



# Interventionsforschung

MSc Klinische Psychologie und Psychotherapie

MSc Umweltpsychologie / Mensch-Technik-Interaktion

SoSe 2026

Prof. Dr. Dirk Ostwald

## (5) Multizentren-Designs

---

Motivation

Multizentren-Parallelgruppendesigns

Multizentren-Regressionsdesigns

Ein Multizentren-Simpson's Paradox-Beispiel

Selbstkontrollfragen

---

## Motivation

Multizentren-Parallelgruppendesigns

Multizentren-Regressionsdesigns

Ein Multizentren-Simpson's Paradox-Beispiel

Selbstkontrollfragen

# Motivation

---

## Multizentrenstudie

### Definition

- Zentrum = Klinik, Hochschulambulanz, Institut, Praxis
- Durchführung einer Studie an zwei oder mehr Zentren nach einem einheitlichen Protokoll

### Motivation

- Rekrutierung höherer Proband:innenzahlen in kürzerer Zeit als bei einem Zentrum
- Höhere Übertragbarkeit auf verschiedene regionale Patient:innengruppen

### Herausforderungen

- Höhere "Ähnlichkeit" von Daten innerhalb eines Zentrums als zwischen Zentren
- Variabilität in der Umsetzung des Studienprotokolls zwischen Zentren
- Variabilität in den Komorbiditäten der Patient:innen zwischen Zentren

Zu den Praktikalitäten einer Multizentrenstudie, siehe z.B. Friedman et al. (2015), Kapitel 21

# Motivation

---

## Anwendungsbeispiel

Li et al. (2025) Supported Mindfulness-Based Self-Help Intervention as an Adjunctive Treatment for Rapid Symptom Change in Emotional Disorders: A Practice-Oriented Multicenter Randomized Controlled Trial

### Research question

- Rapid symptom relief is crucial for individuals with emotional disorders.
- Does a mindfulness-based self-help (MBSH) intervention provide rapid improvement?

### Methods

- A randomized controlled trial was conducted on a sample of  $n = 302$  patients from four centers.
- Participants were randomly assigned to either MBSH+TAU ( $n = 152$ ) or TAU-only ( $n = 150$ ).
- Assessments were conducted at baseline, week 3, week 5, after intervention and at 3-month follow-up.
- Primary outcomes included self-reported and clinician-reported anxiety and depression symptoms.

### Results

- The MBSH+TAU group achieved significantly greater improvements in all primary outcome measures as compared with TAU-only immediately after intervention (Cohen's  $d = 0.19-0.51$ ).

### Conclusion

- MBSH offers a scalable and effective adjunctive treatment option for patients with emotional disorders.

# Motivation

---

Die Datenanalyse von Multizentrenstudien ist ein aktives Forschungsfeld

Localio et al. (2001)

- An overview of data-analytic adjustments for center in multicenter studies
- Disadvantages e.g. underpowered study designs, incorrect p-values

Tangri et al. (2010)

- Localio et al. (2001) and others have advocated controlling for center effects
- Only  $\approx 20\%$  of reviewed multicenter studies adjust for center effects

Brown and Prescott (2015)

- Center effect control by means of fixed center effects or fixed center-treatment interactions
- Center effect control by means of random center effects or random center-treatment interactions

Kahan and Morris (2012)

- Comparison of fixed vs. random center effects (not center-treatment interactions)
- Simulation of a wide variety of scenarios, random effects control generally superior

Edgar (2021)

- Reanalysis of the CRASH-2 multicenter study using random center effects models
- Increase in power and decrease in bias in mean and standard error estimates

---

Motivation

## **Multizentren-Parallelgruppendesigns**

Multizentren-Regressionsdesigns

Ein Multizentren-Simpson's Paradox-Beispiel

Selbstkontrollfragen

## Überblick

Treatmentgruppe und Kontrollgruppe in jedem Zentrum

### (1) Random-Center-Effect-Modell

- Der Treatmenteffekt wird als Fixed-Effect modelliert
- Der Zentreneffekt wird als additiver Random-Effect modelliert
- $\Rightarrow$  Treatment- und Kontrollgruppe können einen Zentren-spezifischen Intercept haben
- $\Rightarrow$  Der Treatmenteffekt wird in jedem Zentrum als identisch angenommen

### (2) Random-Center-by-Treatment-Interaction-Modell

- Der Treatmenteffekt wird als Fixed-Effect modelliert
- Der Zentreneffekt wird als additiver Random-Effect modelliert
- Der Zentreneffekt wird zusätzlich als Zentren-spezifischer Treatmenteffekt modelliert
- $\Rightarrow$  Treatment- und Kontrollgruppe haben einen Zentren-spezifischen Intercept
- $\Rightarrow$  Treatment- und Kontrollgruppe haben einen Zentren-spezifischen Unterschied

# Multizentren-Parallelgruppendesigns

## (1) Random-Center-Effect-Modell

Strukturelle Form

Für Zentren  $i = 1, \dots, k$  und Proband:innen  $j = 1, \dots, n_i$

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{0i} + \varepsilon_{ij} \quad (1)$$

mit

- $x_{ij} := 0$  für Proband:in  $j$  in Zentrum  $i$  in Kontrollgruppe
- $x_{ij} := 1$  für Proband:in  $j$  in Zentrum  $i$  in Treatmentgruppe
- $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$
- $b_{0i} \sim N(0, \sigma_b^2)$

Parameterbedeutungen

$\beta_0$	Erwartungswert der Kontrollgruppe über Zentren
$\beta_1$	Erwartungswertunterschied zwischen Kontrollgruppe und Treatmentgruppe
$b_{0i}$	Erwartungswertunterschiede der Kontrollgruppen über Zentren
$\sigma_\varepsilon^2$	Varianz zwischen Proband:innen
$\sigma_b^2$	Varianz zwischen Zentren



# Multizentren-Parallelgruppendesigns

## (1) Random-Center-Effect-Modell

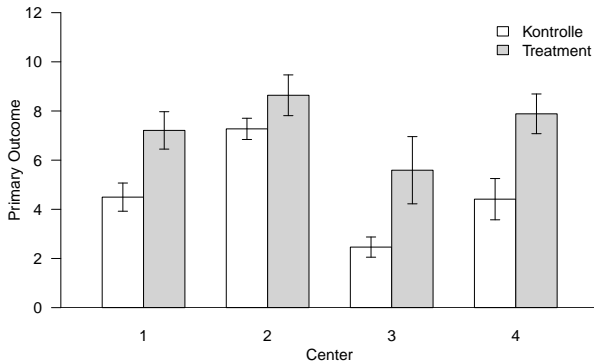
Datengeneration für ein Szenario mit  $k = 4$ ,  $m = 4$  und  $n_i = 8$  für  $i = 1, \dots, k$

```
library(MASS) # Normalverteilung
set.seed(0) # Zufallszahlengeneratorzustand
k = 4 # Anzahl Zentren
m = 4 # Patient:innen pro Zentrum und Bedingung
n_i = 2*m # Anzahl Patient:innen pro Zentrum
n = k*2*m # Gesamtanzahl an Patient:innen
X_i = kronecker(matrix(c(1,1,0,1), ncol = 2), rep(1,m)) # Treatment-Control Designmatrix
X = kronecker(rep(1,k), X_i) # Fixed-Effects-Designmatrix
Z = kronecker(diag(k), rep(1,n_i)) # Random-Effects-Designmatrix
beta = matrix(c(5,2), nrow = 2) # Fixed-Effects-Parameter
b = matrix(c(0,2,-2,0), nrow = k) # Random-Effects-Parameter (E(b) = 0!)
s_eps = 1 # Varianzkomponente
eps = mvrnorm(1, rep(0,n), s_eps*diag(n)) # Fehlervektor
y = X %*% beta + Z %*% b + eps # Datengeneration
TRM = kronecker(rep(1,k), kronecker(c(1:2), rep(1,m))) # Treatmentfaktor
CTR = kronecker(c(1:k), rep(1,n_i)) # Centerfaktor
D = data.frame(TRM = as.factor(TRM), CTR = as.factor(CTR), OUT = y) # Dataframe
write.csv(D, "5-Daten/mz-parallelgruppendesign-1.csv", row.names = FALSE) # Speichern
```

# Multizentren-Parallelgruppendesigns

## (1) Random-Center-Effect-Modell

Deskriptivstatistik für ein Szenario mit  $k = 4$ ,  $m = 4$  und  $n_i = 8$  für  $i = 1, \dots, k$



# Multizentren-Parallelgruppendesigns

## (1) Random-Center-Effect-Modell

Parameterschätzung für ein Szenario mit  $k = 4$ ,  $m = 4$  und  $n_i = 8$  für  $i = 1, \dots, k$

```
library(nlme) # nlme R Paket
D = read.csv("5-Daten/mz-parallelgruppendesign-1.csv", head = T) # Dataframe
D$TRM = as.factor(D$TRM) # R Faktor Kodierung
D$CTR = as.factor(D$CTR) # R Faktor Kodierung
M = lme(OUT ~ TRM, data = D, random = ~ 1 | CTR) # LMM Schätzung
X = model.matrix(M,D) # Fixed-Effects-Designmatrix
Z = model.matrix(~ M$groups[[1]] - 1) # Random-Effects-Designmatrix
beta_hat = M$coefficients$fixed # Fixed-Effects-Parameterschätzer
b_hat = M$coefficients$random$CTR # Random-Effects-Parameterschätzer
s_eps_hat = M$sigma**2 # Varianzkomponentenschätzer
s_b_hat = diag(getVarCov(M)) # Varianzkomponentenschätzer

beta_hat      : 4.66 2.67
b_hat         : -0.14 1.89 -1.9 0.15
sigsqr_b_hat  : 2.49
sigsqr_eps_hat : 0.76

              Value Std.Error lower upper    T    pval
(Intercept)  4.66     0.819  2.98  6.34  5.69  4.79e-06
TRM2         2.67     0.309  2.04  3.30  8.64  2.96e-09
```

# Multizentren-Parallelgruppendesigns

## (2) Random-Center-by-Treatment-Interaction-Modell

Strukturelle Form

Für Zentren  $i = 1, \dots, k$  und Proband:innen  $j = 1, \dots, n_i$

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{0i} + b_{1i} x_{ij} + \varepsilon_{ij} \quad (3)$$

mit

- $x_{ij} := 0$  für Proband:in  $j$  in Zentrum  $i$  in Kontrollgruppe
- $x_{ij} := 1$  für Proband:in  $j$  in Zentrum  $i$  in Treatmentgruppe
- $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$
- $b_{0i} \sim N(0, \sigma_b^2)$
- $b_{1i} \sim N(0, \sigma_b^2)$

