Experiment Design:

Experiment Design I:

- ***Basale Designs:***
  - ***Completely Randomized***
  - ***Randomized Complete Blocks***
  - ***Latin Squares***
- ***Faktorielle Experimente***

Experiment Design II:

- ***Confounding***
- ***Split Plot und Split Block Design***
- ***Incomplete Block Designs, Gitter***

Basale Designs: Completely Randomized

Layout:

- **Zufällige (/keine) Anordnung der Gruppen**
- **Wenn möglich: gleiche Anzahl Individuen pro Gruppe**

Beispiel: **4 verschiedene Futterzusätze**

- **Jeweils 5 Schafe bekommen ein Präparat**
- **Schafe jeder Gruppe werden markiert**
- **Alle Schafe grasen auf derselben Weide**
- **Im Frühjahr werden die Schafe geschoren und Wollerträge gemessen.**

Beispiel: Completely Randomized

Layout:

treatment: v

- **1 *treatment*: Futterzusätze**
  - **4 Unterschiedliche *level*: A-D**
- **5 Individuen pro Gruppe**

replicates: r

ANOVA:

Varianzquelle	df	ss
treatment	(v-1) = 3	$\sum_{i=1}^{v} X_{i.}^2/r - X_{..}^2/rv$
residual	v(r-1) = 16	$\sum_{i=1}^{v} (\sum_{j=1}^{r} X_{ij}^2/r - X_{i.}^2/r)$

```
> aov(yield ~ treatment, data=CR)
```

Completely Randomized

Vorteile:

- **Einfache Konzeption**
- **Einfache Statistik, maximale Anzahl Freiheitsgrade für den Fehler**
- **Unterschiedliche Anzahl Wiederholungen pro Gruppe möglich**

Nachteile:

- **Nur für homogenes Versuchsmaterial geeignet**
- **Insofern ungeeignet für große Anzahl *treatments***

Randomized Complete Block Design

Layout:

- **Blöcke von jeweils ähnlichen Individuen**
- **Zufällige Anordnung der *treatment level* <u>innerhalb</u> der Blöcke**
- **Im Block müssen alle *treatments* vorkommen**

ANOVA:

Varianzquelle	df	ss
treatments	(v-1) = 4	$\sum_i X_{i.}^2/r - X_{..}^2/rv$
blocks	(r-1) = 5	$\sum_j X_{.j}^2/v - X_{..}^2/rv$
residual	(v-1)(r-1) = 20	$\sum_i\sum_j X_{ij}^2 - \sum X_{i.}^2/r - \sum X_{.j}^2/v + X_{..}^2/rv$

```
> aov(yield ~ treatment+blocks, data=RCB)
```

Randomized Complete Block Design

Vorteile:

- **Wenn Versuchsmaterial inhomogen: Eliminieren der Blockvarianz reduziert Fehler**
- **Einfache Auswertung**
- **Einzelne Blöcke können eliminiert werden**

Nachteile:

- **Ungeeignet bei großer Varianz innerhalb eines Blocks**
- **Alle *treatment level* müssen im Block vorkommen**

Randomized Complete Blocks

Beispiel:

- 5 Futter mit unterschiedlichem Proteingehalt
- 6 Schafrassen: I bis VI

ANOVA RCB:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
treatment	4	4314	1078.6	9.065	0.00024	***
blocks	5	466	93.3	0.784	0.57312	
Residuals	20	2380	119.0			

ANOVA CR:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
treatment	4	4314	1078.6	9.474	8.37e-05	***
Residuals	25	2846	113.8			

Randomized Complete Blocks

Missing Data

- Da in jedem Block jedes *treatment* vorkommen muss, müssen fehlende Werte *abgeschätzt* werden.

$$X_{ij} = \frac{v\,X_{i.} + r\,X_{.j} - X_{..}}{(v-1)(r-1)}$$

- Die Abschätzung erfolgt durch lineare Regression
- Die Anzahl der Freiheitsgrade für den Fehler sinkt um 1.
- Wenn mehrere Werte fehlen, muss iterativ ermittelt werden

