



# Evaluation und Metaanalyse

MSc Klinische Psychologie und Psychotherapie

SoSe 2024

Prof. Dr. Dirk Ostwald

## Überblick

Datum	Einheit	Thema
12.04.2024	Einführung	(1) Evaluationsforschung
19.04.2024	Metanalyse	(2) Metaanalysen
26.04.2024	Metanalyse	(3) Standardisierte Effektgrößenschätzung
03.05.2024	Metaanalyse	(4) Fixed- und Random-Effects-Modelle
10.05.2024	Linear Mixed Models	(5) Linear Mixed Models
17.05.2024	Linear Mixed Models	(5) Linear Mixed Models
24.05.2024	Präsentationen	
03.06.2024	Präsentationen	
07.06.2024	Linear Mixed Models	(5) Linear Mixed Models
17.06.2024	Linear Mixed Models	(6) Mehrebenenanalyse
24.06.2024	Präsentationen	
28.06.2024	Linear Mixed Models	(7) Longitudinalanalyse
08.07.2024	Linear Mixed Models	(7) Longitudinalanalyse
12.07.2024	Q&A	
19.07.2024	Klausur	
Februar 2025	Klausurwiederholungstermin	

## (8) Longitudinalanalyse

---

Motivation

Zweistichproben-T-Test bei abhängigen Stichproben

Einfaktorielle Varianzanalyse mit Messwiederholung

Trendanalyse mit Linear Mixed Models

Selbstkontrollfragen

---

## Motivation

Zweistichproben-T-Test bei abhängigen Stichproben

Einfaktorielle Varianzanalyse mit Messwiederholung

Trendanalyse mit Linear Mixed Models

Selbstkontrollfragen

# Motivation

---

## Longitudinalanalyse

Mehrere Messungen ("Repeated Measures") an einer experimentellen Einheit

Annahme der Zufallsfehlerkorrelation innerhalb einer experimentellen Einheit

Typische Beispiele

- Messungen aller Studienbedingungen an einer Proband:in ("Within-Designs")
- Mehrere Messungen an einer Patient:in über den Verlauf einer Psychotherapie

Die Standardtheorie des Allgemeinen Linearen Modells deckt diesen Fall nicht ab

- Annahme der Abwesenheit von Zufallsfehlerkorrelationen zwischen Messpunkten
- Keine mehrfachen Messungen an einer experimentellen Einheit ("Between-Designs")

Aktives Methodenforschungsfeld spätestens seit den 1970er (vgl. Huynh and Feldt (1970))

In der Anwendungen scheinen LMMs momentan Mittel der Wahl (vgl. Detry and Ma (2016))

## Longitudinalanalyse

### Zweistichproben-T-Test bei abhängigen Stichproben

- Rückführung eines Within-Designs auf eine Between-Design-Analyse

### Einfaktorielle Varianzanalyse mit Messwiederholung

- Rückführung eines Within-Designs auf eine (korrigierte) Between-Design-Analyse

### Trendanalyse mit Linear Mixed Models

- Basale Anwendung von LMMs zur Analyse von Zeitseriendaten

---

Motivation

## **Zweistichproben-T-Test bei abhängigen Stichproben**

Einfaktorielle Varianzanalyse mit Messwiederholung

Trendanalyse mit Linear Mixed Models

Selbstkontrollfragen



### Anwendungsszenario

- Zwei Messungen an einer Gruppe experimenteller Einheiten
- Annahme “abhängiger Normalverteilungen” der Gruppen
- Absicht der messungsbezogenen Erwartungswertparameterinferenz

### Anwendungsbeispiel

- Pre- und Posttherapie BDI-II Daten einer Gruppe von  $n = 5$  Patient:innen
- Patient:innen indiziert mit  $i = 1, \dots, n$
- Messungen indiziert mit  $j = 1, 2$  für Pre-Therapie bzw. Post-Therapie

# Zweistichproben-T-Test bei abhängigen Stichproben

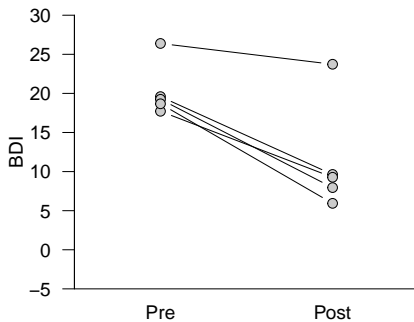
## Beispieldatensatz im Wide Format

```
D = read.csv("../7_Daten/ztt_as_w.csv") # Dateneinlesen
print(D, digits = 1)                  # Print
```

		X	Pre	Post
1	Patient.in 1	1	20	10
2	Patient.in 2	2	18	9
3	Patient.in 3	3	26	24
4	Patient.in 4	4	19	8
5	Patient.in 5	5	19	6

## Zweistichproben-T-Test bei abhängigen Stichproben

### Anwendungsbeispiel



# Zweistichproben-T-Test bei abhängigen Stichproben

## Beispieldatensatz im Long Format

```
D = read.csv("../7_Daten/ztt_as_1.csv") # Dateneinlesen
print(D, digits = 1) # Print
```

	Patient.in	Messung	BDI
1	1	Pre	20
2	1	Post	10
3	2	Pre	18
4	2	Post	9
5	3	Pre	26
6	3	Post	24
7	4	Pre	19
8	4	Post	8
9	5	Pre	19
10	5	Post	6

### Definition (ZTT-AS-Modell in struktureller Form)

$y_{ij}$  mit  $i = 1, \dots, n$  und  $j = 1, 2$  sei die Zufallsvariable, die den Datenpunkt der  $i$ ten experimentellen Einheit in der  $j$  Messung eines Zweistichproben-T-Test bei abhängigen Stichproben Szenarios modelliert. Dann hat das Modell des Zweistichproben-T-Test bei abhängigen Stichproben die strukturelle Form

$$y_{ij} = \mu_j + \pi_i + \varepsilon_{ij} \text{ mit } \pi_i \sim N(0, \sigma_\pi^2) \text{ und } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \text{ u.i.v. f\"ur } i = 1, \dots, n \text{ und } j = 1, 2, \quad (1)$$

wobei

- $\mu_j$  den festen Erwartungswertparameter für die  $j$ te Messung,
- $\pi_i \sim N(0, \sigma_\pi^2)$  den zufälligen Effekt der  $i$ ten experimentellen Einheit mit Varianzparameter  $\sigma_\pi^2 > 0$ ,
- $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$  den Zufallsfehler mit Varianzparameter  $\sigma_\varepsilon^2 > 0$

bezeichnen.

Bemerkungen

- Eine alternative Parameterisierung mit Referenzbedingung und Erwartungswertdifferenzeffekt ist möglich.

## Theorem (ZTT-AS-Modell in Linear Mixed Model-Form)

Gegeben sei die strukturelle Form des ZTT-AS-Modells. Dann hat das ZTT-AS-Modell die Linear-Mixed-Model-Form

$$y = X_f \beta_f + X_r \beta_r + \varepsilon \Leftrightarrow \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ \vdots \\ y_{n1} \\ y_{n2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} 1 & \dots & 0 \\ 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \\ 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_n \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \vdots \\ \varepsilon_{n1} \\ \varepsilon_{n2} \end{pmatrix}, \quad (2)$$

wobei

- $y$  und  $\varepsilon$  jeweils  $2n$ -dimensionale Zufallsvektoren sind,
- $X_f = \mathbf{1}_n \otimes I_2 \in \mathbb{R}^{2n \times 2}$  und  $X_r = I_n \otimes \mathbf{1}_2 \in \mathbb{R}^{2n \times 2}$  sind,
- $\beta_f \in \mathbb{R}^2$  mit Einträgen  $\mu_1$  und  $\mu_2$  ist und  $\beta_r$  ein  $n$ -dimensionaler Zufallsvektor mit Einträgen  $\pi_1, \dots, \pi_n$  ist,
- $\beta_r \sim N(0_n, \sigma_\pi^2 I_n)$  und  $\varepsilon \sim N(0_{2n}, \sigma_\varepsilon^2 I_{2n})$  sind.

