

## Flerfaktorforsøg og faktordiagrammer

### Statistisk Dataanalyse 2

Anders Tolver

Uge 2, tirsdag d. 12/9-2017



## Eksempel 3.2: 2-sidet variansanalyse

Vi har diskuteret den **fulde model med vekselvirkning**

$$Y_i = \gamma(\text{TREAT} \times \text{TIME}_i) + e_i,$$

hvor  $e_i$  er uafh. normalfordelte  $N(0, \sigma^2)$  og den **additive model** for tosidet variansanalyse givet ved

$$Y_i = \alpha(\text{TREAT}_i) + \beta(\text{TIME}_i) + e_i,$$

hvor  $e_i$  er uafh. normalfordelte  $N(0, \sigma^2)$ .

**Spm:** Hvornår kan man ikke teste den additive model mod modellen med vekselvirkning? (-se opgave 2.3)



## Dagens program

Slide 1-10 introducerer begrebet **faktordiagram** og er i høj grad identisk med udvalgte slides til torsdag d. 7/9-2017

- Tosidet variansanalysemodel (repetition)
- Faktordiagram

Dagens hovedprogram

- Firkfaktorforsøg: holdbarhed af afskårne roser
  - udvælg passende hoved- og vekselvirkninger
  - faktordiagram
  - reduktion af model (rækkefølge)
  - konklusioner

Anders Tolver — Flersidet ANOVA — SD2 12/9-2017  
Dias 2/28



## 2-way ANOVA: faktordiagram I

**Faktordiagrammer** benyttes til at skabe sig overblik over strukturen i et forsøgsdesign.

Mandagens gennemgående eksempel indholdt 3 egentlige faktorer

**TREAT, TIME, TREAT  $\times$  TIME**

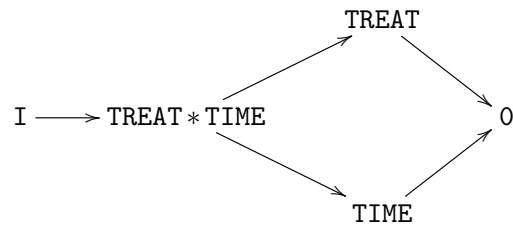
samt de trivielle faktorer

**I, 0**



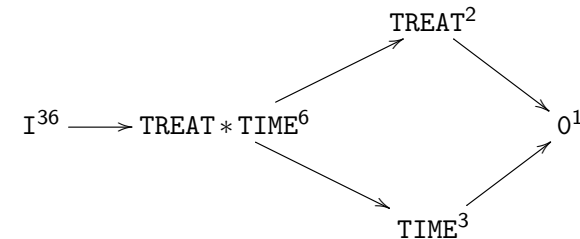
### Eksempel 3.2: balanceret tofaktorforsøg

Der tegnes pile fra **finere** faktorer til **grovere**. Fineste faktor (-enhedsfaktoren I) placeres til venstre på tegningen.



### Eksempel 3.2: balanceret tofaktorforsøg

Der tegnes pile fra **finere** faktorer til **grovere**. Fineste faktor (-enhedsfaktoren I) placeres til venstre på tegningen.

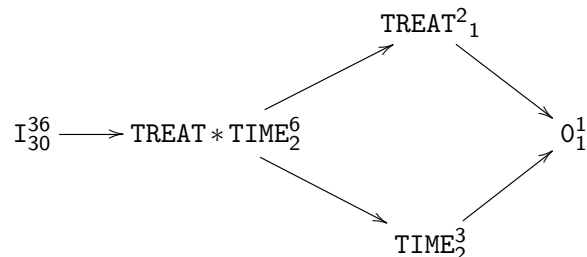


Antallet af **niveauer** skrives i øverste højre hjørne.



### Eksempel 3.2: balanceret tofaktorforsøg

Der tegnes pile fra **finere** faktorer til **grovere**. Fineste faktor (-enhedsfaktoren I) placeres til venstre på tegningen.



