

# Forsøgsplanlægning 1

## Statistisk Dataanalyse 2

Anders Tolver



# Ugens program

## Motivation (hovedspørgsmål)

- Hvordan skal man med givne ressourcer planlægge sit forsøg, for at optimere sandsynligheden for at opdage en effekt af behandlingen, hvis den er der?

## Tirsdag

- Fuldstændigt randomiserede forsøg
- Fuldstændigt randomiserede blokforsøg
- Balancerede ufuldstændige blokforsøg

## Torsdag

- Konfundering, partiel konfundering,  $2^n$ -te forsøg
- Brug af R i forbindelse med konfundering



# Motivation: tostikprøve-problemet

Betragt følgende forsøgsdesign

- 2 behandlinger: kontrol og beh
- $n_1$  : antal forsøgsheder i kontrolgruppen
- $n_2$  : antal forsøgsheder der behandles
- $n = n_1 + n_2$  : ressourceforbrug ved udførelsen af forsøget

**Spørgsmål:** Hvor mange personer ( $n_1, n_2$ ) skal der være i de to grupper?

Svaret på spørgsmålet afhænger af

- vores ressourcer (maksimalt antal forsøgsheder  $n$ )
- hvor stor en forskel mellem behandlingsgrupperne ønsker vi at kunne opdage?
- hvor sikre ønsker vi at være på at opdage en eventuel forskel?
- hvor stor er variationen blandt forsøgsheder som får den samme behandling?
- hvilken statistisk metode (test og signifikansniveau) skal benyttes ved analysen?



## Motivation: tostikprøve-problemet

Lad os lægge en statistisk model ned over forsøget

- kontrolgruppe:  $X_1, \dots, X_{n_1}$  uafh.  $\sim N(\mu_X, \sigma^2)$
- behandlingsgruppe:  $Y_1, \dots, Y_{n_2}$  uafh.  $\sim N(\mu_Y, \sigma^2)$
- hypotese,  $H_0 : \mu_X = \mu_Y$
- teststørrelse,

$$T = \frac{\hat{\mu}_X - \hat{\mu}_Y}{s \cdot \sqrt{1/n_1 + 1/n_2}},$$

hvor  $\hat{\mu}_X, \hat{\mu}_Y$  og  $s$  er estimater for  $\mu_X, \mu_Y$  og  $\sigma$ .

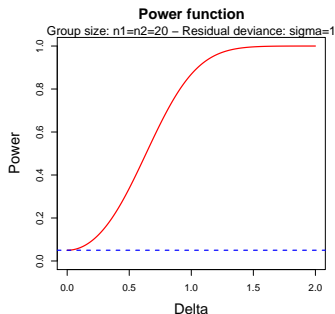
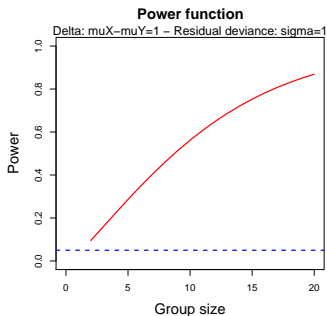
Udkendte størrelser

- størrelsen af behandlingseffekten,  $\mu_X - \mu_Y$
- residual spredningen,  $\sigma$
- signifikansniveau,  $\alpha$
- gruppestørrelser  $n_1, n_2$
- sandsynlighed,  $\beta$ , for at opdage behandlingseffekten (styrken)

Hvis man angiver 4 af størrelserne kan den sidste beregnes.



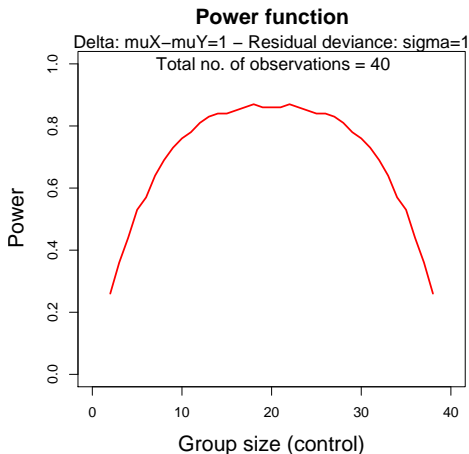
# Motivation: styrkefunktioner



- For  $\sigma = 1$  vil der være over 80 % ssh. for at afsløre en forskel på  $\mu_X - \mu_Y = 1$ , med 17 personer i hver gruppe.
- For  $\sigma = 1$  vil der med 20 personer i hver gruppe være 63.7 % ssh. for at afsløre en forskel på  $\mu_X - \mu_Y = 0.75$ .



# Motivation: tostikprøveproblemet



Højeste styrke opnås ved samme antal personer i de to grupper!



