Tutorium

Allgemeines Lineares Modell

BSc Psychologie SoSe 2023

10. Termin: (9) T-Tests

Sean Mulready

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-α_0-Einstichproben-T-Test wieder.
12. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level-α_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 zweiseitigen Einstichproben-T-Tests bei fester Stichprobengröße.
16. Skizzieren Sie die Powerfunktion des zweiseitigen 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. Geben Sie das Theorem zur Testumfangkontrolle im zweiseitigen Level-α_0-Zweistichproben-T-Test wieder.
24. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level-α_0-Zweistichproben-T-Tests.
25. Geben Sie die Definition des p-Wertes Werts für einen zweiseitigen Zweistichproben-T-Test wieder.

1. Erläutern Sie die Extremszenarien im Kontinuum von ALM Designs.

Extremszenario (1) Die Erwartungswerte aller Datenvariablen sind identisch.

$$\upsilon_i \sim N(\mu, \sigma^2) \text{ u.i.v. für } i = 1, \ldots, n \Leftrightarrow$$
$$\upsilon = X\beta + \varepsilon, X := 1_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R}, \varepsilon \sim N(0_n, \sigma^2 I_n) \qquad (1)$$

$\Rightarrow$ Jegliche Datenvariabilität wird dem Fehlerterm zugeschrieben.

Extremszenario (2) Die Erwartungswerte aller Datenvariablen sind paarweise verschieden

$$\upsilon_i \sim N(\mu_i, \sigma^2) \text{ u.v. für } i = 1, \ldots, n \Leftrightarrow$$
$$\upsilon = X\beta + \varepsilon \text{ mit } X := I_n \in \mathbb{R}^{n \times n}, \beta := (\mu_1, \ldots, \mu_n)^T \in \mathbb{R}^n, \varepsilon \sim N(0_n, \sigma^2 I_n) \qquad (2)$$

$\Rightarrow$ Jegliche Datenvariabilität wird dem Erwartungswertparameter zugeschrieben.

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

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

Anmerkung:

- Mit Datenvariablen sind hier die Komponenten des Datenvektors υ, also die υ_i gemeint

Wiederholung: ALM Modelle u.i.v. und für $p = 2$ ausgeschrieben

ALM für u.i.v. Zufallsvektor mit Erwartungswertparameter $\mu \in \mathbb{R}$ und Varianzparameter σ^2. D.h., für jede Komponente des Datenvektors gilt $\upsilon_i \sim N(\mu, \sigma^2)$ für $i = 1, \dots, n$. Äquivalent dazu ($\Leftrightarrow$) in Matrixschreibweise

$$\upsilon \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := 1_n \in \mathbb{R}^{n \times 1}, \beta := \mu \in \mathbb{R}^1, \varepsilon \sim N(0_n, \sigma^2 I_n).$$

$$\upsilon = X\beta + \varepsilon \Leftrightarrow \begin{pmatrix} \upsilon_1 \\ \vdots \\ \upsilon_n \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \mu + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} \mu + \varepsilon_1 \\ \vdots \\ \mu + \varepsilon_n \end{pmatrix}$$

ALM für u.v. Zufallsvektor mit "individuellen" Erwartungswertparametern μ_i und Varianzparameter σ^2. D.h., für jede Komponente des Datenvektors gilt $y_i \sim N(\mu_i, \sigma^2)$ für $i = 1, \dots, n$. Für ein Beispiel mit p = 2 ($\beta = \begin{pmatrix} \beta_1 & \beta_2 \end{pmatrix}^T$) gilt äquivalent dazu in Matrixschreibweise

$$\upsilon \sim N(X\beta, \sigma^2 I_n) \text{ mit } X \in \mathbb{R}^{n \times 2}, \beta \in \mathbb{R}^2, \varepsilon \sim N(0_n, \sigma^2 I_n).$$

$$\upsilon = X\beta + \varepsilon \Leftrightarrow \begin{pmatrix} \upsilon_1 \\ \vdots \\ \upsilon_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_5 \end{pmatrix} = \begin{pmatrix} x_{11}\beta_1 + x_{12}\beta_2 + \varepsilon_1 \\ \vdots \\ x_{n1}\beta_1 + x_{n2}\beta_2 + \varepsilon_n \end{pmatrix}$$

