

# Eksamen i Statistisk Dataanalyse 2

## (kursusnr.: 210006)

8. april 2006

Alle sædvanlige hjælpemidler, herunder bøger og lommeregner men *ikke* PC, er tilladt. Der er syv sider med i alt tre opgaver der alle ønskes besvaret og indgår med samme vægt i bedømmelsen. Til opgave 1 og 3 er vedlagt udskrift af R-kørsler som kan benyttes i besvarelsen (det er ikke sikkert at alle dele af udskriften skal benyttes). Husk at det er vigtigt at specificere de statistiske modeller og hypoteser du bruger, og at komme med konklusioner på analyserne.

### Opgave 1 (4 spørgsmål)

For at undersøge riboflavins indflydelse på vækst udførte man et rotteforsøg hvor riboflavin blev givet i fire forskellige doser ( $2.5 \mu g$ ,  $5.0 \mu g$ ,  $10.0 \mu g$  hhv.  $20.0 \mu g$  per dag). 16 hanrotter og 16 hunrotter blev tilfældigt inddelt i de fire dosisgrupper med fire rotter af hvert køn i hver gruppe. Rotterne fik den angivne dosis dagligt i fire uger, og efter de fire uger blev deres ugentlige vækstrate (vægtforøgelse i g per uge) målt.

Vækstraterne for de 32 rotter er angivet i tabellen nedenfor.

Group	dose	log(dose)	Growth rate							
			Male				Female			
1	2.5	0.916	11.1	6.0	3.4	7.9	6.2	8.2	10.0	6.0
2	5.0	1.609	9.3	14.5	9.1	14.2	14.3	13.1	13.2	10.5
3	10.0	2.303	16.6	13.0	13.8	18.5	15.9	16.3	17.0	19.4
4	20.0	2.996	22.8	25.8	23.0	21.5	21.2	23.8	23.5	24.2

I besvarelsen kan du benytte udskriften fra R-kørslen sidst i opgaven.

I de to første spørgsmål skal dose benyttes som faktor, ikke som kovariat.

1. Analysér data, særligt med henblik på at undersøge om forskellige doser af riboflavin har forskellig indflydelse på rotters vækst. Angiv herunder estimater for parametrene i den endelige model. Husk også at formulere en konklusion i ord.
2. Angiv et estimat og et 95%-konfidensinterval for forskellen i vækstrate mellem rotter der har fået  $2.5 \mu g$  riboflavin per dag og rotter der har fået  $20.0 \mu g$  riboflavin per dag.
3. I forsøg af denne type ser man ofte at vækstraten er en lineær (retlinet) funktion af logaritmen til dosis. Undersøg om dette også kan antages at være tilfældet for dette forsøg.

4. Angiv et nyt estimat og et nyt 95%-konfidensinterval for forskellen i vækstrate mellem rotter der har fået 2.5  $\mu\text{g}$  riboflavin per dag og rotter der har fået 20.0  $\mu\text{g}$  riboflavin per dag.

Udskrift af R-kørsel (letterer redigeret):

```
##### Indlæsning af data mm.
```

```
> growthdata = read.table("growth.txt",header=T)
> attach(growthdata)
```

```
> growthdata
      sex dose growth
1  Male  2.5   11.1
2  Male  2.5    6.0
.
.
32 Female 20.0   24.3
```

```
> dosefac = factor(dose)
> logdose = log(dose)
```

```
##### Diverse modeller:
```

```
> modelA = lm(growth ~ sex + dosefac + sex:dosefac)
> modelB = lm(growth ~ dosefac + sex)
> modelC = lm(growth ~ dosefac)
> modelD = lm(growth ~ 1)
> modelE = lm(growth ~ logdose)
```

```
##### Diverse anova-tabeller:
```

```
> anova(modelB,modelA)
Analysis of Variance Table
```

```
Model 1: growth ~ dosefac + sex
Model 2: growth ~ sex + dosefac + sex:dosefac
  Res.Df    RSS Df Sum of Sq    F Pr(>F)
1     27 122.077
2     24 118.760  3     3.318 0.2235 0.8791
```

```
> anova(modelC,modelB)
Analysis of Variance Table
```

```
Model 1: growth ~ dosefac
Model 2: growth ~ dosefac + sex
  Res.Df    RSS Df Sum of Sq    F Pr(>F)
1     28 126.882
2     27 122.077  1     4.805 1.0627 0.3117
```

```

> anova(modelD,modelC)
Analysis of Variance Table

Model 1: growth ~ 1
Model 2: growth ~ dosefac
  Res.Df    RSS Df Sum of Sq    F    Pr(>F)
1      31 1209.74
2      28  126.88  3   1082.86 79.654 8.025e-14

> anova(modelE,modelC)
Analysis of Variance Table

Model 1: growth ~ logdose
Model 2: growth ~ dosefac
  Res.Df    RSS Df Sum of Sq    F Pr(>F)
1      30  140.58
2      28  126.88  2     13.70 1.5117  0.238

##### Diverse summary- og confint-kommandoer:

> summary(modelC)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.3500      0.7526   9.766 1.62e-10
dosefac5       4.9250      1.0644   4.627 7.68e-05
dosefac10      8.9625      1.0644   8.420 3.70e-09
dosefac20     15.8875      1.0644  14.927 7.38e-15

Residual standard error: 2.129 on 28 degrees of freedom
Multiple R-Squared: 0.8951,    Adjusted R-squared: 0.8839
F-statistic: 79.65 on 3 and 28 DF,  p-value: 8.025e-14

> confint(modelC)
              2.5 %    97.5 %
(Intercept)  5.808324  8.891676
dosefac5     2.744741  7.105259
dosefac10    6.782241 11.142759
dosefac20   13.707241 18.067759

> summary(modelE)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.2044      1.0389   0.197  0.845
logdose       7.4587      0.4938  15.105 1.46e-15

Residual standard error: 2.165 on 30 degrees of freedom
Multiple R-Squared: 0.8838,    Adjusted R-squared: 0.8799
F-statistic: 228.2 on 1 and 30 DF,  p-value: 1.457e-15

> confint(modelE)
              2.5 %    97.5 %
(Intercept) -1.917378  2.326141
logdose      6.450262  8.467205

