller og teststørrelser ud af R-analyser vha. `summary`, `intervals`, og `anova`.
- udfra faktordiagrammer at afgøre hvilke tilfældige effekter de enkelte systematiske faktorer skal testes mod
- at fortolke resultaterne

Som sædvanlig er det i alle opgaver, både computeropgaverne og de teoretiske opgaver, vigtigt at være omhyggelig med at specificere modellerne ordentligt og at konkludere på passende vis.

Data til de enkelte opgaver kan findes på ugeplanen for uge 4 under Absalon.

Et par bemærkninger vedr. installation af R-pakker:

- `nlme`-pakken som indeholder `lme` loades med kommandoen `library(nlme)`.
- For at bruge `estimable` for modeller med tilfældige effekter skal man hente `gmodels`-pakken

Computerøvelser

Opgave 5.1

I et forsøg til undersøgelse af effekten af E-vitamin på marsvins vækst blev 15 forsøgsdyr fulgt i en 7-ugers periode. Vi skal i denne opgave se på en del af data fra dette forsøg. Dyrene blev vejret i uge 5 og uge 7. Et stykke tid før uge 5 blev dyrenes foder tilsat forskellige mængder (dosis 0, 1 og 2 i uspecificerede enheder) af vitamin E, og denne fodring blev anvendt gennem hele forsøget. Der var 5 dyr i hver gruppe. I nedenstående tabel er vægten (i g) for de 15 marsvin angivet.

Dosis	Uge	Hver celle repræsenterer et dyr				
0	5	504	596	597	583	528
	7	466	587	619	612	576
1	5	524	484	585	637	605
	7	597	569	677	702	675
2	5	622	557	555	601	524
	7	670	609	605	649	583

(Data til denne opgave er blevet brugt i opgave 2 i eksamenssættet i Statistisk Forsøgsplanlægning, juni 2000.)

1. Opstil en statistisk model for forsøget, hvor det er muligt at undersøge om en eventuel effekt af vitamin E ændrer sig over tid.

Tegn også faktordiagrammet og afgør hvilken type forsøg, der er tale om.

Vink:

- Hvilken faktor beskriver det fænomen, at vægtændringen fra uge 5 til uge 7 afhænger af indholdet af vitamin E i foderet?
 - Hvilke faktorer skal indgå i modellen som systematiske faktorer? Hvilke som tilfældige?
 - NB. Vekselvirkninger med dyr ønskes ikke medtaget i modellen.
2. Indlæs data. Installér og load R-pakken `nlme` på computeren (hvis det ikke allerede er gjort).
 3. Analysér data, dvs. fit modellen og reducer den systematiske del mest muligt. Hvad er konklusionen?
 4. Overvej hvordan resultaterne kan præsenteres grafisk, og lav en skitse i hånden.
 5. Angiv estimatorne for variansparametrene.
 6. Estimer den forventede forskel i vægt for to marsvin i uge 7, hvor det ene marsvin har fået vitamin E i dosis 1 og det andet i dosis 0.

Opgave 5.2

Denne opgave går ud på at analysere data fra opgave 2.4 i kompendiet.

1. Tegn faktordiagrammet hvor de tre hovedvirkninger (eye, lens and subj) og de tre tofaktorvekselvirkninger (eye \times lens, subj \times eye, subj \times lens) indgår. Hvilke af faktorerne skal være systematiske? Hvilke skal være tilfældige?
2. Opskriv modellen svarende til faktordiagrammet (på papir, fit den ikke endnu).
3. Undersøg om nogle af varianskomponenterne $\sigma_{\text{subj} \times \text{eye}}^2$ og $\sigma_{\text{subj} \times \text{lens}}^2$ kan sættes til 0. Tegn faktordiagrammet hørende til den reducerede model.

Vink: Da forsøget er pænt balanceret, kan vi teste for de tilfældige effekter vha. MS-størrelserne fra den tilsvarende lineære model, dvs. modellen hvor alle faktorer i modellen (også dem der egentlig er systematiske) er systematiske. Fit derfor denne model vha. `lm`.

