

# Lidt repetition og mere om gentagne målinger

## Statistisk Dataanalyse 2

Anders Tolver



# Dagens program

Vi repeterer og uddyber en del af pensum fra kursusuge 1-5 gennem analyse af glukosemålinger på lam.

Vi kommer især ind på følgende emner

- faktorer, faktordiagrammer og modeller med tilfældige effekter
- værktøjer til undersøgelse af korrelationsstruktur (afhængighed) i forbindelse med analyse af gentagne målinger
  - empirisk korrelationsmatrixmodel og modelkontrol
  - AIC: til valg af passende variansstruktur
  - grafisk modelkontrol: semi-variogram, residualplot og qq-plot
- modelreduktion...
  - ...når en variabel kan indgå både som faktor og kovariat
  - ...når man vil slå udvalgte behandlingsgrupper sammen



## Glukose-indhold i blodet hos lam

20 får som var drægtige med tvillinger blev tilfældigt inddelt i 2 lige store behandlingsgrupper

- Får i gruppen **prenatal=LOW** fik foder med lavt kalorieindhold
- Får i gruppe **prenatal=NORM** fik normalt foder

Kort efter nedkomsten blev hvert par af tvillinge-lam separeret, og den ene tvilling fik i den næste periode normalt foder (**postnatal=LOW**), mens den anden tvilling fik en kalorierig diæt (**postnatal=HCHF**).

36 ud af 40 lam overlevede de første 6 måneder.

Da lammene var 6 måneder gamle måltet glukose-indholdet i blodet på et passende tidspunkt i forhold til fodring.

Datasættet er stillet til rådighed af Anna Hauntoft Kongsted.



# Glukose-indhold i blodet hos lam

```
head(data0,8)
```

##	lamb	sheep	prenatal	postnatal	glu
## 1	1114	2055	NORM	CONV	2.9066
## 6	1115	2055	NORM	HCHF	3.3730
## 11	1116	640	NORM	CONV	3.9439
## 16	1117	640	NORM	HCHF	2.8303
## 21	1118	729	NORM	HCHF	3.5611
## 26	1119	729	NORM	CONV	3.7856
## 31	1120	712	NORM	HCHF	3.3362
## 36	1121	712	NORM	CONV	3.0686

- Hvilke faktorer indgår i datasættet, og hvor mange niveauer har de enkelte faktorer?
- Hvilke faktorer bør indgå med tilfældig effekt i analysen?
- Tegn et faktordiagram og diskuter, hvilken type forsøg der er tale om.



# Statistisk analyse af glukose i 6 mdr gamle lam

Lad os i fællesskab...

- opskrive en statistisk model til analyse af forsøget
- fitte modellen i R
- kontrollere modelforudsætningerne
- reducere modellen mest muligt
- formidle konklusionerne og angive parameterestimer under slutmodellen



# Glukose-tolerance test på 6 mdr gamle lam

Da lammene var 6 måneder gamle udførtes en glukose-tolerance test, hvor glukoseindholdet i blodet blev målt ved baseline samt 2.5,10,20,30 og 60 minutter efter intravenøs injektion med en stor dosis glukose.

```
data<-read.table("../data/lam.txt",header=T)
data$timefac<-factor(data$time)
head(data,8)
```

##	lamb	sheep	prenatal	postnatal	time	glu	baseline	timefac
## 1	1114	2055	NORM	CONV	2.5	7.1661	2.9066	2.5
## 2	1114	2055	NORM	CONV	10.0	5.9055	2.9066	10
## 3	1114	2055	NORM	CONV	20.0	5.1310	2.9066	20
## 4	1114	2055	NORM	CONV	30.0	4.5036	2.9066	30
## 5	1114	2055	NORM	CONV	60.0	3.4464	2.9066	60
## 6	1115	2055	NORM	HCHF	2.5	8.6642	3.3730	2.5
## 7	1115	2055	NORM	HCHF	10.0	6.8156	3.3730	10
## 8	1115	2055	NORM	HCHF	20.0	4.7946	3.3730	20



# Middelværdistruktur

Følgende overvejelser kan være relevante, når vi skal beslutte os for middelværdistrukturen i modellen

- Hvilken responsvariabel skal benyttes?
- Hvilken/hvilke behandlingseffekt(er) skal indgå i modellen?
- Hvilke vekselvirkninger med behandlingen kan være interessante?
- Skal analysen justeres for andre faktorer / kovariater (typisk blot additivt som hovedeffekt)?
- Hvilke faktorer bør indgå med tilfældig effekt?



