

Sprechstunde
jederzeit nach
Vereinbarung und
nach der Vorlesung

Wallstr. 3, 6. Stock,
Raum 06-206



Mathematische und statistische Methoden I

Dr. Malte Persike



persike@uni-mainz.de



lordsofthebortz.de



twitter.com/methodenlehre



tinyurl.com/gplusmethodenlehre

WiSe 2011/2012

Fachbereich Sozialwissenschaften
Psychologisches Institut
Johannes Gutenberg Universität Mainz

Einführung

Annahmen

QS-Zerlegung

Prüfgröße &
Ergebnstab.

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)

- ⊕ **Ziel:** Analyse des Einflusses einer **unabhängigen Variablen** (UVn) auf eine **abhängige Variable** (AV)
- ⊕ Die AV muss dabei **stetig** sein (intervallskaliert) die UVn sind i.d.R **nominal-** oder **ordinalskaliert**
- ⊕ Anders als bei der einfaktoriellen ANOVA werden in der rmANOVA **dieselben Merkmalsträger** in allen Treatmentstufen beobachtet („Messwiederholung“)



Einführung

Annahmen

QS-Zerlegung

Prüfgröße &
Ergebnstab.

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)

- ⊕ **Vorteil:** Die Fehlervarianz wird i.d.R. kleiner, da sich die Unterschiedlichkeit zwischen den Versuchspersonen auf alle Stufenmittelwerte gleich auswirkt → **schärfere Testung**
- ⊕ Es braucht also nicht auf Varianzhomogenität zwischen den Versuchspersonen geachtet zu werden.
- ⊕ **Nachteil:** Die rmANOVA ist durch **anspruchsvollere Voraussetzungen** gekennzeichnet als die ANOVA
- ⊕ Zudem wirkt sich der Ausfall eines Merkmalsträgers auf alle Stufen aus.

Wenn anwendbar, ist die rmANOVA trotzdem zumeist das wünschenswertere Design.



Einführung

Annahmen

QS-Zerlegung

Prüfgröße &
Ergebnstab.

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)

- ⊕ Die Beobachtungen müssen **abhängig** sein
- ⊕ Die **Fehler** innerhalb einer Person sollen **normalverteilt** sein mit einem Erwartungswert von 0.
- ⊕ Die Varianzen der Differenzen zwischen den Treatmentstufen müssen gleich sein
→ **Sphärizitätsannahme**

V_p	A_1	A_2	...	A_n
1	3.8	5.2	...	6.9
2	4.4	6.6	...	7.7
3	3.9	5.4	...	5.8
...
n	5.1	7.1	...	7.3

identische Varianzen



Einführung

Annahmen

QS-Zerlegung

Prüfgröße &
Ergebnstab.

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)

- ⊕ Die Sphärizitätsannahme ist eine wichtige Voraussetzung
- ⊕ Verletzungen führen zu **progressiveren** Entscheidungen bei der F-Prüfgröße, d.h. es stellen sich schneller Signifikanzen ein (unerwünscht)
- ⊕ Die Sphärizitätsannahme wird mit dem **Mauchly Test** auf Sphärizität geprüft
- ⊕ Ist die Sphärizität verletzt, so ist eine **Korrektur der Freiheitsgrade** vorzunehmen (aber: es empfiehlt sich, die Freiheitsgradkorrektur **immer** durchzuführen).
- ⊕ Korrekturmethode sind die **Greenhouse-Geisser** Korrektur sowie die **Huynh-Feldt** Korrektur



Einführung

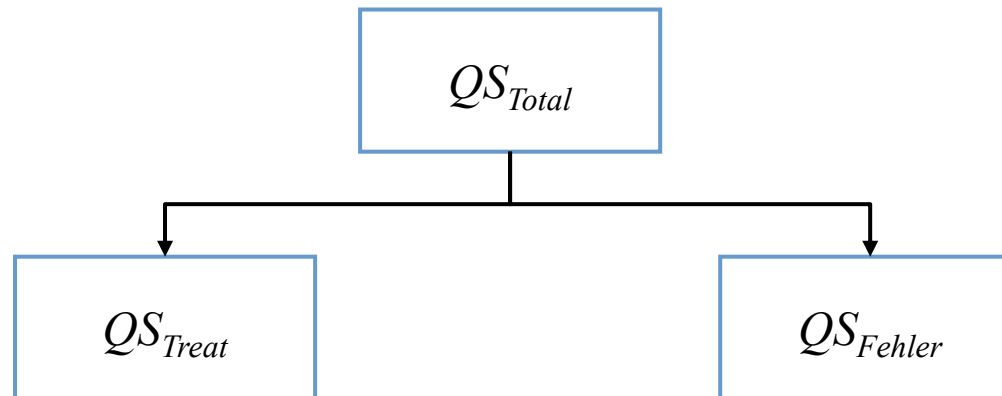
Annahmen

QS-Zerlegung

Prüfgröße &
Ergebnstab.