Extremszenarien aus SKF 1 ausgeschrieben

Extremszenario (1) Die Erwartungswerte aller Datenvariablen sind identisch $\Rightarrow$ wir haben nur ein μ für alle Datenvariablen. $\Rightarrow$ Jegliche Datenvariabilität wird dem Fehlerterm zugeschrieben

$$\upsilon \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := 1_n \in \mathbb{R}^{n\times 1}, \beta := \mu \in \mathbb{R}^1, \varepsilon \sim N(0_n, \sigma^2 I_n).$$

$$\upsilon = X\beta + \varepsilon \Leftrightarrow \begin{pmatrix} \upsilon_1 \\ \upsilon_2 \\ \vdots \\ \upsilon_n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \mu + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} \mu + \varepsilon_1 \\ \mu + \varepsilon_2 \\ \vdots \\ \mu + \varepsilon_n \end{pmatrix}$$

Extremszenario (2) Die Erwartungswerte aller Datenvariablen sind paarweise verschieden $\Rightarrow$ jede Datenvariable υ_i hat ein "individuelles" μ_i. $\Rightarrow$ Jegliche Datenvariabilität wird dem Erwartungswertparameter zugeschrieben. $\Rightarrow$ Es gilt $\hat{\beta} = \upsilon$ und $\hat{\sigma}^2 = 0$.

$$\upsilon \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := I_n \in \mathbb{R}^{n\times n}, \beta := (\mu_1, \dots, \mu_n)^T \in \mathbb{R}^2, \varepsilon \sim N(0_n, \sigma^2 I_n).$$

$$\upsilon = X\beta + \varepsilon \Leftrightarrow \begin{pmatrix} \upsilon_1 \\ \upsilon_2 \\ \vdots \\ \upsilon_n \end{pmatrix} = \begin{pmatrix} 1 & \dots & 0 \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 1\mu_1 + 0\mu_2 + \dots + 0\mu_n + 0 \\ 0\mu_1 + 1\mu_2 + \dots + 0\mu_n + 0 \\ \vdots \\ 0\mu_1 + 0\mu_2 + \dots + 1\mu_n + 0 \end{pmatrix}$$

Überblick - faktoriell vs. parametrisch

2. Erläutern Sie die Begriffe der faktoriellen und parametrischen ALM Designs.

Faktorielle ALM Designs

- Designmatrizen mit 1en und 0en, manchmal -1en.
- Betaparameter repräsentieren Gruppenerwartungswerte.
- Betaparameterschätzer repräsentieren Gruppenstichprobenmittel.

Parametrische ALM Designs

- Designmatrizen besitzen Spalten mit kontinuierlichen reellen Werten.
- Die Designmatrixsspalten werden *Regressoren*, *Prädiktoren*, oder *Kovariaten* genannt.
- Betaparameter repräsentieren Steigungsparameter.
- Betaparameterschätzer ergeben sich als normalisierte Regressor-Daten Kovarianzen.
- Es besteht ein enger Bezug zur Theorie der Korrelation.

Faktoriell-parametrische ALM Designs

- Designmatrizen mit mehreren faktoriellen und parametrischen Werten.
- Die parametrischen Regressoren werden oft als kontrollierte Kovariaten betrachtet.

3. Nennen Sie Beispiele für faktorielle, parametrische, und faktoriell-parametrische ALM Designs.

Faktorielle ALM Designs

- T-Tests, Einfaktorielle Varianzanalyse, Mehrfaktorielle Varianzanalyse
- Anwendungsbeispiel: Gruppenvergleich Treatment vs. No-Treatment

Parametrische ALM Designs

- Einfache lineare Regression, Multiple lineare Regression
- Anwendungsbeispiel: Effekt von Medikamentendosierung und/oder Anzahl Therapiestunden auf Symptomreduktion

Faktoriell-parametrische ALM Designs

- Kovarianzanalyse
- Anwendungsbeispiel: Kombination Treatment vs. No-Treatment und Medikamentendosierung

4. Erläutern Sie das Anwendungsszenario eines Einstichproben-T-Tests.

Im Anwendungsszenario eines Einstichproben-T-Tests betrachten wir **eine Gruppe** (Stichprobe) randomisierter experimenteller Einheiten. Wir nehmen an, dass die Datenpunkte unabhängig und identisch normalverteilte Realisierungen von ZVen ($\upsilon_i \sim N(\mu, \sigma^2)$) sind, wobei μ und σ^2 unbekannt sind.

