



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(12) T-Tests

Die Darstellung der basiert im Wesentlichen auf Kapitel 9 in DeGroot and Schervish (2012).

T-Tests

- T-Teststatistik
- Einstichproben-T-Tests
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - Unabhängige Stichproben
 - Abhängige Stichproben
- Selbstkontrollfragen

T-Tests

- **T-Teststatistik**
- Einstichproben-T-Tests
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - Unabhängige Stichproben
 - Abhängige Stichproben
- Selbstkontrollfragen

Definition (T-Teststatistik)

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungmodells, $\bar{X}_n$ bezeichne das Stichprobenmittel, S_n bezeichne die Stichprobenstandardabweichung, und es sei $\mu_0 \in \mathbb{R}$. Die *T-Teststatistik* ist definiert als

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right). \quad (1)$$

Bemerkungen

- Im Gegensatz zur T-Statistik muss bei der T-Teststatistik nicht $\mu_0 = \mu$ gelten.
- Intuitiv kann die T-Teststatistik als mit der Stichprobengröße (Evidenz) gewichtetes Verhältnis von Signal (systematischer Variabilität) zu Rauschen (unsystematischer Variabilität) verstanden werden:

$$\sqrt{\text{Stichprobengröße}} \left(\frac{\text{Signal}}{\text{Rauschen}} \right) = \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (2)$$

- Die T-Teststatistik ist eine skalare Deskription des Effekt vs. Variabilität Verhältnisses eines Datensatzes.
- In der T-Teststatistik wird die Effektgröße in Einheiten der Stichprobenstandardabweichung gemessen:
 - $T = 1 \Leftrightarrow \sqrt{n}(\bar{X}_n - \mu_0) = 1S_n$
 - $T = 2 \Leftrightarrow \sqrt{n}(\bar{X}_n - \mu_0) = 2S_n$

Definition (Nichtzentrale t -Zufallsvariable)

T sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}$ und WDF

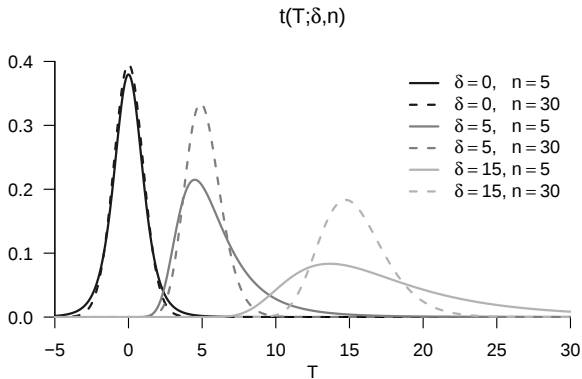
$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, t \mapsto p(t) := \frac{1}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n}{2}\right) (n\pi)^{\frac{1}{2}}} \times \int_0^\infty \tau^{\frac{n-1}{2}} \exp\left(-\frac{\tau}{2}\right) \exp\left(-\frac{1}{2} \left(t \left(\frac{\tau}{n}\right)^{\frac{1}{2}} - \delta\right)^2\right) d\tau. \quad (3)$$

Dann sagen wir, dass T einer nichtzentralen t -Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparameter n unterliegt und nennen T eine *nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter δ und Freiheitsgradparameter n* . Wir kürzen dies mit $t(\delta, n)$ ab. Die WDF einer nichtzentralen t -Zufallsvariable bezeichnen wir mit $t(T; \delta, n)$. Die KVF und inverse KVF einer nichtzentralen t -Zufallsvariable bezeichnen wir mit $\psi(\cdot; \delta, n)$ und $\psi^{-1}(\cdot; \delta, n)$, respektive.

Bemerkungen

- Eine nichtzentrale t -Zufallsvariable mit $\delta = 0$ ist eine t -Zufallsvariable.
- Es gilt also $t(T; 0, n) = t(T; n)$.
- Die funktionale Form der WDF findet sich zum Beispiel in Lehmann (1986), Seite 254, Gl. (80).

Wahrscheinlichkeitsdichtefunktionen nichtzentraler t -Verteilungen



Theorem (Nichtzentrale T-Transformation)

$X \sim N(\mu, 1)$ sei eine normalverteilte Zufallsvariable, $U \sim \chi^2(n)$ sei eine χ^2 Zufallsvariable mit Freiheitsgradparameter n , und X und U seien unabhängige Zufallsvariablen. Dann ist die Zufallsvariable

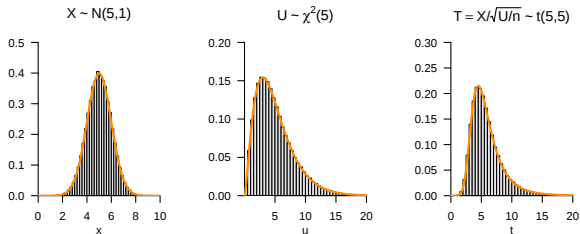
$$T := \frac{X}{\sqrt{U/n}} \quad (4)$$

eine nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter μ und Freiheitsgradparameter n , es gilt also $T \sim t(\mu, n)$

Bemerkung

- Wir verzichten auf einen Beweis.

Nichtzentrale T-Transformation



Theorem (Verteilung der T-Teststatistik)

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungmodells, $\bar{X}_n$ sei das Stichprobenmittel, S_n sei die Stichprobenstandardabweichung, und es sei $\mu_0 \in \mathbb{R}$. Dann ist die T-Teststatistik

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (5)$$

eine nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter

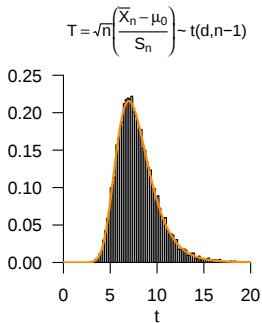
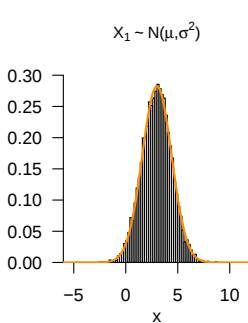
$$d = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (6)$$

und Freiheitsgradparameter $n - 1$, es gilt also $T \sim t(d, n - 1)$

Bemerkung

- Wir verzichten auf einen Beweis.