Parameterbedeutungen

$\beta_0$	Erwartungswert der Kontrollgruppe über Zentren
$\beta_1$	Erwartungswertunterschied zwischen Kontrollgruppe und Treatmentgruppe
$b_{0i}$	Erwartungswertunterschied der Kontrollgruppen über Zentren
$b_{1i}$	Zentren-spezifische Abweichung des Erwartungswertunterschieds zwischen Kontrollgruppe und Treatmentgruppe
$\sigma_\varepsilon^2$	Varianz zwischen Proband:innen
$\sigma_b^2$	Varianz zwischen Zentren

# Multizentren-Parallelgruppendesigns

## (2) Random-Center-by-Treatment-Interaction-Modell

Designmatrixform für  $k = 4, n_1 = n_2 = n_3 = n_4 = 8, m = \frac{1}{2}n_i = 4$  mit  $i = 1, 2, 3, 4$

$$y = X\beta + Zb + \varepsilon \Leftrightarrow \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{16} \\ y_{17} \\ y_{18} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{25} \\ y_{26} \\ y_{27} \\ y_{28} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \\ y_{35} \\ y_{36} \\ y_{37} \\ y_{38} \\ y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \\ y_{45} \\ y_{46} \\ y_{47} \\ y_{48} \end{pmatrix} = \begin{pmatrix} 10 \\ 10 \\ 10 \\ 10 \\ 11 \\ 11 \\ 11 \\ 11 \\ 10 \\ 10 \\ 10 \\ 10 \\ 11 \\ 11 \\ 11 \\ 11 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 11 \\ 11 \\ 11 \\ 11 \\ 10 \\ 10 \\ 10 \\ 10 \\ 11 \\ 11 \\ 11 \\ 11 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 10000000 \\ 10000000 \\ 10000000 \\ 10000000 \\ 11000000 \\ 11000000 \\ 11000000 \\ 11000000 \\ 00100000 \\ 00100000 \\ 00100000 \\ 00100000 \\ 00110000 \\ 00110000 \\ 00110000 \\ 00110000 \\ 00110000 \\ 00001000 \\ 00001000 \\ 00001000 \\ 00001000 \\ 00001000 \\ 00001100 \\ 00001100 \\ 00001100 \\ 00001100 \\ 00000010 \\ 00000010 \\ 00000010 \\ 00000010 \\ 00000011 \\ 00000011 \\ 00000011 \\ 00000011 \end{pmatrix} \begin{pmatrix} b_{01} \\ b_{11} \\ b_{02} \\ b_{12} \\ b_{03} \\ b_{13} \\ b_{04} \\ b_{14} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \\ \varepsilon_{16} \\ \varepsilon_{17} \\ \varepsilon_{18} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{25} \\ \varepsilon_{26} \\ \varepsilon_{27} \\ \varepsilon_{28} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \\ \varepsilon_{35} \\ \varepsilon_{36} \\ \varepsilon_{37} \\ \varepsilon_{38} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{43} \\ \varepsilon_{44} \\ \varepsilon_{45} \\ \varepsilon_{46} \\ \varepsilon_{47} \\ \varepsilon_{48} \end{pmatrix}$$

mit

$$\varepsilon \sim N(0_{32}, \sigma_\varepsilon^2 I_{32}) \text{ und } b \sim N(0_8, \Sigma_b) \quad (4)$$

# Multizentren-Parallelgruppendesigns

## (2) Random-Center-by-Treatment-Interaction-Modell

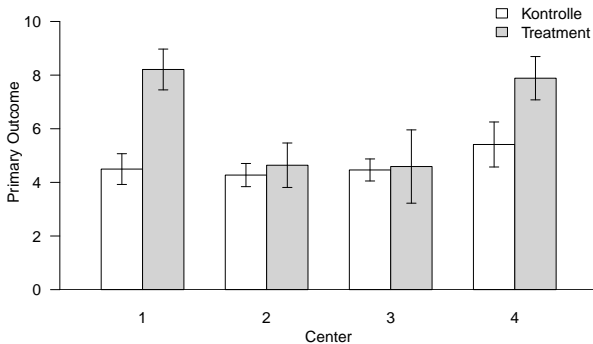
Datengeneration für ein Szenario mit  $k = 4$ ,  $m = 4$  und  $n_i = 8$  für  $i = 1, \dots, k$

```
library(MASS) # multivariate Normalverteilung
set.seed(0) # Zufallszahlengeneratorzustand
k = 4 # Anzahl Zentren
m = 4 # Patient:innen pro Zentrum und Bedingung
n_i = 2*m # Anzahl Patient:innen pro Zentrum
n = k*n_i # Gesamtanzahl an Patient:innen
X_i = kronecker(matrix(c(1,1,0,1), ncol = 2), rep(1,m)) # Treatment-Control Designmatrix
X = kronecker(rep(1,k), X_i) # Fixed-Effects-Designmatrix
Z = kronecker(diag(k), X_i) # Random-Effects-Designmatrix
beta = matrix(c(5,1), nrow = 2) # Fixed-Effects-Parameter
b = matrix(c(0,2,-1,0,0,-2,1,0), nrow = 2*k) # Random-Effects-Parameter (E(b) = 0!)
s_eps = 1 # Varianzkomponente
eps = mvrnorm(1, rep(0,n), s_eps*diag(n)) # Fehlervektor
y = X %>% beta + Z %>% b + eps # Datengeneration
TRM = kronecker(rep(1,k), kronecker(c(1:2), rep(1,m))) # Treatmentfaktor
CTR = kronecker(c(1:k), rep(1,2*m)) # Zentrumfaktor
D = data.frame(TRM = as.factor(TRM), CTR = as.factor(CTR), OUT = y) # Dataframe
write.csv(D, "5-Daten/mz-parallelgruppendesign-2.csv", row.names = FALSE) # Speichern
```

# Multizentren-Parallelgruppendesigns

## (2) Random-Center-by-Treatment-Interaction-Modell

Deskriptivstatistik für ein Szenario mit  $k = 4$ ,  $m = 4$  und  $n_i = 8$  für  $i = 1, \dots, k$



# Multizentren-Parallelgruppendesigns

## (2) Random-Center-by-Treatment-Interaction-Modell

Parameterschätzung für ein Szenario mit  $k = 4$ ,  $m = 4$  und  $n_i = 8$  für  $i = 1, \dots, k$

```
library(nlme) # nlme R Paket
ctrl         = lmeControl(opt = "optim", maxIter = 200) # Anzahl an REML-Iterationen
D            = read.csv("5-Daten/mz-parallelgruppendesign-2.csv", head = T) # Dataframe
D$TRM       = as.factor(D$TRM) # R Faktor Kodierung
D$CTR       = as.factor(D$CTR) # R Faktor Kodierung
M           = lme(OUT ~ TRM, data = D, random = ~ TRM | CTR, control = ctrl) # LMM Schätzung
X           = model.matrix(M,D) # Fixed-Effects-Designmatrix
Z           = kronecker(diag(k), X_i) # Random-Effects-Designmatrix
beta_hat    = M$coefficients$fixed # Fixed-Effects-Parameterschätzer
b_hat      = M$coefficients$random$CTR # Random-Effects-Parameterschätzer
s_eps_hat  = M$sigma**2 # Varianzkomponentenschätzer
s_b_hat     = diag(getVarCov(M)) # Varianzkomponentenschätzer

beta_hat    : 4.66 1.67
b_hat       : 0.27 1.46 -0.27 -1.37 -0.27 -1.39 0.27 1.3
sigsqr_b_hat : 0.11 2.65
sigsqr_eps_hat : 0.64

              Value Std.Error  lower upper    T    pval
(Intercept)  4.66      0.258  4.1315  5.19 18.06 1.33e-16
TRM2         1.67      0.862 -0.0992  3.44  1.94 6.33e-02
```

# Multizentren-Parallelgruppendesigns

Random-Center-by-Treatment-Interaction-Modell in Populationsparameterform

Designmatrixform für  $k = 4$ ,  $n_1 = n_2 = n_3 = n_4 = 8$ ,  $m = \frac{1}{2}n_i = 4$  mit  $i = 1, 2, 3, 4$

Für  $A := 1_k \otimes I_2 \in \mathbb{R}^{2k \times 2}$  gilt

$$X = ZA \Leftrightarrow \begin{pmatrix} 10 \\ 10 \\ 10 \\ 10 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 11 \\ 11 \\ 11 \\ 11 \end{pmatrix} = \begin{pmatrix} 10000000 \\ 10000000 \\ 10000000 \\ 10000000 \\ 11000000 \\ 11000000 \\ 11000000 \\ 11000000 \\ 11000000 \\ 00100000 \\ 00100000 \\ 00100000 \\ 00100000 \\ 00100000 \\ 00110000 \\ 00110000 \\ 00110000 \\ 00110000 \\ 00110000 \\ 00001000 \\ 00001000 \\ 00001000 \\ 00001000 \\ 00001000 \\ 00001000 \\ 00001000 \\ 00001000 \\ 00001000 \\ 00001000 \\ 00001000 \\ 00000010 \\ 00000010 \\ 00000010 \\ 00000010 \\ 00000010 \\ 00000010 \\ 00000010 \\ 00000010 \\ 00000011 \\ 00000011 \\ 00000011 \\ 00000011 \end{pmatrix} \begin{pmatrix} 10 \\ 01 \\ 10 \\ 01 \\ 10 \\ 01 \\ 10 \\ 01 \end{pmatrix}$$

