

# Grundlagen: Versuchsdesign faktorieller Versuche

---

Vortrag im Rahmen des BPP am  
12.04.2023 von Julian Siebert im  
Modul „Angewandte Statistik und  
Versuchswesen“ (SoSe 2023)

1

# Versuchsfrage

## Beobachtung

Etwas verhält sich anders als bekannt.

## Hypothese

Aus diesem Grund verhält es sich anders.

## ➤ Faktorieller Versuch

## 2

# Faktorieller Versuch

Definition:

Mit faktoriellen Versuchsplänen können die  
**Einflüsse von Faktoren** (Temperatur, Licht, Chemikalie,  
...)  
**auf eine oder mehrere Outcomes / Zielgrößen** (Masse,  
Menge, ...)  
**bestimmt werden.**

Gleichzeitig können Aussagen über **Wechselwirkungen**  
getroffen werden.

## 2

# Faktorieller Versuch

## Struktur:

- **Einflussgrößen (Faktoren)** müssen einstellbar und messbar (kontrollierbar) sein
- **Zielgrößen / Outcomes** müssen geeignet sein die Auswirkung der Faktoren zu erfassen. (werden beeinflusst)

Wo möglich sollten sie gezählt, gemessen oder gewogen werden.

Bekanntermaßen normalverteilte Zielgrößen bevorzugen.

Schätzungen (Bonituren) nur wenn nicht anders möglich.

	A	B	C	D	E
1	genotype	trt	number	height_cm	fm_veg_g
2	nil	ww	1	58,5	7,79
3	scarlett	ww	2	54	5,85
4	nil	ww	3	56,2	5,96
5	scarlett	ww	4	57,2	10,75
6	scarlett	ww	5	46,4	6,92
7	nil	ww	6	56	9,43
8	nil	ww	7	57,8	8,92
9	nil	ww	8	63,6	8,07
10	nil	ww	9	64	8,86
11	nil	ww	10	59,7	7,04

Beispielhafte Exceltabelle eines faktoriellen Versuches mit den Faktoren Genotyp und Treatment und den messbaren Zielgrößen / Outcomes Höhe in cm und Frischmasse in gramm.

## 2

# Faktorieller Versuch



### Struktur:

- Einflussgrößen (Faktoren)
- Zielgrößen / Outcomes
- Begleitparamter dienen zur Überprüfung der gleichen Bedingungen (Homogenität) innerhalb des Experimentes

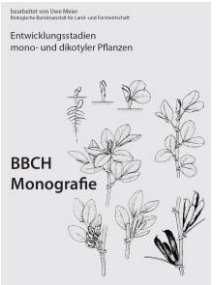
Für alle Pflanzen sollten bis auf die Einflussgrößen gleiche Bedingungen vorliegen.

Basiert häufig auf Bonituren (Beispiel BBCH)

	A	B	C	D	E	F
1	genotype	trt	number	height_cm	fm_veg_g	bbch
2	nil	ww	1	58,5	7,79	32,00
3	scarlett	ww	2	54	5,85	33,00
4	nil	ww	3	56,2	5,96	34,00
5	scarlett	ww	4	57,2	10,75	30,00
6	scarlett	ww	5	46,4	6,92	31,00
7	nil	ww	6	56	9,43	34,00
8	nil	ww	7	57,8	8,92	32,00
9	nil	ww	8	63,6	8,07	31,00
10	nil	ww	9	64	8,86	35,00
11	nil	ww	10	59,7	7,04	36,00

Beispielhafte Exceltabelle eines faktoriellen Versuches mit den Faktoren Genotyp und Treatment und den messbaren Zielgrößen / Outcomes Höhe in cm und Frischmasse in gramm. Begleitparameter phänologisches Entwicklungsstadium von Pflanzen (BBCH)





