

# Statistik II

---

## Einheit 9: Mixed Designs

Wintersemester 2025 | Prof. Dr. Stephan Goerigk

## Kurzvorstellung

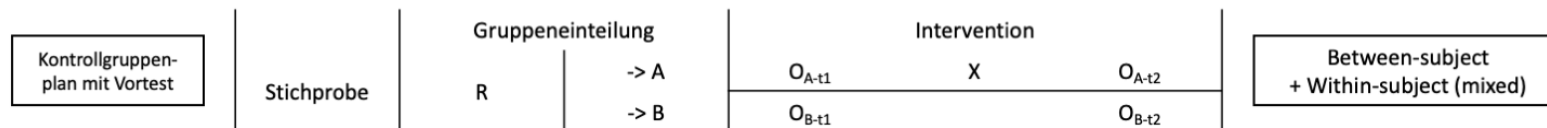
### Mixed Designs:

- Mixed = Mischung aus Messwiederholung und Gruppenvergleich
- Within-Person + Between-Person Faktoren im selben Modell
- Zeit + Gruppenzugehörigkeit (und deren Interaktion) als unabhängige Variablen (UVs)

### Verwendung:

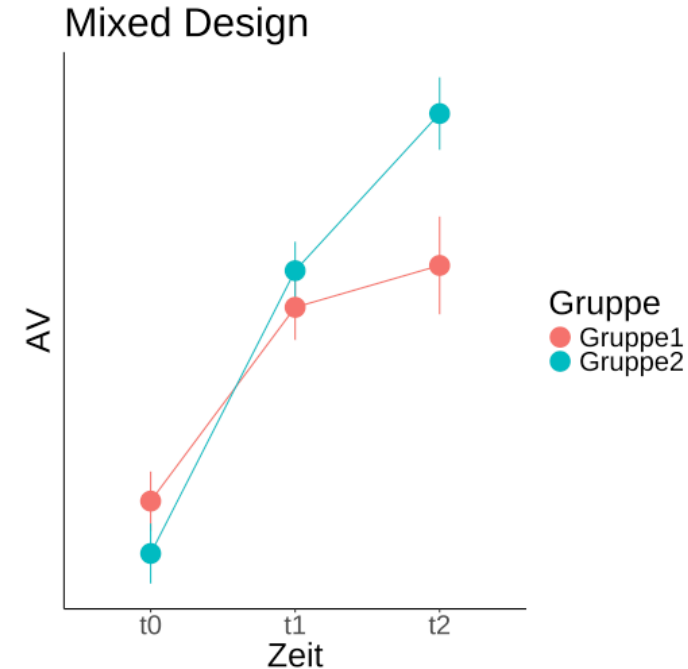
- Klassischer Experimentaufbau in der Psychologie
- Zeit: Prä vs. Post; Gruppe: Experimentalgruppe vs. Kontrollgruppe
- Klinische Studien: Randomisiert-kontrolliertes Trial (RCT)

Wir kennen diesen Aufbau von den Studiendesigns:



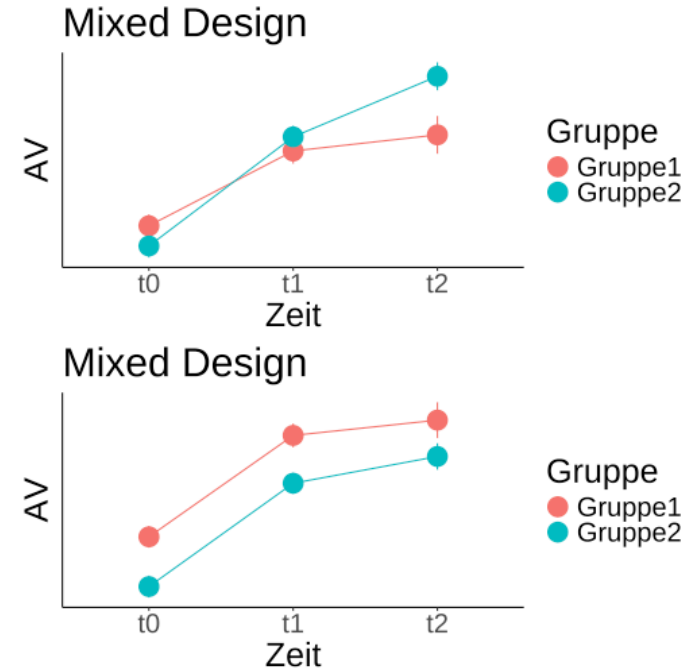
## Kurzvorstellung

- Design:
  - Jede Person wird zu mehreren Zeitpunkten erfasst (Zeit)
  - Einige Personen sind einer Gruppe zugeordnet, andere einer anderen Gruppe
- Gründe:
  - Unterschiedliche Entwicklung über Zeit in unterschiedlichen Gruppen
  - Durch Vorher-Nachher Vergleich kann Veränderung abgebildet werden
- Messwiederholte Daten sind **abhängig** voneinander
- Gruppenvergleiche sind **unabhängig** voneinander



## Kurzworstellung

- Graphische Darstellung i.d.R. mittels gruppenspezifischem Line-Graph
  - AV auf Y-Achse, Zeit auf X-Achse (UV1), Gruppenzugehörigkeit z.B. als Farben (UV2)
  - Punkte = Mittelwert der Gruppe  $G_j$  zu Zeitpunkt  $t_i$
  - Eine Linie pro Gruppe
- Interpretation:
  - parallele Linien → gleiche Veränderung der Gruppen über Zeit
  - nicht parallele Linien → Gruppen verändern sich unterschiedlich