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)



- ⊕ QS_{Treat} = Wirkung der Faktorstufen (interessant)
- ⊕ QS_{Fehler} = Streuung zwischen Personen (Prüfstreuung)

Einführung

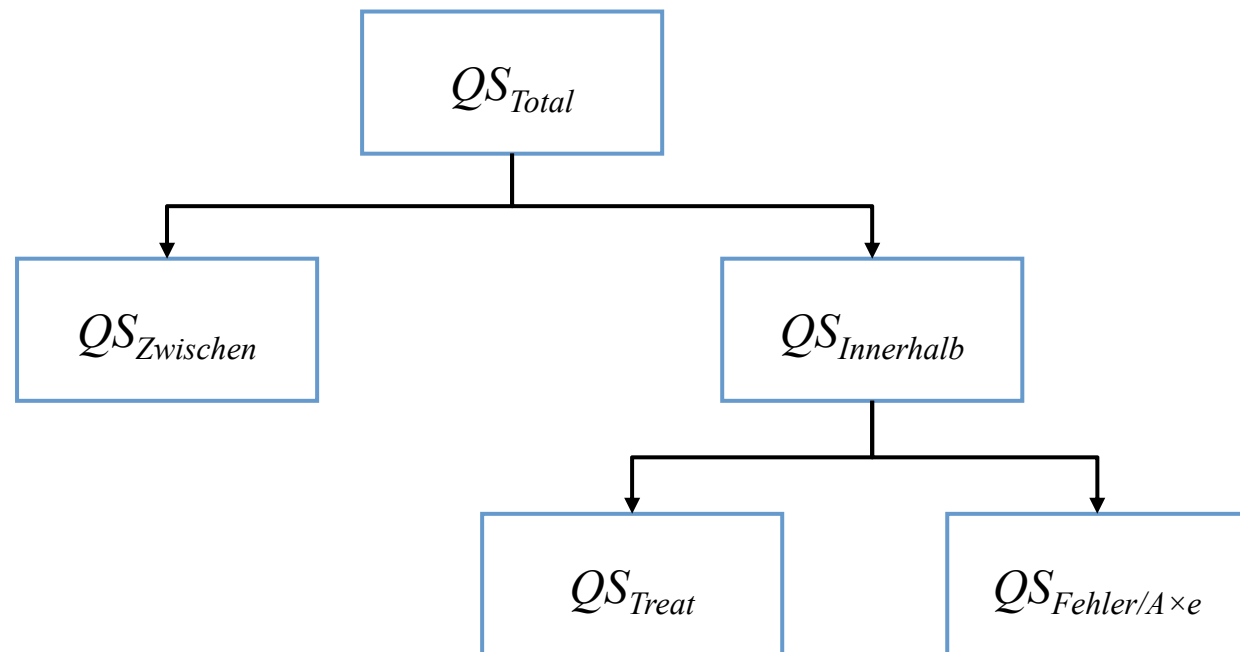
Annahmen

QS-Zerlegung

Prüfgröße &
Ergebnstab.

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)



- ⊕ $QS_{Zwischen}$ = Streuung zwischen Personen (uninteressant)
- ⊕ $QS_{Innerhalb}$ = Streuung innerhalb von Personen
- ⊕ QS_{Treat} = Wirkung der Faktorstufen (interessant)
- ⊕ QS_{Fehler} = Streuung innerhalb einer Person (Prüfstreuung)



Einführung

Annahmen

QS-Zerlegung

Prüfgröße &
Ergebnstab.

