

## Vejledende besvarelser til udvalgte opgaver i kursusuge 6

### Opgave 6.1

```
library(nlme)

goatdata = read.table("goats1.txt",header=T)

### Construction of variables: y is the factor that weight has
### increased by from day 0 to day 91.

goat91 = subset(goatdata, day==91)
attach(goat91)
y = weight/w0

### Fit of model with feed effect and initial weight:
### Modell:  $y_i = a(\text{feed}_i) + b \cdot \text{start}_i + e_i$ 

modell1 = lm(y ~ w0 + factor(feed))
plot(predict(modell1), rstandard(modell1))

### The residual plot looks quite nice.

summary(modell1)

### Model without feed effect
### Modell2:  $y_i = a + b \cdot \text{start}_i + e_i$ 

modell2 = lm(y ~ w0)
anova(modell2, modell1)

### Modell2 is rejected ( $p < 0.0001$ ).

### From the estimates from modell1 it follows that the LSD95-value for
### feed type comparisons is  $qt(0.975, 23) \cdot 0.0273 = 0.057$ . It follows
### that the expected weight gain is significantly lower for feed4
### than for the others feed types. Also, the weight gain is
### significant lower for feed2 than for feed1 whereas there is no
### significant difference between feed1 and feed3.
###
### Also note that the initial weight is significant ( $p = 0.0009$ ), and that
### goats with high initial weight gains less weight than do goats with
### low initial weight.
```

```
### Estimation of expected day91-weights for goat with initial weight
### equal to 12 kg:
```

```
library(gmodels)
```

```
### First the expected values of y:
```

```
est1 = c(1,12,0,0,0)
est2 = c(1,12,1,0,0)
est3 = c(1,12,0,1,0)
est4 = c(1,12,0,0,1)
est = rbind(est1, est2, est3, est4)
```

```
estimable(model1, est, conf.int=0.95)
```

```
### And for the weight:
```

```
12*estimable(model1, est, conf.int=0.95)
```

```
### The expected values are 13.9, 13.1, 14.1 and 11.4, respectively.
```

## Opgave 6.2

```
### Comments (R-code follows below):
```

```
### Spg 2: Diggle-model:
```

```
###      y_i = gamma(dose_i,week_i)+eta(animat_i)+delta_i+e_i
###      where the eta's, delta's and e's are as in BMS, page 190.
###
###      The variogram looks okay. The variogram obviously
###      increases, so it seems necessary to use some serial
###      correlation structure; the Diggle model seems to catch
###      the structure reasonably well.
```

```
### Spg 3: With notation as in BMS, we get estimates from the R-output:
```

```
###      nu^2:      11.49038^2 = 132.03
###      sigma^2 + tau^2 = 39.9255^2 = 1594.0
###      phi:      6.625
###      sigma^2/(sigma^2+tau^2): 0.1629
###
###      From this we get:
###      nu^2:      132.03
###      sigma^2: 259.7
###      phi:      6.625
###      tau^2:      1334
```

```
### Spg 4: The interaction is significant (p=0.01). Likewise, the model
### cannot be reduced to a model with a linear relationship (p<0.0001)
```

```
### Spg 5: Actually, the design is a problem for the analysis. The model
### allows for different means for the three treatment groups -
### even in weeks 1,3,4 where there is no difference between the
### groups.
```

```
###
### Alternatives:
### (a) drop the week 1-3 measurements and use week4 as baseline
### (b) make a new factor "realdose" which for each observation
### codes the actual dose (0 for all obs. in week 1-4)
### (c) summary analysis
```

```
###
### Note that since the week1-observations are not affected
### by the treatment, they can be used as baseline measurements
### (covariates). This would improve chances of getting
### significances.
```

```
### Spg 6: Model for summary measure:  $inc_i = a(dose_i) + e_i$ 
### The hypothesis is  $a(0)=a(low)=a(high)$ , which cannot be rejected
### (p=0.35). Conclusion: no evidence of a dose effect.
```

```
### Spg 7: This analysis is indeed valid, despite the design.
```

```
#####
```

```
### spg 1: Data into R etc.:
```

```
guinea = read.table("bms_opgl0_1.txt",header=T)
attach(guinea)
```

```
weekfac = factor(week)
animal = factor(animal)
dose = factor(dose)
```

```
library(nlme)
```

```
### spg 2: Repeated measures model
```

```
modell1 = lme(wgt ~ dose*weekfac,
              random = ~ 1 | animal,
              method = "ML",
```

```

        corr = corGaus(form = ~ week | animal, nugget=T))

###      Semi-variogram (technical stuff due to my Linux-version of R)
modella = lme(wgt ~ factor(dose)*factor(week),
              random = ~ 1 | animal, data=guinea,
              method = "ML",
              corr = corGaus(form = ~ week | animal, nugget=T))

plot(Variogram(modella, form= ~ week))

### Spg 3: Estimation of variance parameters: (This is the
###      ML-estimates, should actually rather use REML-estimates):

summary(model1)

### spg 4: Test for reduction in systematic part of the model:

###      Additive model:
model3 = lme(wgt ~ dose+weekfac,
            random = ~ 1 | animal,
            method = "ML",
            corr = corGaus(form = ~ week | animal, nugget=T))

anova(model3,model1)

###      Model with linear relationship:
model4 = lme(wgt ~ factor(dose)*week,
            random = ~ 1 | animal,
            method = "ML",
            corr = corGaus(form = ~ week | animal, nugget=T))

anova(model4,model1)

###      Interaction plot:

interaction.plot(week,dose,wgt)

### Spg 5: Individual profiles

interaction.plot(week,animal,wgt)

### spg 6: Analysis of summary measure: construction of increment and
###      oneway ANOVA. Conclusion: dose effect not significant
###      (p=0.35).

guinea7 = subset(guinea, week==7)
guinea1 = subset(guinea, week==1)

