



# Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

## (14) Zweistichproben-T-Tests

---

Zweistichproben-T-Tests bei unabhängigen Stichproben

Zweistichproben-T-Tests bei abhängigen Stichproben

Anwendungsbeispiele

Selbstkontrollfragen

---

## **Zweistichproben-T-Tests bei unabhängigen Stichproben**

Zweistichproben-T-Tests bei abhängigen Stichproben

Anwendungsbeispiele

Selbstkontrollfragen

# Zweistichproben-T-Test bei unabhängigen Stichproben

---

## Anwendungsszenario

- **Zwei Stichproben** experimenteller Einheiten
- Annahme unabhängiger identischer Normalverteilungen  $N(\mu_1, \sigma_1^2)$  und  $N(\mu_2, \sigma_2^2)$
- $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$  unbekannt.
- Quantifizieren der Unsicherheit beim inferentiellen Vergleich von  $\mu_1$  mit  $\mu_2$  beabsichtigt.

## Anwendungsbeispiele

- BDI Score Datenanalyse bei zwei Gruppen psychiatrischer Patient:innen
  - Gruppe 1 Psychoanalyse, Gruppe 2 Kognitive Verhaltenstherapie
  - $\mu_1 \neq \mu_2$  ?  $\Leftrightarrow$  Welche Therapie ist besser?
- **Forcierte Schwimmtestdatenanalyse bei zwei Gruppen genmanipulierter Mäuse**
  - Gruppe 1 Wildtyp, Gruppe 2 Serotoninrezeptormutation
  - $\mu_1 \neq \mu_2$  ?  $\Leftrightarrow$  Trägt Serotoninrezeptor zum Schwimmtestverhalten bei?

## Modellszenarien

- Annahme identischer Varianzen
- Annahme eines bekannten Varianzverhältnisses
- Keine Annahmen zu Varianzen

## Hypothesenszenarien

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

Wir betrachten hier exemplarisch  $H_0 : \mu_1 = \mu_2$  und  $H_1 : \mu_1 \neq \mu_2$

## Modellszenarien

- **Annahme identischer Varianzen**
- Annahme eines bekannten Varianzverhältnisses
- Keine Annahmen zu Varianzen

## Hypothesenszenarien

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

### (1) Statistisches Modell in klassischer Form

$X_{11}, \dots, X_{1n_1} \sim N(\mu_1, \sigma^2)$  sei eine Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter  $\mu_1$  und unbekanntem Varianzparameter  $\sigma^2 > 0$ .  $X_{21}, \dots, X_{2n_2} \sim N(\mu_2, \sigma^2)$  sei eine weitere Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter  $\mu_2$  und unbekanntem Varianzparameter  $\sigma^2 > 0$ . Die Varianzparameter beider Stichproben werden also als identisch vorausgesetzt. Der Parameter von Interesse ist  $(\mu_1, \mu_2)$ , der Parameterraum des Modells ist  $\Theta := \mathbb{R}^2$ .

### (2) Statistisches Modell in generativer Form

Es sei

$$X_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ f\"ur } i = 1, 2 \text{ und } j = 1, \dots, n_i \quad (1)$$

wobei

- $i$  die Stichproben indiziert,
- $j$  die experimentellen Einheiten indiziert,
- $n_i$  die Stichprobengrößen sind,
- $X_{ij}$  beobachtbare Zufallsvariablen sind,
- $\mu_i$  feste Erwartungswertparameter der Stichprobenvariablen sind, und
- $\varepsilon_{ij}$  unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen sind.



## (3) Testhypothesen, Teststatistik, Test

Wir betrachten die einfache Nullhypothese

$$H_0 : \mu_1 = \mu_2 \Leftrightarrow \Theta_0 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 \mid \mu_1 = \mu_2\} \quad (2)$$

und die zusammengesetzte Alternativhypothese

$$H_1 : \mu_1 \neq \mu_2 \Leftrightarrow \Theta_1 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 \mid \mu_1 \neq \mu_2\}. \quad (3)$$