4. Fit den reducerede model vha. `lme`, og reducer den systematiske del af modellen mest muligt.

Vink: Husk at tilfældige vekselvirkninger ikke kan indgå i random-delen med `:` hvis man vil udnytte at to tilfældige faktorer er nestede. Produktfaktorerne skal derimod defineres som separate faktorer først.

Bemærk: Der kan være lidt forskel på p-værdierne afhængigt af, om man benytter likelihood ratio test, F-test eller om man simulerer p-værdien vha. `simulate.lme`. Lad derfor være med at hænge jer alt for meget i, om p-værdien er lige over eller lige under 5 %.

5. Skriv slutmodellen op (på papir), og estimer parametrene i modellen. Bestem også konfidensintervaller for middelværdiparametrene.
6. Det var ikke nødvendigt at fitte startmodellen (den fra spg. 2), men gør det alligevel. Estimer varianskomponenterne i modellen og sammenlign med testene i spg. 3.
7. Antag nu at eksperimentet blev udført således at der kun blev foretaget målinger på venstre øje for individerne 1–3, og kun på højre øje for individerne 4–7. Tegn faktordiagrammet for dette eksperiment.

Opgave 5.3 (ekstra, fortsættelse af opgave 5.1)

Denne opgave er en fortsættelse af opgave 5.1.

7. Bestem også et konfidensinterval for forskellen i forventet vægt i spørgsmål 6 (opgave 5.1).
8. Lav figuren omtalt i spørgsmål 4. Du kan forsøge at benytte følgende R-kode (overvej hvad der sker i de enkelte trin) eller du kan eksperimentere dig frem med brug af `interaction.plot()`:

```
modelA1=lme(vgt~dosis:uge-1, random=~1|dyr, method="REML")
est = fixef(modelA1)
```

```
uge1 = c(5,5,5,7,7,7)
```

```
dosis1 = c(0,1,2,0,1,2)

plot(dosis1,est,type="n")
points(dosis1[uge1==5],est[uge1==5],type="b",pch=16)
points(dosis1[uge1==7],est[uge1==7],type="b",pch=2)
```

9. Udfør modelkontrol for modellen i opgave 5.1.

Det er ikke så ligetil at lave modelkontrol for modeller med tilfældige effekter. Man plejer at gøre det i den tilsvarende lineære model, altså den model hvor de tilfældige effekter flyttes over i den systematiske del.

Opgaven går altså ud på at lave residualplot og evt. qq-plot for den lineære model hvor alle faktorer i modellen (også dem der egentlig er tilfældige) indgår som systematiske faktorer.

Teoretiske øvelser

Opgave 5.4

Denne opgave går ud på at analysere tallene i kompendiets opgave 8.1, evt. udfra nedestående delspørgsmål. Du kan benytte resultatet af R-kørslen nedenfor.

1. I forsøget indgår faktorerne *block*, *treatment* and *variety*. Ved analysen nedenfor ses helt bort fra vekselvirkningen $\text{block} \times \text{variety}$ mens $\text{block} \times \text{treatment}$ skal indtages. Angiv hvilken type forsøg der er tale om? Hvordan udfører man randomisering i et sådant eksperiment?
2. Tegn faktordiagrammet, og bestem hvilke tilfældige effekter de enkelte systematiske effekter skal testes imod. Opstil en statistisk model.
3. Analysér data, dvs. fit modellen og reducer den mest muligt.
NB. I forsøg som disse er vi normalt ikke interesserede i at teste de tilfældelige effekter, da de er en del af designet. Du kan således nøjes med at reducere den systematiske del af modellen.
4. Estimér varianskomponenterne i slutmodellen.
5. Drag konklusioner. Overvej også hvordan man kunne præsentere resultaterne grafisk.

