



Den generelle lineære model og vekselvirkninger

Claus Ekstrøm

E-mail: ekstrom@life.ku.dk



Program

- Den generelle lineære model
- Estimation og parametrisering
- Vekselvirkninger
- Successive tests, modelreduktioner
- Vekselvirkninger mellem kategoriske og kvantitative variable



Den generelle lineære model

Den **generelle lineære model** er givet ved

$$y_i = \sum_{j=1}^d \beta_j \cdot x_{ij} + e_i, \quad i = 1, \dots, n,$$

hvor β 'erne er et sæt af parametre, og x 'erne er forklarende variable — enten kvantitative eller dummyvariable.

Brug `lm()`, `summary()` og `drop1()` som sædvanlig:

Eksempel:

```
lm(blodtryk ~ køn + alder)
```

Bemærk at man godt kan "blande" kategoriske og kvantitative variable i kaldet til `lm()`. Og at man *ikke* behøver at definere dummyvariablene selv — det gør R automatisk.



Inferens i den generelle lineære model

Har allerede lært alt det, som vi skal bruge

Vi kan bruge hele maskineriet fra de tidligere uger til estimation (least squares), test af hypoteser (F -tests), konfidens- og prædiktionsintervaller og modelkontrol.

Det eneste, der er lidt anderledes er antallet af parametre/frihedsgrader, men det kan programmerne let regne ud for en.



Eksempel — sygdomme i agurker

Climate	Dose		
	2.0	3.5	4.0
A	51.5573	47.9937	57.9171
	51.6001	48.3387	51.3147
B	48.8981	48.2108	55.4369
	60.1747	51.0017	51.1251

Sygdomsspredningen er den samme i begge klimaer:

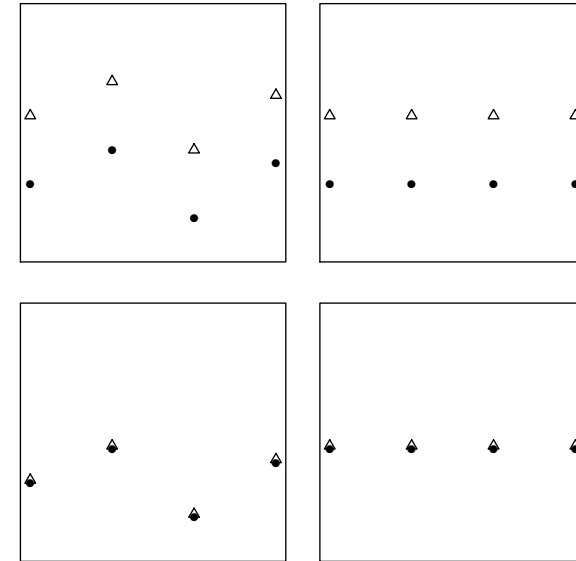
$$H_0 : \alpha_A = \alpha_B$$

Sygdomsspredningen afhænger ikke af dosis:

$$H_1 : \beta_{2.0} = \beta_{3.5} = \beta_{4.0}$$



Grafisk fremstilling af hypoteserne



Men hvad nu hvis ...

... at effekten af dosis *afhæng af hvilket miljø vi var i?*

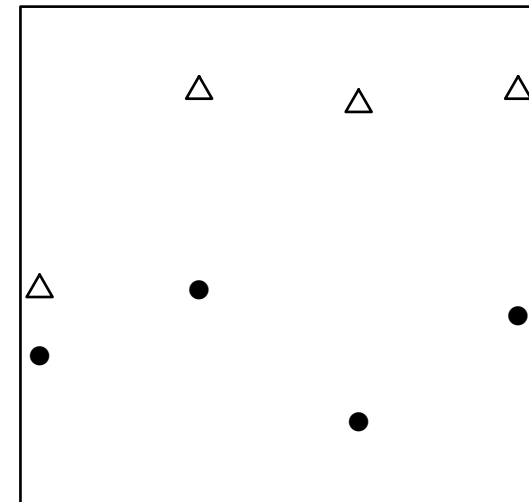
Det betyder jo, at kontrasten/forskellen mellem fx dosis 2.0 og dosis 3.5 har en værdi for miljø A men en anden værdi for miljø B.

Den model kan vores *additive* tosidede variansanalyse ikke håndtere. I den model er kontrasten altid den samme.

