

Statistisk Dataanalyse 2, april 2006

Vejledende besvarelse

Opgave 1

1. Vi opstiller en model med dosis, køn og deres vekselvirkning:

$$A : Y_i = \gamma(\text{dose}_i, \text{sex}_i) + e_i$$

hvor Y_i er vækstraten for i 'te rotte og e_1, \dots, e_{32} er uafhængige, $N(0, \sigma^2)$ -fordelte. Dette er modelA i R-udskriften.

Vi tester først hypotesen om at der ikke er vekselvirkning mellem dosis og køn:

$$H_0 : \gamma(\text{dose}_i, \text{sex}_i) = \alpha(\text{dose}_i) + \beta(\text{sex}_i)$$

Dette svarer til model modelB i udskriften. Fra R-udskriften får vi F -teststørrelsen $F_{A \rightarrow B} = 0.22$ og $p = 0.88$. Hypotesen accepteres altså.

Vi tester derefter hypotesen om at køn ingen effekt har, dvs. $H_0 : \beta(\text{Male}) = \beta(\text{Female})$. Dette svarer til modelC, og vi ser at $F_{B \rightarrow C} = 1.06$, $p = 0.31$. Vi accepterer således hypotesen.

Vi tester endelig hypotesen om at dosis ingen effekt har, dvs. $H_0 : \alpha(2.5) = \dots = \alpha(20.0)$. Dette svarer til modelD, og vi ser at $F_{C \rightarrow D} = 79.7$, $p = < 0.0001$. Hypotesen må således forkastes.

Den endelige model er således modelC:

$$C : Y_i = \alpha(\text{dose}_i) + e_i$$

og vi konkluderer således vækstraten er ens for hanner og hunner, men at vækstraten er forskellig i de fire dosisgrupper.

Fra `summary(modelC)` finder vi estimerne:

$$\hat{\alpha}(2.5) = 7.35$$

$$\hat{\alpha}(5.0) = 7.35 + 4.93 = 12.28$$

$$\hat{\alpha}(10.0) = 7.35 + 8.96 = 16.31$$

$$\hat{\alpha}(20.0) = 7.35 + 15.89 = 23.24$$

og ser at vækstraten altså vokser med voksende indtag af riboflavin. Desuden er estimeret for spredningen lig $s = 2.129$.

2. Forskellen i forventet vækstrate ved dosis $2.5 \mu g$ og $20.0 \mu g$ er

$$\hat{\alpha}(20.0) - \hat{\alpha}(2.5) : 15.89 \quad [13.71, 18.07].$$

(Både estimat og konfidensinterval aflæses direkte fra R-udskriften.)

3. Antagelsen om at vækstraten vokser lineært i logaritmen til dosis svarer til hypotesen

$$H_0 : \alpha(\text{dose}_i) = \delta + \phi \cdot \log(\text{dose}_i)$$

eller modellen

$$E : Y_i = \delta + \phi \cdot \log(\text{dose}_i) + e_i$$

Hypotesen testes i modelC og vi ser fra R-udskriften at $F_{C \rightarrow E} = 1.51$, $p = 0.24$. Hypotesen om linearitet i log-dosis kan således ikke afvises.

4. I modelE er forskellen i forventet vækstrate ved dosis 2.5 μg og 20.0 μg givet ved

$$d = (\delta + \phi \cdot \log(20.0)) - (\delta + \phi \cdot \log(2.5)) = \phi \cdot (\log(20.0) - \log(2.5))$$

Fra R-udskriften finder vi estimat ($\hat{\phi} = 7.46$) og konfidensinterval ([6.45, 8.47]) for ϕ . Indsættes disse i udtrykket for d får vi

$$\hat{d} = 7.46 \cdot (\log(20.0) - \log(2.5)) = 15.51$$

og konfidensinterval [13.41, 17.61].

Opgave 2

1. Forsøget udføres som et fuldstændigt balanceret blokforsøg, hvor hver af de 3 gødningstyper (GØDN) afprøves en gang på hver blok (BLOK). Inden for de enkelte blokke randomiseres gødningstyperne på de 3 plots.

Data analyseres ved den statistiske model, hvor GØDN indgår som systematisk faktor og BLOK som tilfældig faktor:

$$Y_i = \alpha(\text{GØDN}_i) + \beta(\text{BLOK}_i) + e_i,$$

hvor e_1, \dots, e_{12} er uafhængige og normalfordelte $N(0, \sigma^2)$ og $\beta(b_j), j = 1, \dots, 4$, er uafhængige og normalfordelte $N(0, \sigma_{\text{BLOK}}^2)$.