Wir sind daran interessiert die *Unsicherheit*, die mit dem inferentiellen Vergleich von μ und μ_0 verbunden ist, im Sinne eines Hypothesentests zu *quantifizieren*.

5. Erläutern Sie mögliche Hypothesenszenarien eines Einstichproben-T-Tests.

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

Einstichproben-T-Tests - Modellformulierung

6. Geben Sie die Definition des Einstichproben-T-Test-Modells wieder.

Definition (Einstichproben-T-Test-Modell)

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

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

die Datenverteilungsform

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

und für den Datenvektor $\upsilon = (\upsilon_1, \ldots, \upsilon_n)^T$ die Designmatrixform

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

Anmerkung

- Das Modell ist identisch mit dem Modell unabhängiger und identisch normalverteilter Zufallsvariablen.

7. Geben Sie das Theorem zur Parameterschätzung im Einstichproben-T-Test-Modell wieder.

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

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

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} \upsilon_i =: \bar{\upsilon}, \tag{6}$$

und für den Varianzparameterschätzer

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n} (\upsilon_i - \bar{\upsilon})^2 =: s_\upsilon^2 \tag{7}$$

8. Geben Sie das Theorem zur T-Teststatistik des Einstichproben-T-Tests wieder.

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

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

$$c := 1 \text{ und } c^T \beta_0 =: \mu_0, \tag{8}$$

dass

$$T = \sqrt{n} \left(\frac{\bar{\upsilon} - \mu_0}{s_\upsilon} \right) \tag{9}$$

und es gilt

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

Theorem (T-Teststatistik)

Es sei

$$\upsilon = X\beta + \varepsilon \text{ mit } \varepsilon \sim N(0_n, \sigma^2 I_n) \tag{11}$$

das ALM. Weiterhin seien

$$\hat{\beta} := (X^T X)^{-1} X^T \upsilon \text{ und } \hat{\sigma}^2 := \frac{(\upsilon - X\hat{\beta})^T(\upsilon - X\hat{\beta})}{n-p} \tag{12}$$

die Betaparameter- und Varianzparameterschätzer, respektive. Schließlich sei für einen Kontrastgewichtsvektor $c \in \mathbb{R}^p$ und einen Parameter $\beta_0 \in \mathbb{R}^p$

$$T := \frac{c^T\hat{\beta} - c^T\beta_0}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}. \tag{13}$$

die T-Statistik. Dann gilt

$$T \sim t(\delta, n-p) \text{ mit } \delta := \frac{c^T\beta - c^T\beta_0}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}}. \tag{14}$$

Herleitung (Beweis) der T-Teststatistik des Einstichproben-T-Tests

Mit dem T-Teststatistik Theorem in Einheit (7) T-Statistiken gilt

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

Weiterhin gilt mit demselben Theorem

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

Veranschaulichung Einstichproben-T-Test mit $n = 5$

Anmerkung: Theoretische Aspekte sind grau und mit Daten geschätzte Größen sind orange

Modellformulierung (Wie bei ALM für u.i.v. ZVen)

$$\upsilon \sim N(X\beta, \sigma^2 I_n) \text{ mit } X := 1_5 \in \mathbb{R}^{5\times 1}, \beta := \mu \in \mathbb{R}^1, \varepsilon \sim N(0_5, \sigma^2 I_5).$$

$$\upsilon = X\beta + \varepsilon \Leftrightarrow \begin{pmatrix} \upsilon_1 \\ \upsilon_2 \\ \upsilon_3 \\ \upsilon_4 \\ \upsilon_5 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix} = \begin{pmatrix} \mu + \varepsilon_1 \\ \mu + \varepsilon_2 \\ \mu + \varepsilon_3 \\ \mu + \varepsilon_4 \\ \mu + \varepsilon_5 \end{pmatrix}$$

Modellschätzung

$$\hat{\beta} = \bar{\upsilon}, \text{ und } \hat{\sigma}^2 = s_{\upsilon}^2$$

Veranschaulichung Einstichproben-T-Test mit $n = 5$ (fortgeführt)

Modellevaluation

Für die Modellevaluation wollen wir die T-Teststatistik berechnen. Dafür wählen wir im Fall eines Einstichproben-T-Tests einen Kontrastgewichtsvektor von $c := 1$, wobei $c^T \beta_0 = \mu_0$.