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)

- ⊕ Der einfaktoriellen rmANOVA liegt ein einfaches lineares Modell zugrunde

$$x_{ij} = \mu + a_j + e_i + a_j \times e_i$$

Treatment

*Fehler I (Zwischen-
Person Schwankung)*

*Fehler II (Innerhalb-
Person Schwankung)*

x_{ij} = Messwert von Person i ($i = 1 \dots n$) in Treatmentstufe j ($j = 1 \dots p$)

μ = Populationswert aller n Personen

a_j = Effekt der Treatmentstufe j

e_i = Fehler (oder auch "Effekt") der Person i



Einführung

Annahmen

QS-Zerlegung

Prüfgröße &
Ergebnstab.

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)

- ⊕ Daraus lässt sich nun die Quadratsummenzerlegung in der Varianzanalyse ableiten:

$$\begin{aligned}
 QS_{Tot} &= QS_{Zwischen} + QS_{Innerhalb} \\
 \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{G})^2 &= \sum_{i=1}^n \sum_{j=1}^p (\bar{x}_i - \bar{G})^2 + \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_i)^2
 \end{aligned}$$

Mittelwert der Person i über alle j Faktorstufen von A

- ⊕ Die $QS_{Innerhalb}$ zerfällt in

$$\begin{aligned}
 QS_{Innerhalb} &= QS_{Treat} + QS_{Fehler/Axe} \\
 \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_i)^2 &= \sum_{i=1}^n \sum_{j=1}^p (\bar{A}_j - \bar{G})^2 + \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{A}_j - \bar{x}_i + \bar{G})^2
 \end{aligned}$$



Einführung

Annahmen

QS-Zerlegung

**Prüfgröße &
Ergebnstab.**

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)

- ⊕ Aus den Quadratsummen lassen sich wieder spezielle Varianzen bestimmen, die **Populationsvarianzen** $\hat{\sigma}^2$
- ⊕ Dazu werden wieder die **Freiheitsgrade** (df , degrees of freedom) benötigt.
- ⊕ In der rmANOVA gilt:

$$df_{Tot} = n \cdot p - 1$$

$$df_{Zwischen} = n - 1$$

$$df_{Innerhalb} = n \cdot (p - 1)$$

$$df_{Treat} = p - 1$$

$$df_{Fehler/Axe} = (n - 1) \cdot (p - 1)$$



Einführung

Annahmen

QS-Zerlegung

**Prüfgröße &
Ergebnstab.**

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)

⊕ Damit werden die Populationsvarianzen berechnet als

$$\hat{\sigma}_{Tot}^2 = \frac{QS_{Tot}}{df_{tot}} = \frac{\sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{G})^2}{n \cdot p - 1}$$

$$\hat{\sigma}_{Zwischen}^2 = \frac{QS_{Zwischen}}{df_{Zwischen}} = \frac{\sum_{i=1}^n \sum_{j=1}^p (\bar{x}_i - \bar{G})^2}{n - 1}$$

$$\hat{\sigma}_{Innerhalb}^2 = \frac{QS_{Innerhalb}}{df_{Innerhalb}} = \frac{\sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_i)^2}{n \cdot (p - 1)}$$



Einführung

Annahmen

QS-Zerlegung

**Prüfgröße &
Ergebnstab.**

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)

⊕ Damit werden die Populationsvarianzen berechnet als

$$\hat{\sigma}_{Treat}^2 = \frac{QS_{Treat}}{df_{Treat}} = \frac{\sum_{i=1}^n \sum_{j=1}^p (\bar{A}_j - \bar{G})^2}{p-1}$$

$$\hat{\sigma}_{Fehler/A \times e}^2 = \frac{QS_{Fehler/A \times e}}{df_{Fehler/A \times e}} = \frac{\sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{A}_j - \bar{x}_i + \bar{G})^2}{(n-1) \cdot (p-1)}$$



Einführung

Annahmen

QS-Zerlegung

**Prüfgröße &
Ergebnstab.**

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)

⊕ Hat das Treatment **keinen Effekt**, so sind Unterschiede zwischen Mittelwerten der Treatmentstufen rein zufällig

⊕ Die Streuung der Stufenmittelwerte kann dann nur aus der Residualstreuung entstehen. Es muss gelten:

$$\hat{\sigma}_{Treat}^2 \hat{=} \hat{\sigma}_{Fehler}^2$$

⊕ Daraus konstruiert man die **Prüfgröße**

$$F = \frac{\hat{\sigma}_{Treat}^2}{\hat{\sigma}_{Fehler}^2}$$

⊕ Sie ist **F-verteilt** mit $df_{Treat} = p-1$ Zählerfreiheitsgraden und $df_{Fehler} = (n-1) \cdot (p-1)$ Nennerfreiheitsgraden.