# Forsøgsplanlægning: generelle kommentarer

**styrke** = ssh. for at opdage en forskel, hvis den er der.

- Man kan øge styrken, ved at vælge en god statistisk model, hvor residualspredningen er lille
  - inddrag relevante forklarende variable i modellen
  - sørg for at kovariater indgår på “rigtig” måde
- Der er en grænse for, hvor meget vi kan gøre, når først data er indsamlet
- **Forsøgsplanlægning**: der kan være en yderligere gevinst ved at tænke sig om inden forsøget udføres
- kræver at man har en klar ide om forsøgets formål

Setup i dagens eksempler

- $N$  forsøgsheder, behandlingsfaktor med  $k$  niveauer
- Behandling kan være sammensat (fx.  $\text{treat} = A \times B \times C$ )
- Hvordan opnås høj styrke ved sammenligning af niveauerne for behandlingsfaktoren?



# Fuldstændigt randomiserede forsøg

Eksempel 3.1: Chlorophyll produktion i vinterhvede

- 3 behandlinger
- 12 *ensartede* forsøgsenheder (jordlodder)
- ved lodtrækning/randomisering fordeles de 3 beh. ud på forsøgsenhederne, så hver beh. forekommer 4 gange
- a priori forskelle ml. jordlodder fordeles (i gennemsnit) ligeligt på de tre behandlinger jf. randomisering

**Forsøgsplan:** oversigt over allokering af beh. til forsøgsenheder

1	2	3	4	5	6	7	8	9	10	11	12
B	C	A	A	B	A	A	C	C	B	C	B

## Hvorfor randomisering?

I det lange løb (f.x. hvis forsøget gentages i flere år), vil randomisering af forsøgsplanen føre til et korrekt billede af beh.-effekten, selv hvis der skulle være variable, som ikke er blevet inddraget i den statistiske model.





# Fuldstændigt randomiserede forsøg

Eksempel 3.1: Chlorophyll produktion i vinterhvede

Fuldstændig randomisering virker bedst ved (forholdsvis) få ensartede forsøgsenheder.

Statistisk model (ensidet variansanalyse)

$$Y_i = \alpha(\text{treat}_i) + e_i, \quad e_1, \dots, e_N \text{ uafh.} \sim N(0, \sigma^2).$$

Forslag til, hvordan den statistiske analyse kunne forbedres

- Mål **kovariat** hørende til forskellige jordlodder og inddrag denne i analysen
- Inddel forsøgsenhederne i mindre blokke, som må forventes at have en stor grad af homogenitet (**blocking**)

Dårlig løsning (balancing)

- Sørg for at behandlingsgrupperne ligner hinanden (fx. samme gennemsnitsvægt). Giver i værste fald større residualvariation inden for grupper og dermed lavere styrke.



# Fuldstændige randomiserede blokforsøg

## Eksempel 9.1: Udbytte af løg

- 5 behandlinger (lifting times)
- 15 forsøgsenheder
- 3 blokke hver med 5 forsøgsenheder i hver
- forsøgsenheder inden for samme blok er mere ensartede end forsøgsenheder fra forskellige blokke
- Inden for hver blok knyttes forsøgsenhederne til de 5 behandlinger ved randomisering

**Forsøgsplan:** oversigt over allokering af beh. til forsøgsenheder

b1					b2					b3				
E	C	D	B	A	E	C	B	D	A	A	E	D	C	B

Hvilken statistisk model bør benyttes til at analysere data?

Hvis vi ignorerer blokeffekt og udfører fuldst. randomiseret forsøg kræves flere forsøgsenheder for at opdage forskelle ml. behandlinger! Styrken ved sammenligning af beh. øges ved **blocking**



# Fuldstændige randomiserede blokforsøg

Eksempel 9.2: number of after crop plants

- 4 cover crops (C)
- 3 after crops (A)
- 12 behandlinger ( $C \times A$ )
- Fuldstændigt randomiseret blokforsøg med 4 blokke
- Statistisk model og faktordiagram?

Hvis behandlingen er sammensat af mange faktorer haves ofte for få homogene forsøgsenheder til at lave fuldstændigt blokforsøg.

Man bliver da nødt til at lave ufuldstændige blokforsøg, f.eks

- **balancerede ufuldstændige blokforsøg (BIBD)**
- split-plot forsøg (kendt fra uge 4)
- $2^n$ -forsøg (torsdag)

**ufuldstændig** = ikke alle behandlinger kan afprøves i hver blok!



## Ufuldstændige blokforsøg: notation

Ved ufuldstændige blokforsøg benyttes den additive model for blokforsøg.