## Zweistichproben-T-Test bei abhängigen Stichproben

### Beispiel

Es seien  $n = 5$ . Dann ist das ZTT-AS-Modell in LMM-Form gegeben durch

$$y = X_f \beta_f + X_r \beta_r + \varepsilon \Leftrightarrow \begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \\ y_{41} \\ y_{42} \\ y_{51} \\ y_{52} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{51} \\ \varepsilon_{52} \end{pmatrix} \quad (3)$$

# Zweistichproben-T-Test bei abhängigen Stichproben

## Datengeneration

```
# Modellformulierung
set.seed(0)
library(MASS)
n          = 5
m          = 2
j_n       = matrix(rep(1,n),ncol = 1)
j_m       = matrix(rep(1,m),ncol = 1)
I_n       = diag(n)
I_m       = diag(m)
I_nm      = diag(n*m)
X_f       = kronecker(j_n,I_m)
X_r       = kronecker(I_n,j_m)
s_eps     = 20
s_beta_r  = 5
beta_f    = matrix(c(20,10), nrow = 2)

# Reproduzierbarkeit
# Multivariate Normalverteilung
# Anzahl Patient:innen
# Anzahl Messungen
# 1_n Vektor
# 1_m Vektor
# n x n Identitätsmatrix
# m x m Identitätsmatrix
# nm x nm Identitätsmatrix
# Fixed-Effects-Designmatrix
# Random-Effects-Designmatrix
# Fehlervarianzparameter
# Random-Effects-Varianzparameter
# Fixed-Effects-Parameter

# Datengeneration
beta_r    = mvnrm(1, rep(0,n), s_beta_r*I_n)
eps       = mvnrm(1, rep(0,m*n), s_eps*I_nm)
y         = X_f %*% beta_f + X_r %*% beta_r + eps

# Random-Effects-Parameter
# Zufallsfehler
# Datenvektor
```



# Zweistichproben-T-Test bei abhängigen Stichproben

## Datenrepräsentation im Wide Format

```
# Datenspeicherung im Wide Format
DW = data.frame(Pre      = y[seq.int(1,2*n-1,2)],      # Pre-Therapie-Datenwerte
               Post     = y[seq.int(2,2*n  ,2)],      # Post-Therapie-Datenwerte
               row.names = paste("Patient.in", 1:n))  # Zeilennamen
write.csv(DW, "../7_Daten/ztt_as_w.csv")
```

## Datenrepräsentation im Long Format

```
# Datenspeicherung im Long Format
DL = data.frame("Patient:in" = factor(rep(1:n, each = m)),      # Patient:innenfaktor
               "Messung"    = factor(rep(c("Pre", "Post"), times = n)), # Messungsfaktor
               "BDI"       = y)                                # Datenvektor
write.csv(DL, "../7_Daten/ztt_as_l.csv", row.names = FALSE)
```

## Theorem (Marginale Datenverteilung des ZTT-AS-Modells)

Gegeben sei das ZTT-AS-Modell in LMM-Form. Dann ist die marginale Verteilung des Datenvektors gegeben durch

$$y \sim N(X_f \beta_f, \Sigma) \text{ mit } \Sigma = I_n \otimes \begin{pmatrix} \sigma_\pi^2 + \sigma_\varepsilon^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 + \sigma_\varepsilon^2 \end{pmatrix} \quad (4)$$

### Bemerkungen

- Wir verzichten auf einen allgemeinen Beweis und verdeutlichen das Theorem an einem Beispiel.
- Insbesondere gilt also  $\Sigma \neq \sigma^2 I_n$ .
- Die Inferenztheorie des Allgemeinen Linearen Modells  $y \sim N(X\beta, \sigma^2 I_n)$  ist also *prima facie* nicht anwendbar.
- Die Daten einer experimentellen Einheit  $i = 1, \dots, n$  haben den Kovarianzmatrixparameter

$$\Sigma_i := \begin{pmatrix} \sigma_\pi^2 + \sigma_\varepsilon^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 + \sigma_\varepsilon^2 \end{pmatrix}. \quad (5)$$

- Der Random-Effects-Varianzparameter  $\sigma_\pi^2$  entspricht also der Kovarianz zwischen den Messungen.
- Die Daten einer experimentellen Einheit haben folglich die Korrelationsmatrix

$$P_i := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \text{ mit } \rho = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_\varepsilon^2}. \quad (6)$$

# Zweistichproben-T-Test bei abhängigen Stichproben

## Beispiel

Es sei  $n = 5$ . Dann gilt mit dem Theorem zur Marginalen Datenverteilung des Linear Mixed Models zunächst

$$\Sigma = \sigma_{\pi}^2 X_r I_5 X_r^T + \sigma_{\varepsilon}^2 I_{10} = \sigma_{\pi}^2 X_r X_r^T + \sigma_{\varepsilon}^2 I_{10}. \quad (7)$$

Weiterhin gilt

$$X_r X_r^T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \quad (8)$$
$$= \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

# Zweistichproben-T-Test bei abhängigen Stichproben

## Beispiel

Es ergibt sich also

$$\Sigma = \sigma_{\pi}^2 X_r X_r^T + \sigma_{\varepsilon}^2 I_{10}$$

$$= \begin{pmatrix} \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 \end{pmatrix}$$

$$= I_5 \otimes \begin{pmatrix} \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 \\ \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 \end{pmatrix}$$

## Theorem (Verteilung der Datendifferenzvariablen im ZTT-AS-Modell)

Gegeben sei das ZTT-AS-Modell. Für  $i = 1, \dots, n$  sei

$$y_i := y_{i1} - y_{i2} \quad (9)$$

die Differenzvariable der Zufallsvariablen  $y_{i1}$  und  $y_{i2}$ . Dann gilt

$$y_i \sim N(\mu, \sigma^2) \text{ u.i.v. mit } \mu := \mu_1 - \mu_2 \text{ und } \sigma^2 := 2\sigma_{\varepsilon}^2. \quad (10)$$

### Bemerkungen

- Die Differenzvariablen  $y_i$  sind anhand des Einstichproben-T-Test-Modells verteilt.
- Die Hypothesen des Zweistichproben-T-Test bei abhängigen Stichproben hinsichtlich der Erwartungswertparameter  $\mu_1$  und  $\mu_2$  übertragen dabei sich wie folgt auf die Einstichproben-T-Test Hypothesen in bezug auf den Erwartungswertparameter  $\mu$ :

Zweistichproben-T-Test			Einstichproben-T-Test	
$H_0 : \mu_1 = \mu_2$	$H_1 : \mu_1 \neq \mu_2$	$\Rightarrow$	$H_0 : \mu = 0$	$H_1 : \mu \neq 0$
$H_0 : \mu_1 \leq \mu_2$	$H_1 : \mu_1 > \mu_2$	$\Rightarrow$	$H_0 : \mu \leq 0$	$H_1 : \mu > 0$
$H_0 : \mu_1 \geq \mu_2$	$H_1 : \mu_1 < \mu_2$	$\Rightarrow$	$H_0 : \mu \geq 0$	$H_1 : \mu < 0$

# Zweistichproben-T-Test bei abhängigen Stichproben

## Beweis

Wir halten zunächst fest, dass

$$\begin{aligned}y_i &= y_{i1} - y_{i2} \\ &= \mu_1 + \pi_i + \varepsilon_{i1} - \mu_2 - \pi_i - \varepsilon_{i2} \\ &= \mu_1 - \mu_2 + \varepsilon_{i1} - \varepsilon_{i2} \\ &= \mu + \varepsilon_i,\end{aligned}\tag{11}$$

wobei wir  $\mu := \mu_1 - \mu_2$  und  $\varepsilon_i := \varepsilon_{i1} - \varepsilon_{i2}$  definiert haben.

In Generalisierung des Summentransformationstheorems für normalverteilte Zufallsvariablen halten wir weiterhin fest, dass für unabhängige Zufallsvariablen  $\xi_j \sim N(\mu_j, \sigma_j^2)$ ,  $j = 1, \dots, m$  und Koeffizienten  $a_j$ ,  $j = 1, \dots, m$  gilt, dass

$$\sum_{j=1}^m \xi_j \sim N\left(\sum_{j=1}^m \mu_j, \sum_{j=1}^m a_j^2 \sigma_j^2\right)\tag{12}$$

Mit  $a_1 := 1$  und  $a_2 := -1$  folgt dann aus  $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$  für  $j = 1, 2$  direkt, dass

$$\varepsilon_j \sim N\left(0, 2\sigma_\varepsilon^2\right)\tag{13}$$

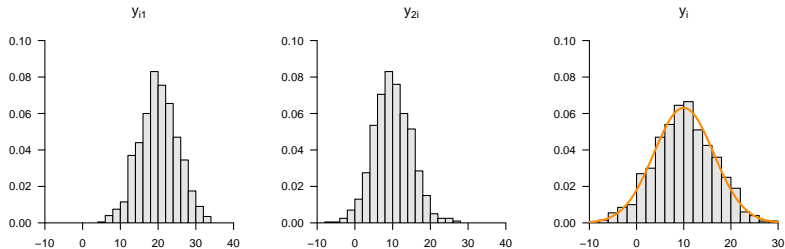
Mit der Definition  $\sigma^2 := 2\sigma_\varepsilon^2$  und dem Transformationstheorem für normalverteilte Zufallsvariablen bei linear-affiner Transformation folgt das Resultat dann direkt.

□

# Zweistichproben-T-Test bei abhängigen Stichproben

## Visualisierung des Theorems zur Verteilung der Datendifferenzvariablen

$$n = 1000, \mu_1 = 20, \mu_2 = 10, \sigma_\beta^2 = 5, \sigma_\varepsilon^2 = 20$$



# Zweistichproben-T-Test bei abhängigen Stichproben

## Praktisches Vorgehen

Äquivalent zu einem Zweistichproben-T-Test mit abhängigen Stichproben führt man einen Einstichproben-T-Test mit unabhängigen Stichproben basierend auf den Datendifferenzvariablen durch.

