



# Allgemeines Lineares Modell

BSc Psychologie, SoSe 2026

Joram Soch

## (10) Zweistichproben-T-Tests

---

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendung/Praxis

Selbstkontrollfragen

Anhang/Details

---

## **Anwendungsszenario**

Modellformulierung

Modellschätzung

Modellevaluation

Anwendung/Praxis

Selbstkontrollfragen

Anhang/Details

**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?

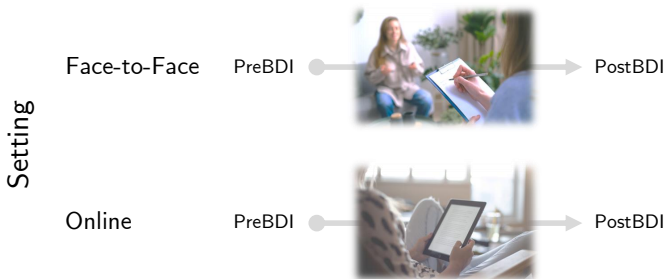
Forcierte Schwimmtest-Datenanalyse bei zwei Gruppen genmanipulierter Mäuse

- 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

Selbstkontrollfragen

Anhang/Details

## 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

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

die Datenverteilungsform

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

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

$$y = X\beta + \varepsilon \quad \text{mit} \quad X := \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \in \mathbb{R}^{n \times 2}, \quad \beta := \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2, \quad \varepsilon \sim N(0_n, \sigma^2 I_n), \quad \text{und} \quad \sigma^2 > 0. \quad (3)$$

### 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] 1.7309172 2.7968874 5.7032697 -2.5064123 0.8918282 -1.4576987
[7] 3.4459620 0.2658066 1.8743341 -2.2704443 -5.4676750 1.5470391
[13] -2.7933376 -4.1134202 0.3507336 3.7247099 -2.5752521 -2.3930230
[19] 6.9067009 3.4553580
```

---

Anwendungsszenario

Modellformulierung

**Modellschätzung**

Modellevaluation

Anwendung/Praxis

Selbstkontrollfragen

Anhang/Details

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

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

$$\hat{\beta} = \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} =: \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \quad (4)$$

und für den Varianzparameterschätzer

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2} =: s_{12}^2 \quad (5)$$

### 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  $y = (y_1^T, y_2^T)^T$  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. Das Konzept der gebündelten Stichprobenvarianz wird uns im Kontext der einfaktoriellen Varianzanalyse erneut begegnen (siehe Einheit (11) in *Allgemeines Lineares Modell*).

## Beweis

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

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T y \\ &= \left( \begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix}^T \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \right)^{-1} \begin{pmatrix} 1_{n_1} & 0_{n_2} \\ 0_{n_1} & 1_{n_2} \end{pmatrix}^T \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^{n_1} y_{1j} \\ \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \\ &= \begin{pmatrix} 1/n_1 & 0 \\ 0 & 1/n_2 \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n_1} y_{1j} \\ \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j} \\ \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j} \end{pmatrix} \\ &=: \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix}.\end{aligned}\tag{6}$$

## Beweis (fortgeführt)

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

$$\begin{aligned}\hat{\sigma}^2 &= \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n - p} \\ &= \frac{1}{n_1 + n_2 - 2} \left( \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \right)^T \left( \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \right) \\ &= \frac{1}{n_1 + n_2 - 2} \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ \vdots \\ y_{2n_2} - \bar{y}_2 \end{pmatrix}^T \begin{pmatrix} y_{11} - \bar{y}_1 \\ \vdots \\ y_{1n_1} - \bar{y}_1 \\ y_{21} - \bar{y}_2 \\ \vdots \\ y_{2n_2} - \bar{y}_2 \end{pmatrix} \\ &= \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2} \\ &=: s_{12}^2.\end{aligned}\tag{7}$$

□

# 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))

hat{beta}      : 3.925 7.4
bar{y}_1, bar{y}_2 : 3.925 7.4
hat{sigsqr}    : 21.595
s_12^2        : 21.595
s_y^2         : 24.378
```