Enten systematisk blokeffekt

$$Y_i = \alpha(\text{beh}_i) + \beta(\text{blok}_i) + e_i, \quad e_i \text{ uafh.} \sim N(0, \sigma^2).$$

eller tilfældig blokeffekt

$$Y_i = \alpha(\text{beh}_i) + b(\text{blok}_i) + e_i, \quad b(j) \text{ uafh.} \sim N(0, \sigma_b^2), e_i \text{ uafh.} \sim N(0, \sigma^2)$$

Notation

- $v_B$  = antal blokke
- $v_T$  = antal behandlinger
- $r_B$  = antal behandlinger per blok



# Ufuldstændige blokforsøg: eksempel

Eksempel på ufuldstændigt blokforsøg

- antal blokke  $v_B = 5$
- antal behandlinger  $v_T = 5$
- antal behandlinger per blok  $r_B = 3$
- antal gange hver behandling optræder i planen  $r_T = 3$

beh	blok				
	1	2	3	4	5
1	x			x	x
2	x	x			x
3	x	x	x		
4		x	x	x	
5			x	x	x

Er dette et godt eller et dårligt design?



## Ufuldstændige blokforsøg: co-incidence matrix

I forbindelse med blokforsøg laves ofte en **co-incidence** matrix

beh	beh				
	1	2	3	4	5
1	3	2	1	1	2
2		3	2	1	1
3			3	2	1
4				3	2
5					3

Diagonalen viser, hvor mange gange hver behandling optræder i planen (-her 3 for alle behandlinger).

Det fremhævede tal (2) for komb. beh=1 og beh=2 viser, at de to behandlinger er afprøvet 2 gange på samme blok i den anførte forsøgsplan.

Ved sammenligning af behandlinger gælder

- højere *co-incidence*  $\Rightarrow$  højere styrke



# Balancerede ufuldstændige blokforsøg: eksempel

beh	dag									
	1	2	3	4	5	6	7	8	9	10
1	x		x	x		x	x			x
2		x	x			x		x	x	x
3				x	x	x	x	x	x	
4	x	x	x		x		x		x	
5	x	x		x	x			x		x

antal blokke  $v_B = 10$

antal behandlinger  $v_T = 5$

antal behandlinger per blok  $r_B = 3$

hver behandling optræder  $r_T = 6$  gange

hvert par optræder  $\lambda = 3$  gange sammen

beh	beh				
	1	2	3	4	5
1	6	3	3	3	3
2		6	3	3	3
3			6	3	3
4				6	3
5					6



## Balancerede ufuldstændige blokforsøg: definition

Et **BIBD** (= Balanced Incomplete Block Design) er et ufuldstændigt blokforsøg karakteriseret ved

- ① hver behandling ( $v_T$  styk) optræder højst en gang i hver blok
- ② alle behandlinger forekommer lige ofte ( $r_T$ )
- ③ alle blokke har lige mange forsøgsheder ( $r_B$ )
- ④ alle par af behandlinger forekommer lige mange gange ( $\lambda$ ) i hver blok

**NB:** Ikke for alle blokstørrelser og antal af beh. findes et BIBD.

Nogle nødvendige men ikke tilstrækkelige krav (jf. Theorem 9.6)

- ① flere blokke end behandlinger  $v_B \geq v_T$
- ②  $v_T \cdot r_T = v_B \cdot r_B$
- ③  $\lambda = \frac{r_T \cdot (r_B - 1)}{v_T - 1} \leftarrow$  dette skal være et helt tal

Selv om de tre krav er opfyldt, er det ikke sikkert man kan lave et BIBD.

Man må undertiden kigge i publicerede forsøgsplaner for at finde ud af, om der kan konstrueres et BIBD, som man kan benytte.





## BIBD: eksempel 9.8 - tomater

- $v_T = 21$  sorter af tomat (a, b, c, ..., u)
- $v_B = 21$  blokke
- $r_B = 5$  sorter i hver blok
- $r_T = 5$  gentagelser af hver sort
- $N = 105$  forsøgsenheder (planter)

Hvert par af behandlinger skal forekomme præcis

$$\lambda = \frac{r_T \cdot (r_B - 1)}{v_T - 1} = \frac{5 \cdot (5 - 1)}{21 - 1} = 1$$

gang inden for samme blok i forsøgsplanen (-se plan i kompendiet).