Latin Square Design

Layout:

- Zwei Varianzquellen werden erfasst:
  sie bilden Reihen und Spalten eines Quadrats
- Jedes *treatment* in jeder Reihe und jeder Spalte

	Alter I	Alter II	Alter III
Rasse I	A	B	C
Rasse II	C	A	B
Rasse III	B	C	A

Latin Square Design

A	B	C
C	A	B
B	C	A

Aufbau:

- Anzahl der *treatment level* definiert Anzahl von Reihen und Spalten
- Anordnung der *level* in Reihen und Spalten ist (bedingt) zufällig
- Bedingung:
  - jede Reihe und jede Spalte enthält alle *treatment level*
  - jedes *level* kommt genau ein mal vor

- *Missing data* im k x k Quadrat:

$$X_{ijh} = \frac{k(X_{i..} + X_{.j.} + X_{..h}) - 2X_{...}}{(k - 1)(k - 2)}$$

Latin Square Design

ANOVA LS (k x k):

Varianzquelle	df	ss	ms=ss/df
Row	(k – 1)	$\Sigma X_{i..}^2/k - X_{...}^2/k^2 = \Sigma (x_i - \bar{x})^2$	R
Column	(k – 1)	$\Sigma X_{.j.}^2/k - X_{...}^2/k^2 = \Sigma (x_j - \bar{x})^2$	C
treatment	(k – 1)	$\Sigma X_{..h}^2/k - X_{...}^2/k^2 = \Sigma (x_h - \bar{x})^2$	T
residuals	(k-1)(k-2)	$\Sigma X_{ijh}^2 - \Sigma X_{i..}^2/k - \Sigma X_{.j.}^2/k - \Sigma X_{..h}^2/k + 2X_{...}^2/k^2$	E

```
> aov(var ~treatment+row+column, data=LS)
```

Latin Square Design

Beispiel:

- **qPCR für 5 Transcriptionsfaktoren A - E (treatment):**
- **5 Zell-Typen (rows),**
  5 Zeitpunkte Day I bis V(*columns*)
- **Angabe des ΔCt Wertes**

	TF									
Celltype	Day_I		Day_II		Day_III		Day_IV		Day_V	
#1	A	4,2	C	4,7	B	5,5	D	5,1	E	4,4
#2	E	4,5	B	5,4	C	5,2	A	4,4	D	5
#3	C	4,1	A	4,6	D	5,7	E	4,7	B	4,8
#4	B	5,6	D	5,2	E	4,9	C	5	A	4,3
#5	D	4,7	E	4,9	A	4,5	B	5,4	C	4,6

```
> aov(deltaCt ~TF+Celltype+Day, data=LS)
```

Latin Square Design

```
> summary(aov(deltaCt~TF+Celltype+Day, data=LS))
            Df Sum Sq Mean Sq F value   Pr(>F)
TF           4 1.0936  0.2734   4.906 0.014105 *
CellType     4 0.1776  0.0444   0.797 0.549839
day          4 2.8616  0.7154  12.836 0.000271 ***
Residuals   12 0.6688  0.0557
```

```
> summary(aov(deltaCt~TF, data=LS))
            Df Sum Sq Mean Sq F value Pr(>F)
TF           4  1.094  0.2734   1.475  0.247
Residuals   20  3.708  0.1854
```

Latin Square Design

A	B	C
C	A	B
B	C	A

Vorteile:

- Hohe Präzision, da viel Varianz erklärt wird
- Noch immer recht einfache Auswertung
- Rechnung für *missing data* recht einfach

Nachteile:

- geeignet nur für 5 bis 12 *treatments*:
- Bei 2x2, 3x3, 4x4 nur 0, 2, 6 DF für Fehler
- Immer so viele Replikate wie *treatments* nötig

Wiederholtes Latin Square Design

A	B	C
C	A	B
B	C	A

A	C	B
B	A	C
C	B	A

Anwendung:

- für 2x2, 3x3, 4x4: mehr Freiheitsgrade für den Fehler
- um einen zusätzlichen Blockeffekt zu berücksichtigen (z.B. Standort etc.)
- 2 x 2 → 2 Möglichkeiten, 3x3 → 12, 4x4 →576, 5x5 →161.280; zufällige Auswahl