# Multizentren-Parallelgruppendesigns

## Random-Center-by-Treatment-Interaction-Modell in Populationsparameterform

```
library(MASS) # multivariate Normalverteilung
set.seed(0) # Zufallszahlengeneratorzustand
k = 4 # Anzahl Zentren
m = 4 # Patient:innen pro Zentrum & Bedingung
n = k*2*m # Gesamtanzahl an Patient:innen
X_i = kronecker(matrix(c(1,1,0,1), ncol = 2), rep(1,m)) # Treatment-Control Designmatrix
X = kronecker(rep(1,k), X_i) # Fixed-Effects-Designmatrix
Z = kronecker(diag(k), X_i) # Random-Effects-Designmatrix
A = kronecker(rep(1,k), diag(2)) # Random-2-Fixed-Matrix
beta = matrix(c(5,1), nrow = 2) # Populationsparameter
s_b = 1 # Populationsparameter
u = mvrnorm(1, A %*% beta, s_b*diag(k*2)) # Zentren-spezifische Effekte
s_eps = 1 # Varianzkomponente
eps = mvrnorm(1, rep(0,n), s_eps*diag(n)) # Fehlervektor
y = Z %*% u + eps # Datengeneration
TRM = kronecker(rep(1,k), kronecker(c(1:2), rep(1,m))) # Treatmentfaktor
CTR = kronecker(c(1:k), rep(1,2*m)) # Centerfaktor
D = data.frame(TRM = TRM, CTR = CTR, OUT = y) # Dataframe
write.csv(D, "5-Daten/mz-parallelgruppendesign-3.csv", row.names = FALSE) # Speichern
```

# Multizentren-Parallelgruppendesigns

## Random-Center-by-Treatment-Interaction-Modell in Populationsparameterform

Deskriptivstatistik für ein Szenario mit  $k = 4$ ,  $m = 4$  und  $n_i = 8$  für  $i = 1, \dots, k$

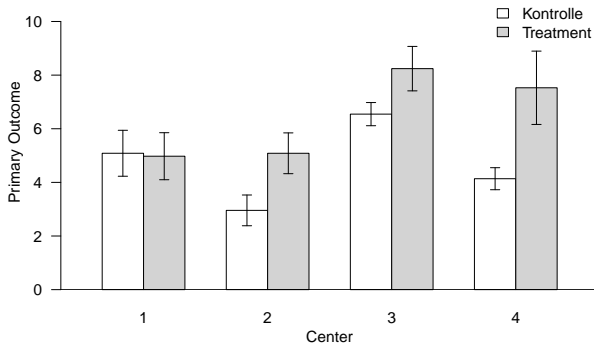
```
library(dplyr) # dplyr für einfache Datengruppierung
D = read.csv("5-Daten/mz-parallelgruppendesign-3.csv") # Dateneinlesen
DS = D %>% group_by(CTR, TRM) %>% summarise(av = mean(OUT, na.rm = TRUE), # Group mean
                                             sd = sd(OUT, na.rm = TRUE), # Group standard deviation
                                             .groups = "drop") # Gruppierungsaufhebung
print(DS) # Ausgabe
```

```
# A tibble: 8 x 4
  CTR TRM av sd
<int> <int> <dbl> <dbl>
1 1 1 5.09 0.857
2 1 2 4.98 0.877
3 2 1 2.96 0.573
4 2 2 5.09 0.761
5 3 1 6.55 0.432
6 3 2 8.24 0.829
7 4 1 4.14 0.411
8 4 2 7.53 1.37
```

# Multizentren-Parallelgruppendesigns

## Random-Center-by-Treatment-Interaction-Modell in Populationsparameterform

Deskriptivstatistik für ein Szenario mit  $k = 4$ ,  $m = 4$  und  $n_i = 8$  für  $i = 1, \dots, k$



## Dokumentation

### (1) Random-Center-Effect-Modell

*Deskriptiv fiel das Primary Outcome in der Treatmentgruppe höher aus ( $M = 7.10$ ,  $SD = 1.80$ ) als in der Kontrollgruppe ( $M = 5.01$ ,  $SD = 1.86$ ). Die Multizentrenstudie wurde mittels eines Linear Mixed Models mit zufälligen Zentrumsseffekten analysiert; dabei zeigte sich ein signifikanter positiver Treatmenteffekt,  $b = 2.10$ ,  $SE = 0.28$ , 95%-KI [1.54, 2.66],  $t(35) = 7.57$ ,  $p < .001$ .*

### (2) Random-Center-by-Treatment-Interaction-Modell

*Deskriptiv fiel das Primary Outcome in der Treatmentgruppe höher aus ( $M = 6.10$ ,  $SD = 1.90$ ) als in der Kontrollgruppe ( $M = 5.01$ ,  $SD = 1.25$ ). Die Multizentrenstudie wurde mittels eines Linear Mixed Models mit zufälligen Zentrumsseffekten und zufälligen Zentren-Treatment-Interaktionen analysiert; dabei ergab sich kein signifikanter Treatmenteffekt,  $b = 1.10$ ,  $SE = 0.72$ , 95%-KI [-0.37, 2.56],  $t(35) = 1.52$ ,  $p = .137$ .*

---

Motivation

Multizentren-Parallelgruppendesigns

**Multizentren-Regressionsdesigns**

Ein Multizentren-Simpson's Paradox-Beispiel

Selbstkontrollfragen

## Anwendungsbeispiel

- Nachweis einer Dosis-Wirkungs-Beziehung in der Depressionstherapie
- $k = 4$  Zentren (Hochschulambulanzen)
- $n_i = 5$  Proband:innen pro Zentrum  $i = 1, 2, 3, 4$
- 24 Stunden Kognitive Verhaltenstherapie pro Patient:in in jedem Zentrum
- 1 bis 5 ACT-Therapiekomponenten pro Patient:in als unabhängige Variable (UV)
- BDI-II-Reduktion als Primary Outcome / abhängige Variable (AV)

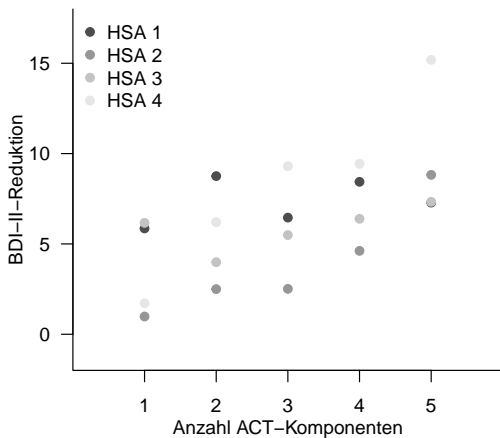
## Bemerkungen

- Die unabhängige Variable modelliert hier explizit *keine* zeitabhängige Variable
- Regression bei zeitabhängigen Variablen betrachtet man als Longitudinaldesigns

## Datengeneration

```
library(MASS) # multivariate Normalverteilung
library(Matrix) # Blockdiagonalmatrizen
set.seed(0) # Zufallszahlengeneratorzustand
k = 4 # Anzahl Zentren
n_i = 5 # Patient:innen pro Zentrum
n = k*n_i # Gesamtanzahl an Patient:innen
x = 1:5 # Anzahl ACT-Komponenten
Xi = matrix(c(rep(1,n_i), x), ncol = 2) # Zentren-spezifische Regressionsmatrizenarray
X = kronecker(rep(1,k), Xi) # Fixed-Effects-Designmatrix
beta = matrix(c(2,1), nrow = 2) # Fixed-Effects-Parameter
Z = kronecker(diag(k), Xi) # Random-Effects-Designmatrix
s_b = 2 # Random-Effects-Varianzparameter
b = c(6,-1,-2,0.5,-2,1,-4,2) # Random-Effects Parameter
s_eps = 3 # Varianzkomponente
eps = mvrnorm(1, rep(0,n), s_eps*diag(n)) # Fehlervektor
y = X %*% beta + Z %*% b + eps # Datengeneration
HSA = kronecker(c(1:k), rep(1,n_i)) # Hochschulambulanzfaktor
ACT = X[,2] # ACT-Anzahl-Regressor
D = data.frame(HSA = HSA, ACT = ACT, BDI = y) # Dataframe
write.csv(D, "5-Daten/mz-regression-1.csv", row.names = FALSE) # Speichern
```

# Multizentren-Regressionsdesigns



## Überblick

Alle Multizentren-Regressionsmodelle modellieren Intercept und Slope als Fixed-Effects

### (1) Random-Intercept-Only-Modell

- Zusätzlich wird nur ein zufälliger Zentren-spezifischer Intercept modelliert

### (2) Random-Slope-Only-Modell

- Zusätzlich wird nur ein zufälliger Zentren-spezifischer Slope modelliert

### (3) Random-Intercept-and-Slope-Modell

- Es werden zusätzlich zufällige Zentren-spezifische Intercepts und Slopes modelliert
- Es ergibt sich eine Populationsparameterinterpretation für die Fixed-Effects

## (1) Random-Intercept-Only-Modell

Strukturelle Form

Für Zentren  $i = 1, \dots, k$  und Proband:innen  $j = 1, \dots, n_i$

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{0i} + \varepsilon_{ij} \quad (5)$$

mit

- $x_{ij}$  := Wert der unabhängigen Variable für Proband:in  $j$  in Zentrum  $i$
- $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$
- $b_{0i} \sim N(0, \sigma_b^2)$

Parameterbedeutungen

$\beta_0$	Erwartungswert der AV bei Wert der UV gleich 0
$\beta_1$	Erwartungswert der Änderung der AV pro UV-Änderung um 1 Einheit
$b_{0i}$	Zentren-spezifische Abweichung der AV bei Wert der UV gleich 0
$\sigma_b^2$	Varianz der $b_{0i}$ zwischen Zentren
$\sigma_\varepsilon^2$	Varianz der Datenpunkte innerhalb der Zentren

# Multizentren-Regressionsdesigns

## (1) Random-Intercept-Only-Modell

Designmatrixform für  $k = 4, n_1 = n_2 = n_3 = n_4 = 5, n = 20$

$$y = X\beta + Zb + \varepsilon \Leftrightarrow \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{25} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \\ y_{35} \\ y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \\ y_{45} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{13} \\ 1 & x_{14} \\ 1 & x_{15} \\ 1 & x_{21} \\ 1 & x_{22} \\ 1 & x_{23} \\ 1 & x_{24} \\ 1 & x_{25} \\ 1 & x_{31} \\ 1 & x_{32} \\ 1 & x_{33} \\ 1 & x_{34} \\ 1 & x_{35} \\ 1 & x_{41} \\ 1 & x_{42} \\ 1 & x_{43} \\ 1 & x_{44} \\ 1 & x_{45} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_{01} \\ b_{02} \\ b_{03} \\ b_{04} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{25} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \\ \varepsilon_{35} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{43} \\ \varepsilon_{44} \\ \varepsilon_{45} \end{pmatrix}$$

mit

$$b \sim N(0_4, \sigma_b^2 I_4) \text{ und } \varepsilon \sim N(0_{20}, \sigma_\varepsilon^2 I_{20}) \quad (6)$$