## Einstichproben-T-Test mit unabhängigen Stichproben basierend auf den Datendifferenzvariablen

```
D = read.csv("../Daten/ztt_as_w.csv") # Dateneinlesen
y = D$Pre - D$Post # Differenzvariablen
n = length(y) # Anzahl Patient:innen
T = sqrt(n)*mean(y)/sd(y) # T-Wert für  $H_0 : \mu = 0$ 
p = 2*(1-pt(abs(T), n-1)) # p-Wert für  $H_0 : \mu = 0$ 
cat("T =", round(T,3), "p =", round(p,3)) # Ausgabe
```

T = 5.18 p = 0.007

## Zweistichproben-T-Test mit abhängigen Stichproben in R

```
R = t.test(x = D$Pre, y = D$Post, paired = T, var.equal = TRUE) # R's T-Testfunktion
cat("T =", round(R$statistic,3), "p =", round(R$p.value,3)) # Ausgabe
```

T = 5.18 p = 0.007



---

Motivation

Zweistichproben-T-Test bei abhängigen Stichproben

**Einfaktorielle Varianzanalyse mit Messwiederholung**

Trendanalyse mit Linear Mixed Models

Selbstkontrollfragen

## Anwendungsszenario

- Messungen von zwei oder mehr Studienbedingungen an einer Gruppe experimenteller Einheiten
- Annahme "abhängiger Normalverteilungen" der Gruppen
- Absicht der bedingungsabhängigen Erwartungswertparameterinferenz
- Balancierte Reihung der Studienbedingungen
- Ausschluss bzw. Vernachlässigung von Reihen-, Lern- oder Zeiteffekten

## Anwendungsbeispiel

- Perzeptuelle Entscheidungsgeschwindigkeit von  $n = 5$  Patient:innen
- Kategorisierung eines visuellen Stimulus einer von  $m = 4$  Objektkategorien
- Patient:innen indiziert mit  $i = 1, \dots, n$ , Studienbedingungen indiziert mit  $j = 1, \dots, m$

## Definition (EVA-MW-Modell in struktureller Form)

$y_{ij}$  mit  $i = 1, \dots, n$  und  $j = 1, \dots, m$  sei die Zufallsvariable, die den Datenpunkt der  $i$ ten experimentellen Einheit in der  $j$ ten Messung im Szenario einer einfaktoriellen Varianzanalyse mit Messwiederholung modelliert. Dann hat das Modell der Einfaktoriellen Varianzanalyse mit Messwiederholung (EVA-MW) die strukturelle Form

$$y_{ij} = \mu_0 + \alpha_j + \pi_i + \varepsilon_{ij} \text{ mit } \pi_i \sim N(0, \sigma_\pi^2) \text{ und } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \text{ u.i.v. f\"ur } i = 1, \dots, n \text{ und } j = 2, \dots, m \quad (14)$$

wobei

- $\mu_0$  den festen Erwartungswertparameter der Referenzbedingung,
- $\alpha_j$  mit  $j = 2, \dots, m$  den festen Effektparameter der  $j$ ten Bedingung,
- $\pi_i \sim N(0, \sigma_\pi^2)$  den zufälligen Effekt der  $i$ ten experimentellen Einheit mit Varianzparameter  $\sigma_\pi^2 > 0$ ,
- $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$  den Zufallsfehler mit Varianzparameter  $\sigma_\varepsilon^2 > 0$ .

bezeichnen.

Bemerkungen

- Wie bei Varianzanalysen üblich enkodieren die  $\alpha_j$  Erwartungswertdifferenzen bezüglich der Referenzbedingung.

## Theorem (EVA-MW-Modell in Linear Mixed Model-Form)

Gegeben sei die strukturelle Form des EVA-MW-Modells. Dann hat dieses Modell die Linear-Mixed-Modell-Form

$$y = X_f \beta_f + X_r \beta_r + \varepsilon \Leftrightarrow \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1m} \\ \vdots \\ y_{n1} \\ y_{n2} \\ \vdots \\ y_{nm} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} + \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_n \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1m} \\ \vdots \\ \varepsilon_{n1} \\ \varepsilon_{n2} \\ \vdots \\ \varepsilon_{nm} \end{pmatrix}$$

wobei

- $y$  und  $\varepsilon$  jeweils  $nm$ -dimensionale Zufallsvektoren,

- $X_f = 1_n \otimes \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{nm \times m}$  und  $X_r = I_n \otimes 1_m \in \mathbb{R}^{nm \times n}$ ,

- $\beta_f \in \mathbb{R}^m$  ein fester Vektor mit Einträgen  $\mu_0, \alpha_2, \dots, \alpha_m$ ,
- $\beta_r$  ein  $n$ -dimensionaler Zufallsvektor mit Einträgen  $\pi_1, \dots, \pi_n$  und
- $\beta_r \sim N(0_n, \sigma_\pi^2 I_n)$  und  $\varepsilon \sim N(0_{nm}, \sigma_\varepsilon^2 I_{nm})$

sind.

# Einfaktorielle Varianzanalyse mit Messwiederholung

## Beispiel

Es seien  $n = 5$  und  $m = 4$ . Dann ist das EVA-MW-Modell in Linear-Mixed-Model-Form gegeben durch

$$y = X_f \beta_f + X_r \beta_r + \varepsilon \Leftrightarrow \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \\ y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \\ y_{51} \\ y_{52} \\ y_{53} \\ y_{54} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{43} \\ \varepsilon_{44} \\ \varepsilon_{51} \\ \varepsilon_{52} \\ \varepsilon_{53} \\ \varepsilon_{54} \end{pmatrix}$$

# Einfaktorielle Varianzanalyse mit Messwiederholung

```
# Modellformulierung
set.seed(0)
library(MASS)
n          = 5
m          = 4
j_n       = matrix(rep(1,n),ncol = 1)
j_m       = matrix(rep(1,m),ncol = 1)
I_n       = diag(n)
X_f_i     = matrix(c(1,0,0,0,
                    1,1,0,0,
                    1,0,1,0,
                    1,0,0,1),
                  byrow = T, nrow = m)

X_f       = kronecker(j_n,X_f_i)
X_r       = kronecker(I_n,j_m)
s_eps     = 100
s_beta_r  = 200
beta_f    = matrix(c(450,50,-150,0), nrow = m)

# Reproduzierbarkeit
# multivariate Normalverteilung
# Anzahl Patient:innen
# Anzahl Bedingungen
# 1_n Vektor
# 1_m Vektor
# n x n Identitätsmatrix
# proband:innenspezifische FFX Designmatrix

# Fixed-Effects-Designmatrix
# Random-Effects-Designmatrix
# Fehlervarianzparameter
# Random-Effects-Varianzparameter
# Fixed-Effects-Parameter

# Datengeneration
beta_r    = mvrnorm(1, rep(0,n) , s_beta_r*I_n)
eps       = mvrnorm(1, rep(0,m*n), s_eps*diag(m*n))
y         = X_f %*% beta_f + X_r %*% beta_r + eps

# Random-Effects-Parameter
# Zufallsfehler
# Datenvektor
```

# Einfaktorielle Varianzanalyse mit Messwiederholung

## Datenrepräsentation im Wide Format

```
# Datenspeicherung im Wide Format
D = data.frame(Condition.1 = y[seq.int(1,m*n,m)],      # Datenwerte Bedingung 1
               Condition.2 = y[seq.int(2,m*n,m)],      # Datenwerte Bedingung 2
               Condition.3 = y[seq.int(3,m*n,m)],      # Datenwerte Bedingung 3
               Condition.4 = y[seq.int(4,m*n,m)],      # Datenwerte Bedingung 4
               row.names = paste("Patient", 1:n))      # Zeilennamen
write.csv(D, "../Daten/eva_mw_w.csv")                 # Speichern
```

## Datenrepräsentation im Long Format

```
# Datenspeicherung im Long Format
D = data.frame("Patient" = factor(rep(1:n, each = m)), # Patient:innenfaktor
               "Condition" = factor(rep(1:m, times = n)), # Messungsfaktor
               "ReactionTime" = y)                    # Datenvektor
write.csv(D, "../Daten/eva_mw_l.csv", row.names = FALSE) # Speichern
```

# Einfaktorielle Varianzanalyse mit Messwiederholung

## Beispieldatensatz im Wide Format

	X	Condition.1	Condition.2	Condition.3	Condition.4
Patient 1		455	514	307	460
Patient 2		466	506	322	459
Patient 3		471	515	316	466
Patient 4		434	487	303	469
Patient 5		468	515	309	452

## Beispieldatensatz im Long Format

Patient	Condition	ReactionTime
1	1	455
1	2	514
1	3	307
1	4	460
2	1	466
2	2	506
2	3	322
2	4	459
3	1	471
3	2	515
3	3	316
3	4	466
4	1	434
4	2	487
4	3	303
4	4	469
5	1	468
5	2	515
5	3	309
5	4	452



## Theorem (Marginale Datenverteilung des EVA-MW-Modells)

Gegeben sei das EVA-MW-Modell in LMM-Form. Dann ist die marginale Verteilung des Datenvektors gegeben durch

$$y \sim N(X_f \beta_f, \Sigma) \quad (15)$$

mit

$$\Sigma = I_n \otimes \Sigma_i \text{ und } \Sigma_i := \begin{pmatrix} \sigma_\pi^2 + \sigma_\varepsilon^2 & \sigma_\pi^2 & \dots & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 + \sigma_\varepsilon^2 & \dots & \sigma_\pi^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\pi^2 & \dots & \dots & \sigma_\pi^2 + \sigma_\varepsilon^2 \end{pmatrix} \in \mathbb{R}^{m \times m} \text{ f\"ur } i = 1, \dots, n. \quad (16)$$

Bemerkungen

- Wir verzichten auf einen allgemeinen Beweis und verdeutlichen das Theorem am Beispiel  $n = 2$  und  $m = 4$ .
- $\Sigma_i \in \mathbb{R}^{m \times m}$  enkodiert die Kovarianzen der Bedingungen  $j = 1, \dots, m$  innerhalb einer experimentellen Einheit
- Für die Korrelationsmatrix der Bedingungen innerhalb einer experimentellen Einheit gilt folglich

$$P_i = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \vdots & 1 \end{pmatrix} \text{ mit } \rho := \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_\varepsilon^2} \quad (17)$$

- Im Kontext der EVA-MW werden Matrizen der Form von  $\Sigma_i$  und  $\Pi_i$  als *sphärisch* bezeichnet.
- Man spricht in diesem Zusammenhang auch von der *Compound Symmetry* von  $\Sigma_i$  und  $P_i$ .