## Bonituren / Schätzungen:

- Verlangen besondere Genauigkeit und Reflexion
- Meist hat schon einmal jemand einen Boniturschlüssel entworfen! Bsp: (2001): BBCH Monografie : Entwicklungsstadien mono- und dikotyle Pflanzen. 2. Auflage. Biologische Bundesanstalt für Land und Forstwirtschaft.
- Ansonsten jede Ausprägung fotografieren und Boniturschlüssel dokumentieren
- Alle Beobachtungen wenn möglich durch dieselbe Person und am selben Tag
- Ansonsten gemeinsames Üben der Schätzung
- Randpflanzen nicht schätzen

Boniturnote	Ausprägung
1	Fehlend oder sehr gering
2	Sehr gering bis gering
3	Gering
4	Gering bis Mittel
5	Mittel
6	Mittel bis Stark
7	Stark
8	Stark bis sehr stark
9	Sehr stark

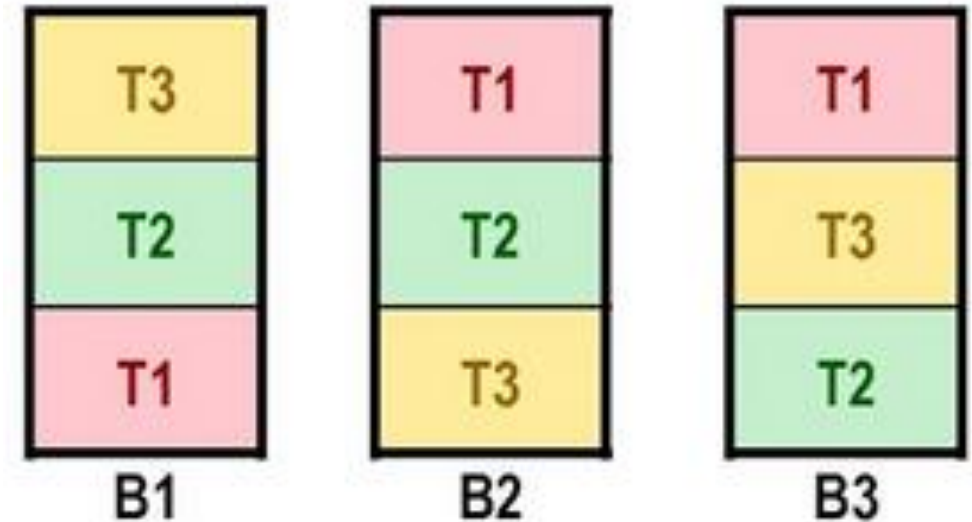
Exemplarischer Boniturschlüssel für die Bonitur eines Merkmales

## 2

# Faktorieller Versuch

### Randomized Complete Block Design (RCBD):

- Die Versuchsgruppen werden in kleinere Gruppen aufgeteilt (Blöcke)
- Jede Treatmentgruppe ist mindestens einmal in jedem Block enthalten
- Soll die Auswirkungen bekannter (aber nicht kontrollierbarer) Größen (Heterogenitäten im Boden, Mikroklima) minimieren um die Effekte im Versuch auf die Treatments / Faktoren eingrenzen zu können. (Erhöhung des Bestimmtheitsmaßes / Anteil erklärte Varianz im stat. Modell)



Konzeptuelle Abbildung eines Randomized Complete Block Design (RCBD), Jeder Block enthält jedes Versuchsglied (Treatmentgruppe) einmal.

## 2

# Faktorieller Versuch

### Operationalisierung

#### **Genauigkeit**

- Stichprobengröße (hängt ab von Variabilität des Merkmals)
- Messungsverfahren
- Häufigkeit der Messung



#### **Aufwand**

- Versuchslaufzeit
- Kapazitäten (Arbeitszeit)
- Kosten



## 2

# Faktorieller Versuch

## Operationalisierung:

- Faktoren
- Datum (oder Codierung)
- Begleitparameter
- Schätzungen
- Maßeinheiten

	A	B	C	D	E	F
1	<u>genotype</u>	<u>trt</u>	<u>date</u>	<u>bbch</u>	<u>height cm</u>	<u>fm veg g</u>
2	nil	ts	t1	12	58,5	7,79
3	nil	ts	t1	11	54	5,85
4	nil	ts	t1	10	56,2	5,96
5	scarlett	ts	t1	13	57,2	10,75
6	scarlett	ts	t1	10	46,4	6,92
7	scarlett	ts	t1	11	56	9,43
8	nil	ww	t1	10	57,8	8,92
9	nil	ww	t1	12	63,6	8,07
10	nil	ww	t1	11	64	8,86
11	scarlett	ww	t1	10	59,7	7,04

Beispielhafter Datensatz für einen Boniturtermin an dem Werte erhoben werden.



