



# Allgemeines Lineares Modell

BSc Psychologie, SoSe 2025

Joram Soch

## (9) T-Tests

---

Überblick

Einstichproben-T-Tests

Zweistichproben-T-Tests

Selbstkontrollfragen

---

## Überblick

Einstichproben-T-Tests

Zweistichproben-T-Tests

Selbstkontrollfragen

## Modul B2: Inferenzstatistik | Allgemeines Lineares Modell

Datum	Einheit	Do, 09-13 (ca. 10:00-12:45)	
10.04.2025	Grundlagen	(0) Formalia	(1) Regression
17.04.2025	Grundlagen	(2) Korrelation	(3) Matrizen
24.04.2025	Grundlagen	(3) Matrizen	
01.05.2025	– Feiertag –	– keine Vorlesung –	
08.05.2025	Grundlagen / Theorie	(4) Normalverteilungen	(5) Modellformulierung
15.05.2025	Theorie	(6) Parameterschätzung	
22.05.2025	Theorie	(7) T-Statistiken	(8) F-Statistiken
29.05.2025	– Feiertag –	– keine Vorlesung –	
05.06.2025	Anwendung	(9) T-Tests	
12.06.2025	Anwendung	(10) Einfaktorielle Varianzanalyse	
19.06.2025	Anwendung	(11) Zweifaktorielle Varianzanalyse	
26.06.2025	Anwendung	(12) Partielle Korrelation	
03.07.2025	Anwendung	(13) Multiple Regression	
10.07.2025	Anwendung	(14) Kovarianzanalyse	
Juli 2025	Klausurtermin		
Februar 2026	Klausurwiederholungstermin		

# Überblick

## Kontinuum von ALM-Designs

Extremszenario (1) Die Erwartungswerte aller Datenvariablen sind identisch.

$$\begin{aligned} y_i &\sim N(\mu, \sigma^2) \quad \text{u.i.v. für } i = 1, \dots, n \Leftrightarrow \\ y &= X\beta + \varepsilon, \quad X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \quad \beta := \mu \in \mathbb{R}, \quad \varepsilon \sim N(0_n, \sigma^2 I_n) \end{aligned} \quad (1)$$

⇒ Jegliche Datenvariabilität wird dem Fehlerterm zugeschrieben.

⇒ Es gilt  $\hat{\beta} = (\mathbf{1}_n^T \mathbf{1}_n)^{-1} \mathbf{1}_n^T y = \bar{y}$  und  $\hat{\sigma}^2 = \frac{1}{n-p} (y - \mathbf{1}_n \bar{y})^T (y - \mathbf{1}_n \bar{y}) = s_y^2$ .

Extremszenario (2) Die Erwartungswerte aller Datenvariablen sind paarweise verschieden.

$$\begin{aligned} y_i &\sim N(\mu_i, \sigma^2) \quad \text{u.v. für } i = 1, \dots, n \Leftrightarrow \\ y &= X\beta + \varepsilon, \quad X := I_n \in \mathbb{R}^{n \times n}, \quad \beta := (\mu_1, \dots, \mu_n)^T \in \mathbb{R}^n, \quad \varepsilon \sim N(0_n, \sigma^2 I_n) \end{aligned} \quad (2)$$

⇒ Jegliche Datenvariabilität wird dem Erwartungswertparameter zugeschrieben.

⇒ Es gilt  $\hat{\beta} = (I_n^T I_n)^{-1} I_n^T y = y$  und  $\hat{\sigma}^2 = \frac{1}{n-p} (y - I_n y)^T (y - I_n y) = 0$ .

Beide Extremszenarien sind wissenschaftlich nicht ergiebig, da sie keine theoriegeleitete systematische Abhängigkeit zwischen unabhängigen und abhängigen Variablen (UV, AV) annehmen. Die im weiteren Verlauf betrachteten ALM-Designs liegen zwischen den beiden Extremszenarien und repräsentieren verschiedene Formen der systematischen Abhängigkeit zwischen UV und AV.

## Faktorielle und parametrische ALM-Designs

### Faktorielle ALM-Designs:

- Designmatrizen mit 1en und 0en, manchmal  $-1$ en.
- Betaparameter repräsentieren Erwartungswerte für Gruppen.
- Betaparameterschätzer repräsentieren Stichprobenmittel für Gruppen.
- $\Rightarrow$  T-Tests, einfaktorielle Varianzanalyse, zweifaktorielle Varianzanalyse

### Parametrische ALM-Designs:

- Designmatrizen besitzen Spalten mit kontinuierlichen reellen Werten.
- Die Designmatrixspalten werden *Regressoren*, *Prädiktoren*, oder *Kovariaten* genannt.
- Betaparameter repräsentieren Steigungsparameter.
- Betaparameterschätzer ergeben sich als normalisierte Kovarianzen von Regressoren und Daten.
- Es besteht ein enger Bezug zur Theorie der Korrelation.
- $\Rightarrow$  einfache lineare Regression, multiple lineare Regression

### Faktoriell-parametrische ALM-Designs:

- Designmatrizen mit mehreren faktoriellen und parametrischen Werten.
- Die parametrischen Regressoren werden oft als kontrollierte Kovariaten betrachtet.
- $\Rightarrow$  Kovarianzanalyse, (partielle Korrelation)

## ALM-Designs als Hypothesentestverfahren\*

Testen von Unterschiedshypothesen:

- T-Tests
- einfaktorielle Varianzanalyse
- zweifaktorielle Varianzanalyse
- Kovarianzanalyse

Testen von Zusammenhangshypothesen:

- einfache lineare Regression/Korrelation
- multiple lineare Regression/partielle Korrelation

\* Diese Sichtweise wird durch Prof. Ostwald nicht favorisiert.

## T-Tests

Es gibt viele T-Test-Varianten, jeweils mit eigenen Testgütefunktionen.

Wir fokussieren hier auf die Darstellung von T-Tests als Spezialfällen des ALMs.

Wir behandeln im Detail:

- Einstichproben-T-Tests mit einfacher Nullhypothese und ungerichteter Alternativhypothese.
- Zweistichproben-T-Tests bei unabhängigen Stichproben unter Annahme identischer Varianzen mit einfacher Nullhypothese und ungerichteter Alternativhypothese.

Wir behandeln nicht:

- T-Tests mit gerichteten Hypothesen oder einfachen Null- und Alternativhypothesen.
- Zweistichproben-T-Tests bei Annahme verschiedener Varianzen (Behrens-Fischer-Problem).
- Zweistichproben T-Tests bei abhängigen Stichproben.

---

Überblick

**Einstichproben-T-Tests**

Zweistichproben-T-Tests

Selbstkontrollfragen

---

# Einstichproben-T-Tests

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendung/Praxis

## **Anwendungsszenario**

Modellformulierung

Modellschätzung

Modellevaluation

Anwendung/Praxis

**Eine Gruppe** (Stichprobe) randomisierter experimenteller Einheiten.

Annahme der unabhängigen und identischen Normalverteilung  $N(\mu, \sigma^2)$ .

$\mu$  und  $\sigma^2$  unbekannt.

Quantifizieren der Unsicherheit beim inferentiellen Vergleich von  $\mu$  mit  $\mu_0$  beabsichtigt.

## Anwendungsbeispiele

Pre-Post-Psychotherapie: BDI-Differenzanalyse einer Gruppe von Patient:innen

- $\mu \neq \mu_0 := 0 \Rightarrow$  Evidenz für Depressions symptomatikveränderung

Gruppenanalysen mit *Wechsler Adult Intelligence Scale* (WAIS)

- $\mu \neq \mu_0 := 100 \Rightarrow$  Evidenz für über- oder unterdurchschnittliche WAIS-Performanz

Gruppenanalysen in der funktionellen Kernspintomographie (fMRT)

- $\mu > \mu_0 := 0 \Rightarrow$  Evidenz für regionale Gehirnaktivierung

## Anwendungsbeispiel



Wir nehmen an, dass die Datenpunkte unabhängige und identisch verteilte (u.i.v.) Realisierungen von ZVen  $y_i \sim N(\mu, \sigma^2)$  sind und nehmen weiter an, dass wir sind an der Quantifizierung der Unsicherheit beim inferentiellen Vergleich des wahren, aber unbekanntem Erwartungswertparameters  $\mu$  im Sinne eines Hypothesentests interessiert sind.

Anwendungsszenario

**Modellformulierung**

Modellschätzung

Modellevaluation

Anwendung/Praxis

## Definition (Einstichproben-T-Test-Modell)

$y_i, i = 1, \dots, n$  seien Zufallsvariablen, die die  $n$  Datenpunkte eines Anwendungsszenarios für den Einstichproben-T-Test modellieren. Dann hat das *Einstichproben-T-Test-Modell* die strukturelle Form

$$y_i = \mu + \varepsilon_i \quad \text{mit} \quad \varepsilon_i \sim N(0, \sigma^2) \quad \text{u.i.v. für} \quad i = 1, \dots, n \quad \text{mit} \quad \mu \in \mathbb{R} \quad \text{und} \quad \sigma^2 > 0, \quad (3)$$

die Datenverteilungsform

$$y_i \sim N(\mu, \sigma^2) \quad \text{u.i.v. für} \quad i = 1, \dots, n \quad \text{mit} \quad \mu \in \mathbb{R} \quad \text{und} \quad \sigma^2 > 0, \quad (4)$$

und für den Datenvektor  $y = (y_1, \dots, y_n)^T$  die Designmatrixform

$$y = X\beta + \varepsilon \quad \text{mit} \quad X := \mathbf{1}_n \in \mathbb{R}^{n \times 1}, \quad \beta := \mu \in \mathbb{R}, \quad \varepsilon \sim N(0_n, \sigma^2 I_n), \quad \text{und} \quad \sigma^2 > 0. \quad (5)$$

### Bemerkungen

- Das Modell ist identisch mit dem Modell unabhängiger und identisch normalverteilter Zufallsvariablen.
- Die Äquivalenz der drei Modellformen wurde bereits diskutiert (vgl. Einheit (5) in *Allgemeines Lineares Modell*).
- Die Anzahl der Betaparameter ist  $p = 1$ .

# Modellformulierung

---

Designmatrix des Einstichproben-T-Test-Modells ( $n = 20, p = 1$ )


## Datensimulation (vgl. Einheit (5) in *Allgemeines Lineares Modell*)

```
# Modellformulierung
library(MASS) # multivariate Normalverteilung
n = 20 # Anzahl von Datenpunkten
p = 1 # Anzahl von Betaparametern
X = matrix(rep(1,n), nrow = n) # n x p Designmatrix
I_n = diag(n) # n x n Einheitsmatrix
beta = 5 # wahrer, aber unbekannter Betaparameter
sigsqr = 14 # wahrer, aber unbekannter Varianzparameter

# Datenrealisierung
y = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung des n-dimensionalen ZVs y
print(y)
```

```
> [1] 1.580 7.032 4.803 3.047 -1.971 6.917 -1.103 12.727 7.353
> [10] 2.928 3.942 4.086 4.075 -1.817 9.297 7.326 4.979 -4.119
> [19] 5.955 -0.238
```

Anwendungsszenario

Modellformulierung

**Modellschätzung**

Modellevaluation

Anwendung/Praxis

## Theorem (Parameterschätzung im Einstichproben-T-Test-Modell)

Gegeben sei die Designmatrixform des Einstichproben-T-Test-Modells. Dann ergeben sich für den Betaparameterschätzer

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n y_i =: \bar{y} \quad (6)$$

und für den Varianzparameterschätzer

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 =: s_y^2. \quad (7)$$

### Bemerkungen

- Die Formen von  $\hat{\beta}$  und  $\hat{\sigma}^2$  wurden bereits hergeleitet (siehe Einheit (6) in *Allgemeines Lineares Modell*).
- $\bar{y}$  und  $s_y^2$  bezeichnen das Stichprobenmittel und die Stichprobenvarianz der  $y_1, \dots, y_n$ .