T-Teststatistik bei $n = 12, \mu = 3, \sigma^2 = 2, \mu_0 = 0$



T-Tests

- T-Teststatistik
- **Einstichproben-T-Tests**
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - Unabhängige Stichproben
 - Abhängige Stichproben
- Selbstkontrollfragen

Anwendungsszenario

- **Eine Stichprobe** experimenteller Einheiten.
- Annahme der unabhängigen und identischen Normalverteilung $N(\mu, \sigma^2)$ der Daten.
- μ und σ^2 unbekannt.
- Quantifizieren der Unsicherheit beim inferentiellen Vergleich von μ mit μ_0 beabsichtigt.

Anwendungsbeispiele

- Gruppenanalysen mit Wechsler Adult Intelligence Scale
 - $\mu \neq \mu_0 := 100 \Rightarrow$ Evidenz für über- oder unterdurchschnittliche WAIS Performanz
- BDI Score Datenanalyse einer Gruppe psychiatrischer Patient:innen
 - $\mu > \mu_0 := 18 \Rightarrow$ Klinisch auffällige Gruppe
- Gruppenanalysen in der funktionellen Kernspintomographie
 - $\mu > \mu_0 := 0 \Rightarrow$ Evidenz für regionale Gehirnaktivierung

Hypothesenszenarien

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

- Theoretisch wichtiges Szenario (Neymann-Pearson Lemma)
- Praktische Relevanz eher gering

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

- Zweiseitiger Einstichproben-T-Test mit ungerichteter Hypothese
- Ungerichtete Fragestellung nach einem Unterschied

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

- Einseitiger Einstichproben-T-Test mit gerichteter Hypothese
- Gerichtete Fragestellung nach einem positiven Unterschied

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

- Gerichtete Fragestellung nach einem negativen Unterschied
- Qualitativ äquivalente Theorie zum umgekehrten Fall

Einstichproben-T-Tests

Im Folgenden näher betrachtete Hypothesenszenarien

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

- Theoretisch wichtiges Szenario (Neymann-Pearson Lemma)
- Praktische Relevanz eher gering

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

- Zweiseitiger Einstichproben-T-Test mit ungerichteter Hypothese
- Ungerichtete Fragestellung nach einem Unterschied

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

- Einseitiger Einstichproben-T-Test mit gerichteter Hypothese
- Gerichtete Fragestellung nach einem positiven Unterschied

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

- Gerichtete Fragestellung nach einem negativen Unterschied
- Qualitativ äquivalente Theorie zum umgekehrten Fall

Gliederung

- (1) Statistisches Modell in klassischer Form
- (2) Statistisches Modell in generativer Form
- (3) Testhypothesen, Teststatistik, Test
- (4) Analyse der Testgütefunktion
- (5) Testumfangkontrolle
- (6) p-Werte
- (7) Analyse der Powerfunktion

T-Tests

- T-Teststatistik
- **Einstichproben-T-Tests**
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - Unabhängige Stichproben
 - Abhängige Stichproben

Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(1) Statistisches Modell in klassischer Form

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter μ und unbekanntem Varianzparameter $\sigma^2 > 0$. Als Parameter von Interesse betrachten wir $\theta = \mu$, so dass sich der Parameterraum von Interesse zu $\Theta = \mathbb{R}$ ergibt.

(2) Statistisches Modell in generativer Form

Es sei

$$X_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, \dots, n \quad (7)$$

wobei

- $X_i, i = 1, \dots, n$ beobachtbare Zufallsvariablen,
- $\mu \in \mathbb{R}$ den festen identischen Erwartungswertparameter über Stichprobenvariablen, und
- $\varepsilon_i, i = 1, \dots, n$ unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen

bezeichnen.

(2) Statistisches Modell in generativer Form (fortgeführt)

Die generative Form betont, dass im vorliegenden Modell beobachtete Daten durch einen systematischen deterministischen Prozess (hier μ) unter dem additiven Einfluss einer Vielzahl unabhängiger und deshalb in der Summe normalverteilter Störprozesse (hier $\varepsilon_i, i = 1, \dots, n$) erzeugt konzipiert werden.

Beweis

Die Äquivalenz beider Modellformen folgt direkt aus der Transformation normalverteilter Zufallsvariablen durch linear-affine Funktionen (cf. (6) Transformationen der Normalverteilung). Speziell gilt im vorliegenden Fall für $\varepsilon_i \sim N(0, \sigma^2)$, dass

$$X_i = f(\varepsilon_i) \text{ mit } f : \mathbb{R} \rightarrow \mathbb{R}, \varepsilon_i \mapsto f(\varepsilon_i) := \varepsilon_i + \mu. \quad (8)$$

Mit den WDF Transformationstheorem bei linear-affinen Abbildungen folgt dann

$$\begin{aligned} p_{X_i}(x_i) &= \frac{1}{|1|} p_{\varepsilon_i} \left(\frac{x_i - \mu}{1} \right) \\ &= N(x_i - \mu; 0, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2\sigma^2} (x_i - \mu - 0)^2 \right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2\sigma^2} (x_i - \mu)^2 \right) \\ &= N(x_i; \mu, \sigma^2), \end{aligned} \quad (9)$$

also $X_i \sim N(\mu, \sigma^2)$.

□

(3) Testhypothesen, Teststatistik, Test

Für ein $\mu_0 \in \mathbb{R}$ betrachten wir die einfache Nullhypothese und die zusammengesetzte Alternativhypothese

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \quad (10)$$

respektive. Weiterhin betrachten wir die T-Teststatistik

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (11)$$

und definieren den zweiseitigen kritischen Wert-basierten Test

$$\phi(X) := 1_{\{|T| \geq k\}} = \begin{cases} 1 & |T| \geq k \\ 0 & |T| < k \end{cases}. \quad (12)$$

(4) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

ϕ sei der im obigen Testszenario definierte Test. Dann ist die Testgütefunktion von ϕ gegeben durch

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k; d_\mu, n - 1) + \psi(-k; d_\mu, n - 1) \quad (13)$$

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

$$d_\mu := \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (14)$$

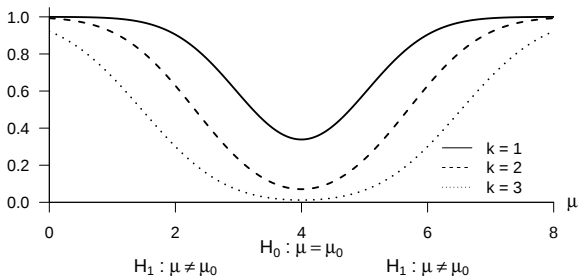
und Freiheitsgradparameter $n - 1$ bezeichnet.

Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(4) Analyse der Testgütefunktion

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

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



(4) Analyse der Testgütefunktion

Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testszenario ist definiert als

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \quad (15)$$

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

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (16)$$

unter der Annahme $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ nach einer nichtzentralen t -Verteilung $t(d, n - 1)$ mit Nichtzentralitätsparameter

$$d = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (17)$$

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

$$A =] - \infty, -k] \cup]k, \infty[. \quad (18)$$

(4) Analyse der Testgütefunktion

Beweis (fortgeführt)

Mit diesem Ablehnungsbereich ergibt sich dann

$$\begin{aligned} q_\phi(\mu) &= \mathbb{P}_\mu(\phi = 1) \\ &= \mathbb{P}_\mu(T \in]-\infty, -k] \cup]k, \infty[) \\ &= \mathbb{P}_\mu(T \in]-\infty, -k]) + \mathbb{P}_\mu(T \in [k, \infty[) \\ &= \mathbb{P}_\mu(T \leq -k) + \mathbb{P}_\mu(T \geq k) \\ &= \mathbb{P}_\mu(T \leq -k) + (1 - \mathbb{P}_\mu(T \leq k)) \\ &= 1 - \mathbb{P}_\mu(T \leq k) + \mathbb{P}_\mu(T \leq -k) \\ &= 1 - \psi(k; d_\mu, n-1) + \psi(-k; d_\mu, n-1), \end{aligned} \tag{19}$$

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

□

(5) Testumfangkontrolle

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

$$k_{\alpha_0} := \psi^{-1} \left(1 - \frac{\alpha_0}{2}; n - 1 \right), \quad (20)$$

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

Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(5) Testumfangkontrolle

Beweis

Damit der betrachtete Test ein Level- α_0 -Test ist, muss bekanntlich $q_\phi(\mu) \leq \alpha_0$ für alle $\mu \in \{\mu_0\}$, also hier $q_\phi(\mu_0) \leq \alpha_0$, gelten. Weiterhin ist der Testumfang des betrachteten Tests durch $\alpha = \max_{\mu \in \{\mu_0\}} q_\phi(\mu)$, also hier durch $\alpha = q_\phi(\mu_0)$ gegeben. Wir müssen also zeigen, dass die Wahl von k_{α_0} garantiert, dass ϕ ein Level- α_0 -Test mit Testumfang α_0 ist. Dazu merken wir zunächst an, dass für $\mu = \mu_0$ gilt, dass

$$\begin{aligned} q_\phi(\mu_0) &= 1 - \psi(k; d_{\mu_0}, n-1) + \psi(-k; d_{\mu_0}, n-1) \\ &= 1 - \psi(k; 0, n-1) + \psi(-k; 0, n-1) \\ &= 1 - \psi(k; n-1) + \psi(-k; n-1), \end{aligned} \tag{21}$$

wobei $\psi(\cdot; d, n-1)$ und $\psi(\cdot; n-1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter d und Freiheitsgradparameter $n-1$ sowie der t -Verteilung mit Freiheitsgradparameter $n-1$, respektive, bezeichnen. Sei nun also $k := k_{\alpha_0}$. Dann gilt

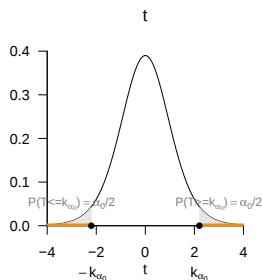
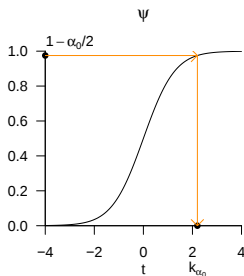
$$\begin{aligned} q_\phi(\mu_0) &= 1 - \psi(k_{\alpha_0}, n-1) + \psi(-k_{\alpha_0}, n-1) \\ &= 1 - \psi(k_{\alpha_0}, n-1) + (1 - \psi(k_{\alpha_0}, n-1)) \\ &= 2(1 - \psi(k_{\alpha_0}, n-1)) \\ &= 2 \left(1 - \psi \left(\psi^{-1}(1 - \alpha_0/2, n-1), n-1 \right) \right) \\ &= 2(1 - 1 + \alpha_0/2) \\ &= \alpha_0, \end{aligned} \tag{22}$$

wobei die zweite Gleichung mit der Symmetrie der t -Verteilung folgt. Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \leq \alpha_0$ ist und der betrachtete Test somit ein Level- α_0 -Test ist. Weiterhin folgt direkt, dass der Testumfang des betrachteten Tests bei der Wahl von $k = k_{\alpha_0}$ gleich α_0 ist.

Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(5) Testumfangkontrolle

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



(5) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz $x_1, \dots, x_n$ eine Realisation von $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ 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} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05$ und $n = 12$, also Freiheitsgradparameter 11, dass $k_{0.05} = \psi^{-1}(1 - 0.05/2; 11) \approx 2.20$ ist.
- Anhand von $n, \mu_0, \bar{x}_n$ und s_n berechnet man die Realisierung der T-Teststatistik

$$t := \sqrt{n} \left(\frac{\bar{x}_n - \mu_0}{s_n} \right) \quad (23)$$

- 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.

(6) p-Werte

Bestimmung des p-Wertes

- Per Definition ist der p-Wert das kleinste Signifikanzlevel α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegendem 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, wie unten gezeigt,

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \quad (24)$$

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

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n - 1)). \quad (25)$$

- 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.