## Definition:

- Englisch für „**Organisiert**“
- In Data-Science in R bedeutet dies, dass der Datensatz nach einem **standardisierten Format** strukturiert ist

1 - Wickham, Hadley. 2014. "Tidy Data." *Journal of Statistical Software* Volume 59 (Issue 10).  
<https://www.jstatsoft.org/index.php/jss/article/view/v059i10/v59i10.pdf>.

## Anforderungen:

- „Organisiert“
- **standardisiertes Format**
- Jede Variable (Faktor) und jedes Outcome bestimmen eine **Spalte**

country	year	cases	population
Afghanistan	1999	1745	19557071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	212666	128028583

variables

Aus „R for Datascience“ (Grolemund und Wickham 2017) Kapitel 12.2 – Tidy data;  
1 - Wickham, Hadley. 2014. “Tidy Data.” *Journal of Statistical Software* Volume 59 (Issue 10).  
<https://www.jstatsoft.org/index.php/jss/article/view/v059i10/v59i10.pdf>.

## Anforderungen:

- „Organisiert“
- standardisiertes Format
- Spalte
- Jede Beobachtung füllt eine Reihe

country	year	cases	population
Pakistan	1999	745	133870
Pakistan	2000	2000	200000
Pakistan	1999	37707	1720000
Pakistan	2000	30400	1740040
China	1999	212200	12720102
China	2000	210700	12604200

observations

Aus „R for Datascience“ (Grolemund und Wickham 2017) Kapitel 12.2 – Tidy data;  
1 - Wickham, Hadley. 2014. “Tidy Data.” *Journal of Statistical Software* Volume 59 (Issue 10).  
<https://www.jstatsoft.org/index.php/jss/article/view/v059i10/v59i10.pdf>.

## 3

The „Tidy“ Dataset nach H.Wickham<sup>1</sup>Anforderungen:

- „Organisiert“
- standardisiertes Format
- Spalte
- Reihe
- Jeder Wert füllt / bekommt eine **Zelle** im Datensatz

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20594360
Brazil	1999	37737	172004362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280425583

values

Aus „R for Datascience“ (Grolemund und Wickham 2017) Kapitel 12.2 – Tidy data;  
 1 - Wickham, Hadley. 2014. “Tidy Data.” *Journal of Statistical Software* Volume 59 (Issue 10).  
<https://www.jstatsoft.org/index.php/jss/article/view/v059i10/v59i10.pdf>.

# 4

## Tidy „Messy“ Datasets



### Problem:

- Die meisten Datensätze folgen nicht dem Konzept von „organisierten Daten“
- Meist orientieren sie sich an Benutzerfreundlichkeit



## 4

# Tidy „Messy“ Datasets



## Wide - Format:

Die zwei häufigsten Probleme sind:

- Ein Faktor / Outcome erstreckt sich über mehrere Spalten -> „**Wide – Format**“
- Einige Spalten sind tatsächlich Werte eines Faktors/Outcome (1999 und 2000 sind Werte des Faktors „year“)
- Das Outcome „cases“ (Anzahl Fälle pro Jahr) erstreckt sich über mehrere Spalten und ist nicht klar erkennbar

country	1999	2000
A	0.7K	2K
B	37K	80K
C	212K	213K

Datensatz im „Wide-Format“, Das Outcome „Fälle“ erstreckt sich über mehrere Spalten,  
Aus „Tidyr – Cheatsheet“ (Rstudio 2023)  
<https://raw.githubusercontent.com/rstudio/cheatsheets/main/tidyr.pdf>

## 4

## Tidy „Messy“ Datasets

Wide - Format:

Die zwei häufigsten Probleme sind:

- „Wide – Format“

- `pivot_longer( data,`  
`cols = c(„1999“, „2000“),`  
`names_to = „year“,`  
`values_to = „cases“ )`

country	1999	2000
A	0.7K	2K
B	37K	80K
C	212K	213K



country	year	cases
A	1999	0.7K
B	1999	37K
C	1999	212K
A	2000	2K
B	2000	80K
C	2000	213K

- Spaltennamen rücken in die Spalte „names\_to“ und die Werte in die Spalte „values\_to“

Datensatz im „Wide-Format,  
 Aus „Tidyr – Cheatsheet“ (Rstudio 2023)  
<https://raw.githubusercontent.com/rstudio/cheatsheets/main/tidyr.pdf>

## 4

## Tidy „Messy“ Datasets

Very – long - Format:

Die zwei häufigsten Probleme sind:

- „**Wide – Format**“
- Mehrere Outcomes sind in einer Spalte gespeichert  
-> „**Very-long-format**“
- Jede Beobachtung / Individuum füllt mehrere Reihen (anstatt einer)
- Eine Spalte enthält mehrere Faktoren / Outcomes („count“ enthält „cases“ und „pop“)

country	year	type	count
A	1999	cases	0.7K
A	1999	pop	19M
A	2000	cases	2K
A	2000	pop	20M
B	1999	cases	37K
B	1999	pop	172M
B	2000	cases	80K
B	2000	pop	174M
C	1999	cases	212K
C	1999	pop	1T
C	2000	cases	213K
C	2000	pop	1T

Datensatz im „Very-Long-Format“,  
Aus „Tidyr – Cheatsheet“ (Rstudio 2023)

<https://raw.githubusercontent.com/rstudio/cheatsheets/main/tidyr.pdf>

## 4

## Tidy „Messy“ Datasets

Very – long - Format:

Die zwei häufigsten Probleme sind:

- „Wide – Format“
- „Very-long-format“
- `pivot_wider( data,`  
`names_from = type,`  
`values_from = count )`

country	year	type	count		country	year	cases	pop
A	1999	cases	0.7K	→	A	1999	0.7K	19M
A	1999	pop	19M		A	2000	2K	20M
A	2000	cases	2K		B	1999	37K	172M
A	2000	pop	20M		B	2000	80K	174M
B	1999	cases	37K		C	1999	212K	1T
B	1999	pop	172M		C	2000	213K	1T
B	2000	cases	80K					
B	2000	pop	174M					
C	1999	cases	212K					
C	1999	pop	1T					
C	2000	cases	213K					
C	2000	pop	1T					

- Werte der Spalte „names\_from“ werden zu eigenen Spalten mit den Werten aus „values\_from“ als Einträge / Beobachtungen

Datensatz im „ Very-Long-Format“,  
 Aus „Tidyr – Cheatsheet“ (Rstudio 2023)  
<https://raw.githubusercontent.com/rstudio/cheatsheets/main/tidyr.pdf>

# 5

# Excel Sheet Generator



## Eine R Shiny App in Entwicklung

- Generiert „Tidy“ Exceltabellen zur Verwendung bei faktoriellen Versuchen.
- Faktoren und Faktorlevel können mit ihrem Namen eingegeben werden
- In R Environment - Ordner kopieren und ausprobieren

➤ <https://netcase.hs-osnabrueck.de/index.php/s/GjmeGQz912JQZzK>

# 6

## Literatur und weiterführende Links

### Eine R Shiny App in Entwicklung

- R for Data Science – Tidy Data (Buch)  
<https://r4ds.had.co.nz/tidy-data.html>
- Moderndiver – Import and Tidy Data (Buch)  
<https://moderndive.com/4-tidy.html>
- Tidy – Cheatsheet (Code – Mappe)  
<https://raw.githubusercontent.com/rstudio/cheatsheets/main/tidy.pdf>
- Shiny – Web App Gallery (Beispiele)  
<https://shiny.rstudio.com/gallery/>
- Mastering Shiny (Buch)  
<https://mastering-shiny.org/>