Weiterhin betrachten wir die *Zweistichproben-T-Teststatistik*

$$T := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\bar{X}_1 - \bar{X}_2}{S_{12}} \right) \quad (4)$$

wobei für  $i = 1, 2$ 

$$\bar{X}_i := \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \text{ und } S_{12} := \sqrt{\frac{\sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)^2 + \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)^2}{(n_1 - 1) + (n_2 - 1)}} \quad (5)$$

die Stichprobenmittelwerte und die *gepoolte Stichprobenstandardabweichung*, respektive, bezeichnen. Schließlich definieren wir den zweiseitigen kritischen Wert-basierten Test

$$\phi(X) := 1_{\{|T| \geq k\}}. \quad (6)$$

## (4) Analyse der Testgütefunktion

## Theorem (Testgütefunktion)

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

$$q_\phi : \mathbb{R}^2 \rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) \\ := 1 - \psi(k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) + \psi(-k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) \quad (7)$$

wobei  $\psi(\cdot; d_{\mu_1, \mu_2}, n_1 + n_2 - 2)$  die KVF der nichtzentralen  $t$ -Verteilung mit Nichtzentralitätsparameter

$$d_{\mu_1, \mu_2} := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\mu_1 - \mu_2}{\sigma} \quad (8)$$

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

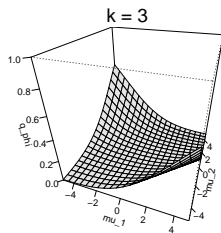
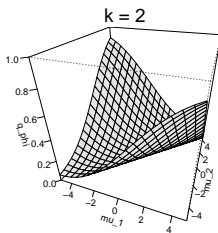
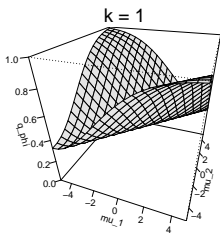
$$n \hookrightarrow n_1 + n_2 - 2, \quad \sqrt{n} \hookrightarrow \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \quad \mu - \mu_0 \hookrightarrow \mu_1 - \mu_2 \quad (9)$$

- Für einen Beweisansatz, siehe DeGroot and Schervish (2012) Seite 591.

(4) Analyse der Testgütefunktion

Testgütefunktion  $q_\phi$  für  $\sigma^2 = 9, n_1 = 12, n_2 = 12$ .

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



## (5) Testumfangkontrolle

### Theorem (Testumfangkontrolle)

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

$$k_{\alpha_0} := \psi^{-1} \left( 1 - \frac{\alpha_0}{2}; n_1 + n_2 - 2 \right), \quad (10)$$

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

$$n - 1 \hookrightarrow n_1 + n_2 - 2. \quad (11)$$

### (5) Testumfangkontrolle

#### Praktisches Vorgehen

- Man nimmt an, dass zwei vorliegende Datensätze  $x_{11}, \dots, x_{1n_1}$  und  $x_{21}, \dots, x_{2n_2}$  Realisationen von  $X_{11}, \dots, X_{1n_1} \sim N(\mu_1, \sigma^2)$  und  $X_{21}, \dots, X_{2n_2} \sim N(\mu_2, \sigma^2)$  mit unbekanntem Parametern  $\mu_1, \mu_2, \sigma^2$  sind.
- Man möchte entscheiden, ob eher  $H_0 : \mu_1 = \mu_2$  oder  $H_1 : \mu_1 \neq \mu_2$  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  $n_1 = 12, n_2 = 12$ , also Freiheitsgradparameter  $12+12-2 = 22$ , dass  $k_{0.05} = \psi^{-1}(1 - 0.05/2; 22) \approx 2.07$  ist.
- Anhand von  $n_1, n_2, \bar{x}_1, \bar{x}_2$  und der gepoolten Stichprobenstandardabweichung  $s_{12}$  berechnet man die Realisierung der Zweistichproben-T-Teststatistik

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

- 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.

