

## Opgaver til kursusuge 5+6

Formålet med nedenstående opgaver er at opnå en forståelse for og fortrolighed med modeller for gentage målinger, altså modeller for summary measures, modellen med random intercepts og Diggle-modellen med seriel korrelationsstruktur. Efter øvelserne bør I kunne

- redegøre for ligheder og forskelle mellem modellerne
- fitte modellerne, estimere parametre, teste hypoteser, foretage konklusioner osv.

Mht. R er det vigtige i denne og næste uge:

- at kunne fitte (og analysere) Diggle-modellen med `lme`; brug optionen `corr`
- at kunne konstruere de relevante variable til summary analyser
- at øge fortroligheden med analyse vha. `lme` (tests, fortolkning af output, konstruktion af kontraster osv.)

Hvis du bruger kald af typen `anova(model2, model1)` til at teste hypoteser (teste `model2` mod `model1`), husk da at modellerne skal fittes med ML snarere end REML. Dette opnås ved at tilføje `method="ML"` til `lme`-kaldet.

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 ugeplanerne for uge 5+6. Bemærk at opgave 6.3 og 6.5 bruger data fra samme forsøg, blot to forskellige responsvariable (hhv. mål for fedtfordeling og glukosekoncentration).

# Opgaver med brug af computer

## Opgave 6.1

Denne opgave går ud på at lave en “summary-analyse” af data fra eksempel 10.1 fra kompendiet (jf. afsnit 10.1 i kompendiet og afsnit 6.2 i R-vejledningen). Formålet med at lave en analyse af et *summary measure* er hurtigt at danne sig et overblik over data f.eks. med henblik på, om man skal forvente, at der er en behandlingseffekt. Det er en langt mindre detaljeret analyse end den mere grundige analyse af hele datasættet, hvor man bruger *random intercept modellen* eller *Diggle modellen*. Ved en “summary-analyse” starter man med at lave et reduceret datasæt, hvori der kun forekommer en måling per måleserie (=subjekt/person/dyr/individ), og man skal derfor ikke inddrage subjekt som tilfældig effekt i modellen. I stedet kan analysen af det valgte summary measure laves ved brug af en af de simple modeller, som vi kender fra undervisningsugerne 1-4. For at lave en god “summary-analyse” skal man benytte et *summary measure*, som opsummerer noget relevant for de enkelte måleserier. For at få en ide til et fornuftigt valg af *summary measure* anbefales det at se på en figur, hvor man optegner et passende udvalg af måleserierne.

1. Indlæs datasættet (fra goats1.txt). Udfør følgende stump R-kode og forklar hvad variablen *y* indeholder. Check også variablene *feed* og *w0*.

```
> goatdata = read.table("../data/goats1.txt", header=T)
> goat91 = subset(goatdata, day==91)
> attach(goat91)
> y = weight/w0
```

2. Udfør en relevant analyse med henblik på at afgøre om vægtudviklingen er forskellig for de forskellige foderbehandlinger. Overvej herunder hvordan startvægten kan inddrages i analysen. Husk som sædvanlig at opskrive modeller og konklusioner og at udføre modelkontrol.
3. Estimér for hver af de fire fodertyper den forventede vægt efter 91 dage for en ged med en startvægt på 12 kg.
4. Udfør evt. en tilsvarende analyse af et andet summary measure (efter eget valg).

## Opgave 6.2

Denne opgave går ud på at analysere data fra opgave 10.1 i kompendiet.

1. Indlæs data og sørg for at R faktisk opfatter alle faktorer som faktorer.
2. Opskriv en Diggle-model for disse data. Fit modellen i R. Lav desuden et semivariogram; ser det rimeligt ud?
3. Estimér variansparametrene i modellen.
4. Lad os interessere os for, om den systematiske del af modellen kan reduceres.
  - Kan den systematiske del af modellen reduceres til en model med additiv effekt af dosis og tid?
  - Kan modellen reduceres til en model, hvor vægten afhænger lineært af tiden (ikke-parallele rette linier)?
  - Brug kommandoen `interaction.plot(week, dose, wgt)` til at vise “gennemsnits-profilerne”. Sammenlign figuren med konklusionerne fra testene (vær dog varsom med lineariteten, hvorfor?)
5. Faktisk blev der først givet forskellige doser i uge 5 (det står også i opgaveteksten). Indtil da var alle dyr behandlet ens!
  - Sammenhold dette med gennemsnitsprofilerne (lav dem vha. `interaction.plot`).
  - Sammenhold også med de individuelle profiler.
  - Overvej om det har nogle konsekvenser for ovenstående analyse; er den “korrekt” alligevel? Kunne/Burde man have analyseret data anderledes (hvordan)?