# Einfaktorielle Varianzanalyse mit Messwiederholung

## Beispiel

Es seien  $n = 2$  und  $m = 4$ . Dann gilt mit dem Theorem zur Marginalen Datenverteilung des Linear Mixed Models zunächst

$$\Sigma = \sigma_{\pi}^2 X_r I_2 X_r^T + \sigma_{\varepsilon}^2 I_8 = \sigma_{\pi}^2 X_r X_r^T + \sigma_{\varepsilon}^2 I_8. \quad (18)$$

Weiterhin gilt

$$X_r X_r^T = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \quad (19)$$
$$= \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

# Einfaktorielle Varianzanalyse mit Messwiederholung

## Beispiel

Es ergibt sich also

$$\Sigma = \sigma_{\pi}^2 X_r X_r^T + \sigma_{\varepsilon}^2 I_8$$

$$= \begin{pmatrix} \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 & 0 & 0 & 0 & 0 \\ \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 & 0 & 0 & 0 & 0 \\ \sigma_{\pi}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 & 0 & 0 & 0 & 0 \\ \sigma_{\pi}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 \\ 0 & 0 & 0 & 0 & \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 \\ 0 & 0 & 0 & 0 & \sigma_{\pi}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 \\ 0 & 0 & 0 & 0 & \sigma_{\pi}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 \end{pmatrix}$$
$$= I_2 \otimes \begin{pmatrix} \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 \\ \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 \\ \sigma_{\pi}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 & \sigma_{\pi}^2 \\ \sigma_{\pi}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 & \sigma_{\pi}^2 + \sigma_{\varepsilon}^2 \end{pmatrix}.$$

## Definition (F-Statistik der EVA-MW)

Gegeben sei das Modell der Einfaktoriellen Varianzanalyse mit Messwiederholung

$$y_{ij} = \mu_0 + \alpha_j + \pi_i + \varepsilon_{ij} \text{ mit } \pi_i \sim N(0, \sigma_\pi^2) \text{ und } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \text{ u.i.v. f\"ur } i = 1, \dots, n \text{ und } j = 2, \dots, m. \quad (20)$$

Weiterhin seien

$$\bar{y}_i := \frac{1}{m} \sum_{j=1}^m y_{ij} \quad \text{das Stichprobenmittel der } i\text{ten experimentellen Einheit,}$$

$$\bar{y}_j := \frac{1}{n} \sum_{i=1}^n y_{ij} \quad \text{das Stichprobenmittel der } j\text{ten Bedingung,}$$

$$\bar{y} := \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m y_{ij} \quad \text{das Gesamtstichprobenmittel,}$$

$$\text{SQB} := n \sum_{j=1}^m (\bar{y}_j - \bar{y})^2 \quad \text{die Between-Sum-of-Squares,}$$

$$\text{SQE} := \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y})^2 \quad \text{die Error-Sum-of-Squares.}$$

Dann ist die F-Statistik der EVA mit Messwiederholung definiert als

$$F := \frac{\text{MSB}}{\text{MSE}} := \frac{\text{SQB}/(m-1)}{\text{SQE}/(m-1)(n-1)}, \quad (21)$$

wobei MSB und MSE als Mean Between Squares und Mean Error Squares bezeichnet werden.

# Einfaktorielle Varianzanalyse mit Messwiederholung

## Bemerkungen

- Das Szenario entspricht einer additiven zweifaktoriellen Varianzanalyse mit Einzelzellbeobachtung.
- SQB quantifiziert die absoluten Abweichungen der Bedingungsstichprobenmittel vom Gesamtstichprobenmittel.
- SQE quantifiziert die residuelle Variabilität nach Anpassung des Gesamtstichprobenmittels für die  $j$ ten Bedingung und  $i$ te experimentelle Einheit.
- Man kann weiterhin eine Total Sum-of-Squares SQT und eine Within Sum-of-Squares SQW definieren.
- Es gilt dann die Varianzzerlegung

$$SQT = SSB + SSW + SSE. \quad (22)$$

- DeGroot and Schervish (2012) Kapitel 11.7. geben einen Überblick und Beweise.
- Man kann also äquivalent eine F-Statistik für den Effekt der experimentellen Einheiten definieren.
- In der Anwendung ist man jedoch primär am Effekt der Studienbedingungen interessiert.
- MSE wird manchmal auch als *mean square of interaction* bezeichnet (vgl. Huynh and Feldt (1970)).

## Theorem (F-Statistik der EVA mit Messwiederholung)

Gegeben sei das Modell der Einfaktoriellen Varianzanalyse mit Messwiederholung

$$y_{ij} = \mu_0 + \alpha_j + \pi_i + \varepsilon_{ij} \text{ mit } \pi_i \sim N(0, \sigma_\pi^2) \text{ und } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \text{ u.i.v. f\"ur } i = 1, \dots, n \text{ und } j = 2, \dots, m. \quad (23)$$

und es sei die F-Statistik der EVA mit Messwiederholung wie oben definiert. Dann gilt unter der Annahme der Nullhypothese

$$\alpha_2 = \dots = \alpha_m = 0, \quad (24)$$

dass

$$F \sim f(m-1, (m-1)(n-1)). \quad (25)$$

### Bemerkungen

- Für einen Beweis verweisen wir auf Huynh and Feldt (1970).
- Das Theorem ermöglicht die Angabe eines p-Wertes  $\mathbb{P}(F \geq \tilde{F})$  für einen Wert  $\tilde{F}$  der F-Statistik.

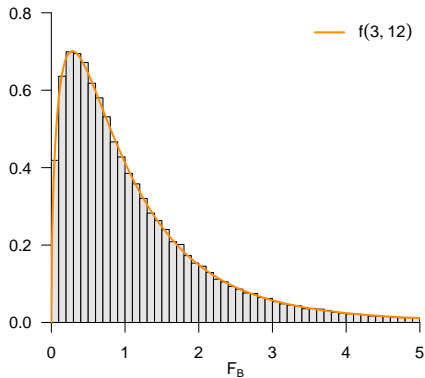
# Einfaktorielle Varianzanalyse mit Messwiederholung

## Verteilung der F-Statistik der EVA mit Messwiederholung bei Zutreffen der Nullhypothese

```
set.seed(0)
library(MASS)
nsim      = 1e5
n         = 5
m         = 4
j_n       = matrix(rep(1,n),ncol = 1)
j_m       = matrix(rep(1,m),ncol = 1)
I_n       = diag(n)
I_m       = diag(m)
I_nm      = diag(n*m)
X_f_i     = matrix(c(1,0,0,0,1,1,0,0,1,0,1,0, 1,0,0,1),
                  byrow = TRUE, nrow = m)
X_f       = kronecker(j_n,X_f_i)
X_r       = kronecker(I_n,j_m)
s_eps     = 20
s_beta_r  = 20
beta_f    = matrix(c(0,0,0,0), nrow = m)
F         = rep(NA,n)
for(s in 1:nsim){
  beta_r   = mvrnorm(1, rep(0,n) , s_beta_r*I_n)
  eps      = mvrnorm(1, rep(0,m*n), s_eps*I_nm)
  y        = X_f %*% beta_f + X_r %*% beta_r + eps
  Y        = t(matrix(y, nrow = m))
  Y_bar    = mean(Y)
  Y_bar_i  = rowMeans(Y)
  Y_bar_j  = colMeans(Y)
  SQB      = n*sum((Y_bar_j - Y_bar)^2)
  MSB      = SQB/(m-1)
  SSE      = 0
  for(i in 1:n){
    for(j in 1:m){
      SQE = SSE + (Y[i,j]-Y_bar_i[i]-Y_bar_j[j]+Y_bar)^2}
    MSE  = SSE/((m-1)*(n-1))
    F[s] = MSB/MSE}
  # Reproduzierbarkeit
  # multivariate Normalverteilung
  # Anzahl Simulatione
  # Anzahl Patient:innen
  # Anzahl Bedingungen
  # 1_n Vektor
  # 1_m Vektor
  # n x n Identitätsmatrix
  # m x m Identitätsmatrix
  # n x m Identitätsmatrix
  # proband:innenspezifische
  # FFX Designmatrix
  # Fixed-Effects-Designmatrix
  # Random-Effects-Designmatrix
  # Fehlervarianzparameter
  # Random-Effects-Varianzparameter
  # Fixed-Effects-Parameter der Nullhypothese
  # F-Statistik Array
  # Simulationsiterationen
  # Random-Effects-Parameter
  # Zufallsfehler
  # Datenvektor
  # Datenmatrix im Wide Format
  # Gesamtstichprobenmittel
  # Zeilenmittelwerte
  # Spaltenmittelwerte
  # Sum-of-Squares Between
  # Mean-Sum-of-Squares Between
  # Sum-of-Error-Squares
  # Zeileniterationen
  # Spalteniterationen
  # Sum-of-Error-Squares
  # Mean Sum-of-Error-Squares
  # F-Statistik
```