# Modellschätzung

```
# Daten einlesen
fname      = "Daten/T-Tests_Daten.csv"           # Dateiname
D          = read.table(fname, sep = ",", header = TRUE) # Dataframe
y          = D$dBDI[D$Setting == "F2F"]         # BDI-Differenzwerte in der F2F-Gruppe

# Modellformulierung
n          = length(y)                          # Anzahl Datenpunkte
p          = 1                                  # Anzahl Betaparameter
X          = matrix(rep(1,n), nrow = n)        # Designmatrix

# Modellschätzung
beta_hat  = solve(t(X) %*% X) %*% t(X) %*% y   # Betaparameterschätzer
eps_hat   = y - X %*% beta_hat                 # Residuenvektor
sigsqr_hat = (t(eps_hat) %*% eps_hat) / (n-p)  # Varianzparameterschätzer

# Ausgabe
cat( "hat{beta}  : ", beta_hat,                # Betaparameterschätzer
     "\nbar{y}   : ", mean(y),                 # Stichprobenmittel
     "\nhat{sigsqr} : ", sigsqr_hat,           # Varianzparameterschätzer
     "\ns_y^2    : ", var(y))                  # Stichprobenvarianz

> hat{beta}  : 3.92
> bar{y}     : 3.92
> hat{sigsqr} : 19.4
> s_y^2      : 19.4
```

Anwendungsszenario

Modellformulierung

Modellschätzung

**Modellevaluation**

Anwendung/Praxis

## Hypothesenszenarien

Einfache Nullhypothese, einfache Alternativhypothese  $H_0 : \mu = \mu_0$ ,  $H_1 : \mu = \mu_1$

- theoretisch wichtiges Szenario (Neymann-Pearson-Lemma)
- praktische Relevanz eher gering

Einfache Nullhypothese, zusammengesetzte Alternativhypothese  $H_0 : \mu = \mu_0$ ,  $H_1 : \mu \neq \mu_0$

- zweiseitiger Einstichproben-T-Test mit ungerichteter Hypothese
- ungerichtete Fragestellung nach einem Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese  $H_0 : \mu \leq \mu_0$ ,  $H_1 : \mu > \mu_0$

- einseitiger Einstichproben-T-Test mit gerichteter Hypothese
- gerichtete Fragestellung nach einem positiven Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese  $H_0 : \mu \geq \mu_0$ ,  $H_1 : \mu < \mu_0$

- gerichtete Fragestellung nach einem negativen Unterschied
- qualitativ äquivalente Theorie zum umgekehrten Fall

## Hypothesenszenarien

Einfache Nullhypothese, einfache Alternativhypothese  $H_0 : \mu = \mu_0, H_1 : \mu = \mu_1$

- theoretisch wichtiges Szenario (Neymann-Pearson Lemma)
- praktische Relevanz eher gering

Einfache Nullhypothese, zusammengesetzte Alternativhypothese  $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$

- zweiseitiger Einstichproben-T-Test mit ungerichteter Hypothese
- ungerichtete Fragestellung nach einem Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese  $H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0$

- einseitiger Einstichproben-T-Test mit gerichteter Hypothese
- gerichtete Fragestellung nach einem positiven Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese  $H_0 : \mu \geq \mu_0, H_1 : \mu < \mu_0$

- gerichtete Fragestellung nach einem negativen Unterschied
- qualitativ äquivalente Theorie zum umgekehrten Fall

Gliederung (vgl. Einheit (11), Folie 28 in *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*)

- (1) Statistisches Modell ✓
- (2) Testhypothesen ✓
- (3) Teststatistik
- (4) Test
- (5) Analyse der Testgütefunktion
- (6) Testumfangkontrolle
- (7) p-Wert
- (8) Analyse der Powerfunktion

### Bezeichnungskonventionen für $t$ -Zufallsvariablen

- Für eine (nichtzentrale)  $t$ -Zufallsvariable  $\xi$  schreiben wir  $\xi \sim t(n)$  (oder  $\xi \sim t(\delta, n)$ ).
- Die WDF einer (nichtzentralen)  $t$ -Zufallsvariable ist  $t(\cdot; n)$  (oder  $t(\cdot; \delta, n)$ ).
- Die KVF einer (nichtzentralen)  $t$ -Zufallsvariable ist  $\psi(\cdot; n)$  (oder  $\psi(\cdot; \delta, n)$ ).
- Die Teststatistik eines T-Test-Designs/Hypothesentests bezeichnen wir mit  $T$ .
- Die Realisierung der T-Teststatistik für einen Datensatz bezeichnen wir mit  $t$ .

### Theorem (T-Teststatistik des Einstichproben-T-Tests)

Gegeben sei die Designmatrixform des Einstichproben-T-Test-Modells. Dann ergibt sich für die T-Teststatistik mit

$$c := 1 \quad \text{und} \quad \beta_0 =: \mu_0, \quad (8)$$

dass

$$T = \sqrt{n} \left( \frac{\bar{y} - \mu_0}{s_y} \right) \quad (9)$$

und es gilt

$$T \sim t(\delta, n - 1) \quad \text{mit} \quad \delta = \sqrt{n} \left( \frac{\mu - \mu_0}{\sigma} \right). \quad (10)$$

#### Bemerkungen

- Das Theorem basiert auf der T-Statistik im Rahmen des ALM.
- Wir erinnern an das verwandte populäre und von der Stichprobengröße unabhängige Effektstärke-Maß *Cohen's d* bei Einstichproben-T-Test-Designs,

$$d := \frac{\bar{y}}{s_y}. \quad (11)$$

- Offenbar gilt für *Cohen's d*, dass mit  $\mu_0 := 0$

$$T = \sqrt{n}d \quad \Leftrightarrow \quad d = T/\sqrt{n}. \quad (12)$$

### Beweis

Mit dem Theorem zur Verteilung der T-Statistik (siehe Einheit (7) in *Allgemeines Lineares Modell*) gilt

$$T = \frac{c^T \hat{\beta} - c^T \beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} = \frac{1^T \bar{y} - 1^T \mu_0}{\sqrt{s_y^2 1^T (1_n^T 1_n)^{-1} 1}} = \sqrt{n} \left( \frac{\bar{y} - \mu_0}{s_y} \right). \quad (13)$$

Weiterhin gilt mit demselben Theorem

$$\delta = \frac{c^T \beta - c^T \beta_0}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}} = \frac{1^T \mu - 1^T \mu_0}{\sqrt{\sigma^2 1^T (1_n^T 1_n)^{-1} 1}} = \sqrt{n} \left( \frac{\mu - \mu_0}{\sigma} \right). \quad (14)$$

□

### Definition (Zweiseitiger Einstichproben-T-Tests)

Gegeben sei das Einstichproben-T-Test-Modell. Für ein  $\mu_0 \in \mathbb{R}$  seien die einfache Nullhypothese und die zusammengesetzte Alternativhypothese als

$$H_0 : \mu = \mu_0 \quad \text{und} \quad H_1 : \mu \neq \mu_0 \quad (15)$$

definiert. Weiterhin sei die T-Teststatistik definiert als

$$T := \sqrt{n} \left( \frac{\bar{y} - \mu_0}{s_y} \right) \quad (16)$$

Dann ist der *zweiseitige Einstichproben-T-Tests* definiert als der kritische-Wert-basierte Test

$$\phi(y) := 1_{\{|T| \geq k\}} = \begin{cases} 1 & |T| \geq k \\ 0 & |T| < k \end{cases}. \quad (17)$$

#### Bemerkungen

- Ausführlicher handelt es sich um den *zweiseitigen Einstichproben-T-Test mit ungerichteter Hypothese*.

### Theorem (Testgütefunktion)

$\phi$  sei der im obigen Testszenario definierte Test. Dann ist die Testgütefunktion von  $\phi$  gegeben durch

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k; \delta, n - 1) + \psi(-k; \delta, n - 1) \quad (18)$$

wobei  $\psi(\cdot; \delta, n - 1)$  die KVF der nichtzentralen  $t$ -Verteilung mit Nichtzentralitätsparameter

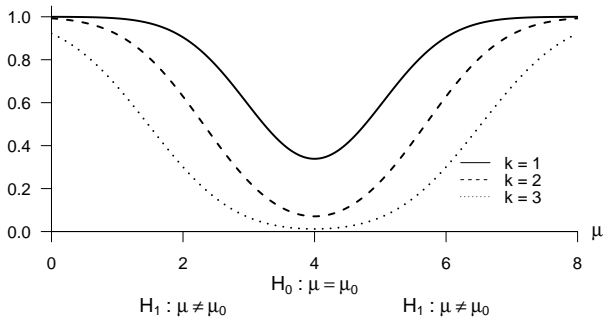
$$\delta := \sqrt{n} \left( \frac{\mu - \mu_0}{\sigma} \right) \quad (19)$$

und Freiheitsgradparameter  $n - 1$  bezeichnet.

## Modellevaluation (5) Analyse der Testgütefunktion

Testgütefunktion  $q_\phi$  für  $\sigma^2 = 9, \mu_0 = 4, n = 12$  und  $k = 1, 2, 3$ .

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi = 1)$$



## Modellevaluation (5) Analyse der Testgütefunktion

### Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testscenario ist definiert als

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \quad (20)$$

Da die Wahrscheinlichkeiten für  $\phi = 1$  und dafür, dass die zugehörige Teststatistik im Ablehnungsbereich des Tests liegt, gleich sind, benötigen wir also zunächst die Verteilung der Teststatistik. Wir haben oben bereits gesehen, dass die T-Teststatistik

$$T := \sqrt{n} \left( \frac{\bar{y} - \mu_0}{s_y} \right) \quad (21)$$

unter der Annahme  $y_i \sim N(\mu, \sigma^2)$  u.i.v. für  $i = 1, \dots, n$  nach einer nichtzentralen  $t$ -Verteilung  $t(\delta, n-1)$  mit Nichtzentralitätsparameter

$$\delta = \sqrt{n} \left( \frac{\mu - \mu_0}{\sigma} \right) \quad (22)$$

verteilt ist. Der Ablehnungsbereich des zweiseitigen T-Tests ergibt sich, wie in ähnlicher Form bei der Betrachtung des zweiseitigen Z-Tests gesehen, zu

$$A = ] -\infty, -k] \cup ]k, \infty[. \quad (23)$$

### Beweis (fortgeführt)

Mit diesem Ablehnungsbereich ergibt sich dann

$$\begin{aligned}q_{\phi}(\mu) &= \mathbb{P}_{\mu}(\phi = 1) \\&= \mathbb{P}_{\mu}(T \in ]-\infty, -k] \cup ]k, \infty[) \\&= \mathbb{P}_{\mu}(T \in ]-\infty, -k]) + \mathbb{P}_{\mu}(T \in [k, \infty[) \\&= \mathbb{P}_{\mu}(T \leq -k) + \mathbb{P}_{\mu}(T \geq k) \\&= \mathbb{P}_{\mu}(T \leq -k) + (1 - \mathbb{P}_{\mu}(T \leq k)) \\&= 1 - \mathbb{P}_{\mu}(T \leq k) + \mathbb{P}_{\mu}(T \leq -k) \\&= 1 - \psi(k; \delta, n - 1) + \psi(-k; \delta, n - 1),\end{aligned}\tag{24}$$

wobei  $\psi(\cdot; \delta, n - 1)$  die KVF der nichtzentralen T-Verteilung mit Nichtzentralitätsparameter  $\delta$  und Freiheitsgradparameter  $n - 1$  bezeichnet.  $\square$

### Theorem (Testumfangkontrolle)

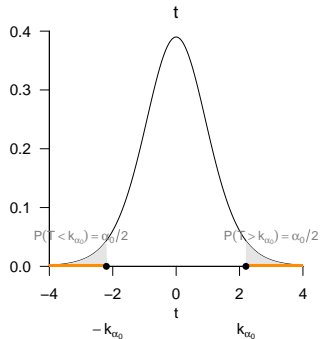
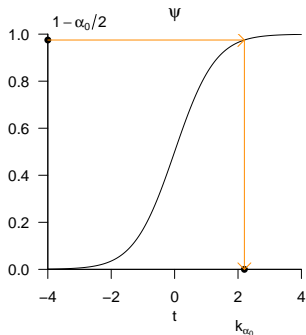
$\phi$  sei der im obigen Testszenario definierte Test. Dann ist  $\phi$  ein Level- $\alpha_0$ -Test mit Testumfang  $\alpha_0$ , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \psi^{-1} \left( 1 - \frac{\alpha_0}{2}; n - 1 \right), \quad (25)$$

wobei  $\psi^{-1}(\cdot; n - 1)$  die inverse KVF der  $t$ -Verteilung mit  $n - 1$  Freiheitsgraden ist.