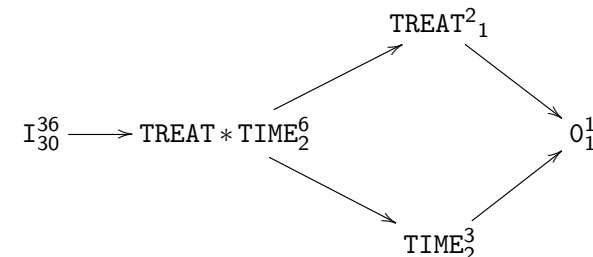
Antallet af **niveauer** skrives i øverste højre hjørne.

Antallet af **df** tilføjes i nederste højre hjørne, ved fra antallet af niveauer at fratrække **df** for grovere faktorer.



### Eksempel 3.2: balanceret tofaktorforsøg

Der tegnes pile fra **finere** faktorer til **grovere**. Fineste faktor (-enhedsfaktoren I) placeres til venstre på tegningen.



Antallet af **niveauer** skrives i øverste højre hjørne.

Antallet af **df** tilføjes i nederste højre hjørne, ved fra antallet af niveauer at fratrække **df** for grovere faktorer.

**Ex:** Ud for  $TREAT \times TIME$  skrives  $6-1-2-1=2!$



## Trefaktorforsøg

Fra noternes eksempel 2.2.

	Lys1				Lys2			
	Kammer 1	Kammer 2	Kammer 3	Kammer 4	Kammer 1	Kammer 2	Kammer 3	Kammer 4
Gødning	*	*	*	*	*	*	*	*
Ingen gødning	*	*	*	*	*	*	*	*

**Faktorer:** G (Gødning), K (Kammer) og L (Lys)

Bemærk: K er finere end L!

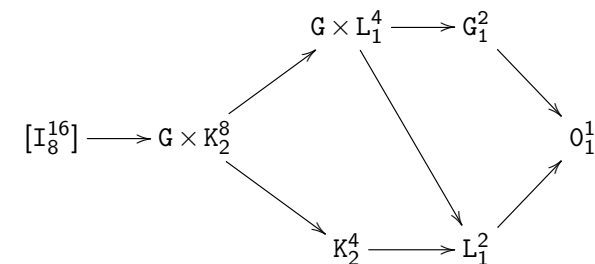
**Vekselvirkninger:**  $G \times K$ ,  $G \times L$ ,  $(K \times L = K)$

**Trivielle faktorer:** I, 0

Lad os forsøge at tegne det tilhørende faktordiagram.



## Trefaktorforsøg



Tilhørende statistiske model:

$$Y_i = \gamma(G \times K_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$

G: Gødning

K: Kammer

L: Lys

**NB:** Vi sætter [...] omkring faktorer som ikke indgår i den systematiske del af modellen!



## Faktordiagrammer: fordele

Et faktordiagram giver overblik over

- forsøgsdesignet/-planen
- hvilken rækkefølge man kan teste for reduktion i modellen

Er designet balanceret (eller blot ortogonalt) kan man endvidere

- afgøre hvilke hoved-/vekselvirkninger der overhovedet kan testes for?
- beregne frihedsgrader og teststørrelser ved 'håndkraft'



## Opgave 2.2: tresidet variansanalyse

Oversigt over datasæt og forsøgsplan

```
##      TEMP  LUC  ADP mineral
## 1      10    1    1    5.30
## 2      10    1    0    2.30
## 3      20    1    1    5.20
## 4      20    1    0    3.59
## 5      10    1    1    4.84
## 6      10    1    0    2.26
## 7      20    1    1    5.60
## 8      20    1    0    3.48
## 9      10    0    1    5.21
## 10     10    0    0    2.27
## 11     20    0    1    5.77
## 12     20    0    0    3.53
## 13     10    0    1    5.56
## 14     10    0    0    2.77
## 15     20    0    1    5.64
## 16     20    0    0    3.47
```