Wir berechnen die T-Teststatistik T mit den Stichproben-Daten mit der Formel

$$T = \sqrt{n} \left(\frac{\bar{\upsilon} - \mu_0}{s_\upsilon} \right),$$

wobei wir eine Theorie darüber haben, wie T verteilt ist. Nämlich

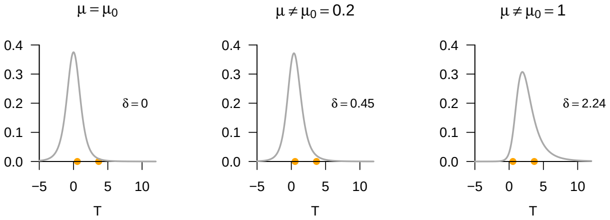
$$T \sim t(\delta, 5 - 1) \text{ mit } \delta = \sqrt{5} \left(\frac{\mu - \mu_0}{\sigma} \right)$$

Veranschaulichung Einstichproben-T-Test mit $n = 5$ (fortgeführt)

Wir ziehen zwei mögliche Szenarien davon, was der wahre, aber unbekannte Parameter und die jeweils damit verbundenen (wahren, aber unbekannten) Verteilungen der T-Teststatistik sein könnten, in Betracht.

Nullhypothesen-Szenario: $c^T\beta = c^T\beta_0 \Leftrightarrow \mu = \mu_0 \Leftrightarrow \delta = 0 \Leftrightarrow T$ ist in Wahrheit zentral-t-verteilt.

Alternativhytpothesen-Szenario: $c^T\beta \neq c^T\beta_0 \Leftrightarrow \mu \neq \mu_0 \Leftrightarrow \delta \neq 0 \Leftrightarrow T$ ist in Wahrheit nicht-zentral-t-verteilt.



Wenn wir basierend auf Daten Parameterwerte geschätzt haben (in diesem Fall also Stichprobenmittel berechnet) und mit den Schätzern die T-Teststatistik berechnen, könnten wir zum Beispiel einen Wert von $T = 0.55$ oder einen Wert von $T = 3.67$ erhalten.

Anmerkung: Theoretische Aspekte sind grau und mit Daten geschätzte Größen sind orange

9. Geben Sie die Definition des zweiseitigen Einstichproben-T-Tests (mit ungerichteter Hypothese) wieder.

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

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \tag{15}$$

definiert. Weiterhin sei die T-Teststatistik definiert als

$$T := \sqrt{n}\left(\frac{\bar{\upsilon} - \mu_0}{s_\upsilon}\right) \tag{16}$$

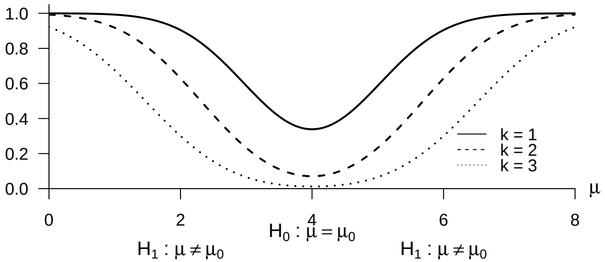
Dann ist der *zweiseitige Einstichproben-T-Test* definiert als der kritische wertbasierte Test

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

10. Skizzieren Sie die Testgütefunktion des zweiseitigen Einstichproben-T-Tests.

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

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



11. Geben Sie das Theorem zur Testumfangkontrolle im zweiseitigen Level-α_0-Einstichproben-T-Test wieder.

Theorem (Testumfangkontrolle)

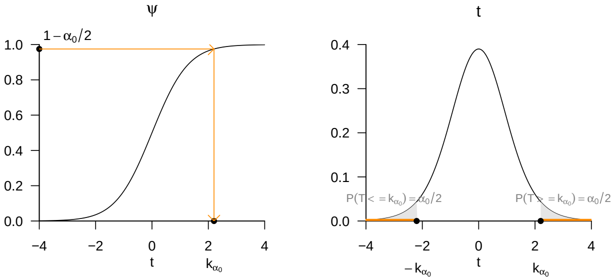
ϕ sei ein zweiseitiger Einstichproben-T-Test. Dann ist ϕ ein Level-α_0-Test mit Testumfang α_0, wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \psi^{-1}\left(1 - \frac{\alpha_0}{2}; n-1\right), \tag{18}$$

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

Veranschaulichung

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



12. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level-α_0-Einstichproben-T-Tests.