---

Anwendungsszenario

Modellformulierung

Modellschätzung

**Modellevaluation**

Anwendung/Praxis

Selbstkontrollfragen

Anhang/Details

## 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 (12) 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 (12), um 01:58:00 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

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

dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \quad (9)$$

und es gilt

$$T \sim t(\delta, n_1 + n_2 - 2) \quad \text{mit} \quad \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right). \quad (10)$$

#### 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 Zweistichproben-T-Test-Designs,

$$d := \frac{\bar{y}_1 - \bar{y}_2}{s_{12}}. \quad (11)$$

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

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} d \Leftrightarrow d = T / \sqrt{\frac{n_1 n_2}{n_1 + n_2}}. \quad (12)$$

## 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

$$c^T \hat{\beta} - c^T \beta_0 = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} - \mu_0 = \bar{y}_1 - \bar{y}_2 - \mu_0 \quad (13)$$

und

$$c^T \beta - c^T \beta_0 = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} - \mu_0 = \mu_1 - \mu_2 - \mu_0. \quad (14)$$

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

$$c^T (X^T X)^{-1} c = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} 1/n_1 & 0 \\ 0 & 1/n_2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1/n_1 & -1/n_2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{n_1} + \frac{1}{n_2}. \quad (15)$$

Außerdem gilt

$$\left( \frac{1}{n_1} + \frac{1}{n_2} \right)^{-\frac{1}{2}} = \left( \frac{n_2}{n_1 n_2} + \frac{n_1}{n_1 n_2} \right)^{-\frac{1}{2}} = \left( \frac{n_1 + n_2}{n_1 n_2} \right)^{-\frac{1}{2}} = \left( \frac{n_1 n_2}{n_1 + n_2} \right)^{\frac{1}{2}}. \quad (16)$$

Zusammengenommen folgt dann, dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \quad \text{und} \quad \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right). \quad (17)$$

□

### 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

$$H_0 : \mu_1 - \mu_2 = \mu_0 \quad (18)$$

und

$$H_1 : \mu_1 - \mu_2 \neq \mu_0 . \quad (19)$$

Weiterhin sei die T-Teststatistik definiert durch

$$T := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \quad (20)$$

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

$$\phi(y) := \mathbf{1}_{\{|T| \geq k\}} = \begin{cases} 1 & |T| \geq k \\ 0 & |T| < k \end{cases} . \quad (21)$$

Bemerkungen

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

Gliederung (vgl. Einheit (12), um 01:58:00 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 (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

$$k_{\alpha_0} := \psi^{-1} \left( 1 - \frac{\alpha_0}{2}; n_1 + n_2 - 2 \right), \quad (22)$$

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

$$n - 1 \hookrightarrow n_1 + n_2 - 2. \quad (23)$$

## 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

$$t := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \quad (24)$$

- 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.

---

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

**Anwendung/Praxis**

Selbstkontrollfragen

Anhang/Details

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
}

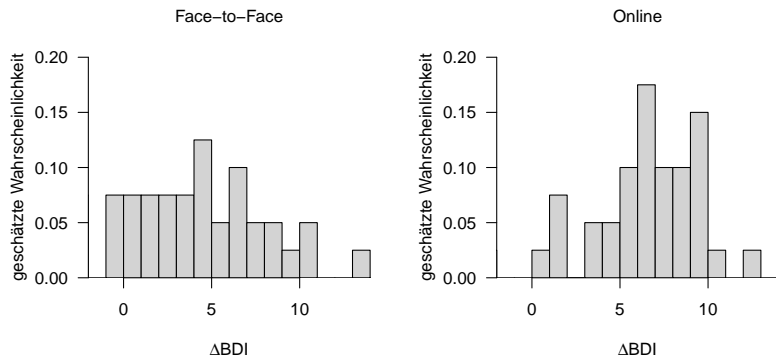
# Speichern
dev.copy2pdf(
  file     = "Abbildungen/F2F_ONL_histogramme.pdf",
  width    = 8,
  height   = 4)
```