Faktorer: **TEMP, LUC, ADP**

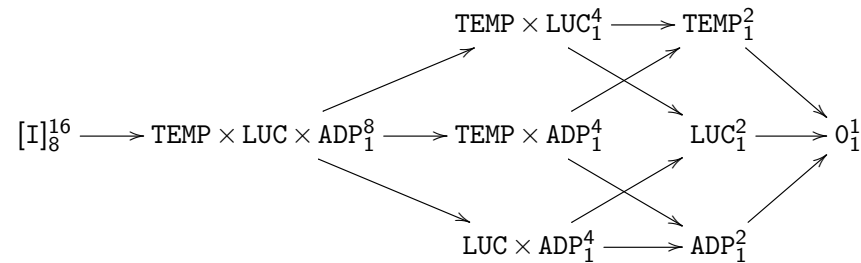
Vekselvirkn: **TEMP×LUC, TEMP×ADP, LUC×ADP**

Andenordens vekselvirkn: **TEMP×LUC×ADP**



## 3-way ANOVA: faktordiagram

Der tegnes pile fra **finere** faktorer til **grovere**.



Antallet af niveauer skrives i øverste højre hjørne.

Antallet af df tilføjes i nederste højre hjørne, ved fra antallet af niveauer at fratrække df for grovere faktorer.

Forsøget kaldes et **balanceret trefaktorforsøg**.



## Dagens hovedeksempel: holdbarhed af afskårne roser

Et forsøg er blevet udført med det formål at undersøge effekten af et middel som forøger holdbarheden af afskårne roser.

Datasættet består af 24 målinger af holdbarheden af bundter af roser, som er fordelt på 3 forskellige flor. Endvidere er registreret om midlet blev tilsat hos gartneren, hos blomsterhandleren og hos kunden.

**Faktorer og niveauer:**

FLOR(1,2,3)

GARTNER(0,1)

HANDLER(0,1)

KUNDE(0,1)



## Afskårne roser: data

```
roses<-read.table(file="../data/Roser.txt",header=T)
roses
```

```
##      Obs gartner handler kunde flor  tid
## 1      1      0      0      0     1 10.1
## 2      2      0      0      0     2  8.9
## 3      3      0      0      0     3 11.4
## 4      4      1      0      0     1 10.9
## 5      5      1      0      0     2  6.9
## 6      6      1      0      0     3 11.2
## 7      7      0      1      0     1  9.6
## 8      8      0      1      0     2  9.9
## 9      9      0      1      0     3 11.3
## 10    10      0      0      1     1 11.8
```

[ ... more datalines here ... ]



## Afskårne roser: spørgsmål

- Virker konserveringsmidlet?
- Betyder det noget, hvor man tilsætter konserveringsmidlet?
- Hvordan opnår man den bedste holdbarhed af roserne?
- Hvor "stor" en effekt har konserveringsmidlet?
- Er der forskelle fra flor til flor?



## Afskårne roser: faktorer i forsøget

Tredjeordens vekselvirkning

$$\text{GARTNER} \times \text{HANDLER} \times \text{KUNDE} \times \text{FLOR}$$

Andenordens vekselvirkninger

$$\text{GARTNER} \times \text{HANDLER} \times \text{KUNDE}, \text{FLOR} \times \text{HANDLER} \times \text{KUNDE}$$

$$\text{GARTNER} \times \text{HANDLER} \times \text{FLOR}, \text{GARTNER} \times \text{FLOR} \times \text{KUNDE}$$

Vekselvirkninger

$$\text{GARTNER} \times \text{HANDLER}, \text{KUNDE} \times \text{HANDLER}, \text{KUNDE} \times \text{FLOR}$$

$$\text{GARTNER} \times \text{KUNDE}, \text{GARTNER} \times \text{FLOR}, \text{FLOR} \times \text{HANDLER}$$