### (6) p-Werte

#### 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,

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \quad (13)$$

- 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

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n_1 + n_2 - 2)). \quad (14)$$

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

## (7) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$\begin{aligned} q_\phi : \mathbb{R}^2 &\rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) \\ &:= 1 - \psi(k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) + \psi(-k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) \quad (15) \end{aligned}$$

als Funktion des Nichtzentralitätsparameters 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

$$\begin{aligned} \pi : \mathbb{R} \times \mathbb{N}^2 &\rightarrow [0, 1], (d, n) \mapsto \\ \pi(d, n) &:= 1 - \psi(k_{\alpha_0}; d, n - 2) + \psi(-k_{\alpha_0}; d, n - 2) \quad (16) \end{aligned}$$

Bei festgelegten  $\alpha_0$  hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekanntem Wert  $d$  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.

### (7) 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 Stichproben auch zur Varianzschätzung in der anderen Stichprobe beitragen, da eine identische Varianz vorausgesetzt wurde.

Die Powerfunktion hängt vom wahren, aber unbekanntem, Parameterwert  $d = \sqrt{n}(\mu_1 - \mu_0)/\sigma$  ab.

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

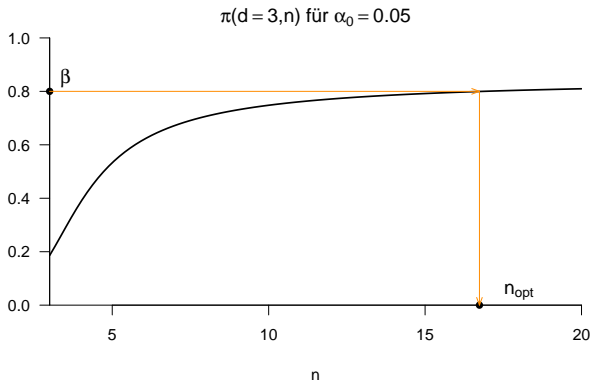
Generell wird folgendes Vorgehen favorisiert

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



## (7) Analyse der Powerfunktion

### Praktisches Vorgehen



## Modellszenarien

- Annahme identischer Varianzen
- **Annahme eines bekannten Varianzverhältnisses**
- Keine Annahmen zu Varianzen

## Hypothesenszenarien

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

## Überblick

- Die obige Theorie kann für eine bekannte Konstante  $k > 0$  auf den Fall des statistischen Modells

$$X_{11}, \dots, X_{1n_1} \sim N(\mu_1, \sigma_1^2) \text{ und } X_{21}, \dots, X_{2n_2} \sim N(\mu_2, k\sigma_1^2) \quad (17)$$

erweitert werden.

- In diesem Fall hat die Zweistichproben-T-Teststatistik definiert durch (cf. DeGroot and Schervish (2012), S. 593)

$$T := \frac{(n_1 + n_2 - 2)^{\frac{1}{2}} (\bar{X}_1 - \bar{X}_2)}{\left(\frac{1}{n_1} + \frac{k}{n_2}\right)^{\frac{1}{2}} \left(S_1^2 + \frac{S_2^2}{k}\right)^{\frac{1}{2}}} \quad (18)$$

mit den jeweiligen Stichprobenmittelwerten  $\bar{X}_1$  und  $\bar{X}_2$ , sowie den Stichprobenvarianzen  $S_1^2$  und  $S_2^2$  eine nichtzentrale  $t$ -Verteilung mit Freiheitsgradparameter  $n_1 + n_2 - 2$ .

- Wir wollen diesen Fall hier nicht weiter vertiefen.

## Modellszenarien

- Annahme identischer Varianzen
- Annahme eines bekannten Varianzverhältnisses
- **Keine Annahmen zu Varianzen**

## Hypothesenszenarien

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

## Überblick

- Wenn im Rahmen des statistischen Modells

$$X_{11}, \dots, X_{1n_1} \sim N(\mu_1, \sigma_1^2) \text{ und } X_{21}, \dots, X_{2n_2} \sim N(\mu_2, \sigma_2^2) \quad (19)$$

alle vier Parameter  $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$  sowie das Varianzverhältnis  $k = \sigma_2^2 / \sigma_1^2$  als unbekannt vorausgesetzt werden, **kann obige Theorie nicht auf diesen Fall angewendet werden**. Das Problem der statistischen Inferenz wird in diesem Fall als **Behrens-Fisher Problem** bezeichnet.