(6) p-Werte

Bestimmung des p-Wertes

- Es bleibt zu zeigen, dass gilt

$$|t| \geq \psi^{-1}(1 - \alpha_0/2; n - 1) \Leftrightarrow \alpha_0 \geq 2\mathbb{P}(T \geq |t|) \quad (26)$$

- Dies aber folgt aus

$$\begin{aligned} |t| &\geq \psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right) \\ \Leftrightarrow \psi(|t|; n - 1) &\geq \psi\left(\psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right); n - 1\right) \\ \Leftrightarrow \psi(|t|; n - 1) &\geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow \mathbb{P}(T \leq |t|) &\geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow \frac{\alpha_0}{2} &\geq 1 - \mathbb{P}(T \leq |t|) \\ \Leftrightarrow \frac{\alpha_0}{2} &\geq \mathbb{P}(T \geq |t|) \\ \Leftrightarrow \alpha_0 &\geq 2\mathbb{P}(T \geq |t|). \end{aligned} \quad (27)$$

(7) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k_{\alpha_0}; d_\mu, n - 1) + \psi(-k_{\alpha_0}; d_\mu, n - 1) \quad (28)$$

bei kontrolliertem Testumfang, also für $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0/2; n - 1)$ mit festem α_0 als Funktion des Nichtzentralitätsparameters und des Stichprobenumfangs. Namentlich hängt hier k_{α_0} auch von n ab.

Es ergibt sich die bivariate reellwertige Funktion

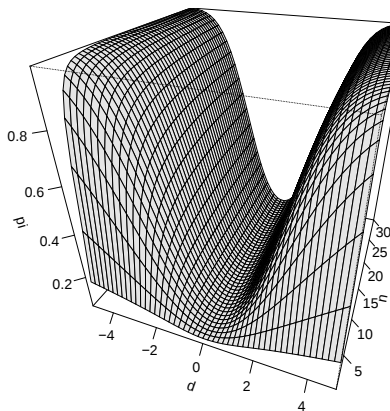
$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (d, n) \mapsto \pi(d, n) := 1 - \psi(k_{\alpha_0}; d, n - 1) + \psi(-k_{\alpha_0}; d, n - 1) \quad (29)$$

Bei festgelegten α_0 hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekannten Wert d und von der Stichprobengröße n ab. Wir visualisieren diese Abhängigkeiten untenstehend.

Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(7) Analyse der Powerfunktion

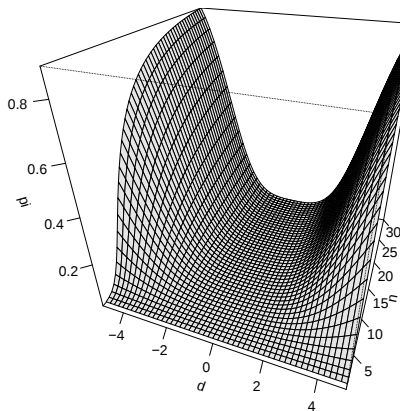
Powerfunktion für $\alpha_0 = 0.05$



Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(7) Analyse der Powerfunktion

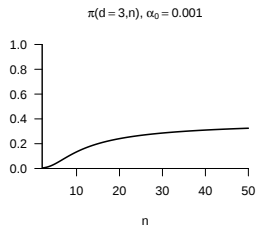
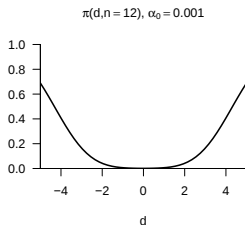
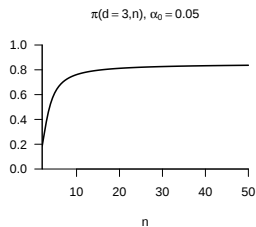
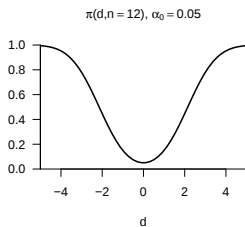
Powerfunktion für $\alpha_0 = 0.001$



Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(7) Analyse der Powerfunktion

Powerfunktionen für $\mu_0 = 0$



(7) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem n steigt die Powerfunktion des Tests an

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

⇒ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

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

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

Generell wird folgendes Vorgehen favorisiert

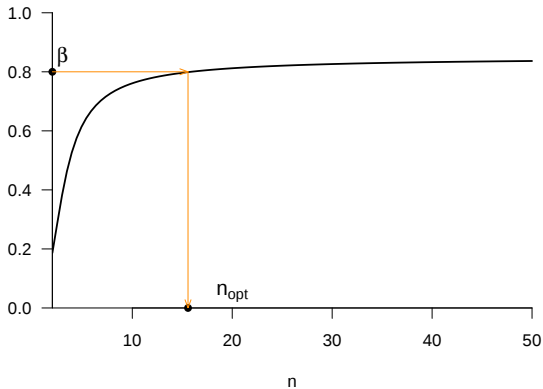
- Man legt das Signifikanzniveau α_0 fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert d^* , den man mit $\pi(d, n) = \beta$ detektieren möchte.
- Ein konventioneller Wert ist $\beta = 0.8$.
- Man liest die für $\pi(d = d^*, n) = \beta$ nötige Stichprobengröße n ab.

Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(7) Analyse der Powerfunktion

Praktisches Vorgehen

$\pi(d=3, n)$ für $\alpha_0 = 0.05$, $\mu_0 = 0$



T-Tests

- T-Teststatistik
- **Einstichproben-T-Tests**
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - Unabhängige Stichproben
 - Abhängige Stichproben

(1) Statistisches Modell in klassischer Form

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter μ und unbekanntem Varianzparameter $\sigma^2 > 0$. Als Parameter von Interesse betrachten wir $\theta = \mu$, so dass sich der Parameterraum von Interesse zu $\Theta = \mathbb{R}$ ergibt.

(2) Statistisches Modell in generativer Form

Es sei

$$X_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, \dots, n \quad (30)$$

wobei

- $X_i, i = 1, \dots, n$ beobachtbare Zufallsvariablen,
- $\mu \in \mathbb{R}$ den festen identischen Erwartungswertparameter über Stichprobenvariablen, und
- $\varepsilon_i, i = 1, \dots, n$ unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen

bezeichnen.