Hovedvirkninger

$$\text{GARTNER}, \text{HANDLER}, \text{KUNDE}, \text{FLOR}, \text{I og O}$$



## Afskårne roser: udgangsmodel

Hjemmeopgave: Lav et overskueligt faktordiagram med alle faktorerne fra foregående slide :-)

Vi tager udgangspunkt i modellen med faktorerne

$$\text{GARTNER} \times \text{HANDLER} \times \text{KUNDE}$$

$$\text{GARTNER} \times \text{HANDLER}, \text{KUNDE} \times \text{HANDLER}$$

$$\text{GARTNER} \times \text{KUNDE}$$

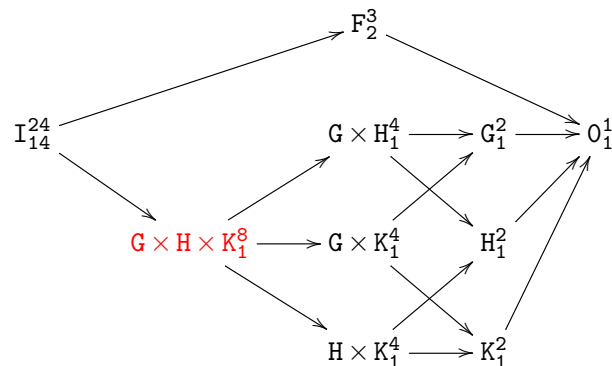
$$\text{GARTNER}, \text{HANDLER}, \text{KUNDE}, \text{FLOR}, \text{I og O}$$

Vi er kun sekundært interesseret i faktoren FLOR.

Faktorer som bidrager til variationen, men som ikke har vores primære interesse, vil man typisk inddrage som tilfældige effekter (-mere om dette i uge 4-5!).



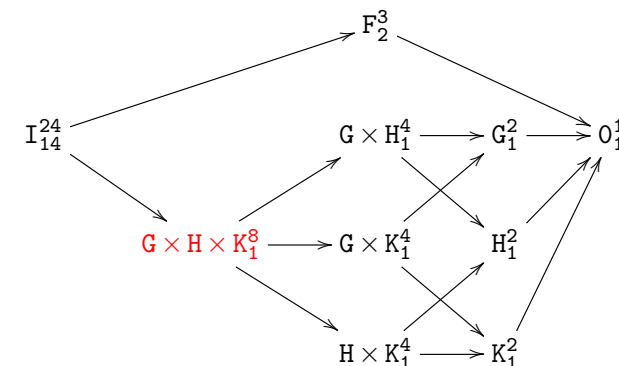
## Afskårne roser: faktordiagram og reduktion



Statistisk model (hørende til faktordiagram) og hypotese:



## Afskårne roser: faktordiagram og reduktion



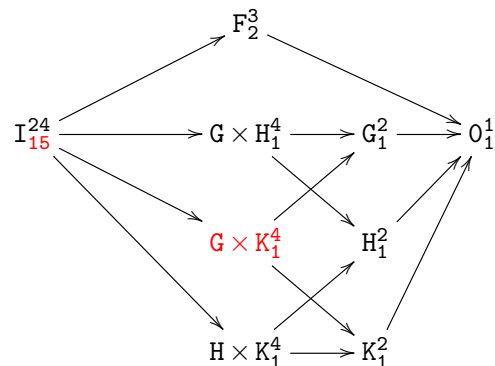
Statistisk model (hørende til faktordiagram) og hypotese:

$$M: Y_i = \gamma(G \times H \times K_i) + \delta(F_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$

$$H_0: Y_i = \alpha(G \times H_i) + \beta(G \times K_i) + \gamma(H \times K_i) + \delta(F_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$