- Eine populäre Methode zur statistischen Inferenz im Behrens-Fisher Problem geht auf Welch (1938) zurück. Diese Methode bedient sich der Teststatistik

$$\tilde{T} := (\bar{X}_1 - \bar{X}_2) / \left( \left( \frac{S_1^2}{n_1(n_1 - 1)} + \frac{S_2^2}{n_2(n_2 - 1)} \right)^{1/2} \right), \quad (20)$$

wobei  $S_1^2$  und  $S_2^2$  die Stichprobenvarianzen bezeichnen. Die Verteilung von  $\tilde{T}$  ist unbekannt, kann aber mit einer (nichtzentralen)  $t$ -Verteilung approximiert werden.

- Der Freiheitsgradparameter dieser approximierenden (nichtzentralen)  $t$ -Verteilung ergibt sich dabei mithilfe der beobachteten Werte  $s_1^2$  und  $s_2^2$  von  $S_1^2$  und  $S_2^2$ , respektive, zu

$$\nu = \left( \frac{s_1^2}{n_1(n_1 - 1)} + \frac{s_2^2}{n_2(n_2 - 1)} \right) / \left( \frac{1}{(n_1 - 1)^3} \left( \frac{s_1^2}{n_1} \right)^2 + \frac{1}{(n_2 - 1)^3} \left( \frac{s_2^2}{n_2} \right)^2 \right). \quad (21)$$

- Wir wollen die theoretische Begründung dieses Vorgehens hier nicht weiter vertiefen.

---

Zweistichproben-T-Tests bei unabhängigen Stichproben

**Zweistichproben-T-Tests bei abhängigen Stichproben**

Anwendungsbeispiele

Selbstkontrollfragen

## Anwendungsszenario

- **Zwei Messungen an einer Gruppe experimenteller Einheiten.**
- Annahme "abhängiger" Normalverteilungen  $N(\mu_1, \sigma_1^2)$  und  $N(\mu_2, \sigma_2^2)$
- $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$  unbekannt.
- Quantifizieren der Unsicherheit beim inferentiellen Vergleich von  $\mu_1$  mit  $\mu_2$  beabsichtigt.

## Anwendungsbeispiele

- BDI Datenanalyse bei Psychotherapie
  - Erste Messung vor Psychotherapie, zweite Messung nach Psychotherapie
  - $\mu_1 \neq \mu_2?$   $\Leftrightarrow$  Hat Psychotherapie einen Einfluss auf BDI?
- Reaktionszeitdatenanalyse zur visuellen Wahrnehmung
  - Bedingung 1: Hoher Kontrast, Bedingung 2: Niedriger Kontrast
  - $\mu_1 \neq \mu_2?$   $\Leftrightarrow$  Hat Stimuluskontrast Einfluss auf perzeptuelle Entscheidungen?

# Zweistichproben-T-Test bei abhängigen Stichproben

## (1) Statistisches Modell in generativer Form

Es sei

$$X_{ij} = \mu_i + \beta_j + \varepsilon_{ij} \text{ mit } \beta_j \sim N(0, \sigma_\beta^2) \text{ und } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \text{ für } i = 1, 2, j = 1, \dots, n, \quad (22)$$

wobei

- $i$  die Messungen (abhängige Stichproben) und  $j$  die experimentellen Einheiten indizieren,
- $X_{ij}$  beobachtbare Zufallsvariablen sind,
- $\mu_i$  feste Erwartungswertparameter der Messungen sind,
- $\beta_j$  unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen sind, und
- $\varepsilon_{ij}$  unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen sind.

Alle  $\beta_i$  und  $\varepsilon_{ij}$  werden als voneinander unabhängig vorausgesetzt.

- Die  $X_{ij}$  modellieren Datenpunkte,
- die  $\mu_i$  modellieren feste Messungs-abhängige Effekte,
- die  $\beta_j$  modellieren zufällige experimentelle Einheits-abhängige Effekte,
- die  $\varepsilon_{ij}$  modellieren (wie immer) zufällige Störeffekte.

Lineare Modelle mit festen Effekten (fixed effects) und zufälligen Effekten (random effects) heißen gemischte Modelle (linear mixed models).