- Man nimmt an, dass ein Datensatz $\upsilon_1, ..., \upsilon_n$ eine Realisation von $\upsilon_i \sim N(\mu, \sigma^2)$ u.i.v. für $i = 1, ..., n$ mit unbekannten Parametern μ 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 α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0}.
- Anhand von $n, \mu_0, \bar{\upsilon}$ und s_υ berechnet man die Realisierung der T-Teststatistik

$$t := \sqrt{n}\left(\frac{\bar{\upsilon} - \mu_0}{s_\upsilon}\right) \tag{19}$$

- Wenn t größer-gleich k_{α_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.

13. Geben Sie die Definition des p-Wertes für einen zweiseitigen Einstichproben-T-Test wieder.

- Der p-Wert ist das kleinste Signifikanzlevel α_0, bei welchem man die Nullhypothese basierend auf einem vorliegenden Wert der Teststatistik ablehnen würde.
- Bei $T = t$ würde H_0 für jedes α_0 mit $|t| \geq \psi^{-1}(1 - \alpha_0/2; n-1)$ abgelehnt werden. Für diese α_0 gilt

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

- 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

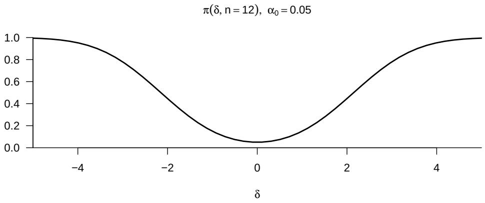
$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n-1)).$$

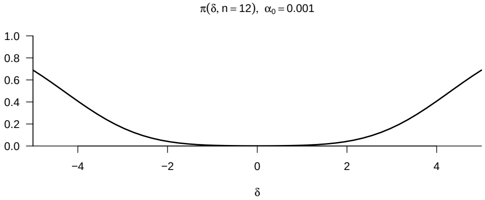
- Im Gegensatz zum Z-Test hängt bei T-Tests der p-Wert auch von der Stichprobengröße ab.
- Zum Beispiel ist für $T = 2.00$ und $n = 10$ der p-Wert 0.076, für $T = 2.00$ und $n = 100$ ist der p-Wert dagegen 0.048.

14. Von welchen Werten hängt die Powerfunktion eines zweiseitigen Einstichproben-T-Tests ab?

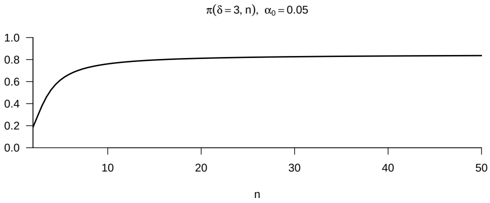
Die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese hängt bei festgelegtem α_0 vom wahren, aber unbekannten, Parameterwert $\delta = \sqrt{n}\frac{\mu - \mu_0}{\sigma}$ und von der Stichprobengröße n ab.

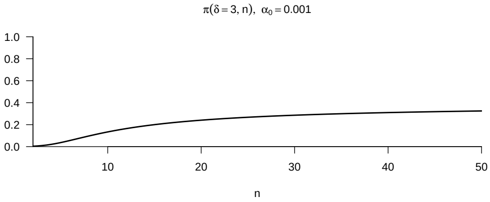
15. Skizzieren Sie die Powerfunktion des zweiseitigen Einstichproben-T-Tests bei fester Stichprobengröße.





16. Skizzieren Sie die Powerfunktion des zweiseitigen Einstichproben-T-Tests bei festem Nichtzentralitätsparameter.





17. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests.

Im Anwendungsszenario eines Zweistichproben-T-Tests betrachten wir **zwei Gruppen** (Stichproben) randomisierter experimenteller Einheiten. Wir nehmen an, dass die Datenpunkte in beiden Gruppen unabhängig und identisch normalverteilte Realisierungen von ZVen sind.