*NB.* To moraler i denne forbindelse: (i) tag forsøgsdesignet alvorligt; (ii) lav så vidt muligt altid tegninger af gentagne målinger.
6. Brug tilvæksten fra uge 1 til uge 7 som “summary measure” til at undersøge om der er en effekt af dosis:
  - konstruér tilvæksten (se evt. eksempel fra mandagsforelæsningerne)
  - Opstil en model for tilvæksten og udfør analysen i denne model. Er der en effekt af dosis?
7. Overvej forsøgsdesignet igen, jf. spørgsmål 7. Er summary analysen “korrekt”, trods designet?

### Opgave 6.3

Man regner med, at det mandlige kønshormon testosteron påvirker foderoptag, fedtfordeling og glukosetolerancen. For at undersøge effekten af testosteronmangel blev 22 hanrotter fordelt tilfældigt i 2 grupper med 11 i hver, hvor rotterne i den ene gruppe blev kastreret, mens rotterne i den anden gruppe fik foretaget et indgreb uden effekt. Alle rotter fik målt koncentrationen af glukose i blodet både 2 og 10 uger efter det operative indgreb. I nedenstående tabel er angivet målinger for fedtfordelingen.

Kastreret			Kontrol		
Rotte	Uge		Rotte	Uge	
	2	10		2	10
1	1206	1677	12	1430	2246
2	1222	1813	13	1268	1870
3	1749	2441	14	1258	2082
4	1204	2090	15	1279	1937
5	1267	1945	16	1300	2227
6	1206	1972	17	1412	2672
7	2139	2581	18	1932	2779
8	1574	2432	19	1256	2234
9	1728	2361	20	1567	2357
10	1417	2332	21	1493	2285
11	1711	2801	22	1537	1790

Data ligger i filen maj04\_2.txt (sammen med responsvariablen hørende til opgave 6.5 fra samme forsøg).

Vi skal i denne opgave undersøge hvordan fedtfordelingen afhænger af behandling og uge.

1. Tegn et faktordiagram for forsøget og opstil en statistisk model (hvis du ikke allerede har gjort det i opgave 6.5).
2. Analysér data. Angiv herunder slutmodellen samt estimater og konfidensintervaller for relevante parametre i slutmodellen. Formuler konklusionerne i ord.

## Opgaver uden brug af computer

### Opgave 6.4

I et kostforsøg ville man undersøge afhængigheden af et bestemt sødemiddel på glukosekoncentrationen i blodet.

20 personer indgik i forsøget. Ved forsøgets start (tid 0) fik de 20 personer et måltid. 10 af personerne fik tilsat sødemidlet, de 10 øvrige fik ikke tilsat sødemidlet. Personerne fik derefter målt glukosekoncentrationen i blodet efter 15, 30, 45, 60, 90 og 120 minutter.

1. Foreslå forskellige summary measures, og opstil en model for sådan et summary measure. Hvordan skulle en sådan model fites i R? (Du behøver ikke overveje hvordan de foreslåede summary measures skal konstrueres; antag blot at de relevante vektorer/variable allerede findes.)
2. Antag nu at vi vil bruge alle observationerne i analysen. Hvilke faktorer bør indgå i analysen? Skal de være systematiske eller tilfældige? Lav faktordiagrammet.
3. Opstil en model med “random intercept” og forklar hvordan variansstrukturen i modellen er (hvad er variansen på en observation? hvilke observationer er uafhængige? hvad er korrelationen/kovariansen mellem observationer der ikke er uafhængige?) Hvordan ville du fitte modellen i R?
4. Opstil en model hvor korrelationen mellem to målinger fra samme individ aftager, jo længere observationerne kommer fra hinanden i tid. Hvordan ville du fitte modellen i R?
5. Antag at man også målte glukosekoncentrationen i blodet hos alle personer umiddelbart inden forsøgets start. Opstil (med udgangspunkt i modellen fra spørgsmål 4.) en model der udnytter disse målinger. Hvordan ville du fitte modellen i R?
6. Antag at tidligere forsøg har antydnet at glukose-profilen (tid-glukose grafen) efter et måltid ligner en parabel. Opstil (igen med udgangspunkt i modellen fra spørgsmål 4.) en model der modellerer denne struktur. Hvordan ville du fitte modellen i R?