# Zweistichproben-T-Test bei abhängigen Stichproben

## (1) Statistisches Modell in generativer Form

### Datengeneration

```
# Modellspezifikation
n           = 12
mu          = c(10,12)
sigsqr_beta = 5
sigsqr_veps = 1

# Stichprobengröße
# feste Effektparameter \mu_1, mu_2
# zufälliger Effektparameter \sigsqr_beta
# Störvariablenparameter \sigsqr_varepsilon

# Datengeneration
X = data.frame(
  mu_1 = rep(NaN,n),
  mu_2 = rep(NaN,n),
  beta_j = rep(NaN,n),
  veps_1_j = rep(NaN,n),
  veps_2_j = rep(NaN,n),
  X_1j = rep(NaN,n),
  X_2j = rep(NaN,n)
)

# Datensatzinitialisierung
# \mu_1
# \mu_2
# \beta_j,
# \varepsilon_{1j},
# \varepsilon_{2j},
# X_{1j}, j = 1, \dots, n
# X_{2j}, j = 1, \dots, n

# Iterationen über experimentelle Einheiten
for(j in 1:n){
  beta_j = rnorm(1,0,sqrt(sigsqr_beta)) # zufälliger experimenteller Einheits-abhängiger Effekt
  X[[j,3]] = beta_j # \beta_j

  # Iterationen über Messungen
  for(i in 1:2){
    eps_ij = rnorm(1,0,sqrt(sigsqr_veps)) # zufälliger Störvariablenwert
    X[[j,i]] = mu[i] # \mu_i
    X[[j,3+i]] = eps_ij # \varepsilon_{ij}
    X[[j,5+i]] = mu[i] + beta_j + eps_ij # X_{ij} = \mu_i + \beta_j + \varepsilon_{ij}
  }
}
```

# Zweistichproben-T-Test bei abhängigen Stichproben

## (1) Statistisches Modell in generativer Form

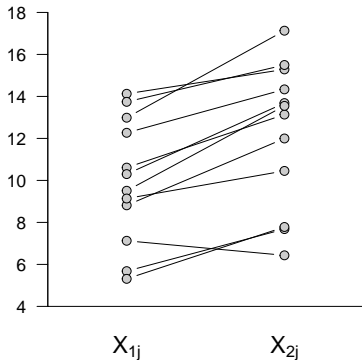
Datenbeispiel

$\mu_{1j}$	$\mu_{2j}$	$\beta_j$	$\text{veps}_{1j}$	$\text{veps}_{2j}$	$X_{1j}$	$X_{2j}$
10	12	-3.131	0.255	-2.437	7.12	6.43
10	12	-0.012	0.622	1.148	10.61	13.14
10	12	-4.074	-0.247	-0.244	5.68	7.68
10	12	-0.632	-0.554	0.629	8.81	12.00
10	12	4.618	-1.631	0.512	12.99	17.13
10	12	-4.166	-0.522	-0.053	5.31	7.78
10	12	1.214	-0.914	0.468	10.30	13.68
10	12	0.812	-1.305	0.738	9.51	13.55
10	12	4.223	-0.097	-0.936	14.12	15.29
10	12	-0.036	-0.827	-1.512	9.14	10.45
10	12	2.092	0.176	0.244	12.27	14.34
10	12	3.630	0.112	-0.134	13.74	15.50

# Zweistichproben-T-Test bei abhängigen Stichproben

## (1) Statistisches Modell in generativer Form

Datenbeispiel



# Zweistichproben-T-Test bei abhängigen Stichproben

## (1) Statistisches Modell in generativer Form

### Theorem (Verteilung der Differenzvariablen)

Das oben formulierte statistische Modell eines Zweistichproben-T-Test bei abhängigen Stichproben sei gegeben und für  $j = 1, \dots, n$  sei  $X_j := X_{j2} - X_{j1}$  die Differenzvariable der Zufallsvariablen  $X_{j1}$  und  $X_{j2}$ . Dann gilt

$$X_1, \dots, X_n \sim N(\mu, \sigma^2) \text{ mit } \mu := \mu_2 - \mu_1 \text{ und } \sigma^2 := 2\sigma_\epsilon^2. \quad (23)$$