## Afskårne roser: faktordiagram og reduktion

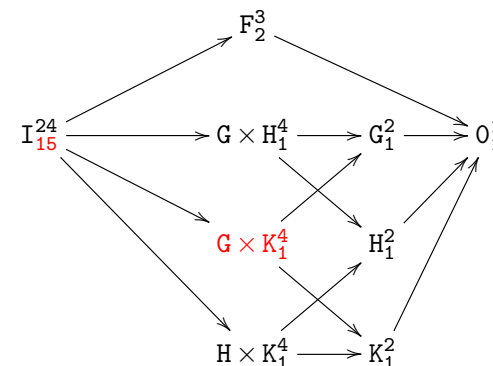


Statistisk model (hørende til faktordiagram) og hypotese:

Anders Tolver — Flersidet ANOVA — SD2 12/9-2017  
Dias 17/28



## Afskårne roser: faktordiagram og reduktion



Statistisk model (hørende til faktordiagram) og hypotese:

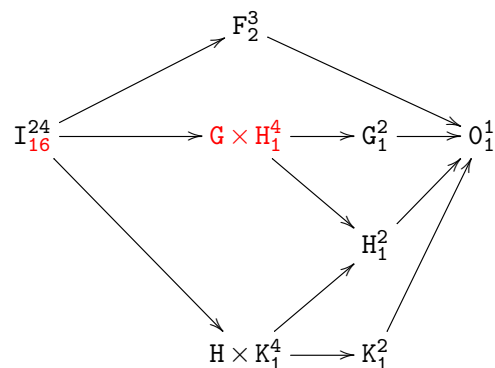
$$M: Y_i = \alpha(G \times H_i) + \beta(G \times K_i) + \gamma(H \times K_i) + \delta(F_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$

$$H_0: Y_i = \alpha(G \times H_i) + \gamma(H \times K_i) + \delta(F_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$

Anders Tolver — Flersidet ANOVA — SD2 12/9-2017  
Dias 17/28



## Afskårne roser: faktordiagram og reduktion

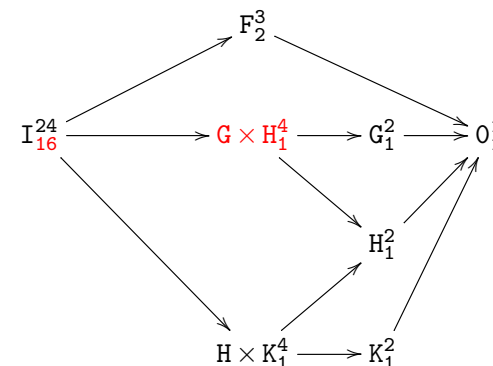


Statistisk model (hørende til faktordiagram) og hypotese:

Anders Tolver — Flersidet ANOVA — SD2 12/9-2017  
Dias 18/28



## Afskårne roser: faktordiagram og reduktion



Statistisk model (hørende til faktordiagram) og hypotese:

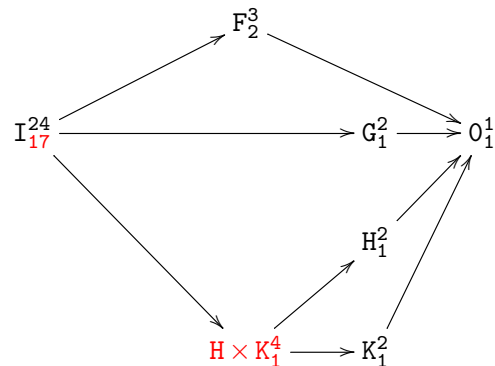
$$M: Y_i = \alpha(G \times H_i) + \gamma(H \times K_i) + \delta(F_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$

$$H_0: Y_i = \alpha(G_i) + \gamma(H \times K_i) + \delta(F_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$