## Modellevaluation (6) Testumfangkontrolle

Wahl von  $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0/2; n - 1)$  mit  $n = 12$ ,  $\alpha_0 := 0.05$  und Ablehnungsbereich



## Modellevaluation (6) Testumfangkontrolle

### Beweis

Damit der betrachtete Test ein Level- $\alpha_0$ -Test ist, muss bekanntlich  $q_\phi(\mu) \leq \alpha_0$  für alle  $\mu \in \{\mu_0\}$ , also hier  $q_\phi(\mu_0) \leq \alpha_0$ , gelten. Weiterhin ist der Testumfang des betrachteten Tests durch  $\alpha = \max_{\mu \in \{\mu_0\}} q_\phi(\mu)$ , also hier durch  $\alpha = q_\phi(\mu_0)$  gegeben. Wir müssen also zeigen, dass die Wahl von  $k_{\alpha_0}$  garantiert, dass  $\phi$  ein Level- $\alpha_0$ -Test mit Testumfang  $\alpha_0$  ist. Dazu merken wir zunächst an, dass für  $\mu = \mu_0$  gilt, dass

$$\begin{aligned}q_\phi(\mu_0) &= 1 - \psi(k; \delta, n - 1) + \psi(-k; \delta, n - 1) \\ &= 1 - \psi(k; 0, n - 1) + \psi(-k; 0, n - 1) \\ &= 1 - \psi(k; n - 1) + \psi(-k; n - 1),\end{aligned}\tag{26}$$

wobei  $\psi(\cdot; \delta, n - 1)$  und  $\psi(\cdot; n - 1)$  die KVF der nichtzentralen  $t$ -Verteilung mit Nichtzentralitätsparameter  $\delta$  und Freiheitsgradparameter  $n - 1$  bzw. der  $t$ -Verteilung mit Freiheitsgradparameter  $n - 1$  bezeichnen. Mit  $k := k_{\alpha_0}$  gilt dann

$$\begin{aligned}q_\phi(\mu_0) &= 1 - \psi(k_{\alpha_0}; n - 1) + \psi(-k_{\alpha_0}; n - 1) \\ &= 1 - \psi(k_{\alpha_0}; n - 1) + (1 - \psi(k_{\alpha_0}; n - 1)) \\ &= 2(1 - \psi(k_{\alpha_0}; n - 1)) \\ &= 2 \left( 1 - \psi \left( \psi^{-1} (1 - \alpha_0/2; n - 1); n - 1 \right) \right) \\ &= 2(1 - (1 - \alpha_0/2)) \\ &= \alpha_0,\end{aligned}\tag{27}$$

wobei die zweite Gleichung mit der Symmetrie der  $t$ -Verteilung folgt. Es folgt also direkt, dass bei der Wahl von  $k = k_{\alpha_0}$ ,  $q_\phi(\mu_0) \leq \alpha_0$  ist und der betrachtete Test somit ein Level- $\alpha_0$ -Test ist. Weiterhin folgt direkt, dass der Testumfang des betrachteten Tests bei der Wahl von  $k = k_{\alpha_0}$  gleich  $\alpha_0$  ist.  $\square$

## Praktisches Vorgehen

- Man nimmt an, dass ein Datensatz  $y_1, \dots, y_n$  eine Realisation von  $y_i \sim N(\mu, \sigma^2)$  u.i.v. für  $i = 1, \dots, n$  mit unbekanntem Parametern  $\mu$  und  $\sigma^2 > 0$  ist.
- Man möchte entscheiden ob für ein  $\mu_0 \in \mathbb{R}$  eher  $H_0 : \mu = \mu_0$  oder  $H_1 : \mu \neq \mu_0$  zutrifft.
- Man wählt ein Signifikanzniveau  $\alpha_0$  und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert  $k_{\alpha_0}$ . Zum Beispiel gilt bei Wahl von  $\alpha_0 := 0.05$  und  $n = 12$ , dass  $k_{0.05} = \psi^{-1}(1 - 0.05/2; 12 - 1) \approx 2.20$  ist.
- Anhand von  $n, \mu_0, \bar{y}$  und  $s_y$  berechnet man die Realisierung der T-Teststatistik

$$t := \sqrt{n} \left( \frac{\bar{y} - \mu_0}{s_y} \right) \quad (28)$$

- Wenn  $t$  größer-gleich  $k_{\alpha_0}$  ist oder wenn  $t$  kleiner-gleich  $-k_{\alpha_0}$  ist, lehnt man die Nullhypothese ab, andernfalls lehnt man sie nicht ab.
- Die oben entwickelte Theorie garantiert dann, dass man in höchstens  $\alpha_0 \cdot 100$  von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

## Bestimmung des p-Wertes

- Per Definition ist der p-Wert das kleinste Signifikanzlevel  $\alpha_0$ , bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.
- Bei  $T = t$  würde  $H_0$  für jedes  $\alpha_0$  mit  $|t| \geq \psi^{-1}(1 - \alpha_0/2; n - 1)$  abgelehnt werden. Für diese  $\alpha_0$  gilt, wie unten gezeigt,

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \quad (29)$$

- Das kleinste  $\alpha_0 \in [0, 1]$  mit  $\alpha_0 \geq 2\mathbb{P}(T \geq |t|)$  ist dann  $\alpha_0 = 2\mathbb{P}(T \geq |t|)$ , also folgt

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n - 1)). \quad (30)$$

- Im Gegensatz zum Z-Test hängt bei T-Tests der p-Wert auch von der Stichprobengröße ab.
- Zum Beispiel ist für  $t = 2.00$  und  $n = 10$  der p-Wert 0.076, für  $t = 2.00$  und  $n = 100$  ist der p-Wert dagegen 0.048.

## Bestimmung des p-Wertes

- Es bleibt zu zeigen, dass gilt

$$|t| \geq \psi^{-1}(1 - \alpha_0/2; n - 1) \Leftrightarrow \alpha_0 \geq 2\mathbb{P}(T \geq |t|) . \quad (31)$$

- Dies aber folgt aus

$$\begin{aligned} |t| &\geq \psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right) \\ \Leftrightarrow \psi(|t|; n - 1) &\geq \psi\left(\psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right); n - 1\right) \\ \Leftrightarrow \psi(|t|; n - 1) &\geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow \mathbb{P}(T \leq |t|) &\geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow \frac{\alpha_0}{2} &\geq 1 - \mathbb{P}(T \leq |t|) \\ \Leftrightarrow \frac{\alpha_0}{2} &\geq \mathbb{P}(T \geq |t|) \\ \Leftrightarrow \alpha_0 &\geq 2\mathbb{P}(T \geq |t|) . \end{aligned} \quad (32)$$

## Modellevaluation (8) Analyse der Powerfunktion

---

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k_{\alpha_0}; \delta, n - 1) + \psi(-k_{\alpha_0}; \delta, n - 1) \quad (33)$$

bei kontrolliertem Testumfang, also für  $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0/2; n - 1)$  mit festem  $\alpha_0$  als Funktion des Nichtzentralitätsparameters und des Stichprobenumfangs. Namentlich hängt hier  $k_{\alpha_0}$  sowohl von  $\alpha_0$  als auch von  $n$  ab.

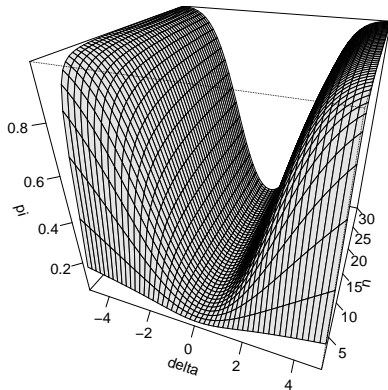
Es ergibt sich die bivariate reellwertige Funktion

$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (\delta, n) \mapsto \pi(\delta, n) := 1 - \psi(k_{\alpha_0}; \delta, n - 1) + \psi(-k_{\alpha_0}; \delta, n - 1) \quad (34)$$

Bei festgelegten  $\alpha_0$  hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekanntem Wert  $\delta$  und von der Stichprobengröße  $n$  ab. Wir visualisieren diese Abhängigkeiten untenstehend.

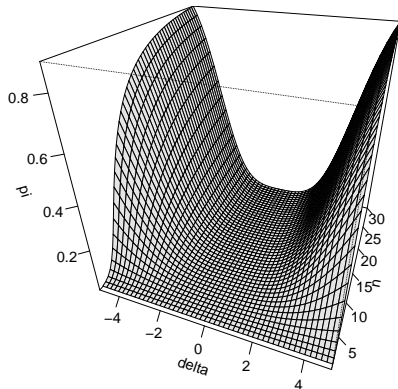
## Modellevaluation (8) Analyse der Powerfunktion

Powerfunktion für  $\alpha_0 = 0.05$



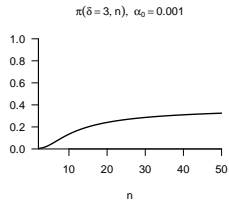
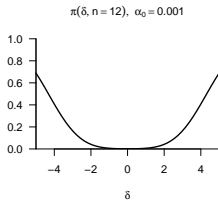
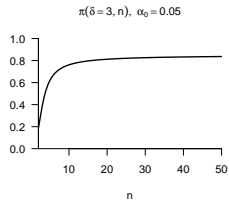
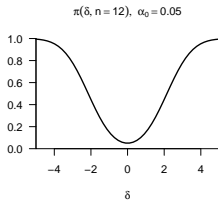
## Modellevaluation (8) Analyse der Powerfunktion

Powerfunktion für  $\alpha_0 = 0.001$



# Modellevaluation (8) Analyse der Powerfunktion

Powerfunktionen für  $\mu_0 = 0$



## Modellevaluation (8) Analyse der Powerfunktion

---

### Praktisches Vorgehen

Mit größerem  $n$  steigt die Powerfunktion des Tests an:

- Ein großer Stichprobenumfang ist besser als ein kleiner Stichprobenumfang.
- Kosten für die Erhöhung des Stichprobenumfangs werden aber nicht berücksichtigt.

⇒ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

Die Powerfunktion hängt vom wahren, aber unbekanntem Wert  $\delta = \sqrt{n} \left( \frac{\mu - \mu_0}{\sigma} \right)$  ab.

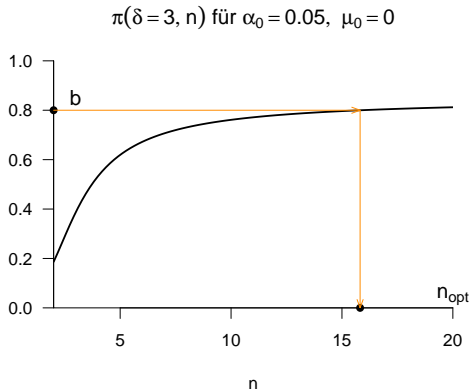
⇒ Wenn man  $\delta$  schon kennen würde, würde man den Test nicht durchführen.

Generell wird folgendes Vorgehen favorisiert:

- Man legt das Signifikanzniveau  $\alpha_0$  fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert  $\delta^*$ , den man mit  $\pi(\delta, n) = b$  detektieren möchte.
- Ein konventioneller Wert ist  $b = 0.8$ .
- Man liest die für  $\pi(\delta = \delta^*, n) = b$  nötige Stichprobengröße  $n$  ab.

## Modellevaluation (8) Analyse der Powerfunktion

### Praktisches Vorgehen



Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

**Anwendung/Praxis**

## Anwendungsbeispiel

Face-to-face

PreBDI



PostBDI

Wir nehmen an, dass die Datenpunkte unabhängige und identisch verteilte (u.i.v.) Realisierungen von ZVen  $y_i \sim N(\mu, \sigma^2)$  sind und nehmen weiter an, dass wir sind an der Quantifizierung der Unsicherheit beim inferentiellen Vergleich des wahren, aber unbekanntem Erwartungswertparameters  $\mu$  im Sinne eines Hypothesentests interessiert sind.