## Statistische Auswertung von Mixed Designs

In der angewandten Statistik sind 2 Arten von Modellen zur Auswertung von Mixed Designs gängig:

### 1. Mixed ANOVA

- Basiert auf Prinzip der mehrfaktoriellen ANOVA
- Messwiederholung: alle Zeitpunkte stehen gleichberechtigt nebeneinander (jede Messung wird mit jeder verglichen)
- Statistisches Konzept: stationäre Mittelwertsvergleiche der Gruppen zu jedem Zeitpunkt
- Dieses Semester

### 2. Linear mixed model (Multilevel-Modell)

- Basiert auf Prinzip der Regression
- Messwiederholung: Zeitpunkte können in lineare Reihenfolge gebracht werden (Wachstumskurve → Veränderung = lineare Steigung)
- Statistisches Konzept: Vergleich der Steigungen einzelner Gruppen
- Erstes Semester Masterstudium

## Anwendungsbespiel

- Datensatz für  $N = 10$  Patient:innen nach Schlaganfall
- 2 Gruppen mit  $n = 5$ : G1 lediglich Zuwendung, G2 erhält Training
- **Forschungsfrage:** Kann kognitives Training Merkfähigkeit stärker verbessern, als reine Zuwendung?
- Es wurden folgende Variablen gemessen:
  - Gedächtnisleistung (AV; 0-50 Punkte) → nach jeder Trainingseinheit gemessen
- Variable im Datensatz
  - Zeitpunkt (UV, 3 Messungen)
  - Gruppe (UV, 2 Stufen)

→ Numerische Frage: Unterschiedlicher Anstieg in Gruppen?

	id	Time	Score	Gruppe
1	1	t0	9	Gruppe1
2	2	t0	10	Gruppe1
6	1	t1	13	Gruppe1
7	2	t1	12	Gruppe1
11	1	t2	16	Gruppe1
12	2	t2	14	Gruppe1
16	6	t0	13	Gruppe2
17	7	t0	10	Gruppe2
21	6	t1	17	Gruppe2
22	7	t1	16	Gruppe2
26	6	t2	18	Gruppe2
27	7	t2	22	Gruppe2

Tabellarische Darstellung je für  $n = 2$  Personen in jeder Gruppe  
(Platzgründe)

## Formalisierung

- "Mixed"-Modelle sind Spezialformen mehrfaktorieller Modelle mit Interaktion
- Besonderheit: eine der beiden UVs ist Zeit

$$Y = Zeit + Gruppe + Zeit \cdot Gruppe + \varepsilon$$

Entsprechend können wir wieder **3 Arten von Effekten** unterscheiden:

- Haupteffekt Zeit
- Haupteffekt Gruppe
- Interaktion Zeit x Gruppe

## Effektarten und prüfbare Hypothesen

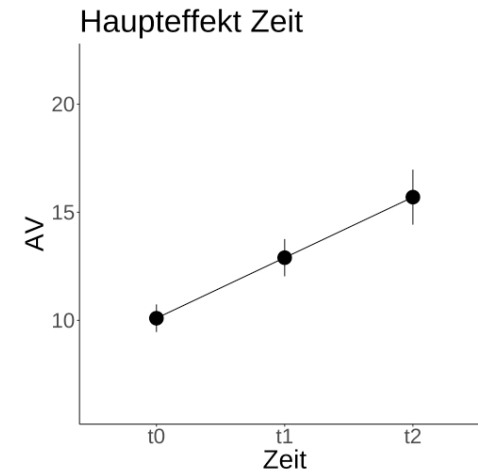
### Haupteffekt Zeit:

$$Y = \text{Zeit} + \text{Gruppe} + \text{Zeit} \cdot \text{Gruppe} + \varepsilon$$



Time	Mittelwert	SD
t0	10.1	2.02
t1	12.9	2.73
t2	15.7	4.03

- $H_1$  : Unterschied zwischen  $t_0$ ,  $t_1$  und  $t_2$  (ungeachtet der Gruppenzugehörigkeit)





## Effektarten und prüfbare Hypothesen

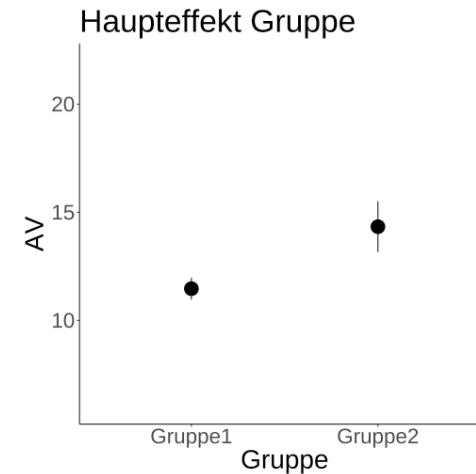
### Haupteffekt Gruppe:

$$Y = \text{Zeit} + \text{Gruppe} + \text{Zeit} \cdot \text{Gruppe} + \varepsilon$$



Gruppe	Mittelwert	SD
Gruppe1	11.47	1.96
Gruppe2	14.33	4.56

- $H_1$  : Unterschied zwischen Gruppe<sub>1</sub> und Gruppe<sub>2</sub> (ungeachtet des Messzeitpunkts)