```
# 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	3.925	19.40449	4.405052
ONL	40	20	-7	7	7.400	23.78462	4.876947

# 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} = 3.925 7.4
fg        = 78
kappa_1   = 2.462508 5.387492
kappa_2   = 5.937508 8.862492
t         = -3.344242
alpha_0   = 0.05
k_alpha_0 = 1.99045
phi       = 1
p-Wert    = 0.001270178
Cohen's d = -0.7477953
```

## 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 = -3.3442, df = 78, p-value = 0.00127
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.543688 -1.406312
sample estimates:
mean of x mean of y
 3.925      7.400
```

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

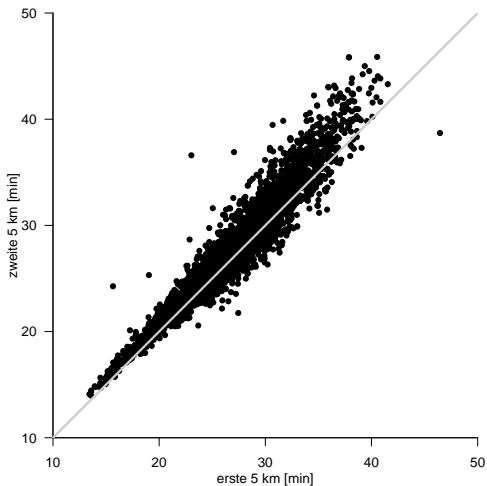
```
[1] "-3.34424213733072"
```

```
# 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      : 27.097 27.805
Varianzparameterschätzer   : 22.202
Zweistichproben-T-Teststatistik : -7.421
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  : 2.197
Zweistichproben-T-Teststatistik : 5.006
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))

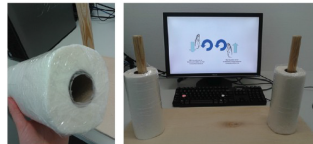
## Turning the hands of time again: a purely confirmatory replication study and a Bayesian analysis

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 Steingroever<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>

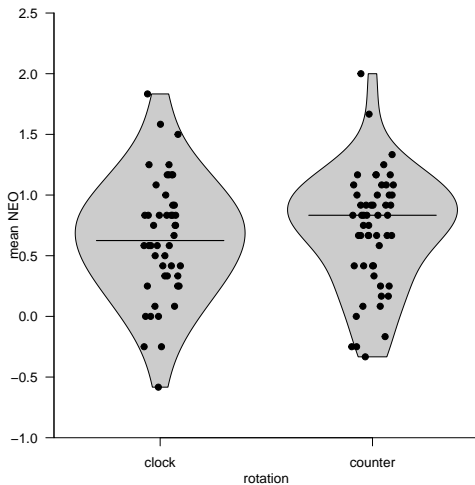
<sup>1</sup> Psychological Methods, University of Amsterdam, Amsterdam, Netherlands, <sup>2</sup> Social Psychology Program, University of Amsterdam, Amsterdam, Netherlands, <sup>3</sup> Brain and Cognition Center, University of Amsterdam, Amsterdam, Netherlands, <sup>4</sup> Cognitive Neuroscience Group,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

```
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),
                    rep(0,n1), rep(1,n2))), ncol = p) # Designmatrix
```