Anders Tolver — Flersidet ANOVA — SD2 12/9-2017  
Dias 18/28



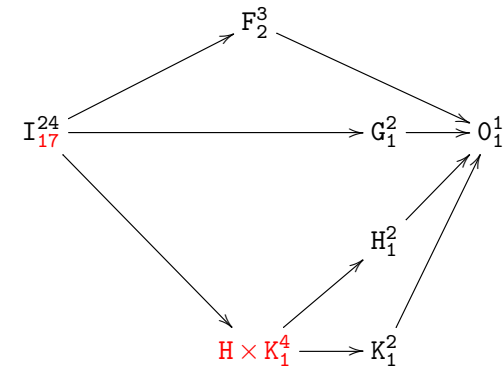
## Afskårne roser: faktordiagram og reduktion



Statistisk model (hørende til faktordiagram):



## Afskårne roser: faktordiagram og reduktion



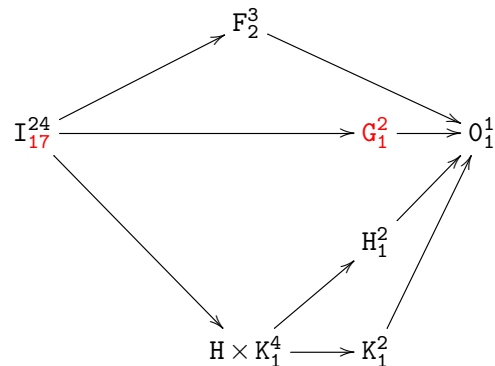
Statistisk model (hørende til faktordiagram):

$$M: Y_i = \alpha(G_i) + \gamma(H \times K_i) + \delta(F_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$

$$H_0: Y_i = \alpha(G_i) + \beta(H_i) + \gamma(K_i) + \delta(F_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$



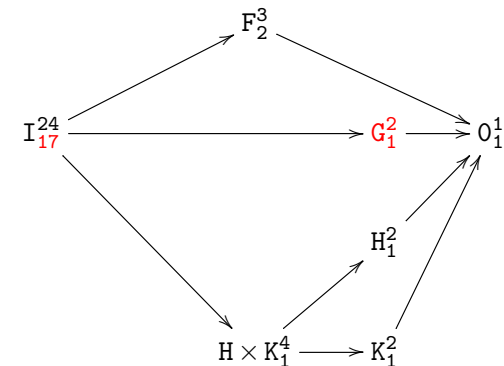
## Afskårne roser: faktordiagram og reduktion



Statistisk model (hørende til faktordiagram):



## Afskårne roser: faktordiagram og reduktion



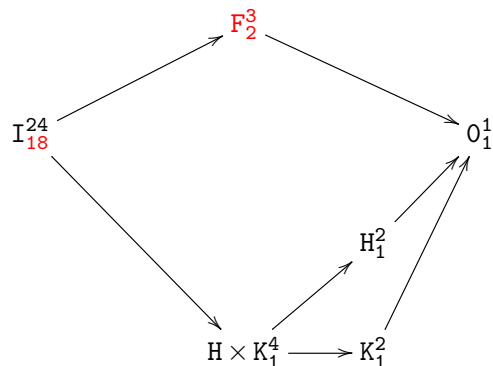
Statistisk model (hørende til faktordiagram):

$$M: Y_i = \alpha(G_i) + \gamma(H \times K_i) + \delta(F_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$

$$H_0: Y_i = \gamma(H \times K_i) + \delta(F_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$



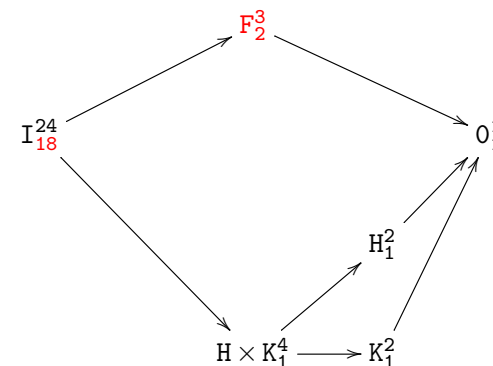
## Afskårne roser: faktordiagram og reduktion