## Daten einlesen

```
fname = "Daten/T-Tests_Daten.csv"  
D = read.table(fname, sep = ",", header = TRUE)
```

X	ID	Setting	PreBDI	PostBDI	dBDI
1	1	F2F	29	25	4
2	2	F2F	32	27	5
3	3	F2F	28	31	-3
4	4	F2F	36	22	14
5	5	F2F	32	29	3
6	6	F2F	28	28	0
7	7	F2F	33	30	3
8	8	F2F	33	26	7
9	9	F2F	33	28	5
10	10	F2F	30	28	2
11	11	F2F	36	25	11
12	12	F2F	32	31	1
13	13	F2F	29	31	-2
14	14	F2F	24	29	-5
15	15	F2F	35	32	3
16	16	F2F	31	29	2
17	17	F2F	31	23	8
18	18	F2F	34	25	9
19	19	F2F	34	23	11
20	20	F2F	33	26	7
21	21	F2F	34	25	9
22	22	F2F	33	27	6
23	23	F2F	31	24	7
24	24	F2F	25	27	-2
25	25	F2F	33	25	8
26	26	F2F	31	33	-2
27	27	F2F	31	29	2
28	28	F2F	26	30	-4
29	29	F2F	29	28	1
30	30	F2F	32	32	0
31	31	F2F	35	25	10
32	32	F2F	31	26	5
33	33	F2F	32	32	0
34	34	F2F	31	25	6
35	35	F2F	27	26	1
36	36	F2F	30	26	4
37	37	F2F	30	26	4
38	38	F2F	31	26	5
39	39	F2F	34	29	5
40	40	F2F	33	26	7

# Anwendung/Praxis

```
# Datensatz von Interesse
BDI_F2F      = D$dBDI[D$Setting == "F2F"]

# Histogrammparameter
h            = 1
b_0         = min(BDI_F2F)
b_k         = max(BDI_F2F)
k           = ceiling((b_k - b_0)/h)
b           = seq(b_0, b_k, by = h)
ylimits     = c(0,0.15)
xlimits     = c(-5,15)

# Abbildungsparameter
par(
  mfcol      = c(1,1),
  family     = "sans",
  pty        = "s",
  bty        = "l",
  las        = 1,
  xaxs       = "i",
  yaxs       = "i",
  font.main  = 1,
  cex        = 1,
  cex.main   = 1)

# Histogramm
hist(BDI_F2F,
     breaks = b,
     freq   = F,
     xlim   = xlimits,
     ylim   = ylimits,
     xlab   = TeX("$\\Delta$ BDI$"),
     ylab   = "geschätzte Wahrscheinlichkeit",
     main   = "")

# Speichern
dev.copy2pdf(
  file      = "Abbildungen/F2F_histogramm.pdf",
  width     = 4,
  height    = 4)

# BDI-Differenzwerte in der F2F-Gruppe

# gewünschte Klassenbreite
# b_0
# b_k
# Anzahl der Klassen
# Klassen [b_0, ..., b_k]
# y-Achsenlimits
# x-Achsenlimits

# für Details siehe ?par
# 1 x 1 Panelstruktur
# Serif-freier Fonttyp
# quadratische Abbildungsregion
# L-förmige Box
# horizontale Achsenbeschriftung
# x-Achse bei y = 0
# y-Achse bei x = 0
# Titel nicht fett
# Textvergrößerungsfaktor
# Titeltextvergrößerungsfaktor

# Delta-BDI_Werte von Therapiebedingung i
# Histogrammklassen
# normierte relative Häufigkeit
# x-Achsenlimits
# y-Achsenlimits
# x-Achsenbeschriftung
# y-Achsenbeschriftung
# Titelbeschriftung
```

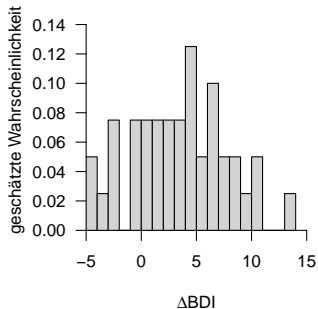
# Anwendung/Praxis

```
# Einlesen der Daten
fname      = "Daten/T-Tests_Daten.csv"
D          = read.table(fname, sep = ",", header = TRUE)

# Initialisierung eines Dataframes
tp         = c("F2F")                # Therapiebedingungen
ntp        = length(tp)              # Anzahl Therapiebedingungen
S          = data.frame(              # Dataframe-Erzeugung
  n         = rep(NaN,ntp),           # Stichprobengrößen
  Max       = rep(NaN,ntp),           # Maxima
  Min       = rep(NaN,ntp),           # Minima
  Median    = rep(NaN,ntp),           # Mediane
  Mean      = rep(NaN,ntp),           # Mittelwerte
  Var       = rep(NaN,ntp),           # Varianzen
  Std       = rep(NaN,ntp),           # Standardabweichungen
  row.names = tp)                    # Therapiebedingungen

# Iterationen über Therapiebedingungen
for(i in 1:ntp){
  data      = D$dBDI[D$Setting == tp[i]] # Daten
  S$n[i]    = length(data)               # Stichprobengröße
  S$Max[i]  = max(data)                  # Maxima
  S$Min[i]  = min(data)                  # Minima
  S$Median[i] = median(data)             # Mediane
  S$Mean[i] = mean(data)                 # Mittelwerte
  S$Var[i]  = var(data)                  # Varianzen
  S$Std[i]  = sd(data)                   # Standardabweichungen
}
```

## Deskriptive Statistiken der negativen PostBDI-PreBDI-Differenzen bei Face-to-Face-Therapie



```
# Ausgabe  
print.AsIs(S)
```

```
>      n Max Min Median Mean  Var  Std  
> F2F 40  14  -5      4 3.92 19.4 4.41
```

# Anwendung/Praxis

```
# Daten einlesen
fname      = "Daten/T-Tests_Daten.csv"
D          = read.table(fname, sep = ",", header = TRUE)
y          = D$dBDI[D$Setting == "F2F"]

# Modellformulierung
n          = length(y)
p          = 1
X          = matrix(rep(1,n), nrow = n)

# Parameterschätzung
beta_hat  = solve(t(X) %*% X) %*% t(X) %*% y
eps_hat   = y - X %*% beta_hat
sigsqr_hat = t(eps_hat) %*% eps_hat) / (n-p)

# Konfidenzintervall
delta     = 0.95
t_delta   = qt((1+delta)/2,n-1)
lambda    = diag(solve(t(X) %*% X))
kappa     = matrix(rep(NA,n*p*2), nrow = p)
for(j in 1:p){
  kappa[j,1] = beta_hat[j]-sqrt(sigsqr_hat*lambda[j])*t_delta
  kappa[j,2] = beta_hat[j]+sqrt(sigsqr_hat*lambda[j])*t_delta
}

# Hypothesentest
c          = matrix(c(1),nrow = p)
mu_0      = 0
alpha_0    = 0.05
k_alpha_0  = qt(1 - (alpha_0/2), n-1)
t_num      = t(c) %*% beta_hat - mu_0
t_den      = sqrt(sigsqr_hat %*% t(c)*solve(t(X) %*% X)%*%c)
t          = t_num/t_den
if(abs(t) >= k_alpha_0){phi = 1} else {phi = 0}
pval       = 2*(1 - pt(abs(t), n-1))
d          = t/sqrt(n)

# Dateiname
# Dataframe
# BDI-Differenzwerte in der F2F-Gruppe

# Anzahl Datenpunkte
# Anzahl Betaparameter
# n x p Designmatrix

# Betaparameterschätzer
# Residuenvektor
# Varianzparameterschätzer

# Konfidenzbedingung
# \Psi^{-1}((1+\delta)/2; n-1)
# \lambda_j Werte
# \beta_j Konfidenzintervall-Array
# Iteration über \beta_j
# untere KI-Grenze
# obere KI-Grenze

# Kontrastgewichtsvektor
# Nullhypothese
# Signifikanzniveau
# kritischer Wert
# T-Teststatistik-Zähler
# T-Teststatistik-Nenner
# T-Teststatistik
# Test 1_{|T(X)| >= k_alpha_0}
# p-Wert
# Cohen's d

> fg          = 39
> kappa_1     = 2.52 5.33
> t           = 5.64
> alpha_0     = 0.05
> k_alpha_0   = 2.02
> phi         = 1
> p-Wert      = 1.66e-06
> Cohen's d   = 0.891
```

## Anwendungsszenario

```
# automatischer Einstichproben-T-Test
varphi = t.test(y,                                # Datensatz
               alternative = c("two.sided"),      # H_1: \mu \neq \mu_0
               mu          = 0,                  # \mu_0 (sic!)
               conf.level  = 1-alpha_0)         # \delta = 1 - \alpha_0

# Ausgabe
print(varphi)
```

```
>
> One Sample t-test
>
> data: y
> t = 6, df = 39, p-value = 2e-06
> alternative hypothesis: true mean is not equal to 0
> 95 percent confidence interval:
>  2.52 5.33
> sample estimates:
> mean of x
>      3.92
```

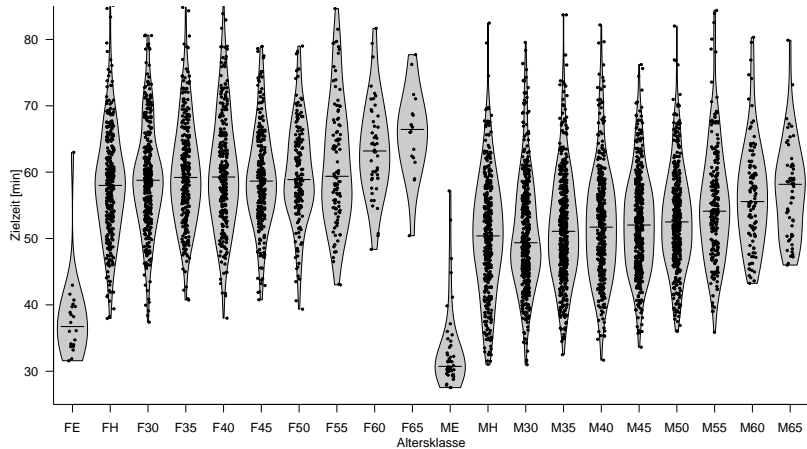
```
# genauere Ausgabe von t
paste(varphi[1]$statistic)
```

```
> [1] "5.63531986397201"
```

```
# genauere Ausgabe von p
paste(varphi[3]$p.value)
```

```
> [1] "1.66216308541e-06"
```

Datensatz: Zielzeiten beim Great 10k (08.10.2017), getrennt nach Altersklasse ( $n = 4883$ )



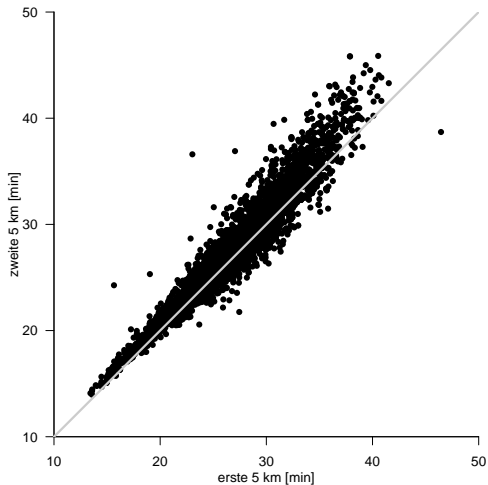
Einstichproben-T-Test: 10-km-Zeit, alle Teilnehmer ( $n = 4883$ )

```
# Daten extrahieren
D$T10k = t10_min # 10-km-Zeit [min]

# Einstichproben-T-Test (falsch)
n      = nrow(D) # Anzahl Datenpunkte
p      = 1        # Anzahl Regressoren
y      = matrix(D$T10k, nrow = n) # Datenvektor
X      = matrix(rep(1,n), ncol = p) # Designmatrix

> Betaparameterschätzer      : 54.9
> Varianzparameterschätzer  : 86.6
> Einstichproben-T-Teststatistik : 412
> Cohen's d                  : 5.9
> p-Wert                     : 0
```

Datensatz: erste vs. zweite 5 km beim Great 10k (08.10.2017), alle Teilnehmer ( $n = 4883$ )



Einstichproben-T-Test: Differenz zweite minus erste 5 km, alle Teilnehmer ( $n = 4883$ )

```
# Daten extrahieren
D$T5k1 = t5_min           # erste 5 km [min]
D$T5k2 = t10_min-t5_min  # zweite 5 km [min]
D$Tdiff= D$T5k2 - D$T5k1 # Differenz [min]

# Einstichproben-T-Test (richtig)
n      = nrow(D)          # Anzahl Datenpunkte
p      = 1                # Anzahl Regressoren
y      = matrix(D$Tdiff, nrow = n) # Datenvektor
X      = matrix(rep(1,n), ncol = p) # Designmatrix
```

```
> Betaparameterschätzer      : 0.708
> Varianzparameterschätzer  : 2.21
> Einstichproben-T-Teststatistik : 33.3
> Cohen's d                  : 0.476
> p-Wert                     : 0
```

---

Überblick

Einstichproben-T-Tests

**Zweistichproben-T-Tests**

Selbstkontrollfragen

---

# Zweistichproben-T-Tests

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendung/Praxis

### **Anwendungsszenario**

Modellformulierung

Modellschätzung

Modellevaluation

Anwendung/Praxis

**Zwei Gruppen** (Stichproben) randomisierter experimenteller Einheiten.

Annahme unabhängiger identischer Normalverteilungen  $N(\mu_1, \sigma^2)$  und  $N(\mu_2, \sigma^2)$ .