Gruppe 1: $\upsilon_{1j} \sim N(\mu_1, \sigma^2)$, und Gruppe 2: $\upsilon_{2j} \sim N(\mu_2, \sigma^2)$ für $j = 1, ..., n$,

wobei μ_1, μ_2 und σ^2 w.a.u. sind. Wir nehmen weiterhin an, dass wir daran interessiert sind, Unsicherheit, die mit dem inferentiellen Vergleich von μ_1 und μ_2 verbunden ist, im Sinne eines Hypothesentests zu quantifizieren.

18. Geben Sie die Definition des Zweistichproben-T-Test-Modells wieder.

Definition (Zweistichproben-T-Test-Modell)

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

$$\upsilon_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, 2, j = 1, ..., n_i \text{ mit } \mu_i \in \mathbb{R} \text{ und } \sigma^2 > 0, \tag{20}$$

die Datenverteilungsform

$$\upsilon_{ij} \sim N(\mu_i, \sigma^2) \text{ u.i.v. für } i = 1, 2, j = 1, ..., n_i \text{ mit } \mu_i \in \mathbb{R} \text{ und } \sigma^2 > 0, \tag{21}$$

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

$$\upsilon = X\beta + \varepsilon \text{ mit } X := \begin{pmatrix} 1_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} \end{pmatrix} \in \mathbb{R}^{n \times 2}, \beta := \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2, \varepsilon \sim N(0_n, \sigma^2 I_n), \sigma^2 > 0. \tag{22}$$

Bemerkungen

- i indiziert die Gruppen, j indiziert die Daten in jeder Gruppe.
- n_1 und n_2 repräsentieren die Gruppengrößen, n repräsentiert die Gesamtanzahl an Datenpunkten.
- Es ist $p = 2$.

Zweistichproben-T-Test - Modellschätzung

19. Geben Sie das Theorem zur Parameterschätzung im Zweistichproben-T-Test-Modell wieder.

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

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

$$\hat{\beta} = \begin{pmatrix} \frac{1}{n_1}\sum_{j=1}^{n_1} \upsilon_{1j} \\ \frac{1}{n_2}\sum_{j=1}^{n_2} \upsilon_{2j} \end{pmatrix} =: \begin{pmatrix} \bar{\upsilon}_1 \\ \bar{\upsilon}_2 \end{pmatrix} \tag{23}$$

und für den Varianzparameterschätzer

$$\hat{\sigma}_{12}^2 = \frac{\sum_{j=1}^{n_1}(\upsilon_{1j} - \bar{\upsilon}_1)^2 + \sum_{j=1}^{n_2}(\upsilon_{2j} - \bar{\upsilon}_2)^2}{n_1 + n_2 - 2} =: s_{12}^2 \tag{24}$$

Bemerkungen

- $\bar{\upsilon}_1$ und $\bar{\upsilon}_2$ bezeichnen die gruppenspezifischen Stichprobenmittel.
- s_{12}^2 wird als *gepoolte Stichprobenvarianz* bezeichnet.
- Für einen Datensatz $\upsilon = (\upsilon_1, \upsilon_2)^T \in \mathbb{R}^{n_1+n_2}$ gilt im Allgemeinen, dass $s_\upsilon^2 \neq s_{12}^2$; die gepoolte Stichprobenvarianz und die Stichprobenvarianz eines konkatenierten Datensatzes sind im Allgemeinen also nicht identisch. Wir wollen das Konzept der gepoolten Stichprobenvarianz hier aber nicht weiter vertiefen.

20. Geben Sie das Theorem zur T-Teststatistik des Zweistichproben-T-Tests wieder.

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

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

$$c := (1, -1)^T \text{ und } c^T\beta_0 =: \mu_0, \tag{25}$$

dass

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{\upsilon}_1 - \bar{\upsilon}_2 - \mu_0}{s_{12}}\right) \tag{26}$$

und es gilt

$$T \sim t(\delta, n_1 + n_2 - 2) \text{ mit } \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\mu_1 - \mu_2 - \mu_0}{\sigma}\right). \tag{27}$$

21. Erläutern Sie mögliche Hypothesenszenarien eines Zweistichproben-T-Tests.

$H_0 : \mu_1 - \mu_2 = \mu_0$ und $H_1 : \mu_1 - \mu_2 \neq \mu_0$

- Zweiseitiger Zweistichproben T-Test mit ungerichteter Hypothese
- Ungerichtete Fragestellungen nach einem Unterschied zwischen μ_1 und μ_2
- Für $\mu_0 := 0$ gelten dabei insbesondere
    - $H_0 : \mu_1 - \mu_2 = 0 \Leftrightarrow H_0 : \mu_1 = \mu_2$
    - $H_1 : \mu_1 - \mu_2 \neq 0 \Leftrightarrow H_1 : \mu_1 \neq \mu_2$