```
Datenpunkte pro Gruppe      : 77 25
Betaparameterschätzer      : 0.592 0.947
Varianzparameterschätzer   : 0.212
Zweistichproben-T-Teststatistik : -3.349
Cohen's d                   : -0.332
p-Wert                       : 0.001
```

---

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendung/Praxis

**Selbstkontrollfragen**

Anhang/Details

# Selbstkontrollfragen

---

1. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests.
2. Geben Sie die Definition des Zweistichproben-T-Test-Modells wieder.
3. Geben Sie das Theorem zur Parameterschätzung im Zweistichproben-T-Test-Modell wieder.
4. Geben Sie das Theorem zur T-Teststatistik des Zweistichproben-T-Tests wieder.
5. Erläutern Sie mögliche Hypothesenszenarien eines Zweistichproben-T-Tests.
6. Geben Sie die Definition des zweiseitigen Zweistichproben-T-Tests mit ungerichteter Alternativhypothese wieder.
7. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level- $\alpha_0$ -Zweistichproben-T-Tests.
8. Gegeben seien zwei Datenvektoren  $y_1$  und  $y_2$ . Unter welchen Bedingungen dürfen diese zu einem Zweistichproben-T-Test herangezogen werden? Unter welchen Bedingungen nicht?

---

Anwendungsszenario

Modellformulierung

Modellschätzung

Modellevaluation

Anwendung/Praxis

Selbstkontrollfragen

**Anhang/Details**

## Theorem (Testgütefunktion)

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

$$q_\phi : \mathbb{R}^2 \rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) := 1 - \psi(k; \delta, n_1 + n_2 - 2) + \psi(-k; \delta, n_1 + n_2 - 2), \quad (25)$$

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

$$\delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right) \quad (26)$$

und Freiheitsgradparameter  $n_1 + n_2 - 2$  bezeichnet.

## Bemerkungen

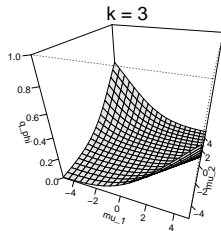
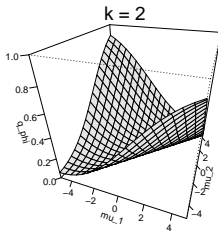
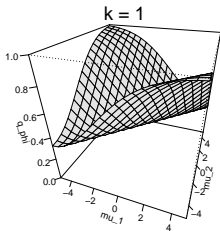
- $q_\phi$  ist eine bivariate reellwertige Funktion.
- $q_\phi$  kann alternativ als univariate reellwertige Funktion von  $\Delta := \mu_1 - \mu_2$  konzipiert werden.
- Im Vergleich zum Einstichprobenszenario gelten

$$n - 1 \hookrightarrow n_1 + n_2 - 2, \quad \sqrt{n} \hookrightarrow \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \quad \mu - \mu_0 \hookrightarrow \mu_1 - \mu_2 - \mu_0 \quad (27)$$

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

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

$$k_{\alpha_0} := \psi^{-1} \left( 1 - \frac{\alpha_0}{2}; n_1 + n_2 - 2 \right), \quad (28)$$

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

$$n - 1 \hookrightarrow n_1 + n_2 - 2. \quad (29)$$

### 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

$$t := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \quad (30)$$

- 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,

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \quad (31)$$

- 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

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n_1 + n_2 - 2)). \quad (32)$$

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

### Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$q_\psi : \mathbb{R}^2 \rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\psi(\mu_1, \mu_2) := 1 - \psi(k; \delta, n_1 + n_2 - 2) + \psi(-k; \delta, n_1 + n_2 - 2) \quad (33)$$

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

$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (\delta, n) \mapsto \pi(\delta, n) := 1 - \psi(k_{\alpha_0}; \delta, n - 2) + \psi(-k_{\alpha_0}; \delta, n - 2). \quad (34)$$

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.

### 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.

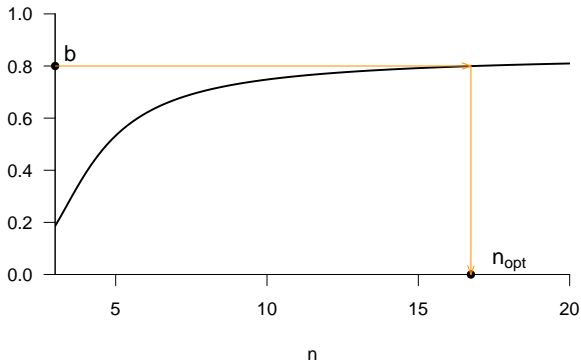
⇒ Wenn man  $\delta$  schon kennen würde, würde man den Test nicht durchführen.

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.

## Praktisches Vorgehen

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



- 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>.