Resultat af R-kørsel som kan benyttes undervejs:

```
> library(nlme)
> blocktreat<-block:treat
> modelA<-lme(yield~variety:treat-1,random=~1|block/blocktreat,method="ML")
> modelB<-lme(yield~variety+treat,random=~1|block/blocktreat,method="ML")

> anova(modelB,modelA)
      Model df      AIC      BIC    logLik    Test  L.Ratio p-value
modelB     1 10 381.9448 402.8882 -180.9724
modelA     2 15 328.7913 360.2065 -149.3956 1 vs 2 63.15347 <.0001

> modelA1<-lme(yield~variety:treat-1,random=~1|block/blocktreat,method="REML")

> summary(modelA1)
Linear mixed-effects model fit by REML
Data: NULL
      AIC      BIC    logLik
299.0572 327.1252 -134.5286

Random effects:
Formula: ~1 | block
(Intercept)
StdDev: 0.7657059

Formula: ~1 | blocktreat %in% block
(Intercept) Residual
```

StdDev: 1.226708 3.041115

Fixed effects: yield ~ variety:treat - 1

	Value	Std.Error	DF	t-value	p-value
varietyan:treatT	63.38	1.505955	39	42.08626	0
varietyci:treatT	79.64	1.505955	39	52.88339	0
varietyko:treatT	75.06	1.505955	39	49.84213	0
varietykr:treatT	65.88	1.505955	39	43.74633	0
varietysl:treatT	78.26	1.505955	39	51.96703	0
varietyvu:treatT	60.72	1.505955	39	40.31993	0
varietyan:treatU	30.50	1.505955	39	20.25293	0
varietyci:treatU	56.04	1.505955	39	37.21227	0
varietyko:treatU	58.52	1.505955	39	38.85907	0
varietykr:treatU	53.68	1.505955	39	35.64516	0
varietysl:treatU	57.12	1.505955	39	37.92942	0
varietyvu:treatU	27.30	1.505955	39	18.12803	0

> intervals(modelA1)

Approximate 95% confidence intervals

Fixed effects:

	lower	est.	upper
varietyan:treatT	60.33392	63.38	66.42608
varietyci:treatT	76.59392	79.64	82.68608
varietyko:treatT	72.01392	75.06	78.10608
varietykr:treatT	62.83392	65.88	68.92608
varietysl:treatT	75.21392	78.26	81.30608
varietyvu:treatT	57.67392	60.72	63.76608
varietyan:treatU	27.45392	30.50	33.54608
varietyci:treatU	52.99392	56.04	59.08608
varietyko:treatU	55.47392	58.52	61.56608
varietykr:treatU	50.63392	53.68	56.72608
varietysl:treatU	54.07392	57.12	60.16608
varietyvu:treatU	24.25392	27.30	30.34608

Opgave 5.5

For at sammenligne frygtsomheden hos transgene og normale mus udførte man følgende forsøg. Ialt 32 mus blev anbragt i 8 bure med 4 mus af samme køn og stamme (transgen eller normal) i hvert bur. Der var 8 mus for hver kombination af køn og stamme. Frygtsomhedstesten blev foretaget i *open field*, her en hvid cirkulær arena (90 cm i diameter) med en indre cirkel (31 cm i diameter) i midten. Området udenfor den indre cirkel blev så opdelt i 12 lige store dele af sorte linier. Hver mus blev derpå anbragt i midten af arenaen og der blev observeret hvor mange gange musen krydsede en linie med alle fire poter i en 10 minutters periode.

		Hunmus		Hanmus	
Normal		1	2	3	4
		306 303	379 429	160 323	256 286
		333 237	476 456	360 316	311 302
Transgen		5	6	7	8
		304 262	324 197	182 267	172 190
		294 178	217 267	209 171	172 212

(Data til denne opgave er blevet brugt i opgave 2 i eksamenssættet i Statistisk Forsøgsplanlægning, maj 2003.)

1. Tegn et faktordiagram for forsøget og opstil en statistisk model.
2. Reducér modellen mest muligt (den systematiske del).
3. Angiv slutmodellen og det tilhørende faktordiagram. Angiv estimaterne for parametrene i modellen (husk også varianskomponenterne).
4. Estimér den forventede forskel mellem normale og transgene mus. Bestem også et konfidensinterval.
5. Tidligere studier har vist at normale mus har 20% højere aktivitet end transgene mus. Hvilken hypotese (udtrykt ved parametrene i modellen) svarer dette til? Hvordan kunne du teste hypotesen i R vha. estimable?
6. Betragt et øjeblik den additive model for tosidet variansanalyse med faktorerne stamme og bur, dvs.

$$Y_i = \alpha(\text{stamme}_i) + \beta(\text{bur}_i) + e_i$$

hvor altså både stamme og bur indgår som systematiske faktorer. Modellen er fittet som tosidet nedenfor.