⊕ Aus der F-Verteilung kann die **Wahrscheinlichkeit** $p(F)$ für das Auftreten dieser Prüfgröße ermittelt werden



Einführung

Annahmen

QS-Zerlegung

**Prüfgröße &
Ergebnstab.**

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)

Beobachtung im Experiment: $\hat{\sigma}_{Treat}^2$ und $\hat{\sigma}_{Residual}^2$

Frage: Sind die Varianzen in Wahrheit gleich?

Geht die Streuung der Mittelwerte der Treatmentstufen nur auf die Fehlerstreuung zurück?

(1) Festlegung eines Signifikanzniveaus α



(2) Berechnung der Prüfgröße: F
deren Häufigkeitsverteilung theoretisch bekannt ist (F-Verteilung)



(3) Berechnung der Wahrscheinlichkeit für
diese Prüfgröße: $p(F)$



(4) Rückschluss: $p(F) = p(\hat{\sigma}_{Treat}^2 = \hat{\sigma}_{Fehler}^2)$



(5) Vergleich von p mit α und
Treffen der Signifikanzaussage

Aber: Bei dieser Aussage irrt man sich mit einer Wahrscheinlichkeit von $\alpha \cdot 100\%$



Einführung

Annahmen

QS-Zerlegung

**Prüfgröße &
Ergebnstab.**

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)

- ⊕ Man erstellt folgende Ergebnistabelle

Faktor	QS	df	Var	F	p
Zwischen	116.800	9	12.978		
Innerhalb	870.667	20	43.533		
A	582.867	2	291.433	18.227	0.000
Fehler	287.800	18	15.989		
Total	987.467	29	34.051		

- ⊕ Dabei wird die zweite Zeile ($QS_{Innerhalb}$) manchmal auch weggelassen, v.a. von Statistikprogrammen.
- ⊕ Die Fehlervarianz wird auch als Residualvarianz bezeichnet, um den „Innerhalb“-Fehler in der rmANOVA vom „Zwischen“-Fehler in der ANOVA abzugrenzen



Einführung

Annahmen

QS-Zerlegung

Prüfgröße &
Ergebnstab.

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)

- ⊕ Zusätzlich ist zu empfehlen, die Varianzaufklärung η^2 (eta²) für das Model zu bestimmen.
- ⊕ Die Varianzaufklärung für $QS_{Zwischen}$ und $QS_{Innerhalb}$ verläuft analog zur ANOVA für unabhängige Daten.

$$\eta_{Zwischen}^2 = \frac{QS_{Zwischen}}{QS_{Tot}}$$

bzw.

$$\eta_{Innerhalb}^2 = \frac{QS_{Innerhalb}}{QS_{Tot}}$$

- ⊕ Für den Messwiederholungsfaktor wird jedoch die $QS_{Innerhalb}$ als Referenz gewählt.

$$\eta_{Treat}^2 = \frac{QS_{Treat}}{QS_{Innerhalb}}$$

bzw.

$$\eta_{Fehler}^2 = \frac{QS_{Fehler}}{QS_{Innerhalb}}$$



Einführung

Annahmen

QS-Zerlegung

Prüfgröße &
Ergebnstab.

Varianzanalyse

Einfaktorielle rmANOVA (repeated measurements)

- ⊕ Sind die Daten zwischen den Messzeitpunkten perfekt **unkorreliert**, ergibt sich ein ähnliches (knapp weniger signifikantes) Ergebnis wie für die ANOVA für unabhängige Stichproben.
- ⊕ Sind die Daten zwischen den Messzeitpunkten **hochkorreliert**, ergibt sich extreme Teststärke der rmANOVA (frühe Signifikanzen).
- ⊕ Sind die Daten **negativ korreliert**, ergibt sich schwächere Teststärke als für die ANOVA für unabhängige Stichproben (späte Signifikanzen)



Einführung

A-Priori
Kontraste

Scheffé Test

Varianzanalyse

Einfaktorielle rmANOVA – Mittelwertevergleiche

- ⊕ **Prinzip:** Die Mittelwertevergleiche in der rmANOVA folgen demselben Prinzip wie in der unabhängigen ANOVA
- ⊕ **Frage:** Wann ist ein Mittelwertsunterschied zwischen Messzeitpunkten verschieden genug von Null, damit das berechnete D als statistisch signifikant bewertet wird?
- ⊕ **Berechnung:** Es gelten praktisch dieselben Berechnungsformeln wie in der einfaktoriellen ANOVA.
- ⊕ Wie auch dort existieren zwei verschiedene Testvarianten:
 1. **A-Priori Tests** zur Prüfung von Hypothesen, die vor der Untersuchung formuliert wurden.
 2. **A-Posteriori Tests** (Post-hoc Tests) zur Prüfung von Hypothesen, die nach Ansehen der Daten gebildet wurden.