## Undersøgelse af den serielle korrelation

For at undersøge, om korrelationen (=afhængigheden) mellem glukose-målinger inden for samme lam, kan man undersøge den empiriske korrelationsmatrix

```
##           glu.2.5    glu.10    glu.20    glu.30    glu.60
## glu.2.5  1.0000000  0.1212765  0.4052189  0.1552730  0.3216885
## glu.10   0.1212765  1.0000000  0.3798282  0.1529574  0.0174670
## glu.20   0.4052189  0.3798282  1.0000000  0.6593987  0.5997320
## glu.30   0.1552730  0.1529574  0.6593987  1.0000000  0.7420988
## glu.60   0.3216885  0.0174670  0.5997320  0.7420988  1.0000000
```

Hvis korrelationen aftager med tidsintervallet mellem målinger, så er det vigtigt, at bruge en model med seriel korrelation.

Dette er tilsyneladende ikke nødvendigt her men lad os se ...

Mere teknisk info: se R-program.





# Sammenligning af korrelationsstrukturer (I)

## Først nogle modeller uden seriel korrelation

```
lm0RIinosheep<-lme(glu~prenatal*postnatal*timefac  
                  +baseline, random=~1|lamb,data)  
lm0RI<-lme(glu~prenatal*postnatal*timefac  
           +baseline, random=~1|sheep/lamb,data)
```

## Dernæst to eksempler på modeller med seriel korrelation

```
lm0diggle<-lme(glu~prenatal*postnatal*timefac  
              +baseline, random=~1|sheep/lamb  
,cor=corGaus(form=~time|sheep/lamb,nugget=T),data)  
lm0exp<-lme(glu~prenatal*postnatal*timefac  
            +baseline, random=~1|sheep/lamb  
,cor=corExp(form=~time|sheep/lamb,nugget=T),data)
```

Diggle-modellen har vi set i sidste uge, men her har vi både tilfældige effekter af **lamb** og den grovere faktor **sheep**.

Modellen **lm0exp** med en såkaldt *spatial exponential correlation* er blot vist som eksempel på en anden korrelationsstruktur i R.



## Sammenligning af korrelationsstrukturer (II)

Modeller med laveste AIC bør foretrækkes.

```
anova(lmORI,lmORInosheep,lm0diggle,lm0exp)
```

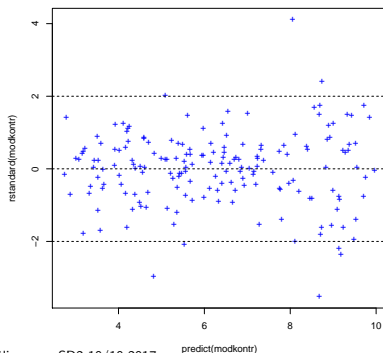
	Model	df	AIC	BIC	logLik
lmORI	1	24	392.7422	466.0921	-172.3711
lmORInosheep	2	23	392.3882	462.6818	-173.1941
lm0diggle	3	26	396.7422	476.2046	-172.3711
lm0exp	4	26	396.7422	476.2046	-172.3711

- Intet behov for at modellere seriel korrelationsstruktur.
- Modellen lmORInosheep er en smule bedre end lmORI målt ud fra AIC.
- Jeg vælger dog at gå videre med lmORI ....