## Opgave 6.5

Man regner med at det mandlige kønshormon testosteron påvirker foderoptag, fedtfordeling og glukosetolerancen. For at undersøge effekten af testosteronmangel blev 22 hanrotter fordelt tilfældigt i 2 grupper med 11 i hver hvor rotterne i den ene gruppe blev kastreret mens rotterne i den anden gruppe fik foretaget et indgreb uden effekt. Alle rotter fik målt koncentrationen af glukose i blodet både 2 og 10 uger efter det operative indgreb. I nedenstående tabel er denne koncentration angivet i mmol/l.

Kastreret			Kontrol		
Rotte	Uge		Rotte	Uge	
	2	10		2	10
1	6.7	5.9	12	5.3	5.1
2	6.0	5.4	13	4.8	5.4
3	6.0	5.8	14	4.3	4.9
4	6.0	5.6	15	5.6	5.2
5	6.3	5.6	16	4.6	4.8
6	5.8	5.7	17	5.5	5.1
7	6.1	6.0	18	6.1	5.2
8	5.4	5.7	19	5.1	5.7
9	7.1	5.5	20	5.6	5.6
10	5.2	4.8	21	5.9	4.8
11	6.1	6.2	22	6.8	5.0

(Data til denne opgave er blevet brugt i opgave 2 i eksamenssættet i Statistisk Forsøgsplanlægning, maj 2004.)

Vi skal i denne opgave undersøge hvordan koncentrationen af glukose afhænger af behandling og uge. Sidst i opgaven er der et uddrag af en R-kørsel, der kan bruges ved besvarelsen af spørgsmål 2. Data ligger i øvrigt i filen maj04\_2.txt (sammen med responsvariablen hørende til opgave 6.3 fra samme forsøg).

1. Tegn et faktordiagram for forsøget og opstil en statistisk model (hvis du ikke allerede har gjort det i opgave 6.2).
2. Analyser data. Angiv herunder slutmodellen samt estimer og konfidensintervaller for relevante parametre i slutmodellen. Formuler konklusionerne i ord.

Vi forestiller os nu at glukosekoncentrationen også blev målt 4, 7 og 13 uger efter det operative indgreb.

3. Opstil en statistisk model for det udvidede forsøg.

Vi forestiller os desuden at glukosekoncentrationen blev målt før det operative indgreb (uge 0).

4. Opstil en statistisk model der inddrager denne måling på passende måde.

R-kode og R-output som kan bruges til besvarelsen af spørgsmål 2:

```
> rottedata = read.table("../data/maj04_2.txt",header=T)
> attach(rottedata)

> rotte = factor(rotte)
> uge = factor(uge)

> library(nlme)

> modelA = lme(glucose ~ uge + behandling + uge:behandling,
               random =~ 1 | rotte, method="ML")
> modelB = lme(glucose ~ uge + behandling, random =~ 1 | rotte, method="ML")

> anova(modelB,modelA)
      Model df      AIC      BIC    logLik    Test  L.Ratio p-value
modelB      1  5 70.15805 79.07900 -30.07903
modelA      2  6 71.82897 82.53411 -29.91448 1 vs 2 0.3290866 0.5662

> modelC = lme(glucose ~ behandling, random =~ 1 | rotte, method="ML")
> modelD = lme(glucose ~ uge, random =~ 1 | rotte, method="ML")

> anova(modelC,modelB)
      Model df      AIC      BIC    logLik    Test  L.Ratio p-value
modelC      1  4 73.48170 80.61845 -32.74085
modelB      2  5 70.15805 79.07900 -30.07903 1 vs 2 5.323642 0.021

> anova(modelD,modelB)
      Model df      AIC      BIC    logLik    Test  L.Ratio p-value
modelD      1  4 78.70029 85.83705 -35.35015
modelB      2  5 70.15805 79.07900 -30.07903 1 vs 2 10.54224 0.0012

modelB1 = lme(glucose ~ uge + behandling, random =~ 1 | rotte, method="REML")

> summary(modelB1)

Random effects:
Formula: ~1 | rotte
(Intercept) Residual
StdDev: 0.1976841 0.4589773

Fixed effects: glucose ~ uge + behandling
              Value Std.Error DF   t-value p-value
(Intercept)  6.025000 0.1338500 21 45.01306 0.0000
uge10        -0.331818 0.1383869 21 -2.39776 0.0259
behandlingkontrol -0.568182 0.1620376 20 -3.50648 0.0022

> intervals(modelB1)
Fixed effects:
              lower      est.      upper
(Intercept)  5.7466436 6.0250000 6.30335637
uge10        -0.6196094 -0.3318182 -0.04402694
behandlingkontrol -0.9061864 -0.5681818 -0.23017721
```