# Einfaktorielle Varianzanalyse mit Messwiederholung

Verteilung der F-Statistik der EVA mit Messwiederholung bei Zutreffen der Nullhypothese





# Einfaktorielle Varianzanalyse mit Messwiederholung

## Anwendung

```
D = read.csv("../Daten/eva_mw_w.csv") # Datensatz im Wide Format
y = as.matrix(D[,2:5]) # numerische Datentabelle
n = nrow(y) # Anzahl Patient:innen
m = ncol(y) # Anzahl Bedingungen
y_bar = mean(y) # Gesamtstichprobenmittel
y_bar_i = rowMeans(y) # Patient:innemittelwerte
y_bar_j = colMeans(y) # Bedingungsittelwerte
SQB = n*sum((y_bar_j - y_bar)^2) # Sum-of-Squares Between
SQW = m*sum((y_bar_i - y_bar)^2) # Sum-of-Squares Between
SQE = 0 # Sum-of-Error-Squares
for(i in 1:n){ # Zeileniterationen
  for(j in 1:m){ # Spalteniterationen
    SQE = SQE + (y[i,j]-y_bar_i[i]-y_bar_j[j]+y_bar)^2} # Sum-of-Error-Squares
MSB = SQB/(m-1) # Mean-Sum-of-Squares Between
MSW = SQW/(n-1) # Mean-Sum-of-Squares Between
MSE = SQE/((m-1)*(n-1)) # Mean Sum-of-Error-Squares
FB = MSB/MSE # F-Statistik Between
FW = MSW/MSE # F-Statistik Within
pvals = c(1-pf(FB,m-1,(m-1)*(n-1)),1-pf(FW,n-1,(m-1)*(n-1)), NA) # p-Werte
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Condition	3	108829	36276.3	392.79	3.06e-12
Patient	4	774	193.4	2.09	1.45e-01
Residuals	12	1108	92.4	NA	NA

## Anwendung mit R's aov() Funktion

```
D           = read.csv("../Daten/eva_mw_1.csv")           # Datensatz im Long Format
D$Condition = as.factor(D$Condition)                   # explizite Faktorenkodierung
D$Patient   = as.factor(D$Patient)                     # explizite Faktorenkodierung
S           = summary(aov(ReactionTime ~ Condition + Patient, data = D)) # additive ZVA mit Einzelzellbeobachtung
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Condition	3	108829	36276	392.790	3.06e-12 ***
Patient	4	774	193	2.094	0.145
Residuals	12	1108	92		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

---

Motivation

Zweistichproben-T-Test bei abhängigen Stichproben

Einfaktorielle Varianzanalyse mit Messwiederholung

**Trendanalyse mit Linear Mixed Models**

Selbstkontrollfragen

## Randomisierte Klinische Studie in Anlehnung an Rief et al. (2018)

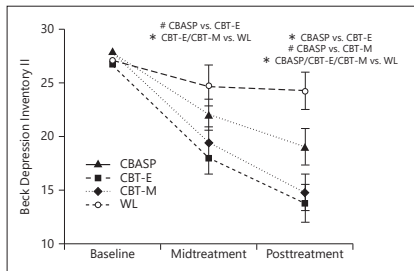
- Effekt von Therapie auf Depressions symptomatik über Therapieverlauf
- BDI-II als primäres Ergebnismaß
- 4 Therapiegruppen
  - Cognitive-Behavioral Analysis System of Psychotherapy (CBASP)
  - Cognitive-Behavioural Therapy with focus on Physical Exercise (CBT-E)
  - Cognitive-Behavioural Therapy with focus on Mindful Activities (CBT-M)
  - Waitlist control group (WL)
- 40 Patient:innen pro Therapiegruppe
- 3 Messzeitpunkte
  - Pretreatment (T = 0)
  - Midtreatment (T = 1)
  - Posttreatment (T = 2)
- Linear-Mixed-Model Analyse "according to current recommendations" (Detry and Ma, 2016)

## Randomisierte Klinische Studie in Anlehnung an Rief et al. (2018)

Daten der ersten zwei Patient:innen jeder Gruppe

Patient	Group	Time	BDI
1	1 CBASP	0	25
2	1 CBASP	1	20
3	1 CBASP	2	17
4	2 CBASP	0	29
5	2 CBASP	1	23
6	2 CBASP	2	19
121	41 CBT-E	0	24
122	41 CBT-E	1	20
123	41 CBT-E	2	12
124	42 CBT-E	0	25
125	42 CBT-E	1	21
126	42 CBT-E	2	11
241	81 CBT-M	0	28
242	81 CBT-M	1	21
243	81 CBT-M	2	16
244	82 CBT-M	0	28
245	82 CBT-M	1	20
246	82 CBT-M	2	15
361	121 WL	0	27
362	121 WL	1	21
363	121 WL	2	23
364	122 WL	0	31
365	122 WL	1	23
366	122 WL	2	26

## Randomisierte Klinische Studie in Anlehnung an Rief et al. (2018)



**Fig. 2.** Depressive symptom severity (adjusted for screening values at study entry) from baseline to week 8 (midtreatment) and to week 16 (posttreatment) by treatment group. Values are estimated marginal means (with SEM) from intention-to-treat analysis using multilevel modeling. Post hoc contrasts for cognitive behavioral analysis system (CBASP), cognitive behavioral therapy emphasizing physical exercises during behavioral activation (CBT-E), and cognitive behavioral therapy emphasizing mindfulness exercises during behavioral activation (CBT-M) versus waitlist (WL). \*  $p < 0.05$ ; #  $p < 0.1$ .

# Trendanalyse mit Linear Mixed Models

Modell in struktureller Form

Es seien

- $g = 4$  die Anzahl an Studiengruppen,
- $k = 40$  die Anzahl an Patient:innen pro Studiengruppe,
- $m = 3$  die Anzahl an Messpunkten pro Patient:in und somit
- $n = gkm = 4 \cdot 40 \cdot 3 = 480$  die Gesamtanzahl an Datenpunkten.

Es sei

$$y_{ijt} \text{ mit } i = 1, \dots, k, j = 1, \dots, g \text{ und } t = 1, \dots, m \quad (26)$$

die Zufallsvariable, die den  $t$ ten Datenpunkt der  $i$ ten experimentellen Einheit in der  $j$ ten Studiengruppe modelliert. Dann hat das hier betrachtete Modell die strukturelle Form

$$y_{ijt} = \mu_0 + \alpha_j + \beta_1 x_t + \gamma_j x_t + \beta_{0i} + \beta_{1i} x_t + \varepsilon_{ijt}, \quad (27)$$

wobei

- $\mu_0$  den Interzeptparameter der Referenzgruppe
- $\alpha_j$  für  $j = 2, \dots, m$  den festen Effektparameter der  $j$ ten Studiengruppe,
- $\beta_1$  den festen Steigungsparameter der Referenzstudiengruppe
- $x_t$  mit  $x_1 := 0, x_2 := 1, x_3 := 2$  den Messzeitpunktregressor,
- $\gamma_j$  den Studiengruppen-spezifischen Steigungsparameter,
- $\beta_{0i} \sim N(0, \sigma_{\beta_r}^2)$  u.i.v. den zufälligen Interzeptparameter der  $i$ ten Patient:in,
- $\beta_{1i} \sim N(0, \sigma_{\beta_r}^2)$  u.i.v. den zufälligen Steigungsparameter der  $i$ ten Patient:in,
- $\varepsilon_{ijt} \sim N(0, \sigma_{\varepsilon}^2)$  u.i.v. den Zufallsfehler des  $t$ -Messzeitpunktes der  $i$ ten Patient:in der  $j$ ten Gruppe

bezeichnen

## Theoretische Studienszenarien

### Szenario (1) Daten ohne Random-Effects

- Analyse (1.1) Keine Modellierung der Random-Effects
- Analyse (1.2) Modellierung der Random-Effects

### Szenario (2) Daten mit Random-Effects

- Analyse (2.1) Keine Modellierung der Random-Effects
- Analyse (2.2) Modellierung der Random-Effects



# Trendanalyse mit Linear Mixed Models

## Datengeneration

```
library(MASS)
set.seed(0)
g      = 4
k      = 40
m      = 3
p      = g*k
n      = g*k*m
T      = matrix(0:2, nrow = 3)
G      = matrix(c(1,0,0,0,1,1,0,0,1,0,1,0,1,0,1,0,0,1),
               byrow = TRUE, nrow = 4)
R      = cbind(rep(1,m), T)
j_km   = matrix(rep(1,k*m), ncol = 1)
j_kg   = matrix(rep(1,p), ncol = 1)
I_kg   = diag(p)
X_g    = kronecker(G, j_km)
X_t    = kronecker(j_kg, T)
X_i    = cbind(X_g[,2]*X_t, X_g[,3]*X_t, X_g[,4]*X_t)
X_f    = cbind(X_g,X_t,X_i)
X_r    = kronecker(I_kg, R)
beta_f = matrix(c(27,0,0,0,-4,-3,-2,3), nrow = ncol(X_f))
s_eps  = 5
s_beta_r = 10
beta_r = mvrnorm(1,rep(0,2*p), s_beta_r*diag(2*p))
eps    = mvrnorm(1,rep(0,n), s_eps*diag(n))
yf     = X_f %>% beta_f + eps
yr     = X_f %>% beta_f + X_r %>% beta_r + eps
```