## Effektarten und prüfbare Hypothesen

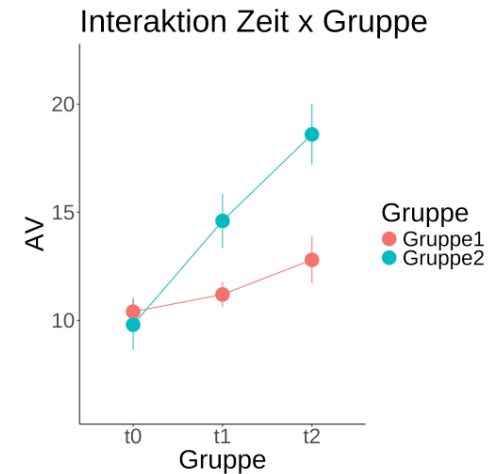
### Interaktionseffekt Zeit x Gruppe:

$$Y = \text{Zeit} + \text{Gruppe} + \text{Zeit} \cdot \text{Gruppe} + \varepsilon$$



Time	Gruppe	Mittelwert	SD
t0	Gruppe1	10.4	1.52
t0	Gruppe2	9.8	2.59
t1	Gruppe1	11.2	1.30
t1	Gruppe2	14.6	2.79
t2	Gruppe1	12.8	2.39
t2	Gruppe2	18.6	3.13

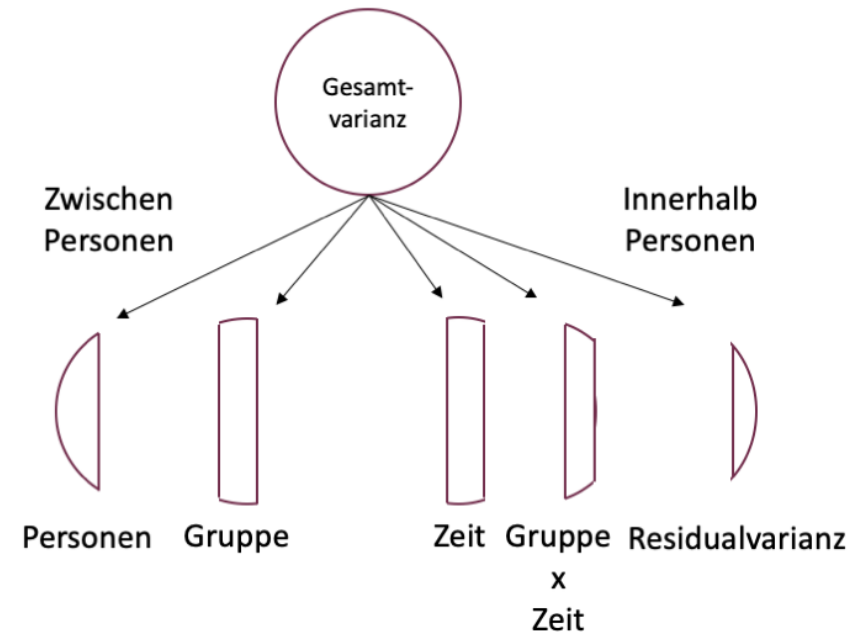
- $H_1$  : Gruppe<sub>1</sub> und Gruppe<sub>2</sub> unterscheiden sich im zeitlichen Verlauf.



## Prinzip der Mixed ANOVA

### Zerlegung der Gesamtvarianz:

- *Personen* = individuelle Unterschiede
- *Gruppe* = durchschnittlicher Gruppenunterschied
- *Zeit* = durchschnittlicher Unterschied über Zeit
- *Zeit x Gruppe* = unterschiedliche Veränderung über Zeit zwischen Gruppen
- *Residuum* = individuelle nicht erklärte Unterschiede über Zeit



## Prinzip der Mixed ANOVA

### Zerlegung der Gesamtvarianz:

Wir müssen uns wiederum fragen, weshalb Messungen der AV unterschiedlich (mit Varianz) ausfallen:

Variation zwischen Personen:

- Eigenschaften der Person selbst (Personenvarianz)
- Eigenschaften/Effekt der Gruppe (Gruppenzugehörigkeit)

Variation Messwiederholung:

- aufgrund in der Zwischenzeit erfolgter Intervention (oder einfach aufgrund des Vergehens von Zeit)
- unterschiedliche Reaktion auf in der Zwischenzeit erfolgter Intervention (oder einfach Vergehen von Zeit) innerhalb der Gruppen

Variation der Messwerte:

$$\sigma_{gesamt}^2 = \sigma_{Vpn}^2 + \sigma_{Zeit}^2 + \sigma_{Gruppe}^2 + \sigma_{Zeit \times Gruppe}^2 + \sigma_{Res}^2$$

## Prinzip der Mixed ANOVA

### Systematische Varianzen (Zeit, Gruppe, Zeit x Gruppe):

Die zu schätzenden, systematischen Varianzkomponenten sind:

- $\hat{\sigma}_{Gruppe}^2$
- $\hat{\sigma}_{Zeit}^2$
- $\hat{\sigma}_{Zeit \times Gruppe}^2$

Formeln nutzen folgende Parameter:

- $p$  = Stufen der Gruppenvariable ( $i$  Stufen)
- $q$  = Stufen der Zeitvariable ( $j$  Messzeitpunkte)
- $n$  = Anzahl der Versuchspersonen

## Prinzip der Mixed ANOVA

### Systematische Varianzen (Zeit, Gruppe, Zeit x Gruppe):

- Schätzung erfolgt analog zur 2-faktoriellen ANOVA ohne Messwiederholung.
- Der Unterschied: Erwartungswerte der Schätzer unterscheiden sich:
- Erwartungswert von  $\hat{\sigma}_{Gruppe}^2$  beinhaltet Effektvarianz, Varianz des Personenfaktors und Fehlervarianz

$$E(\hat{\sigma}_{Gruppe}^2) = n \cdot q \cdot \sigma_{Gruppe}^2 + q \cdot \sigma_{Vpn}^2 + \sigma_{Res}^2$$