## (1) Random-Intercept-Only-Modell

```
library(nlme) # R Paket
D             = read.csv("5-Daten/mz-regression-1.csv") # Einlesen des Datensatzes
D$HSA       = as.factor(D$HSA) # Umwandlung numerischer Werte in R-Faktor
M           = lme(BDI ~ ACT, D, random = ~ 1 | HSA) # Random-Intercept-Only LMM
X           = model.matrix(M, D) # Fixed-Effects-Designmatrix
Z           = model.matrix(~ M$groups[[1]] - 1) # Random-Effects-Designmatrix
beta_hat    = M$coefficients$fixed # Fixed-Effects-Parameterschätzer
b_hat       = M$coefficients$random$HSA # Random-Effects-Parameterschätzer
s_eps_hat   = M$sigma**2 # Varianzkomponentenschätzer
s_b_hat     = diag(getVarCov(M)) # Varianzkomponentenschätzer
```

## (1) Random-Intercept-Only-Modell

Fixed-Effects-Designmatrix

```
(Intercept) ACT
1           1   1
2           1   2
3           1   3
4           1   4
5           1   5
6           1   1
7           1   2
8           1   3
9           1   4
10          1   5
11          1   1
12          1   2
13          1   3
14          1   4
15          1   5
16          1   1
17          1   2
18          1   3
19          1   4
20          1   5
attr(,"assign")
[1] 0 1
```

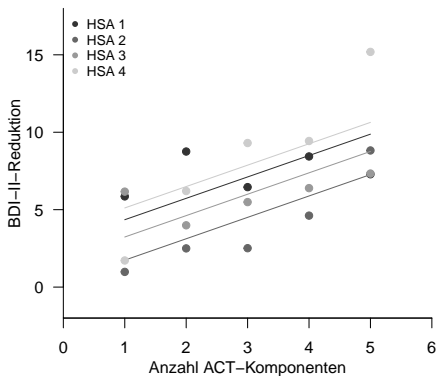
## (1) Random-Intercept-Only-Modell

### Random-Effects-Designmatrix

```
M$groups[[1]]1 M$groups[[1]]2 M$groups[[1]]3 M$groups[[1]]4
1           1           0           0           0
2           1           0           0           0
3           1           0           0           0
4           1           0           0           0
5           1           0           0           0
6           0           1           0           0
7           0           1           0           0
8           0           1           0           0
9           0           1           0           0
10          0           1           0           0
11          0           0           1           0
12          0           0           1           0
13          0           0           1           0
14          0           0           1           0
15          0           0           1           0
16          0           0           0           1
17          0           0           0           1
18          0           0           0           1
19          0           0           0           1
20          0           0           0           1
attr(,"assign")
[1] 1 1 1 1
attr(,"contrasts")
attr(,"contrasts")$`M$groups[[1]]`
[1] "contr.treatment"
```

# Multizentren-Regressionsdesigns

## (1) Random-Intercept-Only-Modell



## (1) Random-Intercept-Only-Modell

```
beta_hat      :  2.23  1.38
sigsqr_eps_hat :  4.71
sigsqr_b_hat   :  2.85
```

b\_hat:

(Intercept)

```
1  0.7415643
2 -1.8691530
3 -0.3745826
4  1.5021712
```

Variabler Intercept  $\hat{\beta}_0 + \hat{b}_{0i}$  der Ausgleichsgerade über Zentren

Identische Steigung  $\hat{\beta}_1$  der Ausgleichsgerade über Zentren

## (2) Random-Slope-Only-Modell

Strukturelle Form

Für Zentren  $i = 1, \dots, k$  und Proband:innen  $j = 1, \dots, n_i$

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{1i} x_{ij} + \varepsilon_{ij} \quad (7)$$

mit

- $x_{ij}$  := Wert der unabhängigen Variable für Proband:in  $j$  in Zentrum  $i$
- $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$
- $b_{1i} \sim N(0, \sigma_b^2)$

Parameterbedeutungen

$\beta_0$	Erwartungswert der AV bei Wert der UV gleich 0
$\beta_1$	Erwartungswert der Änderung der AV pro UV-Änderung um 1 Einheit
$b_{1i}$	Zentren-spezifische Abweichung der erwarteten Änderung der AV pro UV-Änderung um 1 Einheit
$\sigma_b^2$	Varianz der $b_{1i}$ zwischen Zentren
$\sigma_\varepsilon^2$	Varianz der Datenpunkte innerhalb der Zentren

## (2) Random-Slope-Only-Modell

Designmatrixform für  $k = 4, n_1 = n_2 = n_3 = n_4 = 5, n = 20$

$$y = X\beta + Zb + \varepsilon \Leftrightarrow \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{25} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \\ y_{35} \\ y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \\ y_{45} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{13} \\ 1 & x_{14} \\ 1 & x_{15} \\ 1 & x_{21} \\ 1 & x_{22} \\ 1 & x_{23} \\ 1 & x_{24} \\ 1 & x_{25} \\ 1 & x_{31} \\ 1 & x_{32} \\ 1 & x_{33} \\ 1 & x_{34} \\ 1 & x_{35} \\ 1 & x_{41} \\ 1 & x_{42} \\ 1 & x_{43} \\ 1 & x_{44} \\ 1 & x_{45} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} x_{11} & 0 & 0 & 0 \\ x_{12} & 0 & 0 & 0 \\ x_{13} & 0 & 0 & 0 \\ x_{14} & 0 & 0 & 0 \\ x_{15} & 0 & 0 & 0 \\ 0 & x_{21} & 0 & 0 \\ 0 & x_{22} & 0 & 0 \\ 0 & x_{23} & 0 & 0 \\ 0 & x_{24} & 0 & 0 \\ 0 & x_{25} & 0 & 0 \\ 0 & 0 & x_{31} & 0 \\ 0 & 0 & x_{32} & 0 \\ 0 & 0 & x_{33} & 0 \\ 0 & 0 & x_{34} & 0 \\ 0 & 0 & x_{35} & 0 \\ 0 & 0 & 0 & x_{41} \\ 0 & 0 & 0 & x_{42} \\ 0 & 0 & 0 & x_{43} \\ 0 & 0 & 0 & x_{44} \\ 0 & 0 & 0 & x_{45} \end{pmatrix} \begin{pmatrix} b_{11} \\ b_{12} \\ b_{13} \\ b_{14} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{25} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \\ \varepsilon_{35} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{43} \\ \varepsilon_{44} \\ \varepsilon_{45} \end{pmatrix}$$

mit

$$b \sim N(0_4, \sigma_b^2 I_4) \text{ und } \varepsilon \sim N(0_{20}, \sigma_\varepsilon^2 I_{20}) \quad (8)$$

## (2) Random-Slope-Only-Modell

```
library(nlme)
D = read.csv("5-Daten/mz-regression-1.csv")
D$HSA = as.factor(D$HSA)
M = lme(BDI ~ ACT, D, random = ~ ACT - 1 | HSA)
X = model.matrix(M, D)
Z = kronecker(diag(k), matrix(c(1:5), nrow = 5))
beta_hat = M$coefficients$fixed
b_hat = M$coefficients$random$HSA
s_eps_hat = M$sigma**2
s_b_hat = diag(getVarCov(M))
```

# R Paket  
# Einlesen des Datensatzes  
# Umwandlung numerischer Werte in R-Faktor  
# Random-Slope-Only-LMM  
# Fixed-Effects-Designmatrix  
# Random-Effects-Designmatrix  
# Fixed-Effects-Parameterschätzer  
# Random-Effects-Parameterschätzer  
# Varianzkomponentenschätzer  
# Varianzkomponentenschätzer

## (2) Random-Slope-Only-Modell

Fixed-Effects-Designmatrix

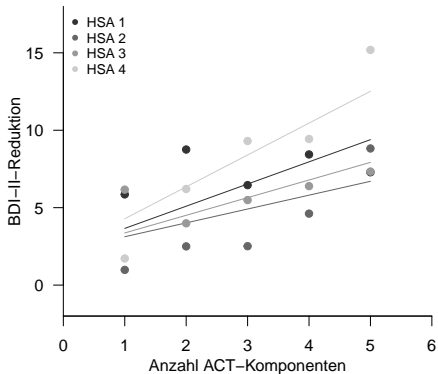
```
(Intercept) ACT
1           1   1
2           1   2
3           1   3
4           1   4
5           1   5
6           1   1
7           1   2
8           1   3
9           1   4
10          1   5
11          1   1
12          1   2
13          1   3
14          1   4
15          1   5
16          1   1
17          1   2
18          1   3
19          1   4
20          1   5
attr(,"assign")
[1] 0 1
```

## (2) Random-Slope-Only-Modell

### Random-Effects-Designmatrix

	[,1]	[,2]	[,3]	[,4]
[1,]	1	0	0	0
[2,]	2	0	0	0
[3,]	3	0	0	0
[4,]	4	0	0	0
[5,]	5	0	0	0
[6,]	0	1	0	0
[7,]	0	2	0	0
[8,]	0	3	0	0
[9,]	0	4	0	0
[10,]	0	5	0	0
[11,]	0	0	1	0
[12,]	0	0	2	0
[13,]	0	0	3	0
[14,]	0	0	4	0
[15,]	0	0	5	0
[16,]	0	0	0	1
[17,]	0	0	0	2
[18,]	0	0	0	3
[19,]	0	0	0	4
[20,]	0	0	0	5

## (2) Random-Slope-Only-Modell



## (2) Random-Slope-Only-Modell

```
beta_hat      :  2.23  1.38
sigsqr_eps_hat :  4.21
sigsqr_b_hat   :  0.31
```

b\_hat:

```
          ACT
1  0.05148833
2 -0.48655804
3 -0.24178981
4  0.67685951
```

Identisches Intercept  $\hat{\beta}_0$  der Ausgleichsgerade über Zentren

Variable Steigung  $\hat{\beta}_1 + \hat{b}_{1i}$  der Ausgleichsgerade über Zentren

## (3) Random-Intercept-and-Slope-Modell

Strukturelle Form

Für Zentren  $i = 1, \dots, k$  und Proband:innen  $j = 1, \dots, n_i$

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{0i} + b_{1i} x_{ij} + \varepsilon_{ij} \quad (9)$$

mit

- $x_{ij}$  := Wert der unabhängigen Variable für Proband:in  $j$  in Zentrum  $i$
- $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$
- $b_{0i} \sim N(0, \sigma_{b_0}^2)$
- $b_{1i} \sim N(0, \sigma_{b_1}^2)$