# multivariate Normalverteilung  
# Zufallszahlengenerator  
# Anzahl Treatmentgruppen  
# Anzahl Patient:innen pro Treatmentgruppe  
# Anzahl Messpunkte pro Patient:in  
# Anzahl Patient:innen  
# Anzahl Datenpunkte  
# Zeitfaktorprädiktor  
# Gruppenfaktordesign  
# Einfache-Lineare-Regressions-Designmatrix  
# 1\_{k\*m} (Datenpunkte pro Gruppe)  
# 1\_{p} (Gesamtzahl Patient:innen)  
# I\_{p} (Gesamtzahl Patient:innen)  
# Gruppenfaktordesignmatrix  
# Zeitfaktordesignmatrix  
# Interaktionsdesignmatrix  
# Fixed-Effects-Designmatrix  
# Random-Effects-Designmatrix  
# Fixed-Effects-Parameter  
# Varianzkomponente  
# Varianzkomponente  
# Random-Effects-Parameter  
# Fehlervektor  
# Daten ohne Random-Effects  
# Datensatz mit Random-Effects

## Szenario (1) Daten ohne Random-Effects

```
groups      = c("CBASP", "CBT-E", "CBT-M", "WL")
Df          = data.frame("Patient" = factor(rep(1:(g*k), each = m)), # Patient:innenfaktor
                        "Group"   = factor(rep(groups, each = k*m)), # Gruppenfaktor
                        "Time"    = rep(c(0:2), times = g*k),       # Messungsregressor
                        "BDI"     = yf)                             # Fixed-Effects-Datensatz
write.csv(Df, "../Daten/ta_ffx.csv", row.names = FALSE)           # Datei
```

## Szenario (2) Daten mit Random-Effects

```
groups      = c("CBASP", "CBT-E", "CBT-M", "WL")
Dr          = data.frame("Patient" = factor(rep(1:(g*k), each = m)), # Patient:innenfaktor
                        "Group"   = factor(rep(groups, each = k*m)), # Gruppenfaktor
                        "Time"    = rep(c(0:2), times = g*k),       # Messungsregressor
                        "BDI"     = yr)                             # Random-Effects-Datensatz
write.csv(Dr, "../Daten/ta_rfx.csv", row.names = FALSE)           # Datei
```

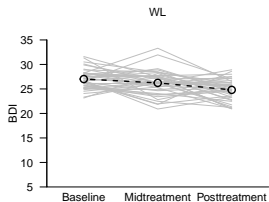
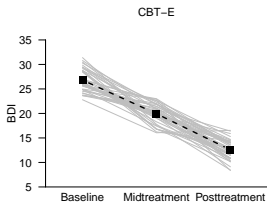
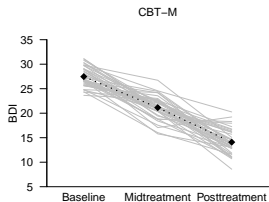
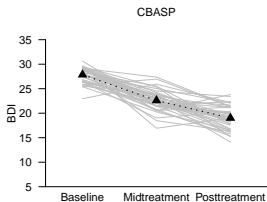
## Szenario (1) Daten ohne Random-Effects

Daten der ersten zwei Patient:innen jeder Gruppe

Patient	Group	Time	BDI
1	1 CBASP	0	25
2	1 CBASP	1	20
3	1 CBASP	2	17
4	2 CBASP	0	29
5	2 CBASP	1	23
6	2 CBASP	2	19
121	41 CBT-E	0	24
122	41 CBT-E	1	20
123	41 CBT-E	2	12
124	42 CBT-E	0	25
125	42 CBT-E	1	21
126	42 CBT-E	2	11
241	81 CBT-M	0	28
242	81 CBT-M	1	21
243	81 CBT-M	2	16
244	82 CBT-M	0	28
245	82 CBT-M	1	20
246	82 CBT-M	2	15
361	121 WL	0	27
362	121 WL	1	21
363	121 WL	2	23
364	122 WL	0	31
365	122 WL	1	23
366	122 WL	2	26

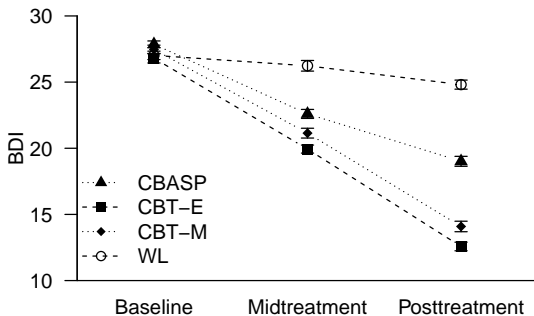
# Trendanalyse mit Linear Mixed Models

## Szenario (1) Daten ohne Random-Effects



# Trendanalyse mit Linear Mixed Models

Szenario (1) Daten ohne Random-Effects



# Trendanalyse mit Linear Mixed Models

## Szenario (1) Daten ohne Random-Effects | Analyse (1.1) Keine Modellierung der Random-Effects

```
D = read.csv("../7_Daten/ta_ffx.csv") # Dateneinlesen
M = lm(BDI ~ Group + Time + Group:Time, data = D) # = X_fb_f + e
```

### Effektparameter

Parameter	R Bezeichner	Interpretation
$\mu_0 = 27$	(Intercept)	Interzept der CBASP Referenzgruppe
$\alpha_2 = 0$	GroupCBT-E	Haupteffekt CBT-E (Differenz zu CBASP)
$\alpha_3 = 0$	GroupCBT-M	Haupteffekt CBT-M (Differenz zu CBASP)
$\alpha_4 = 0$	GroupWL	Haupteffekt WL (Differenz zu CBASP)
$\beta_1 = -4$	Time	Steigung der CBASP Referenzgruppe
$\gamma_2 = -3$	GroupCBT-E:Time	Interaktion CBT-E x Messung (Steigungsdifferenz zu CBASP)
$\gamma_3 = -2$	GroupCBT-M:Time	Interaktion CBT-M x Messung (Steigungsdifferenz zu CBASP)
$\gamma_4 = 3$	GroupWL:Time	Interaktion WL x Messung (Steigungsdifferenz zu CBASP)

# Trendanalyse mit Linear Mixed Models

## Szenario (1) Daten ohne Random-Effects | Analyse (1.1) Keine Modellierung der Random-Effects

```
summary(M)
```

Call:

```
lm(formula = BDI ~ Group + Time + Group:Time, data = D)
```

Residuals:

```
    Min       1Q   Median       3Q      Max
-6.229 -1.440  0.000   1.539   7.283
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  27.579174   0.315589  87.390 < 2e-16 ***
GroupCBT-E   -0.690213   0.446310  -1.546  0.123
GroupCBT-M    0.008261   0.446310   0.019  0.985
GroupWL      -0.452993   0.446310  -1.015  0.311
Time         -4.423279   0.244454 -18.095 < 2e-16 ***
GroupCBT-E:Time -2.692315   0.345710  -7.788 4.36e-14 ***
GroupCBT-M:Time -2.265134   0.345710  -6.552 1.49e-10 ***
GroupWL:Time   3.317774   0.345710   9.597 < 2e-16 ***
```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 2.186 on 472 degrees of freedom

Multiple R-squared: 0.8421, Adjusted R-squared: 0.8397

F-statistic: 359.5 on 7 and 472 DF, p-value: < 2.2e-16

# Trendanalyse mit Linear Mixed Models

## Szenario (1) Daten ohne Random-Effects | Analyse (1.2) Modellierung der Random-Effects

```
library(nlme)
ctrl = lmeControl(opt = "optim", maxIter = 100) # Settings
D = read.csv("../7_Daten/ta_ffx.csv") # Dateneinlesen
M = lme(BDI ~ Group + Time + Group:Time, D, random = ~ Time | Patient, control = ctrl) # y = X_fb_f + X_rb_r + e
```

### Effektparameter

Parameter	R Bezeichner	Interpretation
$\mu_0 = 27$	(Intercept)	Interzept der CBASP Referenzgruppe
$\alpha_2 = 0$	GroupCBT-E	Haupteffekt CBT-E (Differenz zu CBASP)
$\alpha_3 = 0$	GroupCBT-M	Haupteffekt CBT-M (Differenz zu CBASP)
$\alpha_4 = 0$	GroupWL	Haupteffekt WL (Differenz zu CBASP)
$\beta_1 = -4$	Time	Steigung der CBASP Referenzgruppe
$\gamma_2 = -3$	GroupCBT-E:Time	Interaktion CBT-E x Messung (Steigungsdifferenz zu CBASP)
$\gamma_3 = -2$	GroupCBT-M:Time	Interaktion CBT-M x Messung (Steigungsdifferenz zu CBASP)
$\gamma_4 = 3$	GroupWL:Time	Interaktion WL x Messung (Steigungsdifferenz zu CBASP)
$\beta_{0i} \sim N\left(0, \sigma_{\beta_r}^2\right)$		Patient:innenspezifisches Interzept
$\beta_{1i} \sim N\left(0, \sigma_{\beta_r}^2\right)$		Patient:innenspezifische Steigung