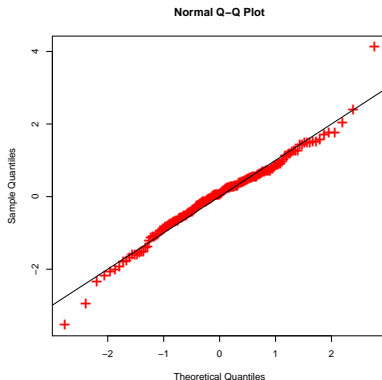
# Modelkontrol: lad alle effekter være systematiske ...

```
modkontr<-lm(glu~baseline+prenatal*postnatal*timefac  
             +factor(sheep)+factor(lamb), data)  
plot(predict(modkontr),rstandard(modkontr),pch="+",col=4)  
abline(h=c(-2,0,2),lty=2)
```



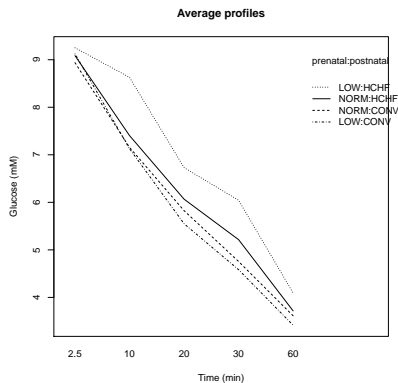
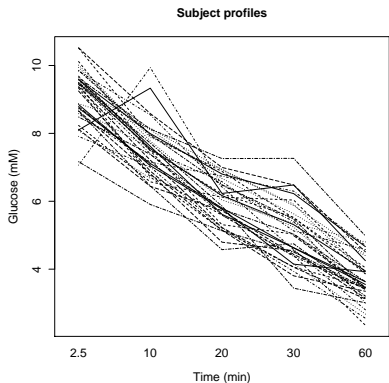
# Modelkontrol: lad alle effekter være systematiske ...

```
modkontr<-lm(glu~baseline+prenatal*postnatal*timefac  
             +factor(sheep)+factor(lamb), data)  
qqnorm(rstandard(modkontr),pch="+",col="red",cex=2)  
abline(0,1)
```



# Model reduktion: tegn først nogle figurer ...

Dette giver tit et godt billede af væsentlige strukturer i data.



# Opskriv modeller og forklar, hvad der testes for (I)

```
data$treat<-data$ prenatal: data$postnatal  
lm0<-lme(glu~baseline+treat*timefac, data  
, random=~1|sheep/lamb, method="ML")  
lm1<-lme(glu~treat*timefac  
, random=~1|sheep/lamb, data, method="ML")
```

```
anova(lm1, lm0)
```

##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
##	lm1	1 23	364.9620	438.1430	-159.4810			
##	lm0	2 24	364.0864	440.4492	-158.0432	1 vs 2	2.875592	0.0899



## Opskriv modeller og forklar, hvad der testes for (II)

```
lm1<-lme(glu~treat*timefac, random=~1|sheep/lamb
          ,data,method="ML")
lm2<-lme(glu~treat+timefac,random=~1|sheep/lamb
          ,data,method="ML")
```

```
anova(lm2,lm1)
```

##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
##	lm2	1 11	364.990	399.9897	-171.495			
##	lm1	2 23	364.962	438.1430	-159.481	1 vs 2	24.02806	0.0202

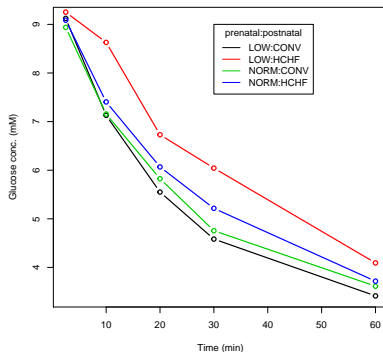
```
lm3<-lme(glu~treat*time,random=~1|sheep/lamb,data,method="ML")
```

```
anova(lm3,lm1)
```

##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
##	lm3	1 11	499.6195	534.6192	-238.8098			
##	lm1	2 23	364.9620	438.1430	-159.4810	1 vs 2	158.6576	<.0001



# Figurer med 'korrekt' tidsakse



Sørg for at inddrage kovariat (**time**) gennem relevant matematisk funktion!