(3) Testhypothesen, Teststatistik, Test

Für ein $\mu_0 \in \mathbb{R}$ betrachten wir die zusammengesetzte Nullhypothese

$$H_0 : \mu \leq \mu_0 \Leftrightarrow \Theta_0 :=] - \infty, \mu_0] \quad (31)$$

und die zusammengesetzte Alternativhypothese

$$H_1 : \mu > \mu_0 \Leftrightarrow \Theta_1 :=]\mu_0, \infty[. \quad (32)$$

Weiterhin betrachten wir die T-Teststatistik

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (33)$$

und definieren den einseitigen kritischen Wert-basierten Test

$$\phi(X) := 1_{\{T \geq k\}} = \begin{cases} 1 & T \geq k \\ 0 & T < k \end{cases}. \quad (34)$$

(4) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

ϕ sei der im obigen Testszenario definierte Test. Dann ist die Testgütefunktion von ϕ gegeben durch

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k; d_\mu, n - 1) \quad (35)$$

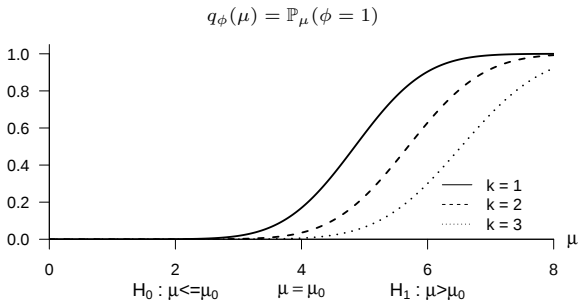
wobei $\psi(\cdot; d_\mu, n - 1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter

$$d_\mu := \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (36)$$

und Freiheitsgradparameter $n - 1$ bezeichnet.

(4) Analyse der Testgütefunktion

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



(4) Analyse der Testgütefunktion

Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testscenario ist definiert als

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \quad (37)$$

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

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (38)$$

unter der Annahme $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ nach einer nichtzentralen t -Verteilung $t(d, n-1)$ mit Nichtzentralitätsparameter

$$d = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (39)$$

verteilt ist. Mit dem Ablehnungsbereich des einseitigen T-Tests $A :=]k, \infty[$ ergibt sich dann

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi = 1) = \mathbb{P}_\mu(T \in]k, \infty[) = \mathbb{P}_\mu(T \geq k) = 1 - \mathbb{P}_\mu(T \leq k) = 1 - \psi(k; d, n-1), \quad (40)$$

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

□

(5) Testumfangkontrolle

Theorem (Testumfangkontrolle)

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

$$k_{\alpha_0} := \psi^{-1}(1 - \alpha_0; n - 1), \quad (41)$$

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

(5) Testumfangkontrolle

Beweis

Damit der betrachtete Test ein Level- α_0 -Test ist, muss bekanntlich $q_\phi(\mu) \leq \alpha_0$ für alle $\mu \in]-\infty, \mu_0]$ gelten. Weiterhin ist der Testumfang des betrachteten Tests durch $\alpha = \max_{-\infty, \mu_0]} q_\phi(\mu)$ gegeben. Mit der Monotonie von q_ϕ , die wir hier als selbstevident voraussetzen, gilt $\alpha = q_\phi(\mu_0)$. Wir müssen also zeigen, dass die Wahl von k_{α_0} garantiert, dass ϕ ein Level- α_0 -Test mit Testumfang α_0 ist. Dazu merken wir zunächst an, dass für $\mu = \mu_0$ gilt, dass

$$\begin{aligned} q_\phi(\mu_0) &= 1 - \psi(k; d_{\mu_0}, n - 1) \\ &= 1 - \psi(k; 0, n - 1) \\ &= 1 - \psi(k; n - 1) \end{aligned} \tag{42}$$

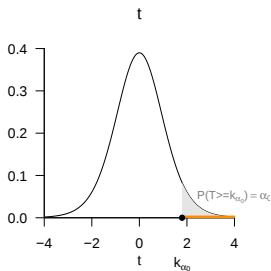
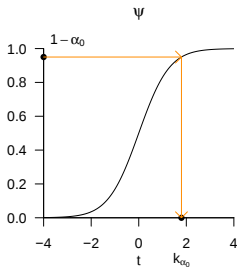
wobei $\psi(\cdot; d, n - 1)$ und $\psi(\cdot; n - 1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter d und Freiheitsgradparameter $n - 1$ sowie der t -Verteilung mit Freiheitsgradparameter $n - 1$, respektive, bezeichnen. Sei nun also $k := k_{\alpha_0}$. Dann gilt

$$\begin{aligned} q_\phi(\mu_0) &= 1 - \psi(k_{\alpha_0}) \\ &= 1 - \psi\left(\psi^{-1}(1 - \alpha_0)\right) \\ &= 1 - 1 + \alpha_0 \\ &= \alpha_0, \end{aligned} \tag{43}$$

Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \leq \alpha_0$ ist und der betrachtete Test somit ein Level- α_0 -Test ist. Weiterhin folgt direkt, dass der Testumfang des betrachteten Tests bei der Wahl von $k = k_{\alpha_0}$ gleich α_0 ist.

(5) Testumfangkontrolle

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



(5) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz $x_1, \dots, x_n$ eine Realisation von $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit unbekannten Parametern μ und $\sigma^2 > 0$ ist.
- Man möchte entscheiden ob für ein $\mu_0 \in \mathbb{R}$ eher $H_0 : \mu \leq \mu_0$ oder $H_1 : \mu > \mu_0$ 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 $n = 12$, also Freiheitsgradparameter 11, dass $k_{0.05} = \psi^{-1}(1 - 0.05; 11) \approx 1.80$ ist.
- Anhand von $n, \mu_0, \bar{x}_n$ und s_n berechnet man die Realisierung der T-Teststatistik

$$t := \sqrt{n} \left(\frac{\bar{x}_n - \mu_0}{s_n} \right) \quad (44)$$

- Wenn t größer-gleich k_{α_0} ist, lehnt man die Nullhypothese ab, andernfalls nicht.
- Die oben entwickelte Theorie 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 α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.
- Bei $T = t$ würde H_0 für jedes α_0 mit $t \geq \psi^{-1}(1 - \alpha_0; n - 1)$ abgelehnt werden. Für diese α_0 gilt, wie unten gezeigt,

$$\alpha_0 \geq \mathbb{P}(T \geq t). \quad (45)$$

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

$$\text{p-Wert} = \mathbb{P}(T \geq t) = 1 - \psi(t; n - 1). \quad (46)$$

- Im Gegensatz zum zweiseitigen T-Tests ergeben sich für negative Werte von t sehr große p-Werte.
- Zum Beispiel ist für $T = 2.00$ und $n = 10$ der p-Wert 0.038, für $T = -2.00$ und $n = 10$ ist der p-Wert dagegen 0.961.