# Trendanalyse mit Linear Mixed Models

## Szenario (1) Daten ohne Random-Effects | Analyse (1.2) Modellierung der Random-Effects

summary(M)

Linear mixed-effects model fit by REML

Data: D  
AIC BIC logLik  
2133.554 2183.438 -1054.777

Random effects:

Formula: ~Time | Patient  
Structure: General positive-definite, Log-Cholesky parametrization  
StdDev Corr  
(Intercept) 0.3739337 (Intr)  
Time 0.3147456 0.907  
Residual 2.0656383

Fixed effects: BDI ~ Group + Time + Group:Time

	Value	Std.Error	DF	t-value	p-value
(Intercept)	27.579174	0.3039549	316	90.73441	0.0000
GroupCBT-E	-0.690213	0.4298572	156	-1.60568	0.1104
GroupCBT-M	0.008261	0.4298572	156	0.01922	0.9847
GroupWL	-0.452993	0.4298572	156	-1.05382	0.2936
Time	-4.423279	0.2362465	316	-18.72315	0.0000
GroupCBT-E:Time	-2.692315	0.3341029	316	-8.05834	0.0000
GroupCBT-M:Time	-2.265134	0.3341029	316	-6.77975	0.0000
GroupWL:Time	3.317774	0.3341029	316	9.93039	0.0000

Correlation:

	(Intr)	GrCBT-E	GrCBT-M	GrpWL	Time	GCBT-E:	GCBT-M:
GroupCBT-E	-0.707						
GroupCBT-M	-0.707	0.500					
GroupWL	-0.707	0.500	0.500				
Time	-0.706	0.499	0.499	0.499			
GroupCBT-E:Time	0.499	-0.706	-0.353	-0.353	-0.707		
GroupCBT-M:Time	0.499	-0.353	-0.706	-0.353	-0.707	0.500	
GroupWL:Time	0.499	-0.353	-0.353	-0.706	-0.707	0.500	0.500

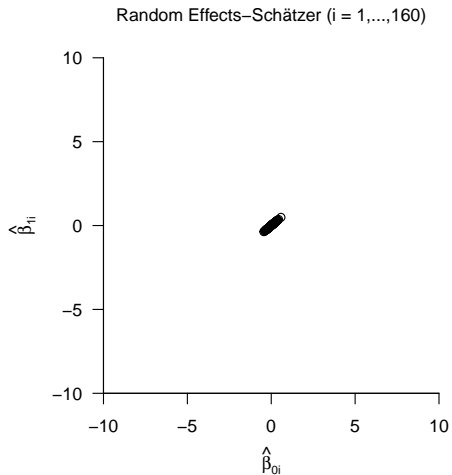
Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-2.84083819	-0.61877795	-0.01508147	0.68239431	3.13956425

Number of Observations: 480  
Number of Groups: 160

# Trendanalyse mit Linear Mixed Models

Szenario (1) Daten ohne Random-Effects | Analyse (1.2) Modellierung der Random-Effects



## Szenario (1) Daten ohne Random-Effects

### Analyse (1.1) Keine Modellierung der Random-Effects

- Gute Schätzung der wahren, aber unbekanntem, Fixed-Effects-Parameter
- Parameterstandardfehler zwischen 0.25 und 0.50

### Analyse (1.2) Modellierung der Random-Effects

- Gute Schätzung der wahren, aber unbekanntem, Fixed-Effects-Parameter
- Parameterstandardfehler zwischen 0.25 und 0.50
- Schätzung der Random-Effects-Parameter durch Werte nahe Null

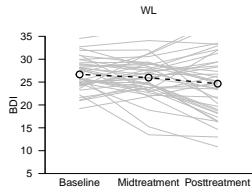
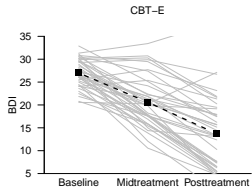
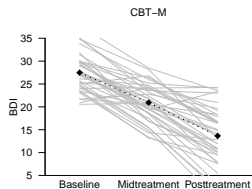
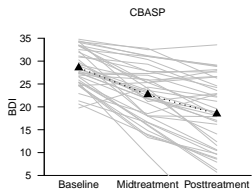
## Szenario (2) Daten mit Random-Effects

Daten der ersten zwei Patient:innen jeder Gruppe

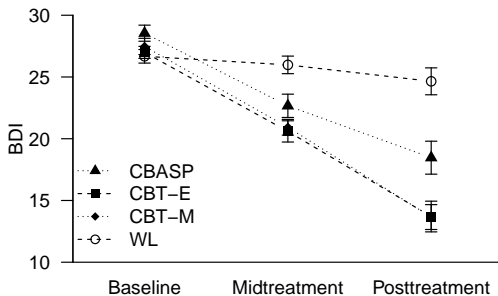
Patient	Group	Time	BDI
1	1 CBASP	0	21
2	1 CBASP	1	14
3	1 CBASP	2	9
4	2 CBASP	0	32
5	2 CBASP	1	25
6	2 CBASP	2	20
121	41 CBT-E	0	24
122	41 CBT-E	1	18
123	41 CBT-E	2	7
124	42 CBT-E	0	24
125	42 CBT-E	1	25
126	42 CBT-E	2	21
241	81 CBT-M	0	27
242	81 CBT-M	1	20
243	81 CBT-M	2	15
244	82 CBT-M	0	30
245	82 CBT-M	1	19
246	82 CBT-M	2	11
361	121 WL	0	22
362	121 WL	1	13
363	121 WL	2	13
364	122 WL	0	31
365	122 WL	1	23
366	122 WL	2	26

# Trendanalyse mit Linear Mixed Models

## Szenario (2) Daten mit Random-Effects



## Szenario (2) Daten mit Random-Effects



## Szenario (2) Daten mit Random-Effects | Analyse (2.1) Keine Modellierung der Random-Effects

```
D = read.csv("../7_Daten/ta_rfx.csv") # Dateneinlesen
M = lm(BDI ~ Group + Time + Group:Time, data = D) # y = X_fb_f + e
```

### Effektparameter

Parameter	R Bezeichner	Interpretation
$\mu_0 = 27$	(Intercept)	Interzept der CBASP Referenzgruppe
$\alpha_2 = 0$	GroupCBT-E	Haupteffekt CBT-E (Differenz zu CBASP)
$\alpha_3 = 0$	GroupCBT-M	Haupteffekt CBT-M (Differenz zu CBASP)
$\alpha_4 = 0$	GroupWL	Haupteffekt WL (Differenz zu CBASP)
$\beta_1 = -4$	Time	Steigung der CBASP Referenzgruppe
$\gamma_2 = -3$	GroupCBT-E:Time	Interaktion CBT-E x Messung (Steigungsdifferenz zu CBASP)
$\gamma_3 = -2$	GroupCBT-M:Time	Interaktion CBT-M x Messung (Steigungsdifferenz zu CBASP)
$\gamma_4 = 3$	GroupWL:Time	Interaktion WL x Messung (Steigungsdifferenz zu CBASP)

# Trendanalyse mit Linear Mixed Models

## Szenario (2) Daten mit Random-Effects | Analyse (2.1) Keine Modellierung der Random-Effects

```
summary(M)
```

Call:

```
lm(formula = BDI ~ Group + Time + Group:Time, data = D)
```

Residuals:

Min	1Q	Median	3Q	Max
-23.6810	-3.4244	-0.1139	3.5064	24.2389

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	28.2660	0.8078	34.991	< 2e-16 ***
GroupCBT-E	-1.1913	1.1424	-1.043	0.2976
GroupCBT-M	-0.6709	1.1424	-0.587	0.5573
GroupWL	-1.4843	1.1424	-1.299	0.1945
Time	-5.0406	0.6257	-8.056	6.52e-15 ***
GroupCBT-E:Time	-1.6054	0.8849	-1.814	0.0703 .
GroupCBT-M:Time	-1.8678	0.8849	-2.111	0.0353 *
GroupWL:Time	4.0272	0.8849	4.551	6.80e-06 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.597 on 472 degrees of freedom

Multiple R-squared: 0.4422, Adjusted R-squared: 0.4339

F-statistic: 53.45 on 7 and 472 DF, p-value: < 2.2e-16



# Trendanalyse mit Linear Mixed Models

## Szenario (2) Daten mit Random-Effects | Analyse (2.2) Modellierung der Random-Effects

```
library(nlme)
ctrl = lmeControl(opt = "optim", maxIter = 100) # Settings
D = read.csv("../7_Daten/ta_rfx.csv") # Dateneinlesen
M = lme(BDI ~ Group + Time + Group:Time, D, random = ~ Time | Patient, control = ctrl) # y = X_fb_f + X_rb_r + e
```