Varianzquelle	df	ms=ss/df
Square	s – 1	S
Row	s(k – 1)	R
Column	s(k – 1)	C
Treatment	k – 1	T
Treatment x Square	(s-1)(k-1)	TS
Residuals	s(k-1)(k-2)	E

Graeco - Latin Square Design

Aα	Bε	Cβ	Dζ	Eγ	Fη	Gδ
Bβ	Cζ	Dγ	Eη	Fδ	Gα	Aε
Cγ	Dη	Eδ	Fα	Gε	Aβ	Bζ
Dδ	Eα	Fε	Gβ	Aζ	Bγ	Cη
Eε	Fβ	Gζ	Aγ	Bη	Cδ	Dα
Fζ	Gγ	Aη	Bδ	Cα	Dε	Eβ
Gη	Aδ	Bα	Cε	Dβ	Eζ	Fγ

Layout:

- Untersuchung von 4 Varianzquellen:
  Row, Column, treatment1, treatment2
- t1 und t2 werden orthogonal überlagert:
  jeder lateinische Buchstabe wird genau ein mal mit jedem griechischen kombiniert
- für k x k gibt es (k-1) orthogonale Quadrate

Varianzquelle	df	ms=ss/df
Latin	k – 1	L
Greek	k – 1	G
Row	k – 1	R
Column	k – 1	C
residuals	(k-1)(k-3)	E

(k-1)(k-2) im Latin square

Das Faktorielle Experiment

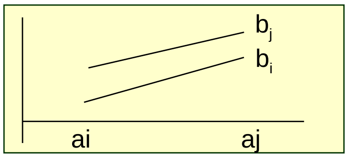
- Bisher wurden Experimente besprochen, die dazu dienten, eventuelle Effekte **eines** *treatments* möglichst deutlich zu machen. Hierzu sollten unbeabsichtigte Varianzquellen möglichst ausgeschlossen werden.
- Wenn mehr als ein *treatment* in einem Experiment untersucht werden soll, liegt meist ein **faktorielles Experiment** vor.
- Es ist möglich, dass Faktoren einander **beeinflussen**!

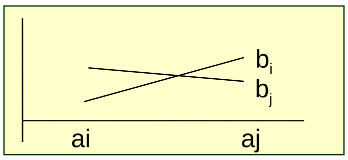
Faktorielle Experimente

Beispiel:

•a: N Düngung
a_i: 50, a_j: 70kg/ha

•b: Bewässerung
b_i: 4, b_j: 8 mm/Tag





- **2 oder mehr Behandlungen (mit je mindestens 2 Stufen) in einem Experiment**
- **Um die Behandlungen unabhängig voneinander untersuchen zu können, müssen ALLE Kombinationen der Faktor-Level im Experiment vorkommen**

2 Behandlungen mit je 2 Stufen: $2^2 = 4$ Kombinationen
3 Behandlungen mit je 2 Stufen: $2^3 = 8$ Kombinationen
2 Behandlungen mit je 3 Stufen: $3^2 = 9$ Kombinationen
2 Behandlg. mit 2 bzw. 3 Stufen: $2*3 = 6$ Kombinationen

- **Interaktion:** **wenn *level* von a für verschiedene *level* von b unterschiedlich angeordnet sind: A x B**

Faktorielle Experimente

Achtung:

- "Faktoriell" bezeichnet kein best. Design, sondern die Art der Daten-Auswertung
- Faktorielle Experimente können als CR, RCB oder LS ausgelegt werden
- für die Auswertung sind äquidistante Level oder logarithmische Abstände vorteilhaft

- **2 oder mehr Behandlungen (mit je mindestens 2 Stufen) in einem Experiment**
- **Um die Behandlungen unabhängig voneinander untersuchen zu können, müssen ALLE Kombinationen der Faktor-Level im Experiment vorkommen**