Parameterbedeutungen

$\beta_0$	Erwartungswert der AV bei Wert der UV gleich 0
$\beta_1$	Erwartungswert der Änderung der AV pro UV-Änderung um 1 Einheit
$b_{0i}$	Zentren-spezifische Abweichung der AV bei Wert der UV gleich 0
$b_{1i}$	Zentren-spezifische Abweichung der erwarteten Änderung der AV pro UV-Änderung um 1 Einheit
$\sigma_{b_0}^2$	Varianz der $b_{0i}$ zwischen Zentren
$\sigma_{b_1}^2$	Varianz der $b_{1i}$ zwischen Zentren
$\sigma_\varepsilon^2$	Varianz der Datenpunkte innerhalb der Zentren

## (3) Random-Intercept-and-Slope-Modell

Designmatrixform für  $k = 4, n_1 = n_2 = n_3 = n_4 = 5, n = 20$

$$y = X\beta + Zb + \varepsilon \Leftrightarrow \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{25} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \\ y_{35} \\ y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \\ y_{45} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{13} \\ 1 & x_{14} \\ 1 & x_{15} \\ 1 & x_{21} \\ 1 & x_{22} \\ 1 & x_{23} \\ 1 & x_{24} \\ 1 & x_{25} \\ 1 & x_{31} \\ 1 & x_{32} \\ 1 & x_{33} \\ 1 & x_{34} \\ 1 & x_{35} \\ 1 & x_{41} \\ 1 & x_{42} \\ 1 & x_{43} \\ 1 & x_{44} \\ 1 & x_{45} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 1 & x_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & x_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & x_{13} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & x_{14} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & x_{15} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{21} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{23} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{24} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{25} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{31} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{32} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{33} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{34} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{35} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & x_{41} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & x_{42} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & x_{43} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & x_{44} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & x_{45} \end{pmatrix} \begin{pmatrix} b_{01} \\ b_{02} \\ b_{12} \\ b_{03} \\ b_{13} \\ b_{04} \\ b_{14} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{25} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \\ \varepsilon_{35} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{43} \\ \varepsilon_{44} \\ \varepsilon_{45} \end{pmatrix}$$

mit

$$b \sim N(0_8, \Sigma_b) \text{ und } \varepsilon \sim N(0_{20}, \sigma_\varepsilon^2 I_{20}) \quad (10)$$

## (3) Random-Intercept-and-Slope-Modell

```
library(nlme) # R Paket
D             = read.csv("5-Daten/mz-regression-1.csv") # Einlesen des Datensatzes
D$HSA       = as.factor(D$HSA) # Umwandlung numerischer Werte in R-Faktor
M           = lme(BDI ~ ACT, D, random = ~ ACT | HSA) # Random-Intercept-and-Slope LMM
X           = model.matrix(M, D) # Fixed-Effects-Designmatrix
Z           = kronecker(diag(k), matrix(c(rep(1,5), c(1:5)), nrow = 5)) # Random-Effects-Designmatrix
beta_hat    = M$coefficients$fixed # Fixed-Effects-Parameterschätzer
b_hat      = M$coefficients$random$HSA # Random-Effects-Parameterschätzer
s_eps_hat  = M$sigma**2 # Varianzkomponentenschätzer
s_b_hat    = diag(getVarCov(M)) # Varianzkomponentenschätzer
```

## (3) Random-Intercept-and-Slope-Modell

Fixed-Effects-Designmatrix

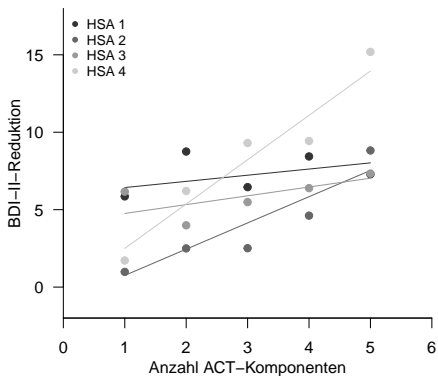
```
(Intercept) ACT
1           1   1
2           1   2
3           1   3
4           1   4
5           1   5
6           1   1
7           1   2
8           1   3
9           1   4
10          1   5
11          1   1
12          1   2
13          1   3
14          1   4
15          1   5
16          1   1
17          1   2
18          1   3
19          1   4
20          1   5
attr(,"assign")
[1] 0 1
```

## (3) Random-Intercept-and-Slope-Modell

### Random-Effects-Designmatrix

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]
[1,]	1	1	0	0	0	0	0	0
[2,]	1	2	0	0	0	0	0	0
[3,]	1	3	0	0	0	0	0	0
[4,]	1	4	0	0	0	0	0	0
[5,]	1	5	0	0	0	0	0	0
[6,]	0	0	1	1	0	0	0	0
[7,]	0	0	1	2	0	0	0	0
[8,]	0	0	1	3	0	0	0	0
[9,]	0	0	1	4	0	0	0	0
[10,]	0	0	1	5	0	0	0	0
[11,]	0	0	0	0	1	1	0	0
[12,]	0	0	0	0	1	2	0	0
[13,]	0	0	0	0	1	3	0	0
[14,]	0	0	0	0	1	4	0	0
[15,]	0	0	0	0	1	5	0	0
[16,]	0	0	0	0	0	0	1	1
[17,]	0	0	0	0	0	0	1	2
[18,]	0	0	0	0	0	0	1	3
[19,]	0	0	0	0	0	0	1	4
[20,]	0	0	0	0	0	0	1	5

## (3) Random-Intercept-and-Slope-Modell



## (3) Random-Intercept-and-Slope-Modell

```
beta_hat      :  2.23  1.38
sigsqr_eps_hat :  1.78
sigsqr_b_hat  : 13.39  1.46
```

b\_hat:

	(Intercept)	ACT
1	3.801795	-0.9829715
2	-3.168384	0.3132948
3	1.958989	-0.8124797
4	-2.592400	1.4821563

Variabler Intercept  $\hat{\beta}_0 + \hat{b}_{0i}$  der Ausgleichsgerade über Zentren

Variable Steigung  $\hat{\beta}_1 + \hat{b}_{1i}$  der Ausgleichsgerade über Zentren

Aber: Populationsparameterschätzung von  $\beta_0$  und  $\beta_1$

---

Motivation

Multizentren-Parallelgruppendesigns

Multizentren-Regressionsdesigns

**Ein Multizentren-Simpson's Paradox-Beispiel**

Selbstkontrollfragen

## Anwendungsbeispiel

- Dosis-Wirkungs-Studie zum Effekt von ACT-Komponentenanzahl auf Depressionssymptomatik
- Multizentrendesign mit  $k = 4$  Hochschulambulanzen mit jeweils  $n_i = 5$  Patient:innen
- Zufällige Zusammenstellung der ACT-Komponentenanzahl in jeder Hochschulambulanz
- Pre-Post-BDI-II-Differenz BDI als primäres Ergebnismaß (Positive Werte = Reduktion)
- HSA: Hochschulambulanz, ACT: ACT-Komponentenanzahl, BDI: Pre-Post-BDI-II-Differenz

# Ein Multizentren-Simpson's Paradox-Beispiel

## Datengeneration

```
# Datengeneration
library(MASS)
library(Matrix)
set.seed(0)
k = 4
n_i = 5
n = k*n_i
x = matrix(rep(NaN,n), ncol = k)
x[,1] = round(runif(n_i, min = 10, max = 20))
x[,2] = round(runif(n_i, min = 15, max = 25))
x[,3] = round(runif(n_i, min = 20, max = 30))
x[,4] = round(runif(n_i, min = 25, max = 35))
Xi = array(rep(NaN,n*2), dim = c(n_i,2,k))
for(i in 1:k){
  Xi[, ,i] = matrix(c(rep(1,n_i),x[,i]), ncol = 2)
}
X = rbind(Xi[, ,1],Xi[, ,2],Xi[, ,3],Xi[, ,4])
beta = matrix(c(5,1), nrow = 2)
Z = as.matrix(bdiag(Xi[, ,1],Xi[, ,2],Xi[, ,3],Xi[, ,4]))
b = matrix(c(15,0,5,0,-5,0,-15,0))
s_eps = 15
eps = mvrnorm(1,rep(0,n), s_eps*diag(n))
y = X %>% beta + Z %>% b + eps
HSA = kronecker(c(1,2,3,4), rep(1,n_i))
D = data.frame(HSA = HSA, ACT = X[,2], BDI = y)
write.csv(D, "5-Daten/mz-regression-2.csv", row.names = FALSE)
```

```
# multivariate Normalverteilung
# Blockdiagonalmatrizen
# Zufallszahlengeneratorzustand
# Anzahl Zentren
# Anzahl Patient:innen pro Zentrum
# Gesamtanzahl an Patient:innen
# Anzahl ACT-Komponenten Array
# Anzahl ACT-Komponenten HSA 1
# Anzahl ACT-Komponenten HSA 2
# Anzahl ACT-Komponenten HSA 3
# Anzahl ACT-Komponenten HSA 4
# Zentren-spezifische Regressionsmatrizenarray
# Zentreniterationen
# Zentren-spezifische Regressionmatrix
# Fixed-Effects-Designmatrix
# Fixed-Effects-Parameter
# Random-Effects-Designmatrix
# Random-Effects-Parameter
# Varianzkomponente
# Fehlervektor
# Datengeneration
# Hochschulambulanzfaktor
# Dataframe
# Speichern
```

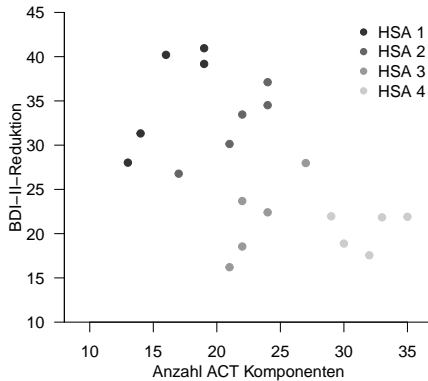
# Ein Multizentren-Simpson's Paradox-Beispiel

## Datensatz

HSA	ACT	BDI
1	19	39
1	13	28
1	14	31
1	16	40
1	19	41
2	17	27
2	24	37
2	24	35
2	22	33
2	21	30
3	21	16
3	22	24
3	22	19
3	27	28
3	24	22
4	33	22
4	30	19
4	32	18
4	35	22
4	29	22

# Ein Multizentren-Simpson's Paradox-Beispiel

## Datendeskription



# Ein Multizentren-Simpson's Paradox-Beispiel

---

Wir wenden auf diesen Datensatz nacheinander folgende Modelle an

- (1) Einfache lineare Regression für den ACT-Effekt
- (2) Einfaktorielle Varianzanalyse für den HSA-Effekt
- (3) Gruppenspezifische einfache lineare Regressionsmodelle für die ACT-Effekte
- (4) Multizentren-Regressionsmodell mit Random-Intercept-Only
- (5) Multizentren-Regressionsmodell mit Random-Intercept-and-Slope

Insbesondere das erste und das letzte Modell bieten unterschiedliche Interpretationen

Dieser inhärente Widerspruch ist ein Beispiel für das Simpson's Paradox (Blyth (1972))

# Ein Multizentren-Simpson's Paradox-Beispiel

## (1) Einfache lineare Regression für den ACT-Effekt

Wir vernachlässigen zunächst die Gruppenstruktur (Hochschulambulanzvariation) des Datensatzes.