(6) p-Werte

Bestimmung des p-Wertes

- Es bleibt zu zeigen, dass gilt

$$t \geq \psi^{-1}(1 - \alpha_0; n - 1) \Leftrightarrow \alpha_0 \geq 2\mathbb{P}(T \geq t) \quad (47)$$

- Dies aber folgt aus

$$\begin{aligned} & t \geq \psi^{-1}(1 - \alpha_0; n - 1) \\ \Leftrightarrow & \psi(t; n - 1) \geq \psi\left(\psi^{-1}(1 - \alpha_0; n - 1); n - 1\right) \\ \Leftrightarrow & \psi(t; n - 1) \geq 1 - \alpha_0 \\ \Leftrightarrow & \mathbb{P}(T \leq t) \geq 1 - \alpha_0 \\ \Leftrightarrow & \alpha_0 \geq 1 - \mathbb{P}(T \leq t) \\ \Leftrightarrow & \alpha_0 \geq \mathbb{P}(T \geq t) \end{aligned} \quad (48)$$

(7) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k_{\alpha_0}; d_\mu, n - 1) \quad (49)$$

bei kontrolliertem Testumfang, also für $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0; n - 1)$ mit festem α_0 als Funktion des Nichtzentralitätsparameters und des Stichprobenumfangs. Namentlich hängt hier k_{α_0} auch von n ab.

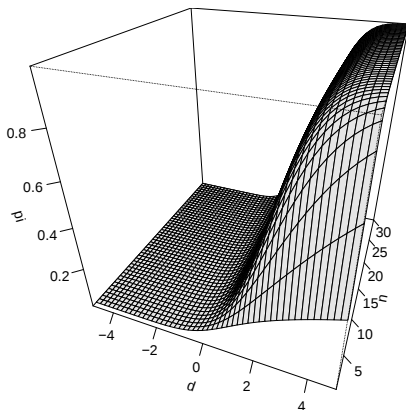
Es ergibt sich die bivariate reellwertige Funktion

$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (d, n) \mapsto \pi(\mu, n) := 1 - \psi(k_{\alpha_0}; d, n - 1) \quad (50)$$

Bei festgelegten α_0 hängt die Powerfunktion des einseitigen T-Tests mit zusammengesetzter Nullhypothese also vom unbekannten Wert d und von der Stichprobengröße n ab. Wir visualisieren diese Abhängigkeiten untenstehend.

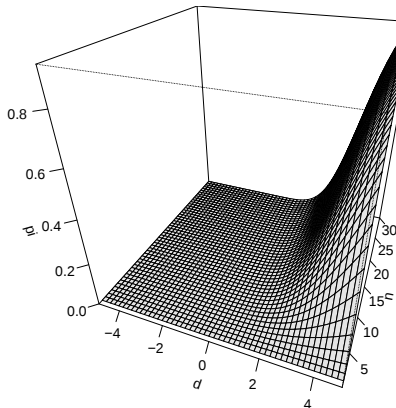
Einstichproben-T-Test mit $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$

Powerfunktion für $\alpha_0 = 0.05$



Einstichproben-T-Test mit $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$

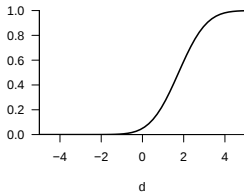
Powerfunktion für $\alpha_0 = 0.001$



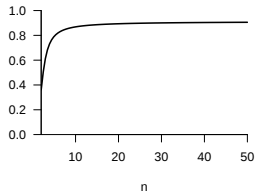
Einstichproben-T-Test mit $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$

Powerfunktionen für $\mu_0 = 0$

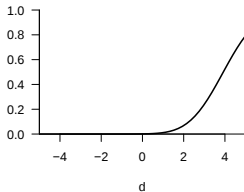
$\pi(d, n = 12), \alpha_0 = 0.05$



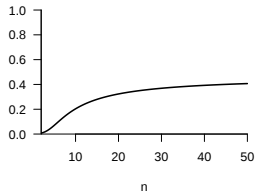
$\pi(d = 3, n), \alpha_0 = 0.05$



$\pi(d, n = 12), \alpha_0 = 0.001$



$\pi(d = 3, n), \alpha_0 = 0.001$



(7) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem n steigt die Powerfunktion des Tests an

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

⇒ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

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

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

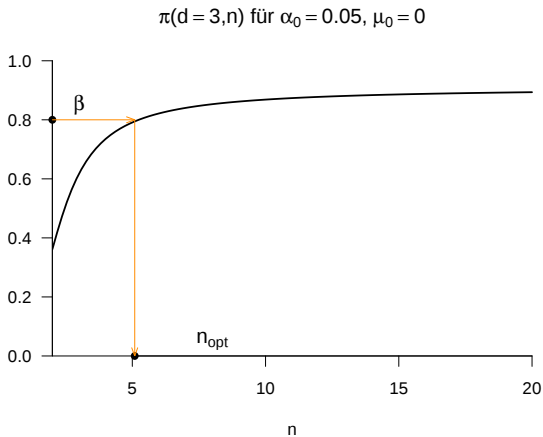
Generell wird folgendes Vorgehen favorisiert

- Man legt das Signifikanzniveau α_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.