$\mu_1, \mu_2$  und  $\sigma^2$  unbekannt.

Annahme eines identischen Varianzparameters für beide Gruppen.

Quantifizieren der Unsicherheit beim inferentiellen Vergleich von  $\mu_1$  mit  $\mu_2$  beabsichtigt.

## Anwendungsbeispiele

BDI-Differenzwert-Datenanalyse bei zwei Gruppen von Patient:innen

- Gruppe 1: Face-to-Face-Therapie; Gruppe 2: Online-Therapie
- $\mu_1 \neq \mu_2 \Leftrightarrow$  Unterscheiden sich die Therapiewirksamkeiten?

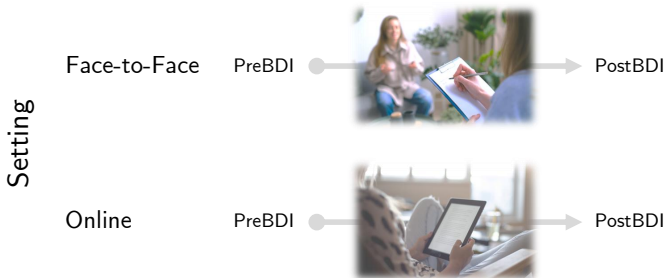
Forcierte Schwimmtest-Datenanalyse bei zwei Gruppen genmanipulierter Mäuse

- Gruppe 1: Wildtyp; Gruppe 2: Serotoninrezeptormutation
- $\mu_1 \neq \mu_2 \Leftrightarrow$  Trägt Serotoninrezeptor zum Schwimmtestverhalten bei?

Analyse von fMRT-Daten in den klinischen Neurowissenschaften

- Gruppe 1: Alzheimer-Patienten; Gruppe 2: gesunde Kontrollen
- $\mu_1 \neq \mu_2 \Leftrightarrow$  Verändert Alzheimer-Erkrankung Hirnaktivität bei Enkodierung ins Gedächtnis?

## Anwendungsbeispiel



Wir nehmen an, dass die Datenpunkte der Face-to-Face-Therapiegruppe u.i.v. Realisierungen von ZVen  $y_{1j} \sim N(\mu_1, \sigma^2)$  und dass die Datenpunkte der Online-Therapiegruppe u.i.v. Realisierungen von ZVen  $y_{2j} \sim N(\mu_2, \sigma^2)$  sind ( $j = 1, \dots, 40$ ). Wir nehmen weiter an, dass wir an der Quantifizierung der Unsicherheit beim inferentiellen Vergleich der wahren, aber unbekanntem Erwartungswertparameter  $\mu_1$  und  $\mu_2$  im Sinne eines Hypothesentests interessiert sind.

Anwendungsszenario

**Modellformulierung**

Modellschätzung

Modellevaluation

Anwendung/Praxis

## Definition (Zweistichproben-T-Test-Modell)

$y_{ij}$  mit  $i = 1, 2$  und  $j = 1, \dots, n_i$  seien Zufallsvariablen, die die Datenpunkte eines Anwendungsszenarios für den Zweistichproben-T-Test modellieren. Dann hat das *Zweistichproben-T-Test-Modell* die strukturelle Form

$$y_{ij} = \mu_i + \varepsilon_{ij} \quad \text{mit} \quad \varepsilon_{ij} \sim N(0, \sigma^2) \quad \text{u.i.v. für} \quad i = 1, 2, j = 1, \dots, n_i \quad \text{mit} \quad \mu_i \in \mathbb{R} \quad \text{und} \quad \sigma^2 > 0, \quad (35)$$

die Datenverteilungsform

$$y_{ij} \sim N(\mu_i, \sigma^2) \quad \text{u.i.v. für} \quad i = 1, 2, j = 1, \dots, n_i \quad \text{mit} \quad \mu_i \in \mathbb{R} \quad \text{und} \quad \sigma^2 > 0, \quad (36)$$

und für den Datenvektor  $y = (y_{11}, \dots, y_{1n_1}, y_{21}, \dots, y_{2n_2})^T$  und  $n := n_1 + n_2$  die Designmatrixform

$$y = X\beta + \varepsilon \quad \text{mit} \quad X := \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \in \mathbb{R}^{n \times 2}, \quad \beta := \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2, \quad \varepsilon \sim N(0_n, \sigma^2 I_n), \quad \text{und} \quad \sigma^2 > 0. \quad (37)$$

### Bemerkungen

- $i$  indiziert die Gruppen,  $j$  indiziert die experimentellen Einheiten in jeder Gruppe.
- $n_1$  und  $n_2$  sind die Gruppengrößen,  $n$  repräsentiert die Gesamtanzahl an Datenpunkten.
- Die Äquivalenz der drei Modellformen ergibt sich mit der Modellformulierung des ALM (vgl. Einheit (5) in *Allgemeines Lineares Modell*) und aus den Regeln der Matrixmultiplikation.
- Die Anzahl der Betaparameter ist  $p = 2$ .



## Datensimulation (vgl. Einheit (5) in *Allgemeines Lineares Modell*)

```
# Modellformulierung
library(MASS) # multivariate Normalverteilung
n_1 = 10 # Anzahl von Datenpunkten Gruppe 1
n_2 = 10 # Anzahl von Datenpunkten Gruppe 2
n = n_1 + n_2 # Gesamtanzahl der Datenpunkte
p = 2 # Anzahl von Betaparametern
X = matrix(c(rep(1,n_1), rep(0,n_1), # n x p Designmatrix
             rep(0,n_2), rep(1,n_2)),
           nrow = n)
I_n = diag(n) # n x n Einheitsmatrix
beta = matrix(c(1,2), nrow = p) # wahre, aber unbekannte Betaparameter
sigsqr = 14 # wahrer, aber unbekannter Varianzparameter

# Datenrealisierung
y = mvrnorm(1, X %>% beta, sigsqr*I_n) # eine Realisierung des n-dimensionalen ZVs y
print(y)

> [1] 4.9935 -0.1728 -0.0448 -6.1469 0.4986 1.4192 7.0748 1.9118
> [9] 1.6604 4.4998 -3.6589 -1.0936 1.9403 -1.5016 1.6354 9.0661
> [17] 4.7605 -2.8812 3.3580 3.7517
```

Anwendungsszenario

Modellformulierung

**Modellschätzung**

Modellevaluation

Anwendung/Praxis

## Theorem (Parameterschätzung im Zweistichproben-T-Test-Modell)

Gegeben sei die Designmatrixform des Zweistichproben-T-Test-Modells. Dann ergeben sich für den Betaparameterschätzer

$$\hat{\beta} = \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} =: \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \quad (38)$$

und für den Varianzparameterschätzer

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2} =: s_{12}^2 \quad (39)$$

### Bemerkungen

- $\bar{y}_1$  und  $\bar{y}_2$  bezeichnen die gruppenspezifischen Stichprobenmittel.
- $s_{12}^2$  wird als *gebündelte Stichprobenvarianz* oder *pooled sample variance* bezeichnet.
- Für einen Datensatz  $y = (y_1, y_2)$  gilt im Allgemeinen, dass  $s_y^2 \neq s_{12}^2$ . Die gebündelte Stichprobenvarianz und die Stichprobenvarianz eines zusammengefügt ("konkatenierten") Datensatzes sind im Allgemeinen also nicht identisch. Wir wollen das Konzept der gebündelten Stichprobenvarianz hier aber nicht weiter vertiefen.

## Beweis

Für  $i = 1, 2$  sei  $y_i := (y_{i1}, \dots, y_{in_i})^T$ . Dann ergibt sich für den Betaparameterschätzer

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T y \\ &= \left( \begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix} \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \right)^{-1} \begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^{n_1} y_{1j} \\ \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \\ &= \begin{pmatrix} 1/n_1 & 0 \\ 0 & 1/n_2 \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n_1} y_{1j} \\ \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \\ &=: \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix}.\end{aligned}\tag{40}$$

## Beweis (fortgeführt)

Gleichsam ergibt sich für den Varianzparameterschätzer mit  $n = n_1 + n_2$  und  $p = 2$

$$\begin{aligned}\hat{\sigma}^2 &= \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \\ &= \frac{1}{n_1 + n_2 - 2} \left( \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \right)^T \left( \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \right) \\ &= \frac{1}{n_1 + n_2 - 2} \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ \vdots \\ y_{2n_2} - \bar{y}_2 \end{pmatrix}^T \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ \vdots \\ y_{2n_2} - \bar{y}_2 \end{pmatrix} \\ &= \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2} \\ &=: s_{12}^2.\end{aligned}\tag{41}$$

□

# Modellschätzung

```
# Daten einlesen
fname      = "Daten/T-Tests_Daten.csv"
D          = read.table(fname, sep = ",", header = TRUE)
y_1       = D$dBDI[D$Setting == "F2F"]
y_2       = D$dBDI[D$Setting == "ONL"]

# Modellformulierung
n_1       = length(y_1)
n_2       = length(y_2)
n         = n_1 + n_2
p         = 2
y         = matrix(c(y_1, y_2), nrow = n)
X         = matrix(c(rep(1, n_1), rep(0, n_1),
                    rep(0, n_2), rep(1, n_2)),
                  nrow = n)

# Modellschätzung
beta_hat  = solve(t(X) %*% X) %*% t(X) %*% y
eps_hat   = y - X %*% beta_hat
sigsqr_hat = (t(eps_hat) %*% eps_hat) / (n-p)
s_sqr_12  = ((n_1-1)*var(y_1) + (n_2-1)*var(y_2)) / (n_1+n_2-2)

# Ausgabe
cat("hat{beta}      : ", round(beta_hat, digits = 3),
    "\nbar{y}_1, bar{y}_2 : ", round(c(mean(y_1), mean(y_2)),
    digits = 3),
    "\nhat{sigsqr}     : ", round(sigsqr_hat, digits = 3),
    "\ns_12^2          : ", round(s_sqr_12, digits = 3),
    "\ns_y^2           : ", round(var(y), digits = 3))

# Betaparameterschätzer
# Residuenvektor
# Varianzparameterschätzer
# gebündelte Stichprobenvarianz
# Betaparameterschätzer
# Stichprobenmittel
# Varianzparameterschätzer
# gebündelte Stichprobenvarianz
# Stichprobenvarianz des konkatenierten Datensatzes

> hat{beta}      : 3.92 7.4
> bar{y}_1, bar{y}_2 : 3.92 7.4
> hat{sigsqr}    : 21.6
> s_12^2         : 21.6
> s_y^2          : 24.4
```

Anwendungsszenario

Modellformulierung

Modellschätzung

**Modellevaluation**

Anwendung/Praxis

## Überblick

- Wir gruppieren frequentistische Konfidenzintervalle und Hypothesentests unter Modellevaluation.
- Wir verzichten an dieser Stelle auf eine Diskussion von Konfidenzintervallen.
- In der Praxis zielt die Evaluation von Zweistichproben-T-Tests-Designs meist auf einen Hypothesentest.
- Die Theorie der Zweistichproben-T-Tests ist umfangreich.
- Ein gutes Verständnis von Hypothesentests wird im Folgenden vorausgesetzt.
- (vgl. Einheit (11) in *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*)

Im Zweistichproben-T-Test ALM Design ergeben sich folgende Hypothesenszenarien:

- $H_0 : \mu_1 - \mu_2 = \mu_0$  und  $H_1 : \mu_1 - \mu_2 \neq \mu_0$
- $H_0 : \mu_1 - \mu_2 \leq \mu_0$  und  $H_1 : \mu_1 - \mu_2 > \mu_0$
- $H_0 : \mu_1 - \mu_2 \geq \mu_0$  und  $H_1 : \mu_1 - \mu_2 < \mu_0$

Wir betrachten hier exemplarisch nur  $H_0 : \mu_1 - \mu_2 = \mu_0$  und  $H_1 : \mu_1 - \mu_2 \neq \mu_0$ .