```

```
myinc = guinea7$wgt - guinea1$wgt
mydose = guinea7$dose
```

```
modelA = lm(inc ~ dose1)
modelB = lm(inc ~ 1)
anova(modelB,modelA)
```

### Opgave 6.3

### Data indlæses i R:

```
rottedata = read.table("maj04_2.txt",header=T)
attach(rottedata)
rotte = factor(rotte)
uge = factor(uge)
```

```
### spg. 1: Den vigtige observation er at rotte er finere end
###         behandling. Der er defor tale om et design af samme
###         struktur som split-plot forsøget.
###
###         Systematiske: uge, behandling, uge:behandling
###         Tilfældig: rotte
###
###         Modellen bliver:
###          $Y_i = a(uge_i, behandling_i) + B(rotte_i) + e_i$ 
###         hvor a'er er parametre, B'er og e'er normalfordelte variable.
```

### spg. 2: Modellen fittes, og der testes for vekselvirkning

```
library(nlme)
```

```
modell1 = lme(fedt ~ uge + behandling + uge:behandling,
              random =~ 1 | rotte, method="ML")
modell2 = lme(fedt ~ uge + behandling, random =~ 1 | rotte, method="ML")

anova(modell2,model1) ## Vekselvirkningen er ikke signifikant (p=0.48).
```

### Tester for hovedvirkninger:

```
modell3 = lme(fedt ~ uge, random =~ 1 | rotte, method="ML")
anova(modell3, modell2) ## behandling ikke sign. (p=0.80)
```

```
modell4 = lme(fedt ~ 1, random =~ 1 | rotte, method="ML")
anova(modell4,model3) ## uge er stærkt sign. (p<0.0001)
```