- Erwartungswerte von  $\hat{\sigma}_{Zeit}^2$  und  $\hat{\sigma}_{Zeit \times Gruppe}^2$  beinhalten zusätzlich Wechselwirkung zwischen Personenfaktor und Zeit

$$E(\hat{\sigma}_{Zeit}^2) = n \cdot p \cdot \sigma_{Zeit}^2 + q \cdot \sigma_{Zeit \times Vpn}^2 + \sigma_{Res}^2$$

$$E(\hat{\sigma}_{Zeit \times Gruppe}^2) = n \cdot \sigma_{Zeit \times Gruppe}^2 + \sigma_{Zeit \times Vpn}^2 + \sigma_{Res}^2$$

## Prinzip der Mixed ANOVA

### Systematische Varianzen (Gruppe):

Berechnung:

$$\hat{\sigma}_{Gruppe}^2 = \frac{QS_{Gruppe}}{df_{Gruppe}}$$

Die Quadratsumme ( $QS$ ) entspricht der quadrierten Abweichung der Mittelwerte ( $\bar{Gruppe}_i$ ) vom Gesamtmittelwert ( $\bar{G}$ ) multipliziert mit der Zellgröße  $n$  und der Anzahl der Stufen der Zeitvariable (Anzahl der Messwiederholungen):

$$\hat{\sigma}_{Gruppe}^2 = \frac{\sum_{i=1}^p n \cdot q \cdot (\bar{Gruppe}_i - \bar{G})^2}{p - 1}$$

## Prinzip der Mixed ANOVA

### Systematische Varianzen (Zeit):

Berechnung:

$$\hat{\sigma}_{Zeit}^2 = \frac{QS_{Zeit}}{df_{Zeit}}$$

Die Quadratsumme ( $QS$ ) entspricht der quadrierten Abweichung der Mittelwerte ( $\bar{Zeit}_j$ ) vom Gesamtmittelwert ( $\bar{G}$ ) multipliziert mit der Zellgröße  $n$  und der Anzahl der Stufen der Gruppenvariable:

$$\hat{\sigma}_{Zeit}^2 = \frac{\sum_{j=1}^q n \cdot p \cdot (\bar{Zeit}_j - \bar{G})^2}{q - 1}$$



## Prinzip der Mixed ANOVA

### Systematische Varianzen (Zeit x Gruppe):

Berechnung:

$$\hat{\sigma}_{Zeit \times Gruppe}^2 = \frac{QS_{Zeit \times Gruppe}}{df_{Zeit \times Gruppe}}$$

Die Quadratsumme entspricht der Differenz der beobachteten Zellmittelwerte und der erwarteten Zellmittelwerte

$$\hat{\sigma}_{Zeit \times Gruppe}^2 = \frac{\sum_{i=1}^p \sum_{j=1}^q n \cdot [Gruppe\bar{Zeit}_{ij} - (Gruppe_i + Zeit_j + \bar{G})]^2}{(p-1) \cdot (q-1)}$$

Laut Haupteffekten von Gruppe und Zeit erwartete Zellmittelwerte:

$$Gruppe\bar{Zeit}_{ij} = Gruppe_i + Zeit_j - \bar{G}$$

## Prinzip der Mixed ANOVA

### Prüfvarianzen:

- Logik F-Bruch: Teilen der Effektvarianz durch die nicht erklärte Varianz

$$F_{(df_1, df_2)} = \frac{\hat{\sigma}_{Effekt}^2}{\hat{\sigma}_{Prüf}^2}$$

→ Zähler und Nenner unterscheiden sich nur durch interessierenden Effekt (je stärker desto größer der Bruch, desto signifikanter der Test)

- Bislang konnte für Haupt- und Interaktionseffekte dieselbe Prüfvarianz im Nenner verwendet werden ( $\hat{\sigma}_{Res}^2$ )
- Problem im mixed Desgin:
  - Erwartungswerte für Faktoren ohne Messwiederholung enthalten  $\sigma_{Vpn}^2$
  - Erwartungswerte für Faktoren mit Messwiederholung enthalten  $\sigma_{Zeit \times Vpn}^2$

→ Für Faktoren mit vs. ohne Messwiederholung müssen unterschiedliche Prüfvarianzen für den Nenner des F-Bruchs gebildet werden

## Prinzip der Mixed ANOVA

### Prüfvarianz für Faktoren ohne Messwiederholung:

- Prüfvarianz für (*Zeit*) spielt hier keine Rolle
- Ziel: Abbilden der Variabilität der Personen in der Stichprobe (zwischen Personen)

Vorgehen:

- Für jede Vpn über alle Messzeitpunkte einen Durchschnitt bilden ( $\bar{GruppeVpn}_{im}$ )
- Dann den Gesamtmittelwert der Einzelgruppen von den gemittelten Werten über alle Zeitpunkte abziehen (unerklärte Varianz der Vpn)

$$\hat{\sigma}_{Prüf(Gruppe)}^2 = \hat{\sigma}_{Vpn}^2 = \frac{QS_{Vpn}}{df_{Vpn}} = \frac{\sum_{i=1}^p \sum_{m=1}^n q \cdot [\bar{GruppeVpn}_{im} - \bar{Gruppe}_i]^2}{p \cdot (n - 1)}$$