Randomisering i forbindelse med BIBD

- på papir laves en oversigt over hvilke sorter, der skal afprøves sammen på de forskellige blokke
- blokkene på papiret allokeres ved randomisering til blokkene som skal indgå i forsøget
- for hver blok randomiseres behandlingerne (sort) ud på de enkelte plots inden for blokken



## Eksempel 9.8: tilfældig blok effekt

### Statistisk model

$$Y_i = \alpha(\text{sort}_i) + b(\text{blok}_i) + e_i, b(j) \text{ uafh.} \sim N(0, \sigma_b^2), e_i \text{ uafh.} \sim N(0, \sigma^2).$$

Hypotese,  $H_0 : \alpha(a) = \dots = \alpha(u) = \mu$

Der er en stærkt signifikant effekt af sort.

```
> slutmodel<-lme(yield~vty-1,random=~1|block,method="REML")
> b.vs.a<-c(-1,1,rep(0,19))
> c.vs.a<-c(0,-1,1,rep(0,18))
> difs=rbind(b.vs.a,c.vs.a)
> estimable(slutmodel,difs,conf.int=0.95)
      Estimate Std. Error  t value DF Pr(>|t|) Lower.CI Upper.CI
b.vs.a  15.5453    4.158238   3.73843  64 3.974e-04   7.2382  23.85230
c.vs.a -17.5352    4.158238  -4.21698  64 7.941e-05 -25.8422 -9.22818
```

LSD-værdien kan beregnes som (-formel s. 169 i kompendiet)

$$\text{LSD} = \frac{\text{Upper.CI} - \text{Lower.CI}}{2} = 8.31$$

**NB:** for et BIBD er LSD ens uanset hvilket par af sorter som sammenlignes.



## Eksempel 9.8: systematisk blok effekt

Statistisk model

$$Y_i = \alpha(\text{sort}_i) + \beta(\text{blok}_i) + e_i, e_i \text{ uafh. } \sim N(0, \sigma^2).$$

Hypotese,  $H_0 : \alpha(a) = \dots = \alpha(u) = \mu$

Der er en stærkt signifikant effekt af sort.

Kan også vise, at der er en signifikant effekt af blok.

LSD-værdien kan beregnes som

$$\text{LSD}_{0.95} = t_{DF_e, 0.975} \cdot s \sqrt{\frac{2 \cdot r_B}{\lambda \cdot v_T}} = 1.998 \cdot 6.077 \sqrt{2 \cdot 5/21} = 8.38.$$

**NB:** for et BIBD er LSD ens uanset hvilket par af sorter som sammenlignes.



## Eksempel 9.8: estimator

Næsten samme resultater bortset fra standard error, som har to forskellige fortolkninger.

sort	systematisk blok effekt		tilfældig blok effekt	
	adj. means	Std.Error	est	Std.Error
a	45.59029	2.953757	45.11323	3.492154
b	60.58410	2.953757	60.65850	3.492154
c	43.48552	2.953757	43.12329	3.492154
⋮	⋮	⋮	⋮	⋮
s	35.49457	2.953757	35.66661	3.492154
t	48.92600	2.953757	48.50759	3.492154
u	56.85457	2.953757	56.89567	3.492154

- Hvordan udregnes adjusted means?
- Hvorfor er der forskel på Std.Error?



# Perspektivering: split-plot forsøg (1)

## Eksempel 8.3: harskhed af svinekød

- behandlingsfaktor sammensat af
  - to typer foder (F)
  - to lagringsmetoder (P)
  - tre lagringstider (S)
- 12 behandlinger ( $F \times P \times S$ )
- cutlets fra samme dyr (A) har størst homogenitet
- blokforsøg med blokke givet ved A
- Kan vi lave fuldstændigt blokforsøg?
- Kan vi lave BIBD?



## Perspektivering: split-plot forsøg (2)

Behandlingsfaktor  $T = F \times (P \times S)$  hvor

- $F$  varierer på helplot-niveau
- $P \times S$  varierer på delplot-niveau

Behandlinger med forskelligt niveau af  $F$  kan ikke gives til forsøgsenheder på samme helplot.

$(f_1, g_1)$  og  $(f_2, g_2)$  kan ikke forekomme på samme helplot  $\Rightarrow$

Fuldstændige- og balancerede ufuldstændige blokforsøg umulige!

Split-plot forsøg er en anden type forsøgsdesign!!!



# Opsummering

Forsøgsplanlægning går ud på at gøre sandsynligheden for at opdage en eventuel forskel mellem behandlinger så stor som muligt med så få ressourcer (observationer) som muligt.

- Fuldstændigt randomiserede forsøg
  - relativt få forsøgsheder
  - randomisering i et trin
- Fuldstændige randomiserede blokforsøg
  - mange forsøgsheder inddelt i mindre homogene blokke
  - randomisering i et trin
- Ufuldstændige blokforsøg
  - balancerede ufuldstændige blokforsøg (BIBD)
  - BIBD giver samme præcision ved sammenligning af alle par af behandlinger
  - split-plot forsøg er et albertivt eksempel på et ufuldstændigt forsøg
  - randomisering i to trin