### Slutmodellen er derfor

```

###          Y_i = c(uge_i) + B(rotte_i) + e_i
###          der fittes med REML

model3a = lme(fedt ~ uge, random =~ 1 | rotte, method="REML")

summary(model3a)
print.matrix = print.default # Technical, for my Linux-version of R
intervals(model3a)

###          Konklusion:
###          Fedtværdien er 762 [663,860] højere i uge 10 end i uge 2.
###          Kastraktion uden betydning for fedtfordelingen.

###          Varianskomponenterne estimeres til
###          sigma: 157
###          sigma_rotte: 245

##### Output fra slutmodel: #####3

## > summary(model3a)
## Random effects:
## Formula: ~1 | rotte
##          (Intercept) Residual
## StdDev:      244.864 156.7615

## Fixed effects: fedt ~ uge
##          Value Std.Error DF  t-value p-value
## (Intercept) 1461.5909  61.98701 21 23.57899      0
## uge10        761.8182  47.26537 21 16.11789      0

## > intervals(model3a)
## Approximate 95% confidence intervals

## Fixed effects:
##          lower      est.      upper
## (Intercept) 1332.6819 1461.5909 1590.4999
## uge10        663.5245  761.8182  860.1119

```

## Opgave 6.4

1. Som summary measures kunne man forestille sig at bruge gennemsnittet af målingerne, den maksimale måling eller lignende. Man burde tegne profilerne for at se hvilke summary measures der beskriver kurverne.

Modellen for ensidet variansanalyse er følgende (hvor  $X_i$  er et summary measure, målt på  $i$ 'te person):

$$X_i = \alpha(B_i) + e_i$$

hvor  $B$  (behandling) er faktoren der angiver om personen har fået sødemiddel eller ej, og  $e_1, \dots, e_{20}$  er uafhængige og  $N(0, \sigma^2)$ -fordelte.

I R: `lm(x ~ Beh)`

2. Faktorer i forsøget er behandling ( $B$ ), person ( $P$ ), og tid ( $T$ ). Også vekselvirkningen  $B \times T$  bør indgå i modellen.  $B$ ,  $T$  og  $B \times T$  skal være systematiske,  $P$  tilfældig.

$P$  er finere end  $B$ , så der er tale om et split-plot design (bortset fra at tidspunkterne naturligvis ikke er randomiseret). Faktordiagrammet bliver derfor lige til kompendiets figur 8.1 (med  $W = P$ ).

3. Hvis responsen er  $Y$  er modellen med random intercept givet ved

$$Y_i = \gamma(B_i, T_i) + A(P_i) + e_i, \quad i = 1, \dots, 120$$

hvor  $A(1), \dots, A(20) \sim N(0, \nu^2)$ ,  $e_1, \dots, e_{120} \sim N(0, \sigma^2)$ ; alle er uafhængige.

Dette svarer præcis til split-plot modellen.

Variansen på et hvert  $Y_i$  er  $\sigma^2 + \nu^2$ . Observationer hørende til forskellige observationer er uafhængige, observationer fra samme person er korrelerede med kovarians  $\nu^2$  eller korrelation  $\nu^2/(\sigma^2 + \nu^2)$ .

I R: `lm(x ~ Beh + Tid + Beh:Tid, random = ~ 1 | Person)`

4. Vil opstille modellen med serial korrelationsstruktur (Diggle-modellen):

$$Y_i = \gamma(B_i, T_i) + A(P_i) + B_i + e_i, \quad i = 1, \dots, 120$$

hvor  $A(1), \dots, A(20) \sim N(0, \nu^2)$ ,  $B_1, \dots, B_{120} \sim N(0, \tau^2)$ ,  $e_1, \dots, e_{120} \sim N(0, \sigma^2)$ . Alle er uafhængige bortset fra at  $\text{Cov}(B_i, B_j) = \tau^2 \exp(-(T_i - T_j)^2/\phi)$  hvis  $P_i = P_j$ .