## Prinzip der Mixed ANOVA

### Prüfvarianz für Faktoren mit Messwiederholung:

- Messwiederholung Zeit und Zeit x Gruppe Interaktion identisch
- Ziel: Abbilden der Variabilität in der Veränderung der Personen (Variabilität innerhalb der Vpn)
- Unterschiede zwischen Vpn müssen auch mitberücksichtigt werden → auch  $\overline{GruppeVpn}_{im}$  und  $(\overline{Gruppe}_i)$  gehen mit ein

$$\hat{\sigma}_{Prüf(Zeit \times Gruppe)}^2 = \hat{\sigma}_{Zeit \times Vpn}^2 = \frac{QS_{Zeit \times Vpn}}{df_{Zeit \times Vpn}} = \frac{\sum_{i=1}^p \sum_{j=1}^q \sum_{m=1}^n [x_{ijm} - [\overline{GruppeZeit}_{ij} + \overline{GruppeVpn}_{im} - \overline{Gruppe}_i]]^2}{p \cdot (q - 1) \cdot (n - 1)}$$

## Prinzip der Mixed ANOVA

### Signifikanztests der Haupteffekte:

Haupteffekt Gruppe:

$$F_{Gruppe(df_{Gruppe}, df_{Vpn})} = \frac{\hat{\sigma}_{Gruppe}^2}{\hat{\sigma}_{Prüf(Gruppe)}^2} = \frac{\hat{\sigma}_{Gruppe}^2}{\hat{\sigma}_{Vpn}^2}$$

mit  $df_{Gruppe} = p - 1$  und  $df_{Vpn} = p \cdot (n - 1)$

Haupteffekt Zeit:

$$F_{Zeit(df_{Zeit}, df_{Zeit \times Vpn})} = \frac{\hat{\sigma}_{Zeit}^2}{\hat{\sigma}_{Prüf(Zeit)}^2} = \frac{\hat{\sigma}_{Zeit}^2}{\hat{\sigma}_{Zeit \times Vpn}^2}$$

mit  $df_{Zeit} = q - 1$  und  $df_{Zeit \times Vpn} = p \cdot (q - 1) \cdot (n - 1)$

## Prinzip der Mixed ANOVA

### Signifikanztest des Interaktionseffekts:

Interaktion:

$$F_{Zeit \times Gruppe(df_{Zeit \times Gruppe}, df_{Zeit \times Vpn})} = \frac{\hat{\sigma}_{Zeit \times Gruppe}^2}{\hat{\sigma}_{Prüf(Zeit \times Gruppe)}^2} = \frac{\hat{\sigma}_{Zeit \times Gruppe}^2}{\hat{\sigma}_{Zeit \times Vpn}^2}$$

mit  $df_{Zeit \times Gruppe} = (p - 1) \cdot (q - 1)$  und  $df_{Zeit \times Vpn} = p \cdot (q - 1) \cdot (n - 1)$

## Prinzip der Mixed ANOVA

### Signifikanztests - Vorgehen:

Händisch:

1. Berechnung der empirischen F-Werte mittels der 3 F-Bruch Formeln für Haupt- und Interaktionseffekte
2. Nachschlagen der kritischen F-Werte in F-Tabelle mittels Zähler- und Nennerfreiheitsgraden
3. Vergleich empirischer F-Wert vs. kritischer F-Wert
4. Wenn  $F_{emp} > F_{krit} \rightarrow$  Test signifikant

In R (folgt auf nächster Folie):

1. Vergleich von  $P$  mit Signifikanzniveau ( $\alpha = .05$ )
2. Wenn  $P < .05 \rightarrow$  Test signifikant

## Anwendungsbespiel R

```
model = aov_ez(dv = "Score", within = c("Time"), between = "Gruppe", id = "id", data = df_long)
summary(model)
```

```
##
## Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
##
##          Sum Sq num Df Error SS den Df  F value    Pr(>F)
## (Intercept) 4992.3      1    92.4      8 432.2338 0.00000003004 ***
## Gruppe       61.6      1    92.4      8  5.3362  0.04969 *
## Time        156.8      2    43.6     16 28.7706 0.0000502007 ***
## Gruppe:Time   52.3      2    43.6     16  9.5902  0.00183 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Mauchly Tests for Sphericity
##
##          Test statistic p-value
## Time          0.43765 0.055453
## Gruppe:Time    0.43765 0.055453
##
##
## Greenhouse-Geisser and Huynh-Feldt Corrections
## for Departure from Sphericity
##
##          GG eps Pr(>F[GG])
## Time          0.64006 0.0001688 ***
## Gruppe:Time 0.64006 0.0081544 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
##          HF eps    Pr(>F[HF])
## Time          0.7084258 0.00008619566
## Gruppe:Time 0.7084258 0.00612281849
```

- **Sum Sq** = Quadratsumme Systematische Varianz
- **num Df** = Zähler-Freiheitsgrade
- **den Df** = Nenner-Freiheitsgrade
- **Error SS** = Quadratsumme Prüfvarianz
- **F value** = Empirischer F-Wert (Ergebnis des F-Bruchs)
- **Pr(>F)** = P-Wert (zu vergleichen mit  $\alpha = .05$ )



## Anwendungsbeispiel R

```
emmeans(model, pairwise ~ Time | Gruppe)$emmeans
```

```
## Gruppe = Gruppe1:
##   Time emmean    SE df lower.CL upper.CL
##   t0      10.4 0.949  8     8.21    12.6
##   t1      11.2 0.975  8     8.95    13.4
##   t2      12.8 1.240  8     9.93    15.7
##
## Gruppe = Gruppe2:
##   Time emmean    SE df lower.CL upper.CL
##   t0       9.8 0.949  8     7.61    12.0
##   t1      14.6 0.975  8    12.35    16.8
##   t2      18.6 1.240  8    15.73    21.5
##
## Confidence level used: 0.95
```