2. På hver af de 4 blokke skal udføres forsøg med 3 kombinationer af produktfaktoren GØDN \times DOSIS, således at hvert par af kombinationer optræder præcis lige mange gange (nemlig 1) inden for samme blok. En mulig løsning fremgår af nedenstående tabel:

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

Ved den praktiske udførelse af forsøget skal de forskellige marker først randomiseres på blok 1, 2, 3 og 4. Derefter randomiseres de tre behandlinger (kombinationer af GØDN og DOSIS) på de tre plots inden for den tildelte mark.

3. For det ønskede design har vi:

$$\text{Antal blokke} = 5$$

$$\text{Antal behandlinger} = 5$$

$$\text{Antal behandlinger pr. blok} = 3$$

Heraf fås umiddelbart, at hver behandling må optræde 3 gange i forsøgsplanen. Antallet af gange hver kombination af behandlingerne skal optræde inden for samme blok bliver:

$$\lambda = \frac{3 \cdot (3 - 1)}{(5 - 1)} = 3/2.$$

Det er således *ikke* muligt at lave et balanceret ufuldstændigt blokforsøg med 5 behandlinger på 5 blokke, som hver er inddelt i 3 plots.

4. Det ønskede forsøg indeholder 3 behandlingsfaktorer (SORT, GØDN og DOSIS) hver med to niveauer, som vi i det følgende blot vil kalde niveau 1 og niveau 2. En mulig forsøgsplan på 2 blokke med hver 4 plots, hvor GØDN \times DOSIS konfunderes med blok, kunne se ud som følger:

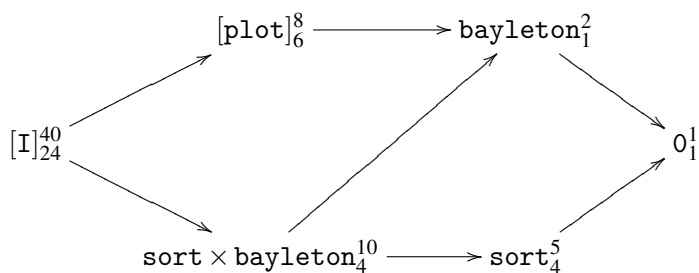
Sort	Gødning	Dosis	Gødning+Dosis	Blok
1	1	1	2	1
1	1	2	3	2
1	2	1	3	2
1	2	2	4	1
2	1	1	2	1
2	1	2	3	2
2	2	1	3	2
2	2	2	4	1

For hver enkelt behandling benyttes summen af niveauerne for GØDN og DOSIS til at afgøre, hvilken blok behandlingen skal udføres på. I ovenstående eksempel er en ulige værdi af Gødning+Dosis blevet knyttet til blok nr. 2.

Opgave 3

1. Forsøget er et split-plot forsøg med plot som helplot, parceller som delplot, bayleton som helplotfaktor og sort som delplotfaktor.

Faktordiagrammet er derfor identisk med faktordiagrammet i kompendiets figur 8.2 med $m = 5$, $k = 2$, $n = 4$.



Den tilhørende model er

$$A : Y_i = \gamma(\text{sort}_i, \text{bayleton}_i) + A(\text{plot}_i) + e_i$$

hvor $A(1), \dots, A(8)$ er $N(0, \sigma_A^2)$ -fordelte og e_1, \dots, e_{40} er $N(0, \sigma^2)$ -fordelte, alle indbyrdes uafhængige. Dette svarer til `modelA` i R-udskriften.

2. Vi tester først for vekselvirkning, dvs. tester hypotesen

$$H_0 : \gamma(\text{sort}_i, \text{bayleton}_i) = \alpha(\text{sort}_i) + \beta(\text{bayleton}_i)$$

Fra R-udskriften ses at $F = 1.30$, $p = 0.30$, så der er ingen tegn på vekselvirkning. Dette svarer til

$$B : Y_i = \alpha(\text{sort}_i) + \beta(\text{bayleton}_i) + A(\text{plot}_i) + e_i$$

Vi tester så hypotesen om at der ingen effekt er af sprøjtning,

$$H_0 : \beta(\text{ja}) = \beta(\text{nej}).$$

Vi ser at $F = 0.19$, $p = 0.68$, så vi kan ikke påvise en effekt af bayleton. Dette svarer til

$$C : Y_i = \alpha(\text{sort}_i) + A(\text{plot}_i) + e_i$$

Vi tester så hypotesen om at der ikke er forskel på sorterne,

$$H_0 : \alpha(\text{Blanding}) = \dots = \alpha(\text{Zita})$$

Vi ser at $F = 9.33$, $p = 0.0001$, så der er klar forskel på sorterne. Slutmodellen er således `modelC`.