Einführung

A-Priori
Kontraste

Scheffé Test

Varianzanalyse

Prinzip des Testens

Beobachtung im Experiment: Mehrere Stufenmittelwerte \bar{A}_j

Frage: Sind (zwei) Mittelwerte in Wahrheit gleich?
Ist ihre Differenz $D=0$?

- (1) Festlegung eines Signifikanzniveaus α
↓
- (2) Berechnung der Prüfgröße: F
deren Häufigkeitsverteilung theoretisch bekannt ist (F-Verteilung)
↓
- (3) Berechnung der Wahrscheinlichkeit für diese Prüfgröße: $p(F)$
↓
- (4) Rückschluss: $p(F) = p(D = 0)$
↓
- (5) Vergleich von p mit α und Treffen der Signifikanzaussage

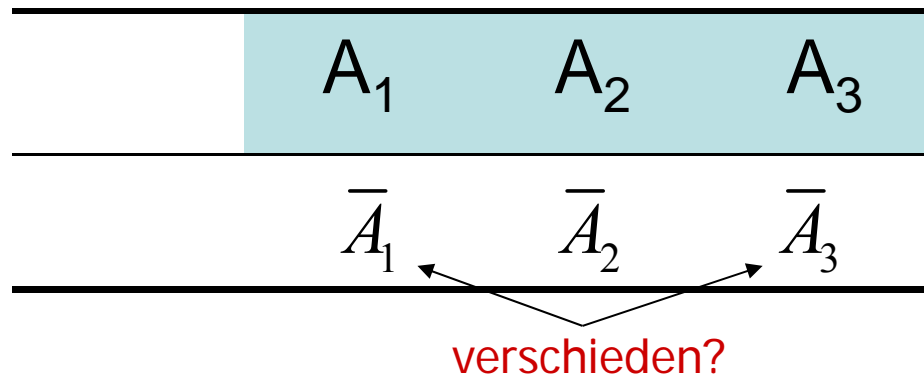
Aber: Bei dieser Aussage irrt man sich mit einer Wahrscheinlichkeit von $\alpha \cdot 100\%$



Varianzanalyse

Einfaktorielle rmANOVA – A-Priori Kontraste

- ⊕ Oft sollen in einer rmANOVA Mittelwerte von Messzeitpunkten auf Unterschiedlichkeit geprüft werden.
- ⊕ Hat a-priori eine Hypothesenbildung stattgefunden, kann dies über **Kontraste** erreicht werden



Einführung

**A-Priori
Kontraste**

Scheffé Test

Varianzanalyse

Einfaktorielle rmANOVA – A-Priori Kontraste

- ⊕ Die Berechnung von a-priori Kontrasten in der rmANOVA verläuft exakt analog zur unabhängigen ANOVA