### Effektparameter

Parameter	R Bezeichner	Interpretation
$\mu_0 = 27$	(Intercept)	Interzept der CBASP Referenzgruppe
$\alpha_2 = 0$	GroupCBT-E	Haupteffekt CBT-E (Differenz zu CBASP)
$\alpha_3 = 0$	GroupCBT-M	Haupteffekt CBT-M (Differenz zu CBASP)
$\alpha_4 = 0$	GroupWL	Haupteffekt WL (Differenz zu CBASP)
$\beta_1 = -4$	Time	Steigung der CBASP Referenzgruppe
$\gamma_2 = -3$	GroupCBT-E:Time	Interaktion CBT-E x Messung (Steigungsdifferenz zu CBASP)
$\gamma_3 = -2$	GroupCBT-M:Time	Interaktion CBT-M x Messung (Steigungsdifferenz zu CBASP)
$\gamma_4 = 3$	GroupWL:Time	Interaktion WL x Messung (Steigungsdifferenz zu CBASP)
$\beta_{0i} \sim N\left(0, \sigma_{\beta_r}^2\right)$		Patient:innenspezifisches Interzept
$\beta_{1i} \sim N\left(0, \sigma_{\beta_r}^2\right)$		Patient:innenspezifische Steigung

# Trendanalyse mit Linear Mixed Models

## Szenario (2) Daten mit Random-Effects | Analyse (2.2) Modellierung der Random-Effects

summary(M)

Linear mixed-effects model fit by REML

Data: D  
AIC BIC logLik  
2724.608 2774.492 -1350.304

Random effects:

Formula: ~Time | Patient  
Structure: General positive-definite, Log-Cholesky parametrization  
StdDev Corr  
(Intercept) 3.081973 (Intr)  
Time 3.112930 0.055  
Residual 2.203029

Fixed effects: BDI ~ Group + Time + Group:Time

	Value	Std.Error	DF	t-value	p-value
(Intercept)	28.266045	0.5818721	316	48.57776	0.0000
GroupCBT-E	-1.191285	0.8228914	156	-1.44768	0.1497
GroupCBT-M	-0.670922	0.8228914	156	-0.81532	0.4161
GroupWL	-1.484315	0.8228914	156	-1.80378	0.0732
Time	-5.040629	0.5503862	316	-9.15835	0.0000
GroupCBT-E:Time	-1.605420	0.7783637	316	-2.06256	0.0400
GroupCBT-M:Time	-1.867817	0.7783637	316	-2.39967	0.0170
GroupWL:Time	4.027242	0.7783637	316	5.17399	0.0000

Correlation:

	(Intr)	GrCBT-E	GrCBT-M	GrpWL	Time	GCBT-E:	GCBT-M:
GroupCBT-E	-0.707						
GroupCBT-M	-0.707	0.500					
GroupWL	-0.707	0.500	0.500				
Time	-0.149	0.105	0.105	0.105			
GroupCBT-E:Time	0.105	-0.149	-0.074	-0.074	-0.707		
GroupCBT-M:Time	0.105	-0.074	-0.149	-0.074	-0.707	0.500	
GroupWL:Time	0.105	-0.074	-0.074	-0.149	-0.707	0.500	0.500

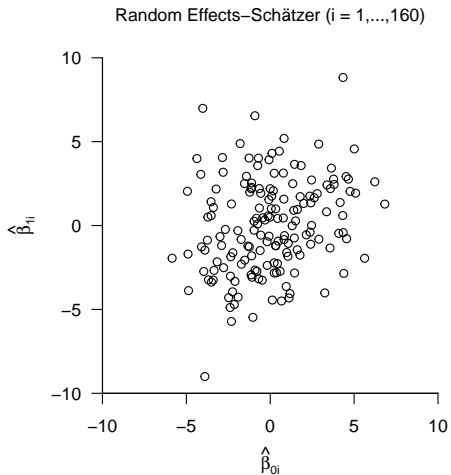
Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.139241949	-0.407949831	-0.009498607	0.441374038	2.140736233

Number of Observations: 480  
Number of Groups: 160

# Trendanalyse mit Linear Mixed Models

Szenario (2) Daten mit Random-Effects | Analyse (2.2) Modellierung der Random-Effects



## Szenario (2) Daten mit Random-Effects

### Analyse (2.1) Keine Modellierung der Random-Effects

- Gute Schätzung der wahren, aber unbekanntem, Fixed-Effects-Parameter
- Parameterstandardfehler zwischen 0.6 und 1.2

### Analyse (2.2) Modellierung der Random-Effects

- Gute Schätzung der wahren, aber unbekanntem, Fixed-Effects-Parameter
- Parameterstandardfehler zwischen 0.5 und 0.80
- Schätzung der Random-Effects-Parameter durch Werte im Bereich -5 bis 5

## Zusammenfassung

- Fixed-Effects-Parameter werden durch Modelle ohne und mit Random-Effects gut geschätzt
- Wenn keine Random-Effects vorliegen, werden sie nahe Null geschätzt
- Wenn Random-Effects vorliegen, reduziert sich der Standardfehler der Fixed-Effects Schätzer

⇒ Anwendung von Linear Mixed Models ist in beiden Szenarien kein Nachteil

---

Motivation

Zweistichproben-T-Test bei abhängigen Stichproben

Einfaktorielle Varianzanalyse mit Messwiederholung

Trendanalyse mit Linear Mixed Models

**Selbstkontrollfragen**

1. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests mit abhängigen Stichproben (ZTT-AS).
2. Geben Sie das Modell eines ZTT-AS in struktureller Form wieder.
3. Geben Sie das Theorem zur marginalen Datenverteilung des ZTT-AS-Modells wieder.
4. Warum ist die Standardinferenztheorie des Allgemeinen Modells für einen ZTT-AS nicht anwendbar?
5. Erläutern Sie die Durchführung eines ZTT-AS mithilfe eines Einstichproben-T-Test.

## Appendix



## Definition (Kroneckerprodukt)

Es seien  $A \in \mathbb{R}^{m \times n}$  und  $B \in \mathbb{R}^{p \times q}$ . Dann ist das *Kroneckerprodukt* von  $A$  und  $B$  definiert als die Abbildung

$$\otimes : \mathbb{R}^{m \times n} \times \mathbb{R}^{p \times q} \rightarrow \mathbb{R}^{mp \times nq}, (A, B) \mapsto \otimes(A, B) := A \otimes B \quad (28)$$

mit

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}. \quad (29)$$

### Bemerkungen

- Das Kroneckerprodukt kann zur Konstruktion von Matrizen mit repetitiver Struktur genutzt werden.
- In der Form  $1_n \otimes M$  setzt das Kroneckerprodukt die Matrix  $M$   $n$ -mal übereinander.
- In der Form  $M \otimes 1_n$  repliziert das Kroneckerprodukt jede Zeile von  $M$   $n$ -mal.
- In der Form  $I_n \otimes M$  erzeugt das Kroneckerprodukt eine Blockdiagonalmatrix  $M$  auf der Diagonale.

## Beispiel (1)

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \quad (30)$$

```
M      = matrix(c(1,1,
                 1,2,
                 1,3),
               byrow = TRUE, nrow = 3)      # Matrix
j_2    = matrix(rep(1,2)
               , nrow = 2)                # 1_2 Vektor
X      = kronecker(j_2,M)                  # Kroneckerprodukt
```

```
      [,1] [,2]
[1,]    1    1
[2,]    1    2
[3,]    1    3
[4,]    1    1
[5,]    1    2
[6,]    1    3
```

## Appendix

### Beispiel (2)

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \quad (31)$$

```
M      = matrix(c(1,0,0,0,                # Matrix
                 1,1,0,0,
                 1,0,1,0,
                 1,0,0,1),
               byrow = TRUE, nrow = 4)
j_2    = matrix(rep(1,2)                 , nrow = 2)   # 1_2 Vektor
X      = kronecker(M,j_2)                # Kroneckerprodukt
```

```
      [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    1    0    0    0
[3,]    1    1    0    0
[4,]    1    1    0    0
[5,]    1    0    1    0
[6,]    1    0    1    0
[7,]    1    0    0    1
[8,]    1    0    0    1
```

## Beispiel (3)

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix} \quad (32)$$

```
M      = matrix(c(2,1,                # Matrix
                 1,2),
               byrow = TRUE, nrow = 2)
I_2    = diag(2)                    # I_2
X      = kronecker(I_2,M)           # Kroneckerprodukt
```

```
      [,1] [,2] [,3] [,4]
[1,]    2    1    0    0
[2,]    1    2    0    0
[3,]    0    0    2    1
[4,]    0    0    1    2
```

- DeGroot, Morris H., and Mark J. Schervish. 2012. *Probability and Statistics*. 4th ed. Boston: Addison-Wesley.
- Detry, Michelle A., and Yan Ma. 2016. "Analyzing Repeated Measurements Using Mixed Models." *JAMA* 315 (4): 407. <https://doi.org/10.1001/jama.2015.19394>.
- Huynh, Huynh, and Leonard S. Feldt. 1970. "Conditions Under Which Mean Square Ratios in Repeated Measurements Designs Have Exact F-Distributions." *Journal of the American Statistical Association* 65 (332): 1582–89. <https://doi.org/10.1080/01621459.1970.10481187>.
- Rief, Winfried, Gabi Bleichhardt, Katharina Dannehl, Frank Euteneuer, and Katrin Wambach. 2018. "Comparing the Efficacy of CBASP with Two Versions of CBT for Depression in a Routine Care Center: A Randomized Clinical Trial." *Psychotherapy and Psychosomatics* 87 (3): 164–78. <https://doi.org/10.1159/000487893>.