```

## Opgave 2 (4 spørgsmål)

I et markforsøg ønsker man at undersøge effekten af to forskellige typer gødning på udbyttet af en bestemt hvedesort. Man råder over 12 forsøgsenheder (plots), som er fordelt på 4 marker (blokke) hver opdelt i 3 plots. Indledningsvis forestiller man sig at anvende gødningstype A på 4 plots, gødningstype B på 4 plots og gødningstype 0 (dvs. ingen gødning) på 4 plots.

1. Foreslå en forsøgsplan og beskriv (kort) hvordan den bør gennemføres med hensyn til randomisering. Opskriv en statistisk model til analyse af resultaterne.

Inden forsøget går i gang, beslutter man sig for i stedet at anvende hver af gødningstyperne A og B i både en høj og en lav dosis, mens ingen plots lades ubehandlede. Man ønsker at lave et balanceret ufuldstændigt blokforsøg (BIBD).

2. Angiv, f.eks. ved at sætte krydser i et skema som det viste, hvordan en forsøgsplan kunne se ud. Forklar endvidere hvordan randomiseringen bør foretages.

Gødning	Dosis	Blok			
		1	2	3	4
A	Lav				
A	Høj				
B	Lav				
B	Høj				

3. Argumentér for, at det ikke er muligt at lave et balanceret ufuldstændigt blokforsøg på 5 blokke, hvis hver kombination af gødningstype og dosis ønskes afprøvet på 3 plots, og der desuden er 3 plots, som slet ikke bliver behandlet med gødning. Det antages stadig at hver blok skal opdeles i 3 plots.

Det følgende år planlægges et mere omfattende markforsøg, hvor gødningstyperne A og B igen skal afprøves i både høj og lav dosis. Desuden ønsker man at inddrage endnu en hvedesort i forsøget, så der i alt indgår to hvedesorter.

4. Beskriv hvordan man kan lave en forsøgsplan, hvor markforsøget udføres på 2 forskellige marker (blokke), som hver inddeles i 4 plots, hvis man samtidig ønsker at konfundere vekselvirkningen mellem gødning og dosis med blok.

### Opgave 3 (5 spørgsmål)

I et markforsøg ønskede man (i) at sammenligne en blanding af bygsorter med sorterne benyttet enkeltvis og (ii) at undersøge effekten af sprøjtning med bayleton.

I forsøget indgik 8 plots, hver inddelt i 5 parceller (40 parceller i alt). Fem sorter (Lami, Zita, Salka, Lofa og en blanding af de fire) blev tilfældigt udlagt, således at hver af dem optrådte på netop en parcel i hvert plot. Endvidere blev samtlige parceller i fire af plottene sprøjtet med bayleton, mens samtlige parceller i de øvrige fire plots forblev usprøjtede.

I tabellen nedenfor er udbyttet angivet for hver af de 40 parceller (i *hkg* per hektar).