Strukturelle Form

Für Proband:innen  $i = 1, \dots, n$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (11)$$

mit

- $x_i :=$  Anzahl der ACT-Komponenten in der Therapie von Proband:in  $i$
- $\varepsilon_i \sim N(0, \sigma^2)$

Parameterbedeutungen

- |            |   |
|------------|---|
| $\beta_0$  | Erwartungswert der BDI-II-Reduktion bei 0 ACT-Komponenten |
| $\beta_1$  | Erwartungswert der BDI-II-Reduktion pro 1 ACT-Komponente  |
| $\sigma^2$ | Varianz zwischen Proband:innen über Zentren               |

# Ein Multizentren-Simpson's Paradox-Beispiel

## (1) Einfache lineare Regression für den ACT-Effekt

Designmatrixform

$$y = X\beta + \varepsilon \Leftrightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{16} \\ y_{17} \\ y_{18} \\ y_{19} \\ y_{20} \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & x_5 \\ 1 & x_6 \\ 1 & x_7 \\ 1 & x_8 \\ 1 & x_9 \\ 1 & x_{10} \\ 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{13} \\ 1 & x_{14} \\ 1 & x_{15} \\ 1 & x_{16} \\ 1 & x_{17} \\ 1 & x_{18} \\ 1 & x_{19} \\ 1 & x_{20} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \\ \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \\ \varepsilon_{16} \\ \varepsilon_{17} \\ \varepsilon_{18} \\ \varepsilon_{19} \\ \varepsilon_{20} \end{pmatrix} \quad \text{mit } \varepsilon \sim N(0_{20}, \sigma^2 I_{20}) \quad (12)$$

# Ein Multizentren-Simpson's Paradox-Beispiel

## (1) Einfache lineare Regression für den ACT-Effekt

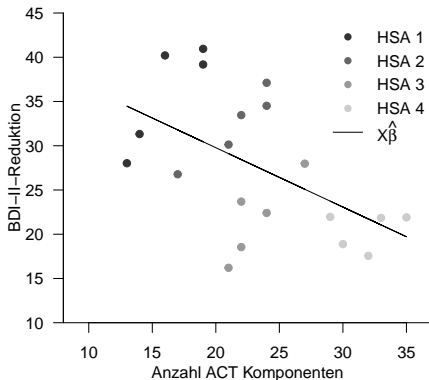
```
D = read.csv("5-Daten/mz-regression-2.csv") # Einlesen des Datensatzes
M = lm(BDI ~ ACT, D) # Einfache lineare Regression
S = summary(M) # Modellschätzungsergebnisse
X = model.matrix(M) # Designmatrix
beta_hat = as.matrix(M$coefficients) # Betaparameterschätzer
sigsqr_hat = S$sigma**2 # Varianzparameterschätzer
```

Designmatrix

```
(Intercept) ACT
1 1 19
2 1 13
3 1 14
4 1 16
5 1 19
6 1 17
7 1 24
8 1 24
9 1 22
10 1 21
11 1 21
12 1 22
13 1 22
14 1 27
15 1 24
16 1 33
17 1 30
18 1 32
19 1 35
20 1 29
attr(,"assign")
[1] 0 1
```

# Ein Multizentren-Simpson's Paradox-Beispiel

## (1) Einfache lineare Regression für den ACT-Effekt



# Ein Multizentren-Simpson's Paradox-Beispiel

---

## (1) Einfache lineare Regression für den ACT-Effekt

beta\_hat : 43.2 -0.67

sigsqr\_hat : 47.88

## Diskussion

- Keine Berücksichtigung der Gruppenstruktur (Hochschulambulanzvariation)
- $\hat{\beta}_1$  legt eine Abnahme der BDI-II-Reduktion mit ACT-Komponentenanzahl nahe
- Daten zeigen aber Anstieg der BDI-II-Reduktion mit ACT-Komponentenanzahl für jede HSA
- Das Modell ist für den betrachteten Anwendungsfall sicherlich nicht ideal

# Ein Multizentren-Simpson's Paradox-Beispiel

## (2) Einfaktorielle Varianzanalyse für den HSA-Effekt

Wir vernachlässigen den Effekt der ACT-Komponentenanzahl und betrachten lediglich den HSA-Effekt.

Strukturelle Form

Für Proband:innen  $j = 1, \dots, n$

$$y_j = \beta_0 + \beta_2 x_{2j} + \beta_3 x_{3j} + \beta_4 x_{4j} + \varepsilon_j \quad (13)$$

mit

- $x_{ij} := 1$  für Proband:in  $j$  in HSA  $i$  und  $x_{ij} := 0$  sonst für  $i = 2, 3, 4$
- $\varepsilon_j \sim N(0, \sigma^2)$

Parameterbedeutungen

$\beta_0$	Erwartungswert für die BDI-II-Reduktion in der Referenzgruppe HSA 1
$\beta_2$	Erwartungswertdifferenz für die BDI-II-Reduktion zwischen HSA 2 und HSA 1
$\beta_3$	Erwartungswertdifferenz für die BDI-II-Reduktion zwischen HSA 3 und HSA 1
$\beta_4$	Erwartungswertdifferenz für die BDI-II-Reduktion zwischen HSA 4 und HSA 1
$\sigma^2$	Varianz zwischen Proband:innen

# Ein Multizentren-Simpson's Paradox-Beispiel

## (2) Einfaktorielle Varianzanalyse für den HSA-Effekt

Designmatrixform

$$y = X\beta + \varepsilon \Leftrightarrow \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{25} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \\ y_{35} \\ y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \\ y_{45} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{25} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \\ \varepsilon_{35} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{43} \\ \varepsilon_{44} \\ \varepsilon_{45} \end{pmatrix} \quad \text{mit } \varepsilon \sim N(0_{20}, \sigma^2 I_{20}) \quad (14)$$

# Ein Multizentren-Simpson's Paradox-Beispiel

## (2) Einfaktorielle Varianzanalyse für den HSA-Effekt

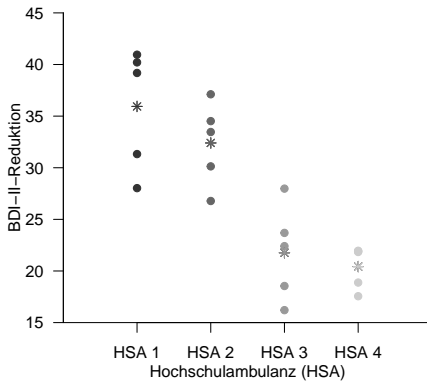
```
D = read.csv("5-Daten/mz-regression-2.csv") # Einlesen des Datensatzes
D$HSA = as.factor(D$HSA) # Umwandlung numerischer Werte in R-Faktor
M = lm(BDI ~ HSA, D) # Einfache lineare Regression
S = summary(M) # Modellschätzungsresultate
X = model.matrix(M) # Designmatrix
beta_hat = as.matrix(M$coefficients) # Betaparameterschätzer
sigsqr_hat = S$sigma**2 # Varianzparameterschätzer
```

Designmatrix

```
(Intercept) HSA2 HSA3 HSA4
1 1 0 0 0
2 1 0 0 0
3 1 0 0 0
4 1 0 0 0
5 1 0 0 0
6 1 1 0 0
7 1 1 0 0
8 1 1 0 0
9 1 1 0 0
10 1 1 0 0
11 1 0 1 0
12 1 0 1 0
13 1 0 1 0
14 1 0 1 0
15 1 0 1 0
16 1 0 0 1
17 1 0 0 1
18 1 0 0 1
19 1 0 0 1
20 1 0 0 1
attr(,"assign")
[1] 0 1 1 1
attr(,"contrasts")
attr(,"contrasts")$HSA
[1] "contr.treatment"
```

# Ein Multizentren-Simpson's Paradox-Beispiel

## (2) Einfaktorielle Varianzanalyse für den HSA-Effekt



(2) Einfaktorielle Varianzanalyse für den HSA-Effekt

beta\_hat : 35.94 -3.54 -14.17 -15.51

sigsqr\_hat : 18.97

Diskussion

- Keine Berücksichtigung der Anzahl an ACT-Komponenten
- Das Modell ist vor dem Hintergrund der Forschungsfrage nicht sinnvoll
- HSA 2, HSA 3 und HSA 4 haben eine geringere BDI-II-Reduktionserwartung als HSA 1

# Ein Multizentren-Simpson's Paradox-Beispiel

## (3) Gruppenspezifische einfache lineare Regressionsmodelle für die ACT-Effekte

Wir betrachten jede der Hochschulambulanzen einzeln.