- Auf der linken Seite sehen wir die Mittelwerte (marginal means)
- Alle Kombinationen der Zeit x Gruppe Interaktion

```
emmeans(model, pairwise ~ Time | Gruppe)$contrasts
```

```
## Gruppe = Gruppe1:
##   contrast estimate    SE df t.ratio p.value
##   t0 - t1      -0.8 0.768  8   -1.042  0.5735
##   t0 - t2      -2.4 1.380  8   -1.741  0.2490
##   t1 - t2      -1.6 0.883  8   -1.812  0.2264
##
## Gruppe = Gruppe2:
##   contrast estimate    SE df t.ratio p.value
##   t0 - t1      -4.8 0.768  8   -6.249  0.0006
##   t0 - t2      -8.8 1.380  8   -6.384  0.0005
##   t1 - t2      -4.0 0.883  8   -4.529  0.0048
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

- Auf der rechten Seite sehen wir die Mittelwertsvergleiche
- Nach Gruppen getrennt dargestellt, um die Interaktion aufzuschlüsseln (unterschiedliche Zeiteffekte innerhalb der Gruppen)
- Interaktionseffekt getragen durch nicht-signifikanten vs. signifikanten Vergleich von t1 vs. t2

## Überprüfung der Sphärizität - Mauchly-Test:

```
model = aov_ez(dv = "Score", within = c("Time"), between = "Gruppe", id = "id", data = df_long)
summary(model)
```

```
##
## Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
##
##              Sum Sq num Df Error SS den Df  F value    Pr(>F)
## (Intercept) 4992.3      1   92.4      8 432.2338 0.00000003004 ***
## Gruppe       61.6      1   92.4      8   5.3362    0.04969 *
## Time        156.8      2   43.6     16 28.7706 0.00000502007 ***
## Gruppe:Time   52.3      2   43.6     16   9.5902    0.00183 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Mauchly Tests for Sphericity
##
##              Test statistic p-value
## Time           0.43765 0.055453
## Gruppe:Time     0.43765 0.055453
##
##
## Greenhouse-Geisser and Huynh-Feldt Corrections
## for Departure from Sphericity
##
##              GG eps Pr(>F[GG])
## Time           0.64006 0.0001688 ***
## Gruppe:Time    0.64006 0.0081544 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              HF eps    Pr(>F[HF])
## Time           0.7084258 0.00008619566
## Gruppe:Time    0.7084258 0.00612281849
```

- Annahme: Homogenität der Varianzen zwischen den Faktorstufen
- Signifikanter Mauchly-Test → Varianzen inhomogen → keine Sphärizität
- Nur für Messwiederholungsfaktoren mit  $> 2$  Messungen pro Person relevant
- Bei Sphärizitätsverletzung Korrekturverfahren anwenden (s.h. Einheit 8)
  - Greenhouse-Geisser Korrektur
  - Huynh-Feldt Korrektur

## Effektstärke partielles $\eta^2$ in R

```
ES = effectsize::eta_squared(model)
```

```
ES
```

```
## # Effect Size for ANOVA (Type III)
##
## Parameter | Eta2 (partial) | 95% CI
## -----
## Gruppe | 0.40 | [0.00, 1.00]
## Time | 0.78 | [0.58, 1.00]
## Gruppe:Time | 0.55 | [0.21, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

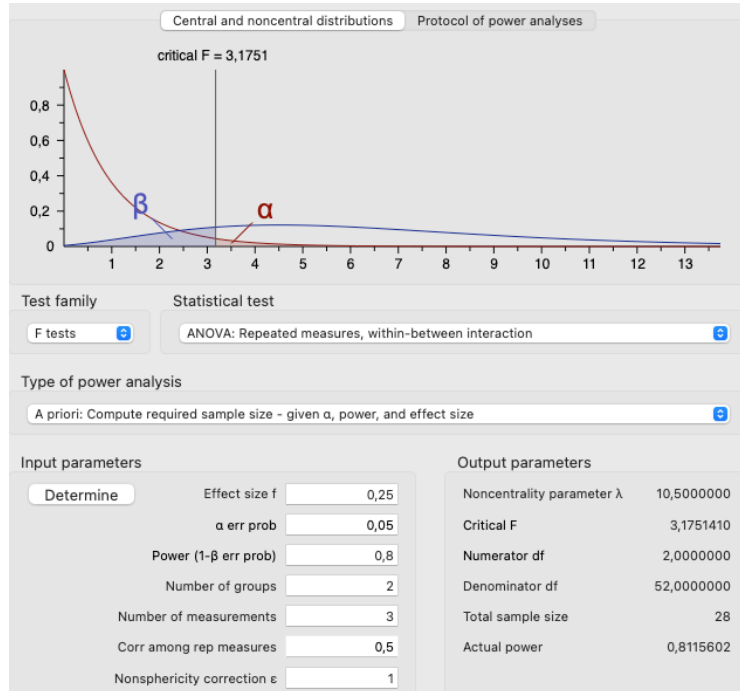
```
# Automatisches Anzeigen der Interpretation
```

```
effectsize::interpret_eta_squared(ES)
```

```
## # Effect Size for ANOVA (Type III)
##
## Parameter | Eta2 (partial) | 95% CI | Interpretation
## -----
## Gruppe | 0.40 | [0.00, 1.00] | large
## Time | 0.78 | [0.58, 1.00] | large
## Gruppe:Time | 0.55 | [0.21, 1.00] | large
##
## - One-sided CIs: upper bound fixed at [1.00].
## - Interpretation rule: field2013
```

- Partielles  $\eta^2$  gibt Anteil aufgeklärter Gesamtvarianz in der Stichprobe für die Modellfaktoren an
- In diesem Fall: alle Effekte werden als "große/starke" Effektstärke eingeordnet

## Stichprobenumfangsplanung



### Folgende Parameter werden benötigt:

- angenommene Effektstärke (in Cohens f)
- Signifikanzniveau ( $\alpha$ )
- gewünschte Teststärke/Power ( $1 - \beta$ )
- Anzahl Gruppen
- Anzahl Messzeitpunkte
- Durchschnittliche angenommene Korrelation zwischen Messzeitpunkten
- Sphärizitätskorrektur (i.d.R. nicht apriori geplant bzw. man hofft auf Sphärizität)

## Berichten der Ergebnisse nach APA

### Studiendesign

Im Rahmen einer randomisiert-kontrollierten Studie (RCT) soll die Wirksamkeit eines neu entwickelten Therapieverfahrens ("**PDD\_Therapie**") für Patient\*innen mit chronischer Depression (PDD = Persistent Depressive Disorder) untersucht werden.