Bestem et estimat og et konfidensinterval for den forventede forskel mellem normale og transgene mus i denne model. Estimater er et andet end det i spg. 4; hvorfor? Konfidensintervallet er smallere end konfidensintervallet fra spg. 4; hvorfor?

R-kode og R-output som kan bruges til besvarelsen:

```
> mus = read.table("maj03_2.txt",header=T)
> attach(mus)
> bur = factor(bur)

> library(nlme)
> modelA = lme(kryds ~ stamme + sex + stamme:sex, random =~ 1 | bur,method="ML")
> modelB = lme(kryds ~ stamme + sex, random =~ 1 | bur,method="ML")
> anova(modelB,modelA)
      Model df      AIC      BIC    logLik    Test    L.Ratio p-value
modelB     1  5 357.5748 364.9035 -173.7874
modelA     2  6 359.4593 368.2537 -173.7296 1 vs 2 0.1154818 0.734
> modelC1 = lme(kryds ~ sex,random =~ 1 | bur,method="ML")
> anova(modelC1,modelB)
```

```

      Model df      AIC      BIC    logLik  Test  L.Ratio p-value
modelC1     1  4 364.3412 370.2041 -178.1706
modelB      2  5 357.5748 364.9035 -173.7874 1 vs 2 8.766401 0.0031
> modelC2 = lme(kryds ~ stamme, random =~ 1 | bur, method="ML")
> anova(modelC2, modelB)
      Model df      AIC      BIC    logLik  Test  L.Ratio p-value
modelC2     1  4 360.6212 366.4842 -176.3106
modelB      2  5 357.5748 364.9035 -173.7874 1 vs 2 5.046455 0.0247
> sim=simulate.lme(modelC2, m2=modelB, nsim=5000, method="ML")
> lr.sim<-2*(sim$alt$ML-sim$null$ML)
> psim<-sum(lr.sim>5.046)/5000
> psim
[1] 0.085
> summary(slutmodel)
Linear mixed-effects model fit by REML
Data: NULL
      AIC      BIC    logLik
343.875 349.4798 -167.9375

Random effects:
Formula: ~1 | bur
(Intercept) Residual
StdDev:    50.61441 50.69752

Fixed effects: kryds ~ stamme
      Value Std.Error DF   t-value p-value
(Intercept)  327.0625  28.30361 24 11.555504 0.0000
stammetransgen -100.9375  40.02735  6 -2.521713 0.0452

> VarCorr(slutmodel)
bur = pdLogChol(1)
      Variance StdDev
(Intercept) 2561.818 50.61441
Residual    2570.238 50.69752

> library(gmodels)
> intervals(slutmodel)
Approximate 95% confidence intervals

Fixed effects:
      lower      est.      upper
(Intercept)  268.6467  327.0625  385.478284
stammetransgen -198.8809 -100.9375  -2.994098
attr(,"label")
[1] "Fixed effects:"

Random Effects:
Level: bur
      lower      est.      upper
sd((Intercept)) 24.85319 50.61441 103.0781

Within-group standard error:
      lower      est.      upper
38.20556 50.69752 67.27395

```



```

> tosidet = lm(kryds ~ stamme + bur)

> summary(tosidet)
Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    294.75      25.35   11.628 2.39e-11 ***
stammetransgen -108.25      35.85   -3.020 0.005922 **
bur2           140.25      35.85    3.912 0.000657 ***
bur3            -5.00      35.85   -0.139 0.890238
bur4            -6.00      35.85   -0.167 0.868481
bur5            73.00      35.85    2.036 0.052895 .
bur6            64.75      35.85    1.806 0.083443 .
bur7            20.75      35.85    0.579 0.568101
bur8              NA         NA      NA      NA

> confint(tosidet)
              2.5 %    97.5 %
(Intercept)  242.4327200 347.06728
stammetransgen -182.2378069 -34.26219
bur2          66.2621931 214.23781
bur3         -78.9878069  68.98781
bur4         -79.9878069  67.98781
bur5         -0.9878069 146.98781
bur6         -9.2378069 138.73781
bur7        -53.2378069  94.73781
bur8              NA      NA

```