Einstichproben-T-Test mit $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$

(7) Analyse der Powerfunktion

Praktisches Vorgehen



T-Tests

- T-Teststatistik
- Einstichproben-T-Tests
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - **Unabhängige Stichproben**
 - Abhängige Stichproben
- 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 μ_1 mit μ_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 μ_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 μ_2 und unbekanntem Varianzparameter $\sigma^2 > 0$. Die Varianzparameter beider Stichproben werden also als identisch vorausgesetzt. Der Parameter von Interesse ist (μ_1, μ_2) , der Parameterraum des Modells ist $\Theta := \mathbb{R}^2$.

(2) Statistisches Modell in generativer Form

Es sei

$$X_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ für } i = 1, 2 \text{ und } j = 1, \dots, n_i \quad (51)$$

wobei

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

(3) Testhypothesen, Teststatistik, Test

Wir betrachten die einfache Nullhypothese

$$H_0 : \mu_1 = \mu_2 \Leftrightarrow \Theta_0 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 \mid \mu_1 = \mu_2\} \quad (52)$$

und die zusammengesetzte Alternativhypothese

$$H_1 : \mu_1 \neq \mu_2 \Leftrightarrow \Theta_1 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 \mid \mu_1 \neq \mu_2\}. \quad (53)$$

Weiterhin betrachten wir die *Zweistichproben-T-Teststatistik*

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

wobei für $i = 1, 2$

$$\bar{X}_i := \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \text{ und } S_{12} := \sqrt{\frac{\sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)^2 + \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)^2}{(n_1 - 1) + (n_2 - 1)}} \quad (55)$$

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

$$\phi(X) := 1_{\{|T| \geq k\}}. \quad (56)$$

(4) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

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

$$\begin{aligned} q_\phi : \mathbb{R}^2 &\rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) \\ &:= 1 - \psi(k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) + \psi(-k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) \end{aligned} \quad (57)$$

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

$$d_{\mu_1, \mu_2} := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\mu_1 - \mu_2}{\sigma} \quad (58)$$

und Freiheitsgradparameter $n_1 + n_2 - 2$ bezeichnet.

Bemerkungen

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

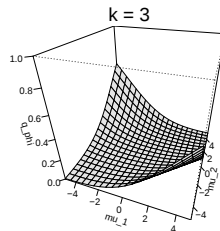
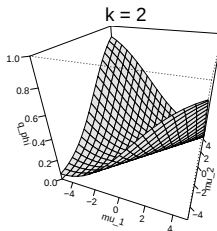
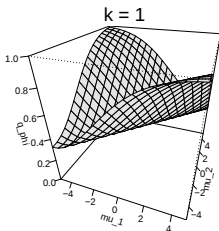
$$n \hookrightarrow n_1 + n_2 - 2, \quad \sqrt{n} \hookrightarrow \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \quad \mu - \mu_0 \hookrightarrow \mu_1 - \mu_2 \quad (59)$$

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

(4) Analyse der Testgütefunktion

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

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



(5) Testumfangkontrolle

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

$$k_{\alpha_0} := \psi^{-1} \left(1 - \frac{\alpha_0}{2}; n_1 + n_2 - 2 \right), \quad (60)$$

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

$$n - 1 \hookrightarrow n_1 + n_2 - 2. \quad (61)$$

(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 unbekannten Parametern μ_1, μ_2, σ^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 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 $n_1, n_2, \bar{x}_1, \bar{x}_2$ und der gepoolten Stichprobenstandardabweichung s_{12} berechnet man die Realisierung der Zweistichproben-T-Teststatistik

$$t := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{x}_1 - \bar{x}_2}{s_{12}} \right) \quad (62)$$

- 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.

(6) p-Werte

Bestimmung des p-Wertes

- Per Definition ist der p-Wert das kleinste Signifikanzlevel α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegendem 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,

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \quad (63)$$

- 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

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n_1 + n_2 - 2)). \quad (64)$$

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

(7) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$\begin{aligned} q_\phi : \mathbb{R}^2 &\rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) \\ &:= 1 - \psi(k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) + \psi(-k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) \end{aligned} \quad (65)$$

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 α_0 .

Es ergibt sich die multivariate reellwertige Funktion

$$\begin{aligned} \pi : \mathbb{R} \times \mathbb{N}^2 &\rightarrow [0, 1], (d, n) \mapsto \\ \pi(d, n) &:= 1 - \psi(k_{\alpha_0}; d, n - 2) + \psi(-k_{\alpha_0}; d, n - 2) \end{aligned} \quad (66)$$

Bei festgelegten α_0 hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekannten 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 unbekannten, Parameterwert $d = \sqrt{n}(\mu_1 - \mu_0)/\sigma$ ab.

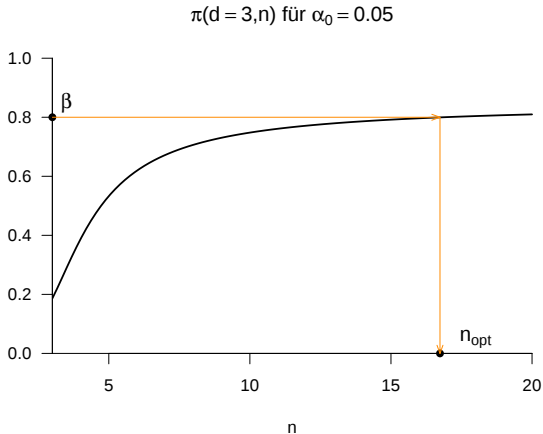
⇒ Wenn man d schon kennen würde, würde man den Test nicht durchführen.

Generell wird folgendes Vorgehen favorisiert

- Man legt das Signifikanzniveau α_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

$$X_{11}, \dots, X_{1n_1} \sim N(\mu_1, \sigma_1^2) \text{ und } X_{21}, \dots, X_{2n_2} \sim N(\mu_2, k\sigma_1^2) \quad (67)$$

erweitert werden.