Statistisk model (hørende til faktordiagram):



## Afskårne roser: faktordiagram og reduktion

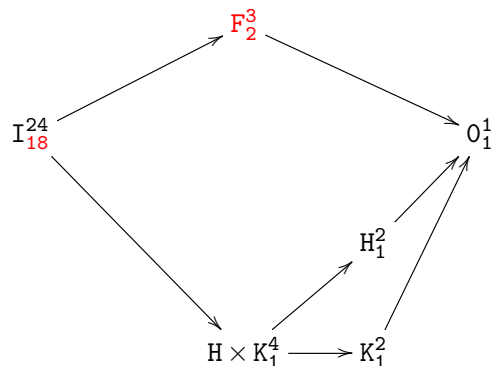


Statistisk model (hørende til faktordiagram):

$$Y_i = \gamma(H \times K_i) + \delta(F_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$



## Afskårne roser: faktordiagram og reduktion



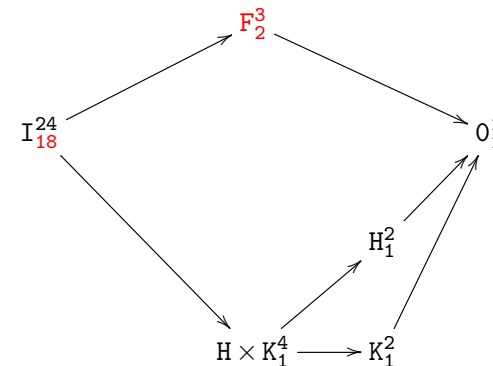
Statistisk model (hørende til faktordiagram):

$$Y_i = \gamma(H \times K_i) + \delta(F_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$

Hypotese,  $H_0$  (svarende til at fjerne markerede faktor):



## Afskårne roser: faktordiagram og reduktion



Statistisk model (hørende til faktordiagram):

$$Y_i = \gamma(H \times K_i) + \delta(F_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$

Hypotese,  $H_0$  (svarende til at fjerne markerede faktor):

$$Y_i = \gamma(H \times K_i) + e_i, \quad e_i \sim N(0, \sigma^2).$$





## Afskârne roser: slutmodel

Vores slutmodel bliver

$$Y_i = \alpha(\text{FLOR}_i) + \phi(\text{HANDLER} \times \text{KUNDE}_i) + e_i,$$

hvor  $e_1, \dots, e_{24}$  er uafhængige og normalfordelte  $N(0, \sigma^2)$ .

Slutmodellen er blot en **additiv model** mellem faktorerne **FLOR** og **HANDLER**  $\times$  **KUNDE**.

Hvordan angives parameterestimer for additiv model?

Hvor mange parameterestimer skal jeg regne med, at R angiver?



## Afskârne roser: parameterestimer for slutmodel

Kontrolgruppe: **FLOR** = 1, **HANDLER** = 0, **KUNDE** = 0

9.7083

Kontraster for **FLOR**

$$\hat{\alpha}(2) - \hat{\alpha}(1) = -1.0375 \quad \hat{\alpha}(3) - \hat{\alpha}(1) = 1.6125$$

Kontraster for **HANDLER**  $\times$  **KUNDE**

$$\begin{aligned} \hat{\phi}(1,0) - \hat{\phi}(0,0) &= 0.7333 & \hat{\phi}(0,1) - \hat{\phi}(0,0) &= 1.0833 \\ \hat{\phi}(1,1) - \hat{\phi}(0,0) &= 3.9500 \end{aligned}$$

Estimat for **residual spredning**:  $s = \hat{\sigma} = 0.9783$ .