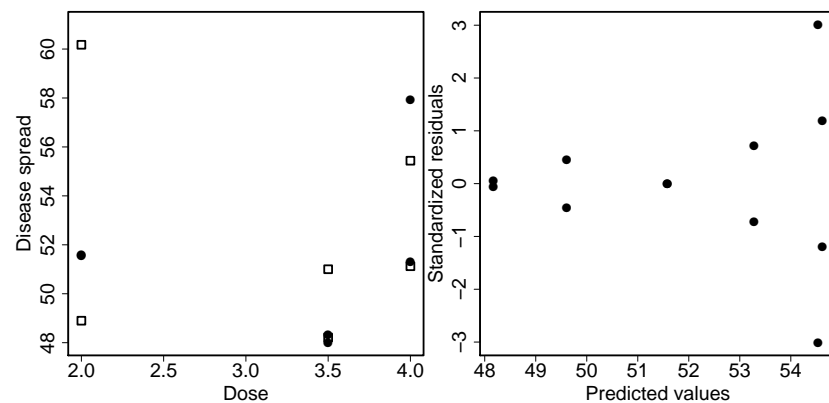
Hvis en variabel påvirker effekten af en anden variabel siges de to variable at **vekselvirke** (engl: interact).



Grafisk fremstilling af vekselvirkning



Agurkeforsøget — interaction plot



Slide 9— Statistisk Dataanalyse 1 (Uge 5-2) — Generelle lineær model og vekselvirkninger



Vekselvirkninger

Når vi benytter vekselvirkninger mellem to kategoriske variable forestiller vi os, at effekten af den ene variabel kan være helt forskellig fra kategori til kategori af den anden variabel.

Bemærk: for overhovedet at kunne regne på modeller med vekselvirkninger skal vi have mindst to observationer pr. "celle".

I R angives vekselvirkninger med en *:

```
lm(spredning ~ dosis + klima + dosis*klima)
```

Vi kan teste hypotesen "ingen vekselvirkning mellem dosis og klima" svarende til at vi kan fjerne vekselvirkningen fra vores model.

Slide 10— Statistisk Dataanalyse 1 (Uge 5-2) — Generelle lineær model og vekselvirkninger



Successive tests, modelreduktioner

Der er to vigtige "resultater" at have i baghovedet, når man vil forsøge at finde den bedste model til at beskrive sine data:

Modelreduktion / successive tests

Hvis vi har lavet en analyse og fundet ud af, at en eller flere parametre ikke er signifikante bør vi tilpasse en model, hvor disse parametre *ikke* indgår så vi får det bedst mulige estimat for de resterende parametre.

Det hierarkiske princip

Hvis en model indeholder vekselvirkninger mellem kategoriske variable giver det ikke mening at fjerne de tilhørende hovedvirkninger eller lavere ordens vekselvirkninger så længe vekselvirkningen stadig er med i modellen.

Slide 11— Statistisk Dataanalyse 1 (Uge 5-2) — Generelle lineær model og vekselvirkninger



Vekselvirkninger mellem kategoriske og kvantitative variable

Kan sammenblende kategoriske og kvantitative variable:

```
birth weight = sex + age
```

Vekselvirkninger mellem kategoriske og kvantitative variable:

```
birth weight = sex + age + sex*age
```

fortolkes som at der findes en regressionshældning (hørende til age) for hvert niveau af sex. Her laver vi altså to regressionslinjer — en for hvert køn.

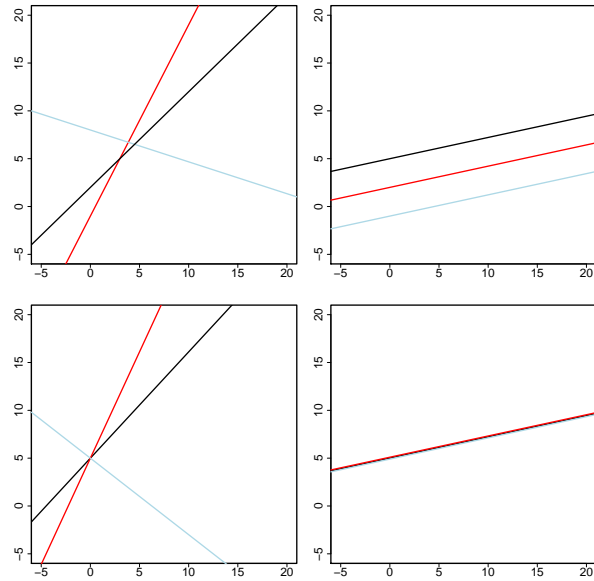
I R angives vekselvirkninger ligesom før med en *:

```
lm(bweight ~ age + sex + age*sex)
```

Slide 12— Statistisk Dataanalyse 1 (Uge 5-2) — Generelle lineær model og vekselvirkninger



Grafisk fremstilling af vekselvirkningsmodeller



Fødselsvægt