$H_0 : \mu_1 - \mu_2 \leq \mu_0$ und $H_1 : \mu_1 - \mu_2 > \mu_0$

- Einseitiger Zweistichproben T-Test mit gerichteter Hypothese
- Gerichtete Fragestellungen nach einem positiven Unterschied zwischen μ_1 und μ_2

$H_0 : \mu_1 - \mu_2 \geq \mu_0$ und $H_1 : \mu_1 - \mu_2 < \mu_0$

- Einseitiger Zweistichproben T-Test mit gerichteter Hypothese
- Gerichtete Fragestellungen nach einem negativen Unterschied zwischen μ_1 und μ_2

Anmerkung: Theoretische Aspekte sind grau und mit Daten geschätzte Größen sind orange

Modellformulierung (Ähnlich wie bei ALM für u.i.v. ZVen, aber mit Aufteilung in 2 Gruppen, indiziert mit $i = 1, 2$ und Gruppengrößen $n_1 = 2$ und $n_2 = 3$)

$$\upsilon_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) u.i.v. \text{ für } i = 1, 2 \text{ und } j = 1, \ldots, n_i$$

$$\upsilon = X\beta + \varepsilon \text{ mit } X := \begin{pmatrix} 1_2 & 0_2 \\ 0_3 & 1_3 \end{pmatrix} \in \mathbb{R}^{5\times 2}, \beta := \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2, \varepsilon \sim N(0_5, \sigma^2 I_5), \sigma^2 > 0.$$

$$\upsilon = X\beta + \varepsilon \Leftrightarrow \begin{pmatrix} \upsilon_{11} \\ \upsilon_{12} \\ \upsilon_{21} \\ \upsilon_{22} \\ \upsilon_{23} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{pmatrix} = \begin{pmatrix} 1\mu_1 + 0\mu_2 + \varepsilon_{11} \\ 1\mu_1 + 0\mu_2 + \varepsilon_{12} \\ 0\mu_1 + 1\mu_2 + \varepsilon_{21} \\ 0\mu_1 + 1\mu_2 + \varepsilon_{22} \\ 0\mu_1 + 1\mu_2 + \varepsilon_{23} \end{pmatrix} = \begin{pmatrix} 1\mu_1 + \varepsilon_{11} \\ 1\mu_1 + \varepsilon_{12} \\ 1\mu_2 + \varepsilon_{21} \\ 1\mu_2 + \varepsilon_{22} \\ 1\mu_2 + \varepsilon_{23} \end{pmatrix}$$

Modellschätzung

$$\hat{\beta} = \begin{pmatrix} \bar{\upsilon}_1 \\ \bar{\upsilon}_2 \end{pmatrix}, \text{ und } \hat{\sigma}^2_{12} = s^2_{12}$$

Anmerkung:

- s^2_{12} ist die gepoolte Stichprobenvarianz

Modellevaluation

Für die Modellevaluation wollen wir die T-Teststatistik berechnen. Dafür wählen wir diesem Fall eines Zweistichproben-T-Tests einen Kontrastgewichtsvektor von $c := (1, -1)^T$, wobei $c^T \beta_0 = \mu_0$.

Wir berechnen die T-Teststatistik T mit den Stichproben-Daten mit der Formel

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

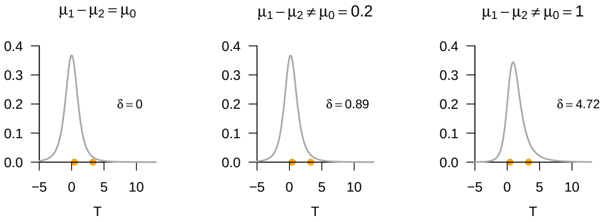
wobei wir eine Theorie darüber haben, wie T verteilt ist. Nämlich

$$T \sim t(\delta, 2 + 3 - 2) \text{ mit } \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right)$$

Wir ziehen zwei mögliche Szenarien davon, was der wahre, aber unbekannte Parameter sein könnte und die damit verbundenen jeweiligen (wahren, aber unbekannten) Verteilungen der T-Teststatistik in Betracht.