Für  $\mu_0 := 0$  gelten dabei insbesondere:

- $H_0 : \mu_1 - \mu_2 = 0 \Leftrightarrow H_0 : \mu_1 = \mu_2$
- $H_1 : \mu_1 - \mu_2 \neq 0 \Leftrightarrow H_1 : \mu_1 \neq \mu_2$

Gliederung (vgl. Einheit (11), Folie 29 in *Wahrscheinlichkeitstheorie und Frequentistische Inferenz*)

- (1) Statistisches Modell ✓
- (2) Testhypothesen ✓
- (3) Teststatistik
- (4) Test
- (5) Analyse der Testgütefunktion
- (6) Testumfangkontrolle
- (7) p-Wert
- (8) Analyse der Powerfunktion

### Theorem (T-Teststatistik des Zweistichproben-T-Tests)

Gegeben sei die Designmatrixform des Zweistichproben-T-Test-Modells. Dann ergibt sich für die T-Teststatistik mit

$$c := (1, -1)^T \quad \text{und} \quad c^T \beta_0 =: \mu_0, \quad (42)$$

dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \quad (43)$$

und es gilt

$$T \sim t(\delta, n_1 + n_2 - 2) \quad \text{mit} \quad \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right). \quad (44)$$

#### Bemerkungen

- Das Theorem basiert auf der T-Statistik im Rahmen des ALM.
- Wir erinnern an das verwandte populäre und von der Stichprobengröße unabhängige Effektstärke-Maß *Cohen's d* bei Zweistichproben-T-Test-Designs,

$$d := \frac{\bar{y}_1 - \bar{y}_2}{s_{12}}. \quad (45)$$

- Offenbar gilt für dieses *Cohen's d*, dass mit  $\mu_0 := 0$

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} d \Leftrightarrow d = T / \sqrt{\frac{n_1 n_2}{n_1 + n_2}}. \quad (46)$$

## Modellevaluation (3) Teststatistik

### Beweis

Mit dem Theorem zur Verteilung der T-Statistik (siehe Einheit (7) in *Allgemeines Lineares Modell*) gilt zunächst für die Zähler von  $T$  und  $\delta$ , dass

$$c^T \hat{\beta} - c^T \beta_0 = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} - \mu_0 = \bar{y}_1 - \bar{y}_2 - \mu_0 \quad (47)$$

und

$$c^T \beta - c^T \beta_0 = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} - \mu_0 = \mu_1 - \mu_2 - \mu_0. \quad (48)$$

Weiterhin gilt für die Nenner von  $T$  und  $\delta$ , dass

$$c^T (X^T X)^{-1} c = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} 1/n_1 & 0 \\ 0 & 1/n_2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1/n_1 & -1/n_2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{n_1} + \frac{1}{n_2}. \quad (49)$$

Außerdem gilt

$$\left( \frac{1}{n_1} + \frac{1}{n_2} \right)^{-\frac{1}{2}} = \left( \frac{n_2}{n_1 n_2} + \frac{n_1}{n_1 n_2} \right)^{-\frac{1}{2}} = \left( \frac{n_1 + n_2}{n_1 n_2} \right)^{-\frac{1}{2}} = \left( \frac{n_1 n_2}{n_1 + n_2} \right)^{\frac{1}{2}}. \quad (50)$$

Zusammengenommen folgt dann, dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \quad \text{und} \quad \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right). \quad (51)$$

□

### Definition (Zweiseitiger Zweistichproben-T-Test)

Gegeben sei das Zweistichproben-T-Test-Modell. Für ein  $\mu_0 \in \mathbb{R}$  seien die Nullhypothese und die Alternativhypothese gegeben durch

$$H_0 : \mu_1 - \mu_2 = \mu_0 \quad (52)$$

und

$$H_1 : \mu_1 - \mu_2 \neq \mu_0 . \quad (53)$$

Weiterhin sei die T-Teststatistik definiert durch

$$T := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \quad (54)$$

Dann ist der *zweiseitige Zweistichproben-T-Test* definiert als der kritische-Wert-basierte Test

$$\phi(y) := 1_{\{|T| \geq k\}} = \begin{cases} 1 & |T| \geq k \\ 0 & |T| < k \end{cases} . \quad (55)$$

Bemerkungen

- Ausführlicher handelt es sich um den *zweiseitigen Zweistichproben-T-Test mit ungerichteter Hypothese*.

## Theorem (Testgütefunktion)

Es sei  $\phi$  der im obigen Modell formulierte Zweistichproben-T-Test. Dann ist die Testgütefunktion von  $\phi$  gegeben durch

$$q_\phi : \mathbb{R}^2 \rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) := 1 - \psi(k; \delta, n_1 + n_2 - 2) + \psi(-k; \delta, n_1 + n_2 - 2), \quad (56)$$

wobei  $\psi(\cdot; \delta, n_1 + n_2 - 2)$  die KVF der nichtzentralen  $t$ -Verteilung mit Nichtzentralitätsparameter

$$\delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right) \quad (57)$$

und Freiheitsgradparameter  $n_1 + n_2 - 2$  bezeichnet.

### Bemerkungen

- $q_\phi$  ist eine bivariate reellwertige Funktion.
- $q_\phi$  kann alternativ als univariate reellwertige Funktion von  $\Delta := \mu_1 - \mu_2$  konzipiert werden.
- Im Vergleich zum Einstichprobenszenario gelten

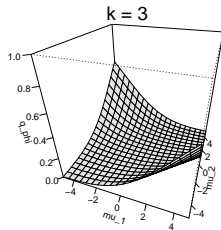
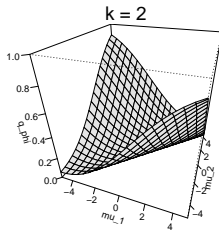
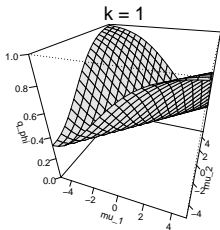
$$n - 1 \hookrightarrow n_1 + n_2 - 2, \quad \sqrt{n} \hookrightarrow \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \quad \mu - \mu_0 \hookrightarrow \mu_1 - \mu_2 - \mu_0 \quad (58)$$

- Wir verzichten auf einen Beweis. Für einen Beweisansatz, siehe DeGroot and Schervish (2012), Seite 591.

## Modellevaluation (5) Analyse der Testgütefunktion

Testgütefunktion  $q_\phi$  für  $\sigma^2 = 9$ ,  $\mu_0 = 0$ ,  $n_1 = 12$ ,  $n_2 = 12$  und  $k = 1, 2, 3$ .

$$q_\phi(\mu_1, \mu_2) = \mathbb{P}_{\mu_1, \mu_2}(\phi = 1)$$



### Theorem (Testumfangkontrolle)

$\phi$  sei der im obigen Testszenario definierte Test. Dann ist  $\phi$  ein Level- $\alpha_0$ -Test mit Testumfang  $\alpha_0$ , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \psi^{-1} \left( 1 - \frac{\alpha_0}{2}; n_1 + n_2 - 2 \right), \quad (59)$$

wobei  $\psi^{-1}(\cdot; n_1 + n_2 - 2)$  die inverse KVF der  $t$ -Verteilung mit  $n_1 + n_2 - 2$  Freiheitsgraden ist.

### Bemerkungen

- Das Resultat folgt in Analogie zum Einstichproben-T-Test.
- Im Vergleich zum Einstichproben-T-Testfall gilt lediglich

$$n - 1 \hookrightarrow n_1 + n_2 - 2. \quad (60)$$

## Praktisches Vorgehen

- Man nimmt an, dass die Daten zweier Gruppen  $y_{11}, \dots, y_{1n_1}$  und  $y_{21}, \dots, y_{2n_2}$  Realisationen von  $y_{1j} \sim N(\mu_1, \sigma^2)$  u.i.v. für  $j = 1, \dots, n_1$  und  $y_{2j} \sim N(\mu_2, \sigma^2)$  u.i.v. für  $j = 1, \dots, n_2$  mit unbekanntem Parametern  $\mu_1, \mu_2$  und  $\sigma^2$  sind.
- Man möchte entscheiden, ob eher  $H_0 : \mu_1 - \mu_2 = \mu_0$  oder  $H_1 : \mu_1 - \mu_2 \neq \mu_0$  zutrifft.
- Man wählt ein Signifikanzniveau  $\alpha_0$  und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert  $k_{\alpha_0}$ . Zum Beispiel gilt bei Wahl von  $\alpha_0 := 0.05$  und  $n_1 = 12, n_2 = 12$ , dass  $k_{0.05} = \psi^{-1}(1 - 0.05/2; 12 + 12 - 2) \approx 2.07$  ist.
- Anhand von  $n_1, n_2, \bar{y}_1, \bar{y}_2$  und der gebündelten Stichprobenstandardabweichung  $s_{12}$  berechnet man die Realisierung der Zweistichproben-T-Teststatistik

$$t := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \quad (61)$$

- Wenn  $t$  größer-gleich  $k_{\alpha_0}$  ist oder wenn  $t$  kleiner-gleich  $-k_{\alpha_0}$  ist, lehnt man die Nullhypothese ab, andernfalls lehnt man sie nicht ab.
- Die oben entwickelte Theorie des Zweistichproben-T-Tests garantiert dann, dass man in höchstens  $\alpha_0 \cdot 100$  von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

### Bestimmung des p-Wertes

- Per Definition ist der p-Wert das kleinste Signifikanzlevel  $\alpha_0$ , bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.
- Bei  $T = t$  würde  $H_0$  für jedes  $\alpha_0$  mit  $|t| \geq \psi^{-1}(1 - \alpha_0/2; n_1 + n_2 - 2)$  abgelehnt werden. Für diese  $\alpha_0$  gilt, wie bereits mehrfach gezeigt,

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \quad (62)$$

- Das kleinste  $\alpha_0 \in [0, 1]$  mit  $\alpha_0 \geq 2\mathbb{P}(T \geq |t|)$  ist dann  $\alpha_0 = 2\mathbb{P}(T \geq |t|)$ , also folgt

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n_1 + n_2 - 2)). \quad (63)$$

- Im Vergleich zum Einstichprobenfall gilt lediglich  $n - 1 \hookrightarrow n_1 + n_2 - 2$ .

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R}^2 \rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) := 1 - \psi(k; \delta, n_1 + n_2 - 2) + \psi(-k; \delta, n_1 + n_2 - 2) \quad (64)$$

als Funktion des Nichtzentralitätsparameters  $\delta := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right)$  und der Summe der Stichprobenumfänge  $n := n_1 + n_2$  bei kontrolliertem Testumfang, also für  $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0/2; n - 2)$  mit festem  $\alpha_0$ .

Es ergibt sich die multivariate reellwertige Funktion

$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (\delta, n) \mapsto \pi(\delta, n) := 1 - \psi(k_{\alpha_0}; \delta, n - 2) + \psi(-k_{\alpha_0}; \delta, n - 2). \quad (65)$$

Bei festgelegten  $\alpha_0$  hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekanntem Wert  $\delta$  und von der Summe der Stichprobengrößen  $n$  ab. De facto handelt es sich also um die gleiche Powerfunktion wie beim zweiseitigen Einstichproben-T-Test mit dem einzigen Unterschied, dass für den Freiheitsgradparameter  $n - 2$  anstelle von  $n - 1$  gilt. Wir verzichten auf eine erneute Visualisierung.