Variansen af et  $Y_i$  er  $\nu^2 + \tau^2 + \sigma^2$ , og kovariansen mellem to observationer fra samme person til tidspunkterne  $t_1$  og  $t_2$  er  $\nu^2 + \tau^2 \exp(-(t_1 - t_2)^2/\phi)$

`lme(y ~ Beh + Tid + Beh:Tid, random = ~ 1 | Person,  
corr = corGaus(form = ~ Tid | Person, nugget=T))`

5. Det er naturligt at udnytte målingen inden forsøgets start som en baselinemåling. Lad  $x_i$  være målingen for den  $i$ 'te person. Modellen fra 5. modificeres til

$$Y_i = \gamma(B_i, T_i) + \beta \cdot x_i + A(P_i) + B_i + e_i, \quad i = 1, \dots, 120$$

med samme antagelser som før, og R-koden modificeres til

`lme(y ~ x + Beh + Tid + Beh:Tid, random = ~ 1 | Person,  
corr = corGaus(form = ~ Tid | Person, nugget=T))`

6. Spørgsmålet er om tiden, og tiden kvadreret, skal indgå som kovariat snarere end som faktor. Modellen er

$$Y_i = \alpha(B_i) + \beta(B_i) \cdot T_i + \gamma(B_i) \cdot T_i^2 + A(P_i) + B_i + e_i, \quad i = 1, \dots, 120$$

med samme antagelser som før.

Hvis Tidnum er en numerisk variabel med tidspunkterne kan modellen fittes således (den kvadrerede tidsvariabel konstrueres først):

```
Tidnumkv = Tidnum*Tidnum
lme(y ~ Beh:Tidnum + Beh:Tidnumkv, random = ~ 1 | Person,
    corr = corGaus(form = ~ Tid | Person, nugget=T))
```

## Opgave 6.5

R-kode og R-output er givet i opgaven.

spg 1: Den vigtige observation er at rotte er finere end behandling. Der er derfor tale om et design af samme struktur som split-plot forsøget.

Systematiske: uge, behandling, uge:behandling  
 Tilfældig: rotte

Modellen bliver:

$$Y_{ij} = a(u_{ij}, b_{ij}) + B(\text{rotte}_{ij}) + e_{ij}$$

hvor  $a$ 'er er parametre,  $B$ 'er og  $e$ 'er normalfordelte variable.

spg 2 Test for vekselvirkning:  $F=0.301$ ,  $p=0.59$ .

Begge hovedvirkninger er signifikante:

uge (mod residualen):  $F = 5.75$ ,  $p=0.026$   
 behandling (mod rotte):  $F=12.3$ ,  $p=0.0022$

ModelB er således slutmodellen:

$$Y_{ij} = b(\text{beh}_{ij}) + c(u_{ij}) + B(\text{rotte}_{ij}) + e_{ij}$$

Konklusioner:

glucoseindholdet er 0.33 [0.04,0.62] højere i uge 2 end i uge 10.  
 glucoseindholdet er 0.57 [0.23,0.91] højere for kastrerede rotter end for kontrolrotter.

Varianskomponenterne estimeres til

sigma: 0.46  
 sigma\_rotte: 0.20

spg 3: Kan opstille model for gentagne målinger (med serieel korrelation):

$$Y_{ij} = \gamma(u_{ij}, b_{ij}) + B(\text{rotte}_{ij}) + C_{ij} + e_{ij}$$

hvor  $B$ 'er er  $N(0, \sigma_B^2)$ ,  $e$ 'er er  $N(0, \sigma^2)$ ,  
 $C$ 'er  $N(0, \sigma_C^2)$ .  $C$ 'erne er korrelerede med  
 $\text{Cov}(C_{ij}, C_{kj}) = 0$  hvis  $\text{rotte}_{ij}$  ikke er lig  $\text{rotte}_{kj}$   
 $\text{Cov}(C_{ij}, C_{kj}) = \sigma_C^2 \exp(-(|u_{ij} - u_{kj}|)^2 / \rho)$



hvis rotte\_i=rotte\_j

spg 4: Kald målingen på i'te rotte før det operative indgreb for  $x_i$ .

Kan bruge den som baseline-måling:

$Y_i = \text{gamma}(\text{uge}_i, \text{beh}_i) + \text{beta} * x_i + B(\text{rotte}_i) + C_i + e_i$   
med samme antagelser som før.