- Insgesamt wurden  $N = 90$  stationär behandelte Personen mit chronischer Depression in die Studie eingeschlossen.
- Die Teilnehmenden wurden **randomisiert** einer von zwei Behandlungsgruppen zugewiesen:
- Die Gruppe **PDD\_Therapy** ( $n = 45$ ) erhielt die neu entwickelte psychotherapeutische Intervention.
- Die Gruppe **TAU** (Treatment as Usual,  $n = 45$ ) erhielt eine etablierte kognitive Verhaltenstherapie (KVT).
- **Forschungsfrage:** Führt PDD\_Therapy im Vergleich zu TAU zu einer stärkeren Reduktion depressiver Symptome im Therapieverlauf?
- Variablen:
  - **Abhängige Variable:** Depressivität, erfasst mit dem Beck Depression Inventory-II (BDI-II; Wertebereich: 0–63).
  - **Unabhängige Variablen:**
    - Gruppe (2 Stufen: PDD\_Therapy, TAU)
    - Messzeitpunkt (3 Zeitpunkte: vor Beginn, nach 3 Wochen und nach 6 Wochen Psychotherapie)

	id	Messzeitpunkt	BDI_II	Gruppe
1	1	t0	28	PDD_Therapy
2	2	t0	35	PDD_Therapy
46	1	t1	30	PDD_Therapy
47	2	t1	25	PDD_Therapy
91	1	t2	14	PDD_Therapy
92	2	t2	21	PDD_Therapy
136	46	t0	43	TAU
137	47	t0	32	TAU
181	46	t1	39	TAU
182	47	t1	32	TAU
226	46	t2	23	TAU
227	47	t2	26	TAU

Tabellarische Darstellung je für  $n = 2$  Personen in jeder Gruppe (Platzgründe)

## Berichten der Ergebnisse nach APA

**Depressivität** gemessen durch BDI-II (Beck Depression Inventory-II)  
**Gruppe:** TAU (treatment as usual); PDD\_Therapy **Messzeitpunkte:** 3 MZP

```
##
## Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
##
##               Sum Sq num Df Error SS den Df  F value    Pr(>F)
## (Intercept)    231324      1   1899.3      88 10717.869 < 2.2e-16 ***
## Gruppe         446      1   1899.3      88   20.663 1.738e-05 ***
## Messzeitpunkt   7023      2  3842.3     176  160.858 < 2.2e-16 ***
## Gruppe:Messzeitpunkt 692      2  3842.3     176   15.857 4.656e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Mauchly Tests for Sphericity
##
##               Test statistic p-value
## Messzeitpunkt      0.90367 0.012204
## Gruppe:Messzeitpunkt 0.90367 0.012204
##
##
## Greenhouse-Geisser and Huynh-Feldt Corrections
## for Departure from Sphericity
##
##               GG eps Pr(>F[GG])
## Messzeitpunkt  0.91214 < 2.2e-16 ***
## Gruppe:Messzeitpunkt 0.91214 1.262e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               HF eps  Pr(>F[HF])
## Messzeitpunkt  0.9304264 7.857475e-38
## Gruppe:Messzeitpunkt 0.9304264 1.025451e-06
```

**Statistischer Bericht: (In Ihrer Klausur)** Wenn Sie in Ihrer Klausur den Output eines mixed Designs berichten sollen, könnte dies so aussehen:

Im Rahmen eines RCTs wurde die Wirksamkeit einer neu entwickelten Therapie für chronisch depressive Patient\*innen (PDD\_Therapy) untersucht.

Es zeigte sich ein signifikanter **Haupteffekt der Gruppe**,  $F(1, 88) = 20.66$ ,  $p < .001$ . Auch der **Faktor Messzeitpunkt** hatte unabhängig von der Gruppenzugehörigkeit einen signifikanten Einfluss auf die Depressivität (BDI-II), Huynh-Feldt korrigiert:  $F(2, 176) = 160.86$ ,  $p < .001$ . Darüber hinaus wurde eine signifikante **Interaktion zwischen Gruppe und Messzeitpunkt** gefunden, Huynh-Feldt  $F(2, 176) = 15.86$ ,  $p < .001$ .

Aufgrund signifikanter Ergebnisse in den Mauchly-Tests ( $ps < .05$ ) wurde bei Effekten mit Messwiederholung die **Huynh-Feldt-Korrektur** angewendet.