## Modellevaluation (8) Analyse der Powerfunktion

### Praktisches Vorgehen

Mit größerem  $n = n_1 + n_2$  steigt die Powerfunktion des Tests an:

- Ein großer Stichprobenumfang ist besser als ein kleiner Stichprobenumfang.
- Kosten für die Erhöhung des Stichprobenumfangs werden aber nicht berücksichtigt.
- Ungleichgewichte zwischen  $n_1$  und  $n_2$  werden durch die Tatsache ausglich, dass Datenpunkte einer Stichprobe auch zur Varianzschätzung in der anderen Stichprobe beitragen, da eine identische Varianz vorausgesetzt wurde.

Die Powerfunktion hängt vom wahren, aber unbekanntem, Wert  $\delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right)$  ab.

⇒ Wenn man  $\delta$  schon kennen würde, würde man den Test nicht durchführen.

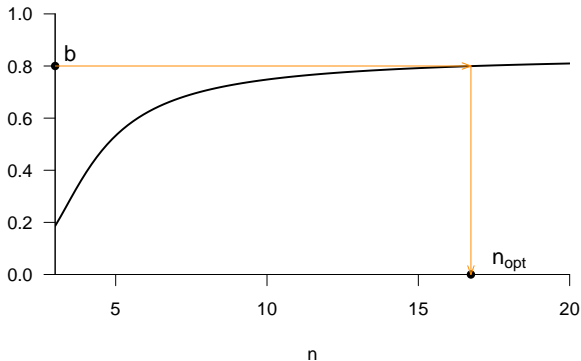
Generell wird folgendes Vorgehen favorisiert:

- Man legt das Signifikanzniveau  $\alpha_0$  fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert  $\delta^*$ , den man mit  $\pi(\delta, n) = b$  detektieren möchte.
- Ein konventioneller Wert ist  $b = 0.8$ .
- Man liest die für  $\pi(\delta = \delta^*, n) = b$  nötige Stichprobengröße  $n$  ab.

# Modellevaluation (8) Analyse der Powerfunktion

## Praktisches Vorgehen

$\pi(\delta = 3, n)$  für  $\alpha_0 = 0.05$



Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

**Anwendung/Praxis**

Daten einlesen:  $j = 1, \dots, 15$  für jede Gruppe

```
fname = "Daten/T-Tests_Daten.csv"  
D = read.table(fname, sep = ",", header = TRUE)
```

	X	ID	Setting	PreBDI	PostBDI	dBDI
1	1	1	F2F	29	25	4
2	2	2	F2F	32	27	5
3	3	3	F2F	28	31	-3
4	4	4	F2F	36	22	14
5	5	5	F2F	32	29	3
6	6	6	F2F	28	28	0
7	7	7	F2F	33	30	3
8	8	8	F2F	33	26	7
9	9	9	F2F	33	28	5
10	10	10	F2F	30	28	2
11	11	11	F2F	36	25	11
12	12	12	F2F	32	31	1
13	13	13	F2F	29	31	-2
14	14	14	F2F	24	29	-5
15	15	15	F2F	35	32	3
41	41	41	ONL	31	24	7
42	42	42	ONL	31	30	1
43	43	43	ONL	34	25	9
44	44	44	ONL	34	25	9
45	45	45	ONL	30	26	4
46	46	46	ONL	30	28	2
47	47	47	ONL	33	26	7
48	48	48	ONL	34	26	8
49	49	49	ONL	32	24	8
50	50	50	ONL	35	25	10
51	51	51	ONL	33	26	7
52	52	52	ONL	30	24	6
53	53	53	ONL	33	28	5
54	54	54	ONL	28	21	7
55	55	55	ONL	37	27	10

# Anwendung/Praxis

```
# Histogrammparameter
h          = 1                                # gewünschte Klassenbreite
b_0        = min(D$dBDI)                      # b_0
b_k        = max(D$dBDI)                      # b_0
k          = ceiling((b_k - b_0)/h)          # Anzahl der Klassen
b          = seq(b_0, b_k, by = h)            # Klassen [b_0, ..., b_k]
ylimits    = c(0, .2)                         # y-Achsenlimits
xlimits    = c(-2, 14)                        # x-Achsenlimits
therapie   = c("F2F", "ONL")                 # Therapiebedingungen
labs       = c("Face-to-Face", "Online")     # Abbildungslabel

# Abbildungsparameter
par(
  mfcol      = c(1,2),                        # für Details siehe ?par
  family     = "sans",                       # 1 x 2 Panelstruktur
  pty        = "m",                           # Serif-freier Fonttyp
  bty        = "l",                           # maximale Abbildungsregion
  las        = 1,                             # L-förmige Box
  xaxs       = "i",                           # horizontale Achsenbeschriftung
  yaxs       = "i",                           # x-Achse bei y = 0
  font.main  = 1,                             # y-Achse bei x = 0
  cex        = 1,                             # Titel nicht fett
  cex.main   = 1)                            # Textvergrößerungsfaktor
                                          # Titeltextvergrößerungsfaktor

# Iteration über Therapiebedingungen
for(i in 1:2){
  hist(
    D$dBDI[D$Setting == therapie[i]],         # Delta-BDI-Werte von Therapiebedingung i
    breaks = b,                               # Histogrammklassen
    freq   = F,                               # normierte relative Häufigkeit
    xlim  = xlimits,                          # x-Achsenlimits
    ylim  = ylimits,                          # y-Achsenlimits
    xlab  = TeX("$\\Delta$ BDI$"),            # x-Achsenbeschriftung
    ylab  = "geschätzte Wahrscheinlichkeit", # y-Achsenbeschriftung
    main  = labs[i])                          # Titelbeschriftung
}

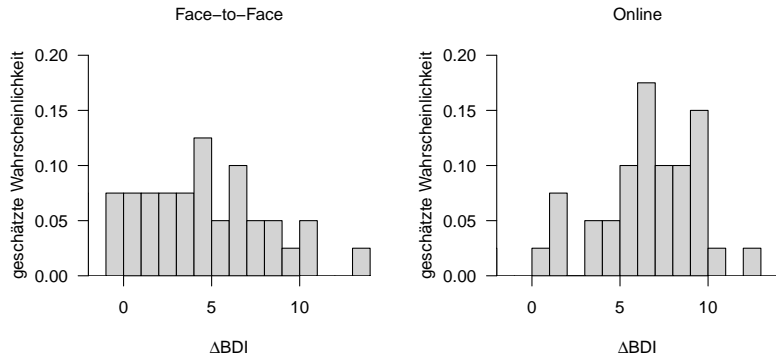
# Speichern
dev.copy2pdf(
  file    = "Abbildungen/F2F_ONL_histogramme.pdf",
  width   = 8,
  height  = 4)
```

```
# Initialisierung eines Dataframes
tp      = c("F2F", "ONL")
ntp     = length(tp)
S       = data.frame(
  n      = rep(NaN,ntp),
  Max    = rep(NaN,ntp),
  Min    = rep(NaN,ntp),
  Median = rep(NaN,ntp),
  Mean   = rep(NaN,ntp),
  Var    = rep(NaN,ntp),
  Std    = rep(NaN,ntp),
  row.names = tp)

# Therapiebedingungen
# Anzahl Therapiebedingungen
# Dataframeerzeugung
# Stichprobengrößen
# Maxima
# Minima
# Mediane
# Mittelwerte
# Varianzen
# Standardabweichungen
# Therapiebedingungen

# Iterationen über Therapiebedingungen
for(i in 1:ntp){
  data = D$BDI[D$Setting == tp[i]]
  S$n[i] = length(data)
  S$Max[i] = max(data)
  S$Min[i] = min(data)
  S$Median[i] = median(data)
  S$Mean[i] = mean(data)
  S$Var[i] = var(data)
  S$Std[i] = sd(data)
}
```

## Deskriptive Statistiken der PostBDI-PreBDI-Differenzen bei Face-to-Face- und Online-Therapie



```
# Ausgabe  
print.AsIs(S)
```

```
>      n Max Min Median Mean Var Std  
> F2F 40 14 -5     4 3.92 19.4 4.41  
> ONL 40 20 -7     7 7.40 23.8 4.88
```

# Anwendung/Praxis

```
# Modellevaluation
fname      = "Daten/T-Tests_Daten.csv"
D          = read.table(fname, sep = ",", header = TRUE)
y_1       = D$dBDI[D$Setting == "F2F"]
y_2       = D$dBDI[D$Setting == "ONL"]
n_1       = length(y_1)
n_2       = length(y_2)
n         = n_1 + n_2
p         = 2
y         = matrix(c(y_1, y_2), nrow = n)
X         = matrix(c(rep(1,n_1), rep(0,n_2),
                    rep(0,n_1), rep(1,n_2)), nrow = n)

beta_hat  = solve(t(X) %*% X) %*% t(X) %*% y
eps_hat   = y - X %*% beta_hat
sigsqr_hat = t(eps_hat) %*% eps_hat) / (n-p)
delta     = 0.95
t_delta   = qt((1+delta)/2,n-1)
lambda    = diag(solve(t(X) %*% X))
kappa     = matrix(rep(NA,n*p*2), nrow = p)
for(j in 1:p){
  kappa[j,1] = beta_hat[j]-sqrt(sigsqr_hat*lambda[j])*t_delta
  kappa[j,2] = beta_hat[j]+sqrt(sigsqr_hat*lambda[j])*t_delta
}
c         = matrix(c(1,-1), nrow = 2)
mu_0     = 0
alpha_0  = 0.05
k_alpha_0 = qt(1 - (alpha_0/2), n-1)
t_num    = t(c) %*% beta_hat - mu_0
t_den    = sqrt(sigsqr_hat*t(c) %*% solve(t(X) %*% X)%*%c)
t        = t_num/t_den
if(abs(t) >= k_alpha_0){phi = 1} else {phi = 0}
pval     = 2*(1-pt(abs(t), n_1+n_2-2))
d        = t/sqrt((n_1*n_2)/(n_1 + n_2))

# Dateiname
# Dataframe
# BDI-Differenzwerte in der F2F-Gruppe
# BDI-Differenzwerte in der ONL-Gruppe
# Anzahl Datenpunkte Gruppe 1 (F2F)
# Anzahl Datenpunkte Gruppe 2 (ONL)
# Gesamtanzahl Datenpunkte
# Anzahl Betaparameter
# Datenvektor
# Designmatrix

# Betaparameterschätzer
# Residuenvektor
# Varianzparameterschätzer
# Konfidenzbedingung
# \Psi^{-1}((1+(delta)/2,n-1)
# \lambda_j Werte
# \beta_j Konfidenzintervall-Array
# Iteration über \beta_j
# untere KI Grenze
# obere KI Grenze

# Kontrastgewichtsvektor
# Nullhypothese H_0
# Signifikanzniveau
# kritischer Wert
# T-Teststatistik Zähler
# T-Teststatistik Nenner
# T-Teststatistik
# Test 1_{|T(X)| >= k_alpha_0}
# p-Wert
# Cohen's d

> hat{beta} = 3.92 7.4
> Ig       = 78
> kappa_1  = 2.46 5.39
> kappa_2  = 5.94 8.86
> t        = -3.34
> alpha_0  = 0.05
> k_alpha_0 = 1.99
> phi      = 1
> p-Wert   = 0.00127
> Cohen's d = -0.748
```

## Anwendungsszenario

```
# automatischer Zweistichproben-T-Test
varphi = t.test(y_1, y_2,
               var.equal = TRUE,
               alternative = c("two.sided"),
               conf.level = 1-alpha_0)

# Datensatz
# \sigma_1^2 = \sigma_2^2
# H_1: \mu_1 \neq \mu_2
# \delta = 1 - \alpha_0

# Ausgabe
print(varphi)
```

```
>
> Two Sample t-test
>
> data: y_1 and y_2
> t = -3, df = 78, p-value = 0.001
> alternative hypothesis: true difference in means is not equal to 0
> 95 percent confidence interval:
>  -5.54 -1.41
> sample estimates:
> mean of x mean of y
>    3.92    7.40
```

```
# genauere Ausgabe von t
paste(varphi[1]$statistic)
```

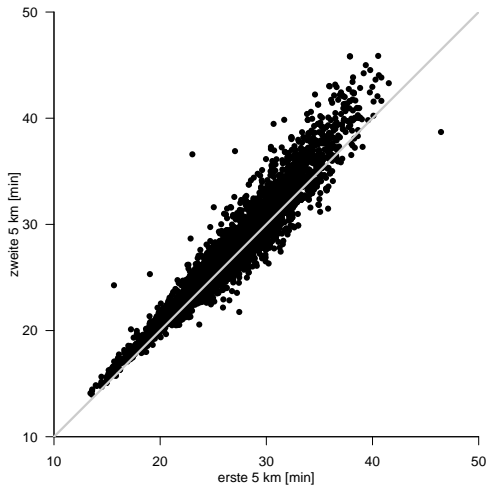
```
> [1] "-3.34424213733072"
```

```
# genauere Ausgabe von p
paste(varphi[3]$p.value)
```

```
> [1] "0.00127017790178135"
```

- R nutzt hier eine alternative Parameterisierung des Zweistichproben-T-Test Szenarios, die sogenannten *Effektdarstellung*.
- Wir werden die Effektdarstellung im Kontext der einfaktoriellen Varianzanalyse ausführlich diskutieren.