Strukturelle Form

Für Hochschulambulanzen  $i = 1, \dots, k$  und Proband:innen  $j = 1, \dots, n_i$

$$y_{ij} = \beta_{i0} + \beta_{i1}x_{ij} + \varepsilon_{ij} \quad (15)$$

mit

- $x_{ij}$  := Anzahl der ACT-Komponenten in der Therapie von Proband:in  $j$  in Hochschulambulanz  $i$
- $\varepsilon_{ij} \sim N(0, \sigma_i^2)$

Parameterbedeutungen für  $i = 1, \dots, k$

- $\beta_{i0}$  Erwartungswert der BDI-II-Reduktion bei 0 ACT-Komponenten in Hochschulambulanz  $i$
- $\beta_{i1}$  Erwartungswert der BDI-II-Reduktion pro 1 ACT-Komponente in Hochschulambulanz  $i$
- $\sigma_i^2$  Varianz zwischen Proband:innen in Hochschulambulanz  $i$

# Ein Multizentren-Simpson's Paradox-Beispiel

## (3) Gruppenspezifische einfache lineare Regressionsmodelle für die ACT-Effekte

Designmatrixform

$$y_1 = X_1\beta_1 + \varepsilon_1 \Leftrightarrow \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{13} \\ 1 & x_{14} \\ 1 & x_{15} \end{pmatrix} \begin{pmatrix} \beta_{10} \\ \beta_{11} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \end{pmatrix} \quad \text{mit } \varepsilon_1 \sim N(0_5, \sigma_1^2 I_5) \quad (16)$$

$$y_2 = X_2\beta_2 + \varepsilon_2 \Leftrightarrow \begin{pmatrix} y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{25} \end{pmatrix} = \begin{pmatrix} 1 & x_{21} \\ 1 & x_{22} \\ 1 & x_{23} \\ 1 & x_{24} \\ 1 & x_{25} \end{pmatrix} \begin{pmatrix} \beta_{20} \\ \beta_{21} \end{pmatrix} + \begin{pmatrix} \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{25} \end{pmatrix} \quad \text{mit } \varepsilon_2 \sim N(0_5, \sigma_2^2 I_5) \quad (17)$$

$$y_3 = X_3\beta_3 + \varepsilon_3 \Leftrightarrow \begin{pmatrix} y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \\ y_{35} \end{pmatrix} = \begin{pmatrix} 1 & x_{31} \\ 1 & x_{32} \\ 1 & x_{33} \\ 1 & x_{34} \\ 1 & x_{35} \end{pmatrix} \begin{pmatrix} \beta_{30} \\ \beta_{31} \end{pmatrix} + \begin{pmatrix} \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \\ \varepsilon_{35} \end{pmatrix} \quad \text{mit } \varepsilon_3 \sim N(0_5, \sigma_3^2 I_5) \quad (18)$$

$$y_4 = X_4\beta_4 + \varepsilon_4 \Leftrightarrow \begin{pmatrix} y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \\ y_{45} \end{pmatrix} = \begin{pmatrix} 1 & x_{41} \\ 1 & x_{42} \\ 1 & x_{43} \\ 1 & x_{44} \\ 1 & x_{45} \end{pmatrix} \begin{pmatrix} \beta_{40} \\ \beta_{41} \end{pmatrix} + \begin{pmatrix} \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{43} \\ \varepsilon_{44} \\ \varepsilon_{45} \end{pmatrix} \quad \text{mit } \varepsilon_4 \sim N(0_5, \sigma_4^2 I_5) \quad (19)$$

# Ein Multizentren-Simpson's Paradox-Beispiel

## (3) Gruppenspezifische einfache lineare Regressionsmodelle für die ACT-Effekte

```
library(nlme)
D           = read.csv("5-Daten/mz-regression-2.csv")      # Einlesen des Datensatzes
M           = lmList(BDI ~ ACT | HSA, D)                  # gruppenspezifische ELRs
S           = list()                                     # Liste für Modellschätzungsresultate
X           = list()                                     # Liste für Designmatrizen
beta_hat    = list()                                     # Liste für beta_hat
sigsqr_hat  = list()                                     # Liste für Varianzparameterschätzer
for(i in 1:length(M)){                                  # Gruppeniterationen
  S[[i]]     = summary(M[[i]])                           # Modellschätzungsresultate
  X[[i]]     = model.matrix(M[[i]])                      # Designmatrix
  beta_hat[[i]] = as.matrix(M[[i]]$coefficients)        # Betaparameterschätzer
  sigsqr_hat[[i]] = S[[i]]$sigma*2}                      # Varianzparameterschätzer
```

BDI ~ ACT | HSA ist eine erweiterte R-Formel, die BDI und ACT anhand der Level des Faktors HSA partitioniert!

# Ein Multizentren-Simpson's Paradox-Beispiel

## (3) Gruppenspezifische einfache lineare Regressionsmodelle für die ACT-Effekte

### Designmatrizen

```
(Intercept) ACT
1          1  19
2          1  13
3          1  14
4          1  16
5          1  19
attr(,"assign")
[1] 0 1

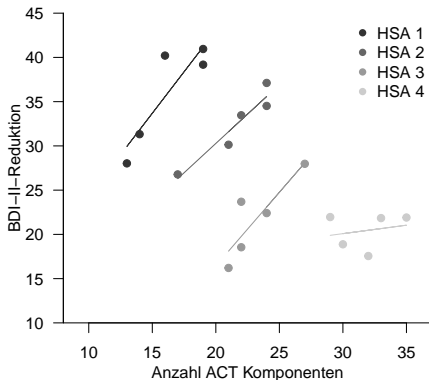
(Intercept) ACT
6          1  17
7          1  24
8          1  24
9          1  22
10         1  21
attr(,"assign")
[1] 0 1

(Intercept) ACT
11         1  21
12         1  22
13         1  22
14         1  27
15         1  24
attr(,"assign")
[1] 0 1

(Intercept) ACT
16         1  33
17         1  30
18         1  32
19         1  35
20         1  29
attr(,"assign")
[1] 0 1
```

# Ein Multizentren-Simpson's Paradox-Beispiel

## (3) Gruppenspezifische einfache lineare Regressionsmodelle für die ACT-Effekte



# Ein Multizentren-Simpson's Paradox-Beispiel

## (3) Gruppenspezifische einfache lineare Regressionsmodelle für die ACT-Effekte

HSA 1	beta_hat	:	5.57	1.87	sigsqr_hat:	9.84
HSA 2	beta_hat	:	3.75	1.33	sigsqr_hat:	2.05
HSA 3	beta_hat	:	-16.8	1.66	sigsqr_hat:	6.98
HSA 4	beta_hat	:	14.35	0.19	sigsqr_hat:	5.45

## Diskussion

- Die BDI-II-Reduktion in jeder HSA mit steigender ACT-Komponentenanzahl wird deutlich
- Dies entspricht der Intuition bei HSA-spezifischer Dateninspektion.
- Es wird allerdings kein HSA-übergreifender Effekt der ACT-Komponentenanzahl bestimmt.
- Es wird auch kein Maß für die Varianz der Effekte zwischen den Hochschulambulanzen bestimmt.

# Ein Multizentren-Simpson's Paradox-Beispiel

## (4) Multizentren-Regressionsmodell mit Random-Intercept-Only

Strukturelle Form

Für Hochschulambulanzen  $i = 1, 2, 3, 4$  und Proband:innen  $j = 1, 2, 3, 4, 5$

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{0i} + \varepsilon_{ij} \quad (20)$$

mit

- $x_{ij}$  := Anzahl ACT-Komponenten für Proband:in  $j$  und Hochschulambulanz  $i$
- $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$
- $b_{0i} \sim N(0, \sigma_b^2)$

Parameterbedeutungen

$\beta_0$	Erwartungswert der BDI-II-Reduktion bei 0 ACT-Komponenten
$\beta_1$	Erwartungswert der BDI-II-Reduktion pro 1 ACT-Komponente
$b_{0i}$	HSA-spezifische Abweichung der BDI-II-Reduktion bei 0 ACT-Komponenten
$\sigma_b^2$	Varianz der $b_{0i}$ zwischen Hochschulambulanzen
$\sigma_\varepsilon^2$	Varianz der Datenpunkte innerhalb der Hochschulambulanzen

# Ein Multizentren-Simpson's Paradox-Beispiel

## (4) Multizentren-Regressionsmodell mit Random-Intercept-Only

Designmatrixform

$$y = X\beta + Zb + \varepsilon \Leftrightarrow \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{25} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \\ y_{35} \\ y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \\ y_{45} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{13} \\ 1 & x_{14} \\ 1 & x_{15} \\ 1 & x_{21} \\ 1 & x_{22} \\ 1 & x_{23} \\ 1 & x_{24} \\ 1 & x_{25} \\ 1 & x_{31} \\ 1 & x_{32} \\ 1 & x_{33} \\ 1 & x_{34} \\ 1 & x_{35} \\ 1 & x_{41} \\ 1 & x_{42} \\ 1 & x_{43} \\ 1 & x_{44} \\ 1 & x_{45} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_{01} \\ b_{02} \\ b_{03} \\ b_{04} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{25} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \\ \varepsilon_{35} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{43} \\ \varepsilon_{44} \\ \varepsilon_{45} \end{pmatrix} \quad (21)$$

mit

$$b \sim N(0_4, \sigma_b^2 I_4) \text{ und } \varepsilon \sim N(0_{20}, \sigma_\varepsilon^2 I_{20}) \quad (22)$$

## (4) Multizentren-Regressionsmodell mit Random-Intercept-Only

```
library(nlme) # R Paket
D = read.csv("5-Daten/mz-regression-2.csv") # Einlesen des Datensatzes
D$HSA = as.factor(D$HSA) # Umwandlung numerischer Werte in R-Faktor
M = lme(BDI ~ ACT, D, random = ~ 1 | HSA) # LMM vom Typ der einfachen linearen Regression
X = model.matrix(M, D) # Fixed-Effects-Designmatrix
Z = model.matrix(~ M$groups[[1]] - 1) # Random-Effects-Designmatrix
beta_hat = M$coefficients$fixed # Fixed-Effects-Parameterschätzer
b_hat = M$coefficients$random$HSA # Random-Effects-Parameterschätzer
s_eps_hat = M$sigma**2 # Varianzkomponentenschätzer
s_b_hat = diag(getVarCov(M)) # Varianzkomponentenschätzer
```

# Ein Multizentren-Simpson's Paradox-Beispiel

## (4) Multizentren-Regressionsmodell mit Random-Intercept-Only

Fixed-Effects-Designmatrix

```
(Intercept) ACT
1           1  19
2           1  13
3           1  14
4           1  16
5           1  19
6           1  17
7           1  24
8           1  24
9           1  22
10          1  21
11          1  21
12          1  22
13          1  22
14          1  27
15          1  24
16          1  33
17          1  30
18          1  32
19          1  35
20          1  29
attr(,"assign")
[1] 0 1
```