2 Behandlungen mit je 2 Stufen: $2^2 = 4$ Kombinationen
3 Behandlungen mit je 2 Stufen: $2^3 = 8$ Kombinationen
2 Behandlungen mit je 3 Stufen: $3^2 = 9$ Kombinationen
2 Behandlg. mit 2 bzw. 3 Stufen: $2*3 = 6$ Kombinationen

- **Interaktion: wenn *level* von a für verschiedene *level* von b unterschiedlich angeordnet sind: A x B**

Faktorielle Experimente

Achtung:

- "Faktoriell" bezeichnet kein best. Design, sondern die Art der Daten-Auswertung
- Faktorielle Experimente können als CR, RCB oder LS ausgelegt werden
- für die Auswertung sind äquidistante Level oder logarithmische Abstände vorteilhaft

Vorteile:

- **hoher Informationsgehalt, da verschiedene Behandlungen unabhängig betrachtet werden**
- **Interaktionen werden nachweisbar**

Nachteil:

- **Vollständigkeit bedingt hohe Zahl von Kombinationen: viele Individuen → hohe Varianz**
- **Lösung: unvollständige Designs (*confounding*)**

Faktorielle Experimente:

"ohne"; "mit"

"Düngung"; "Bewässerung"

Beispiel 2^2:

- **Erbsenanbau ohne / mit NPK-Düngung**
- **ohne / mit künstlicher Bewässerung**
- **Anlage als *Randomized Complete Blocks* mit 4 Blöcken**

Erträge	"00"	"01"	"10"	"11"
I	53.13	28.66	33.21	6.42
II	60.00	21.97	45.00	35.06
III	60.67	20.27	39.82	24.35
IV	58.69	35.06	30.00	18.44

Faktorielle Experimente:

Effects	"00"	"01"	"10"	"11"	S +	S -	S total
	232.49	105.96	148.03	84.27			
B: -/+ Wasser	+	-	+	-	380.52	190.23	190.29
A: -/+ Dünger	+	+	-	-	338.45	232.3	106.15
A X B	+	-	-	+	316.76	253.99	62.77
total	+	+	+	+	570.75		570.75

Erträge	"00"	"01"	"10"	"11"
I	53.13	28.66	33.21	6.42
II	60.00	21.97	45.00	35.06
III	60.67	20.27	39.82	24.35
IV	58.69	35.06	30.00	18.44

Faktorielle Experimente:

Effects	"00"	"01"	"10"	"11"	S +	S -	S total
	232.49	105.96	148.03	84.27			
B: -/+ Wasser	+	-	+	-	380.52	190.23	190.29
A: -/+ Dünger	+	+	-	-	338.45	232.3	106.15
A X B	+	-	-	+	316.76	253.99	62.77
total	+	+	+	+	570.75		570.75

$(190.29)^2/16$

```
               Df  Sum Sq Mean Sq  F value    Pr(>F)
Block           3   208.1    69.4    1.183  0.369572
Dünger          1   704.2   704.2   12.010  0.007098 **
Wasser          1  2263.1  2263.1   38.596  0.000156 ***
Dünger:Wasser   1   246.3   246.3    4.200  0.070693 .
Residuals       9   527.7    58.6
```

Faktorielle Experimente:

Experiment-Design:

- *completely random* und *randomized block* sind immer möglich
- insbesondere bei unterschiedlicher Anzahl der Level (z.B. 2^n x 3^s)
- *Latin square* geht bei 2^2, 2^3 und 3^2.
- 5^2, 3^3, 2^5, 4^3, 5^2 sind meist zu groß:
- Lösung: Überlagerung (*confounding*)

Effects	"00"	"01"	"10"	"11"	S +	S -	S total
	232.49	105.96	148.03	84.27			
B: -/+ Wasser	+	-	+	-	380.52	190.23	190.29
A: -/+ Dünger	+	+	-	-	338.45	232.3	106.15
A X B	+	-	-	+			
total	+	+	+	+			

Zufällige Varianz

"random effect"

```
> aov(yield ~ Dünger*Wasser+Error(Block), data= F2x2)
```

```
> aov(yield ~ Dünger*Wasser+Block, data= F2x2)
```