Datensatz: erste vs. zweite 5 km beim Great 10k (08.10.2017), alle Teilnehmer ( $n = 4883$ )



Zweistichproben-T-Test: erste vs. zweite 5 km, alle Teilnehmer ( $n_1 = n_2 = 4883$ )

```
# Zweistichproben-T-Test (falsch)
y1      = D$T5k1                                # "Gruppe" 1
y2      = D$T5k2                                # "Gruppe" 2
n1      = length(y1)                            # Anzahl Gruppe 1
n2      = length(y2)                            # Anzahl Gruppe 2
n       = n1 + n2                               # Anzahl Datenpunkte
p       = 2                                     # Anzahl Regressoren
y       = matrix(c(y1,y2), nrow = n)           # Datenvektor
X       = matrix(c(rep(1,n1), rep(0,n2),
                  rep(0,n1), rep(1,n2))), ncol = p) # Designmatrix
```

```
> Datenpunkte pro Gruppe      : 4883 4883
> Betaparameterschätzer      : 27.1 27.8
> Varianzparameterschätzer   : 22.2
> Zweistichproben-T-Teststatistik : -7.42
> Cohen's d                   : -0.075
> p-Wert                      : 0
```

Zweistichproben-T-Test: Differenz 5-km-Strecken, Frauen vs. Männer ( $n_1 = 1921$ ,  $n_2 = 2962$ )

```
# Zweistichproben-T-Test (richtig)
y1      = D$Tdiff[D$Sex=="F"]           # Gruppe 1
y2      = D$Tdiff[D$Sex=="M"]           # Gruppe 2
n1      = length(y1)                    # Anzahl Gruppe 1
n2      = length(y2)                    # Anzahl Gruppe 2
n       = n1 + n2                        # Anzahl Datenpunkte
p       = 2                              # Anzahl Regressoren
y       = matrix(c(y1,y2), nrow = n)     # Datenvektor
X       = matrix(c(rep(1,n1), rep(0,n2),  # Designmatrix
                  rep(0,n1), rep(1,n2))), ncol = p)
```

```
> Datenpunkte pro Gruppe      : 1921 2962
> Betaparameterschätzer      : 0.84 0.622
> Varianzparameterschätzer   : 2.2
> Zweistichproben-T-Teststatistik : 5.01
> Cohen's d                   : 0.072
> p-Wert                      : 0
```

Beispiel: Induzieren Bewegungen im Uhrzeigersinn Offenheit für neue Erfahrungen (= "psychische Zustände zeitlichen Fortschreitens und eine Orientierung auf die Zukunft und Neuartigkeit")? (Wagenmakers et al. (2015))

## Turning the hands of time again: a purely confirmatory replication study and a Bayesian analysis

Eric-Jan Wagenmakers<sup>1\*</sup>, Titia F. Beek<sup>2</sup>, Mark Rotteveel<sup>2,3</sup>, Alex Gierholz<sup>2</sup>, Dora Matzke<sup>1</sup>, Helen Steingrover<sup>1</sup>, Alexander Ly<sup>1</sup>, Josine Verhagen<sup>1</sup>, Ravi Selker<sup>1</sup>, Adam Sasiadek<sup>1</sup>, Quentin F. Gronau<sup>1</sup>, Jonathon Love<sup>1</sup> and Yair Pinto<sup>3,4</sup>

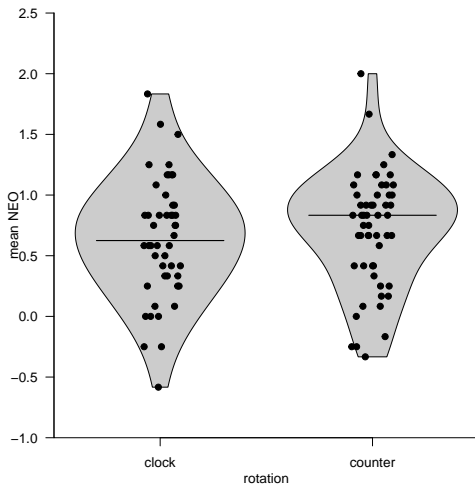
<sup>1</sup> Psychological Methods, University of Amsterdam, Amsterdam, Netherlands, <sup>2</sup> Social Psychology Program, University of Amsterdam, Amsterdam, Netherlands, <sup>3</sup> Brain and Cognition Center, University of Amsterdam, Amsterdam, Netherlands, <sup>4</sup> Cognitive Neuroscience Group, University of Amsterdam, Amsterdam, Netherlands

In a series of four experiments, Topolinski and Sparenberg (2012) found support for the conjecture that clockwise movements induce psychological states of temporal progression and an orientation toward the future and novelty. Here we report the results of a preregistered replication attempt of Experiment 2 from Topolinski and Sparenberg (2012). Participants turned kitchen rolls either clockwise or counterclockwise while answering items from a questionnaire assessing openness to experience. Data from 102 participants showed that the effect went slightly in the direction opposite to that predicted by Topolinski and Sparenberg (2012), and a preregistered Bayes factor hypothesis test revealed that the data were 10.76 times more likely under the null hypothesis than under the alternative hypothesis. Our findings illustrate the theoretical importance and practical advantages of preregistered Bayes factor replication studies, both for psychological science and for empirical work in general.

**Keywords:** preregistration, replication, Bayes factor, statistical evidence, null hypothesis testing



Datensatz: mittlere NEO-PI-R-Scores, getrennt nach Rotationsrichtung ( $n = 102$ )



Zweistichproben-T-Test: mittlere NEO-PI-R-Scores, im UZS vs. gegen den UZS ( $n_1 = 48$ ,  $n_2 = 54$ )

```
# Zweistichproben-T-Test
y1      = D$mean_NEO[D$Rotation=="clock"]      # Gruppe 1
y2      = D$mean_NEO[D$Rotation=="counter"]    # Gruppe 2
n1      = length(y1)                          # Anzahl Gruppe 1
n2      = length(y2)                          # Anzahl Gruppe 2
n       = n1 + n2                             # Anzahl Datenpunkte
p       = 2                                    # Anzahl Regressoren
y       = matrix(c(y1,y2), nrow = n)          # Datenvektor
X       = matrix(c(rep(1,n1), rep(0,n2),
                  rep(0,n1), rep(1,n2))), ncol = p) # Designmatrix
```

```
> Datenpunkte pro Gruppe      : 48 54
> Betaparameterschätzer      : 0.641 0.713
> Varianzparameterschätzer   : 0.234
> Zweistichproben-T-Teststatistik : -0.754
> Cohen's d                   : -0.075
> p-Wert                      : 0.453
```

Zweistichproben-T-Test: mittlere NEO-PI-R-Scores, Frauen vs. Männer ( $n_1 = 77$ ,  $n_2 = 25$ )

```
# Zweistichproben-T-Test
y1      = D$mean_NEO[D$Sex=="F"]           # Gruppe 1
y2      = D$mean_NEO[D$Sex=="M"]           # Gruppe 2
n1      = length(y1)                       # Anzahl Gruppe 1
n2      = length(y2)                       # Anzahl Gruppe 2
n       = n1 + n2                          # Anzahl Datenpunkte
p       = 2                                 # Anzahl Regressoren
y       = matrix(c(y1,y2), nrow = n)       # Datenvektor
X       = matrix(c(rep(1,n1), rep(0,n2),   # Designmatrix
                  rep(0,n1), rep(1,n2))), ncol = p)
```

```
> Datenpunkte pro Gruppe      : 77 25
> Betaparameterschätzer      : 0.592 0.947
> Varianzparameterschätzer   : 0.212
> Zweistichproben-T-Teststatistik : -3.35
> Cohen's d                   : -0.332
> p-Wert                      : 0.001
```

---

Überblick

Einstichproben-T-Tests

Zweistichproben-T-Tests

**Selbstkontrollfragen**

# Selbstkontrollfragen

---

1. Erläutern Sie die Extremszenarien im Kontinuum von ALM-Designs.
2. Erläutern Sie die Begriffe der faktoriellen und parametrischen ALM-Designs.
3. Nennen Sie Beispiele für faktorielle, parametrische und faktoriell-parametrische ALM-Designs.
4. Erläutern Sie das Anwendungsszenario eines Einstichproben-T-Tests.
5. Erläutern Sie mögliche Hypothesenszenarien eines Einstichproben-T-Tests.
6. Geben Sie die Definition des Einstichproben-T-Test-Modells wieder.
7. Geben Sie das Theorem zur Parameterschätzung im Einstichproben-T-Test-Modell wieder.
8. Geben Sie das Theorem zur T-Teststatistik des Einstichproben-T-Tests wieder.
9. Geben Sie die Definition des zweiseitigen Einstichproben-T-Tests (mit ungerichteter Hypothese) wieder.
10. Skizzieren Sie die Testgütefunktion des zweiseitigen Einstichproben-T-Tests.
11. Geben Sie das Theorem zur Testumfangkontrolle im zweiseitigen Level- $\alpha_0$ -Einstichproben-T-Test wieder.
12. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level- $\alpha_0$ -Einstichproben-T-Tests.
13. Geben Sie die Definition des p-Wertes Werts für einen zweiseitigen Einstichproben-T-Test wieder.
14. Von welchen Werten hängt die Powerfunktion eines zweiseitigen Einstichproben-T-Tests ab?
15. Skizzieren Sie die Powerfunktion des Einstichproben-T-Tests bei fester Stichprobengröße.
16. Skizzieren Sie die Powerfunktion des Einstichproben-T-Tests bei festem Nichtzentralitätsparameter.

# Selbstkontrollfragen

---

17. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests.
18. Geben Sie die Definition des Zweistichproben-T-Test-Modells wieder.
19. Geben Sie das Theorem zur Parameterschätzung im Zweistichproben-T-Test-Modell wieder.
20. Geben Sie das Theorem zur T-Teststatistik des Zweistichproben-T-Tests wieder.
21. Erläutern Sie mögliche Hypothesenszenarien eines Zweistichproben-T-Tests.
22. Geben Sie die Definition des zweiseitigen Zweistichproben-T-Tests (mit ungerichteter Hypothese) wieder.
23. Skizzieren Sie die Testgütfunktion des zweiseitigen Zweistichproben-T-Tests.
24. Geben Sie das Theorem zur Testumfangkontrolle im zweiseitigen Level- $\alpha_0$ -Zweistichproben-T-Test wieder.
25. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level- $\alpha_0$ -Zweistichproben-T-Tests.
26. Geben Sie die Definition des p-Wertes für einen zweiseitigen Zweistichproben-T-Test wieder.
27. Von welchen Werten hängt die Powerfunktion eines zweiseitigen Zweistichproben-T-Tests ab?
28. Skizzieren Sie die Powerfunktion des Zweistichproben-T-Tests bei fester Stichprobengröße.
29. Skizzieren Sie die Powerfunktion des Zweistichproben-T-Tests bei festem Nichtzentralitätsparameter.

- DeGroot, Morris H., and Mark J. Schervish. 2012. *Probability and Statistics*. 4th ed. Boston: Addison-Wesley.
- Wagenmakers, Eric-Jan, Titia F. Beek, Mark Rotteveel, Alex Gierholz, Dora Matzke, Helen Steingroever, Alexander Ly, et al. 2015. "Turning the Hands of Time Again: A Purely Confirmatory Replication Study and a Bayesian Analysis." *Frontiers in Psychology* 6 (April). <https://doi.org/10.3389/fpsyg.2015.00494>.