Plot	Bayleton	Sort				
		Blanding	Lami	Lofa	Salka	Zita
1	Ja	49.4397	51.3836	48.6232	48.6232	48.7334
2	Ja	50.8523	53.4953	50.7372	52.1466	50.8523
3	Ja	51.5586	53.4953	45.8045	51.4419	50.1460
4	Ja	53.6774	54.1991	44.3951	54.2607	51.5586
5	Nej	51.9692	56.8205	47.1603	52.7914	54.1376
6	Nej	52.6715	54.7161	47.1603	52.0875	51.3253
7	Nej	48.4578	54.0146	51.3836	49.9758	50.6222
8	Nej	44.2441	53.3131	47.8642	49.9758	52.0284

I besvarelsen kan du benytte udskriften fra R-kørslen sidst i opgaven.

1. Lav et faktordiagram (med frihedsgrader) og opstil en statistisk model. Hvilken type forsøg er der tale om?
2. Analysér data, med henblik på at undersøge om sprøjtemidlet bayleton har en effekt på udbyttet og om de fem sorter giver forskelligt udbytte. Angiv herunder estimer for parametrene i slutmodellen. Husk også at formulere en konklusion i ord.
3. I forhold til sorterne var man først og fremmest interesseret i forskellen mellem udbyttet fra blandingen og det *gennemsnitlige* udbytte af de fire rene sorter. Beregn et estimat for denne forskel.
4. Opstil hypotesen om at blandingen giver samme udbytte som gennemsnittet af de fire rene sorter, og forklar hvordan du ville teste hypotesen i R.
5. Faktisk lå plottene to og to sammen på fire marker, således at der på hver mark var både et sprøjtet og et usprøjtet plot. Opstil en statistisk model der tager hensyn til variationen mellem marker, og lav det tilhørende faktordiagram. Angiv en R-kommando som kan bruges til at analysere modellen.

Udskrift af R-kørsel (lettere redigeret):

```
### Indlæsning af data mm.
```

```
> meldug = read.table("meldug.txt",header=T)
> attach(meldug)
```

```
> meldug
      sort bayleton udbytte plot
1      Zita      ja 48.7334    1
2      Zita      ja 50.8523    2
3      Zita      ja 50.1460    3
4      Zita      ja 51.5586    4
5      Lofa      ja 48.6232    1
.
.
40 Blanding     nej 44.2441    8
```

```
> plot = factor(plot)
> library(nlme)
```

```
### Diverse modeller:
```

```
> modelA = lme(udbytte ~ sort + bayleton + bayleton:sort, random =~ 1| plot)
> modelB = lme(udbytte ~ sort + bayleton, random =~ 1| plot)
> modelC = lme(udbytte ~ sort, random =~ 1| plot)
> modelC1 = lme(udbytte ~ sort-1, random =~ 1| plot)
```

```
### Diverse anova-tabeller:
```

```
> anova(modelA)
              numDF denDF    F-value p-value
(Intercept)      1    24 14544.711  <.0001
sort              4    24    9.734  0.0001
bayleton          1     6    0.186  0.6811
sort:bayleton     4    24    1.301  0.2978
```

```
> anova(modelB)
              numDF denDF    F-value p-value
(Intercept)      1    28 14544.653  <.0001
sort              4    28    9.333  0.0001
bayleton          1     6    0.186  0.6811
```

```
> anova(modelC)
              numDF denDF    F-value p-value
(Intercept)      1    28 16457.626  <.0001
sort              4    28    9.333  1e-04
```

### Diverse summary-kommandoer:

```
> summary(modelC)
```

Linear mixed-effects model fit by REML

Random effects:

Formula: ~1 | plot

(Intercept) Residual

StdDev: 0.6738702 2.009896

Fixed effects: udbytte ~ sort

	Value	Std.Error	DF	t-value	p-value
(Intercept)	50.35882	0.7494817	28	67.19153	0.0000
sortLami	3.57088	1.0049480	28	3.55329	0.0014
sortLofa	-2.46778	1.0049480	28	-2.45562	0.0205
sortSalka	1.05404	1.0049480	28	1.04885	0.3032
sortZita	0.81665	1.0049480	28	0.81263	0.4233

```
> summary(modelC1)
```

Linear mixed-effects model fit by REML

Random effects:

Formula: ~1 | plot

(Intercept) Residual

StdDev: 0.6738702 2.009896

Fixed effects: udbytte ~ sort - 1

	Value	Std.Error	DF	t-value	p-value
sortBlanding	50.35882	0.7494817	28	67.19153	0
sortLami	53.92970	0.7494817	28	71.95599	0
sortLofa	47.89105	0.7494817	28	63.89889	0
sortSalka	51.41286	0.7494817	28	68.59789	0
sortZita	51.17547	0.7494817	28	68.28115	0