## Berichten der Ergebnisse nach APA

**Depressivität** gemessen durch BDI-II (Beck Depression Inventory-II)

**Gruppe:** TAU (treatment as usual); PDD\_Therapy **Messzeitpunkte:** 3 MZP

```
##
## Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
##
##               Sum Sq num Df Error SS den Df  F value    Pr(>F)
## (Intercept)    231324      1   1899.3      88 10717.869 < 2.2e-16 ***
## Gruppe         446      1   1899.3      88   20.663 1.738e-05 ***
## Messzeitpunkt   7023      2   3842.3     176 160.858 < 2.2e-16 ***
## Gruppe:Messzeitpunkt 692      2   3842.3     176   15.857 4.656e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Mauchly Tests for Sphericity
##
##               Test statistic p-value
## Messzeitpunkt      0.90367 0.012204
## Gruppe:Messzeitpunkt 0.90367 0.012204
##
##
## Greenhouse-Geisser and Huynh-Feldt Corrections
## for Departure from Sphericity
##
##               GG eps Pr(>F[GG])
## Messzeitpunkt      0.91214 < 2.2e-16 ***
## Gruppe:Messzeitpunkt 0.91214 1.262e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               HF eps  Pr(>F[HF])
## Messzeitpunkt      0.9304264 7.857475e-38
## Gruppe:Messzeitpunkt 0.9304264 1.025451e-06
```

### Inhaltlich bedeutet dies:

Es zeigten sich signifikante Mittelwertsunterschiede zwischen den beiden Therapiegruppen, über alle Messzeitpunkte hinweg betrachtet. Zudem traten signifikante Mittelwertsunterschiede zwischen den Messzeitpunkten auf, unabhängig von der Gruppenzugehörigkeit. Der **Symptomverlauf** unterschied sich zudem **zwischen den Gruppen**, d.h. die Entwicklung der Depressionswerte verlief in den beiden Interventionsbedingungen unterschiedlich über die Zeit hinweg.

## Post-hoc Vergleich

**Depressivität** gemessen durch BDI-II (Beck Depression Inventory-II)

**Gruppe:** TAU (treatment as usual); PDD\_Therapy **Messzeitpunkte:** 3 MZP

```
## $emmeans
##   Gruppe   Messzeitpunkt emmean    SE df lower.CL upper.CL
##   PDD_Therapy t0          34.8 0.626 88     33.6     36.1
##   TAU        t0          34.5 0.626 88     33.2     35.7
##   PDD_Therapy t1          30.2 0.725 88     28.8     31.6
##   TAU        t1          31.3 0.725 88     29.8     32.7
##   PDD_Therapy t2          18.9 0.729 88     17.5     20.4
##   TAU        t2          25.9 0.729 88     24.5     27.4
##
## Confidence level used: 0.95
##
## $contrasts
##   contrast              estimate    SE df t.ratio p.value
##   PDD_Therapy t0 - TAU t0      0.378 0.886 88    0.427 0.9981
##   PDD_Therapy t0 - PDD_Therapy t1  4.644 0.837 88    5.548 <.0001
##   PDD_Therapy t0 - TAU t1      3.578 0.958 88    3.734 0.0044
##   PDD_Therapy t0 - PDD_Therapy t2 15.933 0.994 88   16.027 <.0001
##   PDD_Therapy t0 - TAU t2      8.911 0.961 88    9.270 <.0001
##   TAU t0 - PDD_Therapy t1      4.267 0.958 88    4.453 0.0003
##   TAU t0 - TAU t1              3.200 0.837 88    3.822 0.0033
##   TAU t0 - PDD_Therapy t2     15.556 0.961 88   16.182 <.0001
##   TAU t0 - TAU t2              8.533 0.994 88    8.583 <.0001
##   PDD_Therapy t1 - TAU t1     -1.067 1.030 88   -1.040 0.9030
##   PDD_Therapy t1 - PDD_Therapy t2 11.289 1.110 88   10.214 <.0001
##   PDD_Therapy t1 - TAU t2      4.267 1.030 88    4.149 0.0011
##   TAU t1 - PDD_Therapy t2     12.356 1.030 88   12.014 <.0001
##   TAU t1 - TAU t2              5.333 1.110 88    4.826 0.0001
##   PDD_Therapy t2 - TAU t2     -7.022 1.030 88   -6.808 <.0001
##
## P value adjustment: tukey method for comparing a family of 6 estimates
```

## Ergebnisse des Post-hoc Tests

Ein **Post-hoc-Test mit Tukey-Korrektur** zeigte, dass in beiden Therapiegruppen eine signifikante Reduktion der Depressivität zwischen t0 und t2 auftrat. In der TAU-Gruppe betrug der mittlere Rückgang 8.53 Punkte,  $t(88) = 8.53, p < .001$ . In der PDD\_Therapie-Gruppe fiel der Rückgang mit 15.93 Punkten deutlich stärker aus,  $t(88) = 16.03, p < .001$ .

Zwischen den Gruppen zeigten sich zu t0 und t1 keine signifikanten Unterschiede ( $ps > .05$ ). Zum Zeitpunkt t2 unterschieden sich die Gruppen jedoch signifikant,  $t(88) = -7.02, p < .001$ , zugunsten der **PDD\_Therapie-Gruppe**, was auf deren höhere Wirksamkeit im Vergleich zur etablierten KVT hinweist.



- **Mixed Designs** erlauben Prüfung von Hypothesen die Gruppenunterschiede und Messwiederholungen kombinieren.
- Von besonderem Interesse ist i.d.R. der Interaktionseffekt **Zeit x Gruppe**.
- Interaktionseffekt sagt aus, ob es **Gruppenunterschiede hinsichtlich der Veränderung** gab (Unterschied in Unterschieden).
- Es werden Variabilität Zwischen-Personen (**between-subject**) und Innerhalb-Personen (**within-subject**) unterscheiden
- Zum Prüfen der Effekte müssen separate **Prüfvarianzen** berechnet werden.
- Als Voraussetzung wird die **Spärizität** geprüft und ggf. korrigiert.
- Wenn Spärizität erfüllt ist, können Post-Hoc Vergleiche mittels **Tukey-Test** geprüft werden.