# Ein Multizentren-Simpson's Paradox-Beispiel

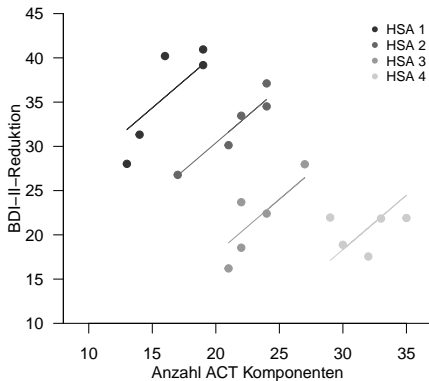
## (4) Multizentren-Regressionsmodell mit Random-Intercept-Only

Random-Effects-Designmatrix

```
M$groups[[1]]1 M$groups[[1]]2 M$groups[[1]]3 M$groups[[1]]4
1          1          0          0          0
2          1          0          0          0
3          1          0          0          0
4          1          0          0          0
5          1          0          0          0
6          0          1          0          0
7          0          1          0          0
8          0          1          0          0
9          0          1          0          0
10         0          1          0          0
11         0          0          1          0
12         0          0          1          0
13         0          0          1          0
14         0          0          1          0
15         0          0          1          0
16         0          0          0          1
17         0          0          0          1
18         0          0          0          1
19         0          0          0          1
20         0          0          0          1
attr(,"assign")
[1] 1 1 1 1
attr(,"contrasts")
attr(,"contrasts")$`M$groups[[1]]`
[1] "contr.treatment"
```

# Ein Multizentren-Simpson's Paradox-Beispiel

## (4) Multizentren-Regressionsmodell mit Random-Intercept-Only



# Ein Multizentren-Simpson's Paradox-Beispiel

---

## (4) Multizentren-Regressionsmodell mit Random-Intercept-Only

```
beta_hat      :  -0.87  1.23
sigsqr_eps_hat :  7.65
sigsqr_b_hat   :  228.4
```

b\_hat:

```
(Intercept)
1  16.792256
2   6.689060
3  -5.828815
4 -17.652501
```

### Diskussion

- Die BDI-II-Reduktion über HSAs mit steigender ACT-Komponentenanzahl wird deutlich
- Die Annahme, dass die BDI-II-Reduktionsrate über HSAs gleich ist, mag etwas stark sein.

# Ein Multizentren-Simpson's Paradox-Beispiel

## (5) Multizentren-Regressionsmodell mit Random-Intercept-and-Slope

Strukturelle Form

Für Hochschulambulanzen  $i = 1, 2, 3, 4$  und Proband:innen  $j = 1, 2, 3, 4, 5$

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{0i} + b_{1i} x_{ij} + \varepsilon_{ij} \quad (23)$$

mit

- $x_{ij} :=$  Anzahl der ACT-Komponenten in der Therapie von Proband:in  $j$  in Hochschulambulanz  $i$
- $b_{0i} \sim N(0, \sigma_{b_0}^2)$  und  $b_{1i} \sim N(0, \sigma_{b_1}^2)$
- $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

Parameterbedeutungen

$\beta_0$	Erwartungswert der BDI-II-Reduktion bei 0 ACT-Komponenten
$\beta_1$	Erwartungswert der BDI-II-Reduktion pro 1 ACT-Komponente
$b_{0i}$	HSA-spezifische Abweichung der BDI-II-Reduktion bei 0 ACT-Komponenten
$b_{1i}$	HSA-spezifische Abweichung der BDI-II-Reduktion pro 1 ACT-Komponente
$\sigma_{b_0}^2$	Varianz der $b_{0i}$ zwischen Hochschulambulanzen
$\sigma_{b_1}^2$	Varianz der $b_{1i}$ zwischen Hochschulambulanzen
$\sigma_\varepsilon^2$	Varianz der Datenpunkte innerhalb der Hochschulambulanzen

# Ein Multizentren-Simpson's Paradox-Beispiel

## (5) Multizentren-Regressionsmodell mit Random-Intercept-and-Slope

Designmatrixform

$$y = X\beta + Zb + \varepsilon \Leftrightarrow \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{25} \\ y_{31} \\ y_{32} \\ y_{33} \\ y_{34} \\ y_{35} \\ y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \\ y_{45} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{13} \\ 1 & x_{14} \\ 1 & x_{15} \\ 1 & x_{21} \\ 1 & x_{22} \\ 1 & x_{23} \\ 1 & x_{24} \\ 1 & x_{25} \\ 1 & x_{31} \\ 1 & x_{32} \\ 1 & x_{33} \\ 1 & x_{34} \\ 1 & x_{35} \\ 1 & x_{41} \\ 1 & x_{42} \\ 1 & x_{43} \\ 1 & x_{44} \\ 1 & x_{45} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 1 & x_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & x_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & x_{13} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & x_{14} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & x_{15} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{21} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{23} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{24} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{25} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{31} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{32} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{33} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{34} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{35} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & x_{41} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & x_{42} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & x_{43} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & x_{44} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & x_{45} \end{pmatrix} \begin{pmatrix} b_{01} \\ b_{11} \\ b_{02} \\ b_{12} \\ b_{03} \\ b_{13} \\ b_{04} \\ b_{14} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{15} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{25} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \\ \varepsilon_{35} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{43} \\ \varepsilon_{44} \\ \varepsilon_{45} \end{pmatrix} \quad (24)$$

mit

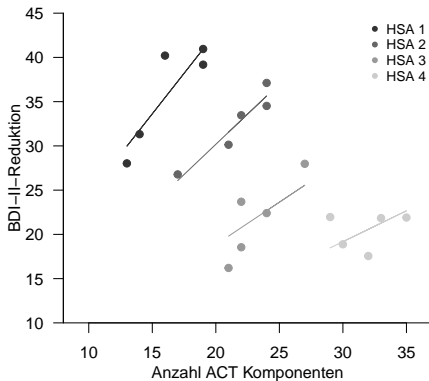
$$b \sim N(0_8, \Sigma_b) \text{ und } \varepsilon \sim N(0_{20}, \sigma_\varepsilon^2 I_{20}) \quad (25)$$

## (5) Multizentren-Regressionsmodell mit Random-Intercept-and-Slope

```
library(nlme) # R Paket
ctrl         = lmeControl(opt = "optim", maxIter = 100) # Anzahl an REML-Iterationen
D            = read.csv("5-Daten/mz-regression-2.csv") # Einlesen des Datensatzes
D$HSA       = as.factor(D$HSA) # Umwandlung numerischer Werte in R-Faktor
M           = lme(BDI ~ ACT, D, random = ~ ACT | HSA, control = ctrl) # LMM einfache lineare Regression
X           = model.matrix(M,D) # Fixed-Effects-Designmatrix
beta_hat    = M$coefficients$fixed # Fixed-Effects-Parameterschätzer
b_hat       = M$coefficients$random # Random-Effects-Parameterschätzer
s_eps_hat   = M$sigma**2 # Varianzkomponentenschätzer
s_b_hat     = diag(getVarCov(M)) # Varianzkomponentenschätzer
```

# Ein Multizentren-Simpson's Paradox-Beispiel

## (5) Multizentren-Regressionsmodell mit Random-Intercept-and-Slope



# Ein Multizentren-Simpson's Paradox-Beispiel

Multizentren-Regressionsmodell mit Random-Intercept-and-Slope

```
beta_hat      : 1.74 1.21
sigsqr_b_hat  : 14.25 0.25
sigsqr_eps_hat : 6.03
```

b\_hat

\$HSA

	(Intercept)	ACT
1	4.438574	0.6167610
2	1.116603	0.1531704
3	-2.062339	-0.2545769
4	-3.492838	-0.5153544

Diskussion

- Die BDI-II-Reduktion über HSAs mit steigender ACT-Komponentenanzahl wird deutlich
- Die Fixed-Effects-Parameter können als Populationsparameter interpretiert werden
- Die HSA-spezifischen Intercepts und Reduktionsraten können kovariieren

---

Motivation

Multizentren-Parallelgruppendesigns

Multizentren-Regressionsdesigns

Ein Multizentren-Simpson's Paradox-Beispiel

**Selbstkontrollfragen**

1. Erläutern Sie die Definition, die Motivation und die Herausforderungen einer Multizentrenstudie.
2. Erläutern Sie das Random-Center-Effect-Modell für Multizentren-Parallelgruppendesigns.
3. Erläutern Sie das Random-Center-by-Treatment-Interaction-Modell für Multizentren-Parallelgruppendesigns.
4. Erläutern Sie das Random-Intercept-Only-Modell für Multizentren-Regressionsdesigns.
5. Erläutern Sie das Random-Slope-Only-Modell für Multizentren-Regressionsdesigns.
6. Erläutern Sie das Random-Intercept-and-Slope-Modell für Multizentren-Regressionsdesigns.

- Blyth, Colin R. 1972. "On Simpson's Paradox and the Sure-Thing Principle." *Journal of the American Statistical Association* 67 (338): 364–66. <https://doi.org/10.1080/01621459.1972.10482387>.
- Brown, Helen, and Robin Prescott. 2015. *Applied Mixed Models in Medicine*. 3rd ed. Wiley.
- Edgar, Kate. 2021. *Including Random Centre Effects in Design, Analysis and Presentation of Multi-Centre Trials*.
- Friedman, Lawrence M., Curt D. Furberg, David L. DeMets, David M. Reboussin, and Christopher B. Granger. 2015. *Fundamentals of Clinical Trials*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-18539-2>.
- Kahan, Brennan C, and Tim P Morris. 2012. *Analysis of Multicentre Trials with Continuous Outcomes: When and How Should We Account for Centre Effects?*
- Li, Yanjuan, Yi Zhang, Chun Wang, et al. 2025. "Supported Mindfulness-Based Self-Help Intervention as an Adjunctive Treatment for Rapid Symptom Change in Emotional Disorders: A Practice-Oriented Multicenter Randomized Controlled Trial." *Psychotherapy and Psychosomatics*, January, 1–11. <https://doi.org/10.1159/000542937>.
- Localio, A. Russell, Jesse A. Berlin, Thomas R. Ten Have, and Stephen E. Kimmel. 2001. "Adjustments for Center in Multicenter Studies: An Overview." *Annals of Internal Medicine* 135 (2): 112–23. <https://doi.org/10.7326/0003-4819-135-2-200107170-00012>.
- Tangri, Navdeep, Georgios D. Kitsios, Shi Hann Su, and David M. Kent. 2010. "Accounting for Center Effects in Multicenter Trials." *Epidemiology* 21 (6): 912–13. <https://doi.org/10.1097/ede.0b013e3181f56fc0>.