#### Bemerkungen

- Hinsichtlich der Differenzvariablen  $X_j$  handelt es sich um das Einstichproben-T-TestszENARIO.
- Die Hypothesen des Zweistichproben-T-Test bei abhängigen Stichproben hinsichtlich der Erwartungswertparameter  $\mu_1$  und  $\mu_2$  übertragen sich wie folgt auf die Einstichproben-T-Test Hypothesen in bezug auf den Erwartungswertparameter  $\mu$ :

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

# Zweistichproben-T-Test bei abhängigen Stichproben

## (1) Statistisches Modell in generativer Form

### Beweis

Wir halten zunächst fest, dass

$$\begin{aligned}X_j &= X_{2j} - X_{1j} \\ &= \mu_2 + \beta_j + \varepsilon_{2j} - \mu_1 - \beta_j - \varepsilon_{1j} \\ &= \mu_2 - \mu_1 + \varepsilon_{2j} - \varepsilon_{1j} \\ &= \mu + \varepsilon_j,\end{aligned}\tag{24}$$

wobei wir  $\mu := \mu_2 - \mu_1$  und  $\varepsilon_j := \varepsilon_{2j} - \varepsilon_{1j}$  definiert haben.

In Generalisierung des Summentransformationstheorems für normalverteilte Zufallsvariablen (cf. (6) Transformationen der Normalverteilung) halten wir weiterhin fest, dass für unabhängige Zufallsvariablen  $X_i \sim N(\mu_i, \sigma_i^2)$ ,  $i = 1, \dots, n$  und Koeffizienten  $a_i$ ,  $i = 1, \dots, n$  gilt, dass

$$\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)\tag{25}$$

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

$$\varepsilon_j \sim N\left(0, 2\sigma_\varepsilon^2\right)\tag{26}$$

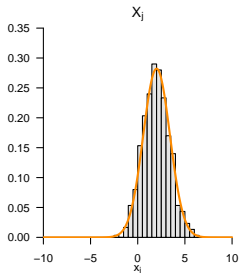
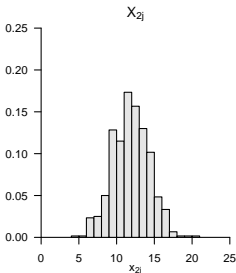
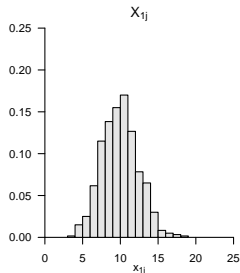
Mit der Definition  $\sigma^2 := 2\sigma_\varepsilon^2$  und dem Transformationstheorem für normalverteilte Zufallsvariablen bei linear-affiner Transformation folgt das Resultat dann direkt.

# Zweistichproben-T-Test bei abhängigen Stichproben

## (1) Statistisches Modell in generativer Form

Visualisierung des Zweistichproben-T-Test Theorems

$$n = 12, \mu_1 = 10, \mu_2 = 12, \sigma_\beta^2 = 5, \sigma_\varepsilon^2 = 1$$



### (3) Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz von Messpaaren  $x_{11}$  und  $x_{21}$ ,  $x_{12}$  und  $x_{22}$ , ...,  $x_{1n}$  und  $x_{2n}$  eine Realisation von  $X_{ij}$ ,  $i = 1, 2, j = 1, \dots, n$  wie in der Beschreibung des generativen Modells definiert ist.
  - Man bildet die Differenzwerte  $x_1 = x_{21} - x_{11}$ ,  $x_2 = x_{22} - x_{12}$ , ...,  $x_n = x_{2n} - x_{1n}$ .
- ⇒ Man nutzt die Theorie des Einstichproben-T-Tests zum Testen der Hypothesen des Zweistichproben-T-Tests mit abhängigen Stichproben.