Fra estimererne ser vi at sorten Lami giver højst udbytte, derefter følger Salka, Zita, blandingen, og endelig giver Lofa lavest udbytte. Hverken Salka eller Zita giver dog signifikant højere udbytte end blandingen.

Desuden finder vi følgende estimer for varianskomponenterne:

$$\hat{\sigma}_A^2 = 0.674^2 = 0.454, \quad \hat{\sigma}^2 = 2.010^2 = 4.040$$

Ovenstående svar vedr. sorterne er fuldt tilstrækkeligt, men faktisk kan vi fra udskriften fra `summary(modelC)` næsten finde LSD-værdien: standard error for forskellene er 1.005, LSD-værdien er derfor cirka $2 \cdot 1.005 = 2.01$. (Præcist: $t_{0.975, 28} \cdot 1.005 = 2.05 \cdot 1.005 = 2.06$). Dette kan bruges til fortolkning som sædvanlig.

3. Den ønskede forskel er

$$d = \alpha(\text{Blanding}) - \frac{1}{4}(\alpha(\text{Lami}) + \dots + \alpha(\text{Zita}))$$

som estimeres ved at indsætte α -estimerne:

$$\hat{d} = 50.36 - \frac{1}{4}(53.93 + 47.89 + 51.41 + 51.18) = -0.74.$$

4. Hypotesen er at $d = 0$, dvs. at

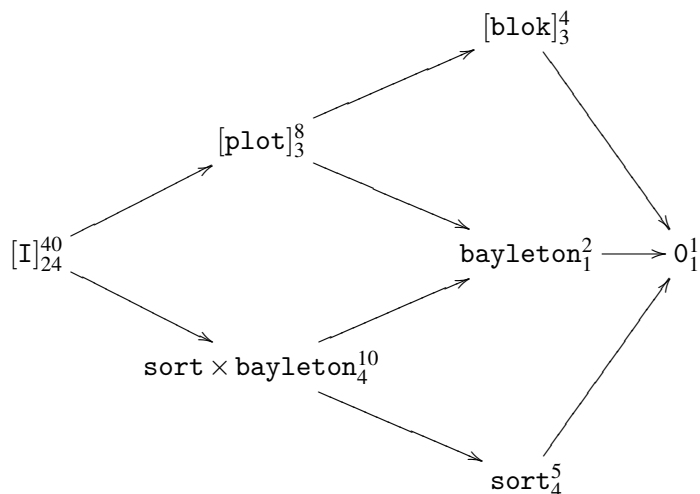
$$\alpha(\text{Blanding}) = \frac{1}{4}(\alpha(\text{Lami}) + \dots + \alpha(\text{Zita}))$$

Man kan bruge `estimable` til at teste hypotesen, fx. vha. følgende kommando (pakken `gmodels` skal være "loaded"):

```
> d = c(1,-0.25,-0.25,-0.25,-0.25)
> estimable(modelC1,d)
```

Dette giver bla. en p -værdi for testet af hypotesen.

- Den vigtige observation er at plot er finere en blok (dvs. mark). Faktordiagrammet svarer således til det i kompendiets figur 8.3.



Den tilhørende model er givet ved

$$A : Y_i = \gamma(\text{sort}_i, \text{bayleton}_i) + A(\text{plot}_i) + B(\text{blok}_i) + e_i$$

hvor $A(1), \dots, A(8)$ er $N(0, \sigma_A^2)$ -fordelt, $B(1), \dots, B(4)$ er $N(0, \sigma_B^2)$ -fordelt, og e_1, \dots, e_{40} er $N(0, \sigma^2)$ -fordelt, alle indbyrdes uafhængige.

Modellen kan fittes med R-kommandoen

```
lme(udbytte ~ sort + bayleton + bayleton:sort, random =~ 1| blok/plot)
```