Die Prüfgröße ist

$$F = \frac{D^2}{\hat{\sigma}_D^2}$$

mit

$$\hat{\sigma}_D^2 = \frac{2}{n} \cdot \hat{\sigma}_{Fehler}^2$$

- ⊕ Für die Prüfgröße F gilt:
 $df_{Zähler} = 1$
 $df_{Nenner} = df_{Fehler}$



Einführung

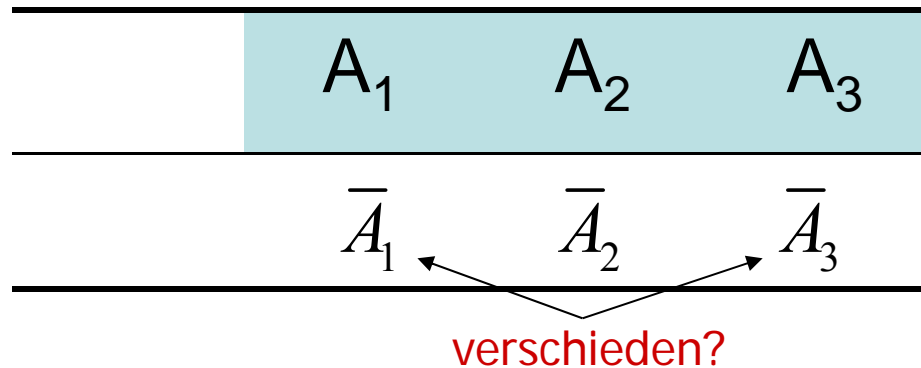
A-Priori
Kontraste

Scheffé Test

Varianzanalyse

Einfaktorielle rmANOVA – A-Posteriori Scheffé Test

- ⊕ Oft sollen in einer rmANOVA Mittelwerte von Messzeitpunkten auf Unterschiedlichkeit geprüft werden.
- ⊕ Hat die Hypothesenbildung erst nach dem Experiment (a-posteriori) stattgefunden, kann der **Scheffé Test** verwendet werden



Einführung

A-Priori
Kontraste

Scheffé Test

Varianzanalyse

Einfaktorielle rmANOVA – A-Posteriori Scheffé Test

- ⊕ Die Prüfgröße für paarweise a-posteriori Mittelwertsunterschiede verändert sich wieder nur um den Faktor $1/(p-1)$ im Vergleich zum a-priori Kontrast

$$F_{corr} = \frac{1}{df_{Zähler}} \cdot F = \frac{1}{df_{Zähler}} \cdot \frac{D^2}{\hat{\sigma}_D^2}$$

wieder mit

$$\hat{\sigma}_D^2 = \frac{2}{n} \cdot \hat{\sigma}_{Fehler}^2$$

- ⊕ Für die Prüfgröße F_{corr} gilt:

$$df_{Zähler} = p - 1$$

$$df_{Nenner} = df_{Fehler}$$



Varianzanalyse

Weitere wichtige Typen von Varianzanalysen

- ⊕ Es können in abhängigen und unabhängigen ANOVA Designs beliebig viele Faktoren mit beliebig vielen Stufen miteinander kombiniert werden
- ⊕ Auch die Kombination von abhängigen und unabhängigen Faktoren ist möglich (**mixed ANOVA**)
- ⊕ **Beispiel:** Therapiestudie mit einem Messwiederholungsfaktor, einem Geschlechtfaktor und einem Faktor „Therapieart“



Varianzanalyse

Weitere wichtige Typen von Varianzanalysen

- ⊕ Die Anzahl von Zellen in der ANOVA wird bei multiplen Faktoren mit vielen Stufen sehr schnell sehr groß (**kombinatorische Explosion**)
- ⊕ Wenn der Experimentator nur an den Haupteffekten interessiert ist, kann der Versuchsplan drastisch reduziert werden, indem nur bestimmte Kombinationen von Stufen der Faktoren realisiert werden

	a1	a2	a3
b1	c1	c2	c3
b2	c2	c3	c1
b3	c3	c1	c2

Lateinische Quadrate



Varianzanalyse

Weitere wichtige Typen von Varianzanalysen

- ⊕ In einigen Untersuchungsdesigns wird ein zweiter Faktor nur für bestimmte Stufen eines anderen Faktors realisiert
- ⊕ **Beispiel:** Therapiestudie mit Placebogruppe, Medikamentengruppe und VT-Gruppe. Die Medikamentengruppe zerfällt in drei Untergruppen (kleine, mittlere, starke Dosis)

Hierarchische ANOVA Designs



Varianzanalyse

Weitere wichtige Typen von Varianzanalysen

- ⊕ Bei vielen Fragestellungen ist es sinnvoll, verschiedene Treatments an denselben Merkmalsträgern zu testen.
- ⊕ **Beispiel:** Strahlungsstärken in der Krebstherapie, Reizstärken bei Tests auf Schmerzempfindlichkeit, Effektivität verschiedener Düngemittel.
- ⊕ **Problem:** Die einer Treatmentstufe ausgesetzte Stelle ist „kontaminiert“ und für weitere Stufen unbrauchbar.
- ⊕ **Lösung:** Man teilt den Merkmalsträger auf in verschiedene, als eigenschaftsgleich definierte Stücke („Plots“) und appliziert hier die verschiedenen Stufen des Treatments.