Nullhypothesen-Szenario: $c^T\beta = c^T\beta_0$ bzw. $\mu_1 - \mu_2 = \mu_0 \Leftrightarrow \delta = 0 \Leftrightarrow T$ ist in Wahrheit zentral-t-verteilt.

Alternativhytpothesen-Szenario: $c^T\beta \neq c^T\beta_0$ bzw. $\mu_1 - \mu_2 \neq \mu_0 \Leftrightarrow \delta \neq 0 \Leftrightarrow T$ ist in Wahrheit nicht-zentral-t-verteilt.



Wenn wir basierend auf Daten Parameterwerte geschätzt haben (in diesem Fall also Stichprobenmittel berechnet) und mit den Schätzern die T-Teststatistik berechenen, könnten wir zum Beispiel einen Wert von $T = 0.43$ oder einen Wert von $T = 4.22$ erhalten.

Anmerkung: Theoretische Aspekte sind grau und mit Daten geschätzte Größen sind orange

22. Geben Sie die Definition des zweiseitigen Zweistichproben-T-Tests (mit ungerichteter Hypothese) wieder.

Definition (Zweiseitiger Zweistichproben-T-Test)

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

$$H_0 : \mu_1 - \mu_2 = \mu_0 \Leftrightarrow \Theta_0 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 | \mu_1 - \mu_2 = \mu_0\} \tag{28}$$

und

$$H_1 : \mu_1 - \mu_2 \neq \mu_0 \Leftrightarrow \Theta_1 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 | \mu_1 - \mu_2 \neq \mu_0\}, \tag{29}$$

respektive. Weiterhin sei die T-Teststatistik definiert durch

$$T := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{\upsilon}_1 - \bar{\upsilon}_2 - \mu_0}{s_{12}} \right) \tag{30}$$

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

$$\phi(\upsilon) := 1_{\{|T| \geq k\}}. \tag{31}$$

23. Geben Sie das Theorem zur Testumfangkontrolle im zweiseitigen Level-α_0-Zweistichproben-T-Test wieder.

Theorem (Testumfangkontrolle)

ϕ sei der im obigen Testszenario definierte Test. Dann ist ϕ ein Level-α_0-Test mit Testumfang α_0, wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \psi^{-1}\left(1 - \frac{\alpha_0}{2}; n_1 + n_2 - 2\right), \tag{32}$$

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

$$n - 1 \hookrightarrow n_1 + n_2 - 2 \text{ und } \sqrt{n} \hookrightarrow \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \mu - \mu_0 \hookrightarrow \mu_1 - \mu_2 - \mu_0 \tag{33}$$

24. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level-α_0-Zweistichproben-T-Tests.

- Man nimmt an, dass die Daten zweier Gruppen $\upsilon_{11}, \ldots, \upsilon_{1n_1}$ und $\upsilon_{21}, \ldots, \upsilon_{2n_2}$ Realisationen von $\upsilon_{1j} \sim N(\mu_1, \sigma^2)$ u.i.v. für $j = 1, \ldots, n_1$ und $\upsilon_{2j} \sim N(\mu_2, \sigma^2)$ u.i.v. für $j = 1, \ldots, n_2$ mit unbekannten Parametern μ_1, μ_2, σ^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 α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_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 $n_1, n_2, \bar{\upsilon}_1, \bar{\upsilon}_2, \mu_0$ und der gepoolten Stichprobenstandardabweichung s_{12} berechnet man die Realisierung der Zweistichproben-T-Teststatistik

$$t := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{\upsilon}_1 - \bar{\upsilon}_2 - \mu_0}{s_{12}} \right) \tag{34}$$

- Wenn t größer-gleich k_{α_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.

25. Geben Sie die Definition des p-Wertes Werts für einen zweiseitigen Zweistichproben-T-Test wieder.

- Per Definition ist der p-Wert das kleinste Signifikanzlevel α_0, bei welchem man die Nullhypothese basierend auf einem vorliegenden Wert der Teststatistik ablehnen würde.

- Bei $T = t$ würde H_0 für jedes α_0 mit $|t| \geq \psi^{-1}(1 - \alpha_0/2; n_1 + n_2 - 2)$ abgelehnt werden. Für diese α_0 gilt, wie bereits mehrfach gezeigt,

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

- 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

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

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