## Afskârne roser: slutmodel i R

```
roses$HK<-relevel(roses$handler:roses$kunde,ref="0:0")
modelfinal<-lm(tid~HK+flor,data=roses)
```

```
summary(modelfinal)
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	9.7083	0.4891	19.848	1.10e-13 ***
HK0:1	1.0833	0.5648	1.918	0.07110 .
HK1:0	0.7333	0.5648	1.298	0.21054
HK1:1	3.9500	0.5648	6.994	1.57e-06 ***
flor2	-1.0375	0.4891	-2.121	0.04806 *
flor3	1.6125	0.4891	3.297	0.00401 **

Residual standard error: 0.9783 on 18 degrees of freedom



## Afskârne roser: Konklusion I

**95 %-konf.int.** for kontrolgp. (**FLOR** = 1, **HANDLER** = 0, **KUNDE** = 0)

$$9.7033 \pm t_{0.975,18} \cdot s \sqrt{\frac{3 + (2 \cdot 2) - 1}{24}},$$

samt LSD-værdier

$$LSD_{\text{FLOR}} = t_{0.975,18} \cdot s \sqrt{\frac{2}{8}} = 1.028$$

$$LSD_{\text{HANDLER} \times \text{KUNDE}} = t_{0.975,18} \cdot s \sqrt{\frac{2}{6}} = 1.119.$$

Benyt formler på s. 34-35 samt at  $t_{0.975,18} = 2.101$ !



## Afskærne roser: Konklusion II

Hvad fortæller analysen os?

- Lige gyldigt om gartneren benytter midlet
- Flor 2 er dårligere end flor 1:  $-1.038$   $[-2.065, -0.010]$
- Flor 3 er bedre end flor 1:  $1.612$   $[0.585, 2.640]$
- KUN hvis konserveringsmidlet tilsættes hos både forhandler og kunde ses en (signifikant) forbedring af rosernes holdbarhed

$$\hat{\phi}(1,1) - \hat{\phi}(0,0) = 3.9500 \quad [2.763, 5.137]$$

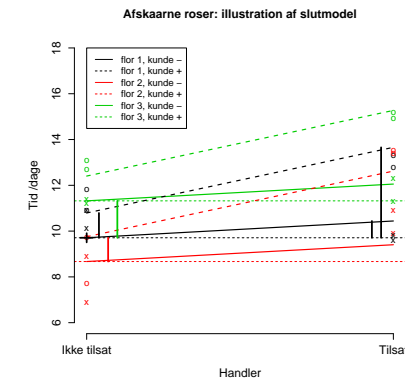
Alle udsagn er baseret på et 95 %-signifikansniveau og følgende R-output

```
##           2.5 %      97.5 %
## (Intercept) 8.6807103 10.735956318
## HK0:1      -0.1032635  2.269930147
## HK1:0      -0.4532635  1.919930147
## HK1:1       2.7634032  5.136596813
## flor2      -2.0651230 -0.009877016
## flor3       0.5848770  2.640122984
```

Anders Tolver — Flersidet ANOVA — SD2 12/9-2017  
Dias 26/28



## Afskærne roser: fortolkning af parameterestimer



Parameterestimerne svarer til niveauet for kontrolgruppen givet ved (FLOR = 1, HANDLER = 0, KUNDE = 0) samt til længden af de lodrette linjestykker på figuren.

Anders Tolver — Flersidet ANOVA — SD2 12/9-2017  
Dias 27/28



## Spørgsmål (til hjemmebrug)

- Forklar forskellen på den additive model og modellen med vekselvirkning i forb. med tosidet variansanalyse.
- Opskriv alle de mulige hypoteser i forb. med et balanceret trefaktorforsøg (-se opgave 2.2/kompendiets exercise 3.4).
- Overvej nogle forskellige muligheder for rækkefølgen af reduktion ved analyse af et balanceret trefaktorforsøg.
- Overvej nogle forskellige muligheder for rækkefølgen af reduktion ved analyse af trefaktorforsøget i kompendiets example 2.2.