## Zweistichproben-T-Test bei abhängigen Stichproben

### (3) Praktisches Vorgehen

#### Datenbeispiel

$\mu_1$	$\mu_2$	$\beta_j$	$\text{veps}_{1_j}$	$\text{veps}_{2_j}$	$X_{1j}$	$X_{2j}$	$X_j$
10	12	-4.271	-0.279	-0.313	5.45	7.42	1.966
10	12	2.387	0.070	-0.639	12.46	13.75	1.291
10	12	-0.112	-0.251	0.445	9.64	12.33	2.696
10	12	6.161	0.047	0.578	16.21	18.74	2.531
10	12	0.264	-1.912	0.862	8.35	13.13	4.774
10	12	-0.544	-0.206	0.019	9.25	11.47	2.225
10	12	0.066	0.550	-2.274	10.62	9.79	-0.824
10	12	5.998	-0.361	0.213	15.64	18.21	2.575
10	12	2.402	-0.665	1.114	11.74	15.52	3.779
10	12	-0.550	-1.178	-0.976	8.27	10.47	2.202
10	12	2.382	0.132	0.489	12.51	14.87	2.357
10	12	-3.800	-1.471	0.284	4.73	8.48	3.755

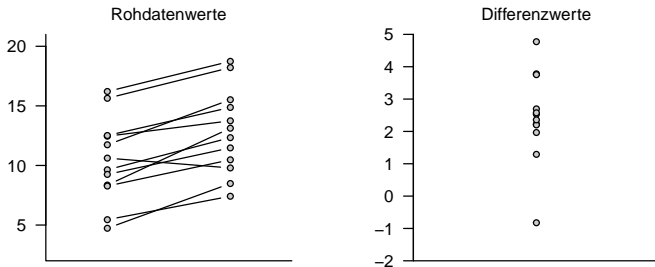
⇒ Einstichproben-T-Test basierend auf  $X_j, j = 1, \dots, n$



# Zweistichproben-T-Test bei abhängigen Stichproben

## (3) Praktisches Vorgehen

### Datenbeispiel



⇒ Einstichproben-T-Test basierend auf  $X_j, j = 1, \dots, n$

---

Zweistichproben-T-Tests bei unabhängigen Stichproben

Zweistichproben-T-Tests bei abhängigen Stichproben

## **Anwendungsbeispiele**

Selbstkontrollfragen

---

Zweistichproben-T-Tests bei unabhängigen Stichproben

Zweistichproben-T-Tests bei abhängigen Stichproben

Anwendungsbeispiele

**Selbstkontrollfragen**

# Selbstkontrollfragen

---

1. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests bei unabhängigen Stichproben.
2. Nennen Sie drei Modellszenarien für Zweistichproben-T-Tests bei unabhängigen Stichproben.
3. Geben Sie das statistische Modell eines Zweistichproben-T-Tests bei unabhängigen Stichproben in klassischer Form an.
4. Geben Sie das statistische Modell eines Zweistichproben-T-Tests bei unabhängigen Stichproben in generativer Form an.
5. Definieren Sie die gepoolte Stichprobenstandardabweichung.
6. Definieren Sie die Zweistichproben-T-Teststatistik.
7. Definieren Sie den kritischen Wert für einen zweiseitigen Level- $\alpha_0$ -Zweistichproben-T-Tests bei unabhängigen Stichproben.
8. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level- $\alpha_0$ -Zweistichproben-T-Tests bei unabhängigen Stichproben.
9. Erläutern Sie das Szenario eines Zweistichproben-T-Tests bei unabhängigen Stichproben mit bekanntem Varianzverhältnis.
10. Definieren Sie das Behrens-Fisher Problem.
11. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests bei abhängigen Stichproben.
12. Geben Sie das statistische Modell eines Zweistichproben-T-Tests bei abhängigen Stichproben in generativer Form an.
13. Geben Sie das Theorem zur Verteilung der Differenzvariablen im Zweistichproben-T-Tests bei abhängigen Stichproben wieder.
14. Erläutern Sie das praktische Vorgehen bei Durchführung eines Zweistichproben-T-Tests bei abhängigen Stichproben.

## References

---

DeGroot, Morris H., and Mark J. Schervish. 2012. *Probability and Statistics*. 4th ed. Boston: Addison-Wesley.

Welch, B. L. 1938. "The Significance of the Difference Between Two Means When the Population Variances Are Unequal." *Biometrika* 29 (3-4): 350–62. <https://doi.org/10.1093/biomet/29.3-4.350>.