# Tid som kovariat: forklar hvad der sker?

```
lm4<-lme(glu~treat*time+treat*I(time^2),random=~1|sheep/lamb
        ,data,method="ML")
anova(lm4,lm1)
```

```
##      Model df      AIC      BIC    logLik    Test  L.Ratio p-value
## lm4      1 15 365.127 412.8538 -167.5635
## lm1      2 23 364.962 438.1430 -159.4810 1 vs 2 16.16503 0.0401
```

```
lm4alt<-lme(glu~treat+treat:time+treat:I(time^2)-1,random=~1|sheep/lamb
            ,data,method="ML")
```

```
summary(lm4alt)$tTable
```

##		Value	Std.Error	DF	t-value	p-value
##	treatLOW:CONV	9.5420	0.2262	13	42.1809	0e+00
##	treatLOW:HCHF	9.8383	0.2680	13	36.7143	0e+00
##	treatNORM:CONV	9.3314	0.2262	13	41.2598	0e+00
##	treatNORM:HCHF	9.3903	0.2377	13	39.5055	0e+00
##	treatLOW:CONV:time	-0.2422	0.0161	135	-15.0879	0e+00
##	treatLOW:HCHF:time	-0.1606	0.0189	135	-8.5063	0e+00
##	treatNORM:CONV:time	-0.2161	0.0158	135	-13.6866	0e+00
##	treatNORM:HCHF:time	-0.1968	0.0166	135	-11.8240	0e+00
##	treatLOW:CONV:I(time^2)	0.0023	0.0002	135	9.7048	0e+00
##	treatLOW:HCHF:I(time^2)	0.0011	0.0003	135	3.8665	2e-04
##	treatNORM:CONV:I(time^2)	0.0020	0.0002	135	8.5303	0e+00
##	treatNORM:HCHF:I(time^2)	0.0017	0.0002	135	6.8384	0e+00



# Forslag til endelig analyse (I)

```
lm1<-lme(glu~treat*timefac,random=~1|sheep/lamb
        ,data,method="ML")
data$treatny<-factor(data$treat=="LOW:HCHF")
lm6<-lme(glu~treatny*timefac,random=~1|sheep/lamb
        ,data,method="ML")
```

```
anova(lm6,lm1)
```

##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
##	lm6	1 13	353.5237	394.8869	-163.7618			
##	lm1	2 23	364.9620	438.1430	-159.4810	1 vs 2	8.561694	0.5741

```
lm7<-lme(glu~timefac-1,random=~1|sheep/lamb
        ,data,method="ML")
```

```
anova(lm7,lm6)
```

##	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
##	lm7	1 8	378.4048	403.8591	-181.2024			
##	lm6	2 13	353.5237	394.8869	-163.7618	1 vs 2	34.88113	<.0001

AT — Repetition og gentagne målinger — SD2 10/10-2017

Dias 18/20



# Parameterestimer fra slutmodel

```
lm6refit<-lme(glu~timefac+treatny:timefac-1
              ,random=~1|sheep/lamb,method="REML",data)
summary(lm6refit)$tTable
```

##	Value	Std.Error	DF	t-value	p-value
## timefac2.5	9.0254	0.1352	133	66.7700	0.0000
## timefac10	7.2009	0.1352	133	53.2725	0.0000
## timefac20	5.7767	0.1387	133	41.6458	0.0000
## timefac30	4.8162	0.1352	133	35.6306	0.0000
## timefac60	3.5540	0.1352	133	26.2928	0.0000
## timefac2.5:treatnyTRUE	0.3228	0.2757	133	1.1708	0.2438
## timefac10:treatnyTRUE	1.5252	0.2757	133	5.5315	0.0000
## timefac20:treatnyTRUE	1.0500	0.2777	133	3.7810	0.0002
## timefac30:treatnyTRUE	1.3236	0.2757	133	4.8005	0.0000
## timefac60:treatnyTRUE	0.6351	0.2757	133	2.3033	0.0228



## Variansestimater fra slutmodel (lm6refit)

Som en del af `summary(lm6refit)` fås desuden

Random effects:

Formula: ~1 | sheep  
(Intercept)

StdDev: 0.3007142

Formula: ~1 | lamb %in% sheep  
(Intercept) Residual

StdDev: 0.3017676 0.5469781

- Opskriv slutmodellen på papir
- Hvordan kunne man teste for effekt af **lamb** og **sheep**?