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

$$T := \frac{(n_1 + n_2 - 2)^{\frac{1}{2}} (\bar{X}_1 - \bar{X}_2)}{\left(\frac{1}{n_1} + \frac{k}{n_2}\right)^{\frac{1}{2}} \left(S_1^2 + \frac{S_2^2}{k}\right)^{\frac{1}{2}}} \quad (68)$$

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

$$X_{11}, \dots, X_{1n_1} \sim N(\mu_1, \sigma_1^2) \text{ und } X_{21}, \dots, X_{2n_2} \sim N(\mu_2, \sigma_2^2) \quad (69)$$

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

$$\tilde{T} := (\bar{X}_1 - \bar{X}_2) / \left(\left(\frac{S_1^2}{n_1(n_1 - 1)} + \frac{S_2^2}{n_2(n_2 - 1)} \right)^{1/2} \right), \quad (70)$$

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

$$\nu = \left(\frac{s_1^2}{n_1(n_1 - 1)} + \frac{s_2^2}{n_2(n_2 - 1)} \right) / \left(\frac{1}{(n_1 - 1)^3} \left(\frac{s_1^2}{n_1} \right)^2 + \frac{1}{(n_2 - 1)^3} \left(\frac{s_2^2}{n_2} \right)^2 \right). \quad (71)$$

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

T-Tests

- T-Teststatistik
- Einstichproben-T-Tests
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - Unabhängige Stichproben
 - **Abhängige Stichproben**
- Selbstkontrollfragen

Zweistichproben-T-Test bei abhängigen Stichproben

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 μ_1 mit μ_2 beabsichtigt.

Anwendungsbeispiele

- Tense Arousal Datenanalyse bei einer Filmclipbedingung
 - Erste Messung vor Filmclipexposition, zweite Messung nach Filmclipexposition
 - $\mu_1 \neq \mu_2? \Leftrightarrow$ Hat Filmclipexposition einen Einfluss auf TA?
- 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

$$X_{ij} = \mu_i + \beta_j + \varepsilon_{ij} \text{ mit } \beta_j \sim N(0, \sigma_\beta^2) \text{ und } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \text{ für } i = 1, 2, j = 1, \dots, n, \quad (72)$$

wobei

- i die Messungen (abhängige Stichproben) und j die experimentellen Einheiten indizieren,
- X_{ij} beobachtbare Zufallsvariablen sind,
- μ_i feste Erwartungswertparameter der Messungen sind,
- β_j unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen sind, und
- ε_{ij} unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen sind.

Alle β_i und ε_{ij} werden als voneinander unabhängig vorausgesetzt.

- Die X_{ij} modellieren Datenpunkte,
- die μ_i modellieren feste Messungs-abhängige Effekte,
- die β_j modellieren zufällige experimentelle Einheits-abhängige Effekte,
- die ε_{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 \sigsgqr_\beta
# Störvariablenparameter    \sigqr_\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,...,n
  # X_{2j}, j = 1,...,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

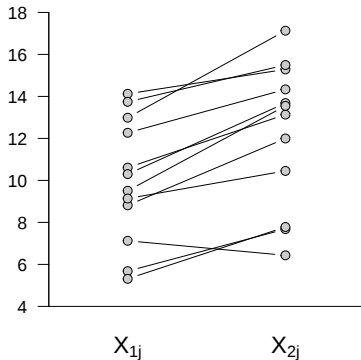
Datenbeispiel

mu_1	mu_2	beta_j	veps_1_j	veps_2_j	X_1j	X_2j
10	12	-0.792	0.796	1.136	10.00	12.34
10	12	-0.538	-0.951	0.166	8.51	11.63
10	12	-3.158	1.011	-0.571	7.85	8.27
10	12	-0.066	0.183	-1.227	10.12	10.71
10	12	0.537	-1.381	0.475	9.16	13.01
10	12	3.592	0.245	1.137	13.84	16.73
10	12	4.749	0.342	-0.610	15.09	16.14
10	12	-1.080	0.311	0.599	9.23	11.52
10	12	0.675	0.503	0.609	11.18	13.28
10	12	-1.062	-1.293	-0.362	7.64	10.58
10	12	2.490	0.432	-0.169	12.92	14.32
10	12	-0.458	-0.697	-0.488	8.85	11.05

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

$$X_1, \dots, X_n \sim N(\mu, \sigma^2) \text{ mit } \mu := \mu_2 - \mu_1 \text{ und } \sigma^2 := 2\sigma_\varepsilon^2. \quad (73)$$

Bemerkungen

- Hinsichtlich der Differenzvariablen X_j handelt es sich um das Einstichproben-T-Testscenario.
- Die Hypothesen des Zweistichproben-T-Test bei abhängigen Stichproben hinsichtlich der Erwartungswertparameter μ_1 und μ_2 übertragen sich wie folgt auf die Einstichproben-T-Test Hypothesen in bezug auf den Erwartungswertparameter μ :

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

$$\begin{aligned}X_j &= X_{2j} - X_{1j} \\&= \mu_2 + \beta_j + \varepsilon_{2j} - \mu_1 - \beta_j - \varepsilon_{1j} \\&= \mu_2 - \mu_1 + \varepsilon_{2j} - \varepsilon_{1j} \\&= \mu + \varepsilon_j,\end{aligned}\tag{74}$$

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

$$\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)\tag{75}$$

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

$$\varepsilon_j \sim N\left(0, 2\sigma_\varepsilon^2\right)\tag{76}$$

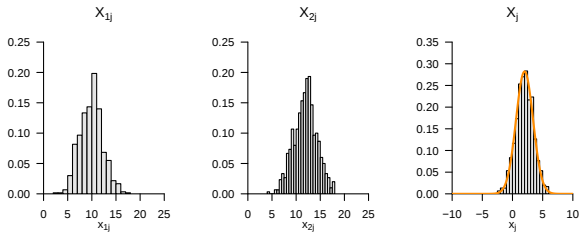
Mit der Definition $\sigma^2 := 2\sigma_\varepsilon^2$ und dem Transformationstheorem für normalverteilte Zufallsvariablen bei linear-affiner Transformation folgt das Resultat dann direkt.

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

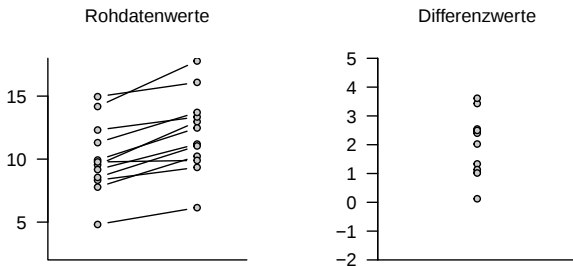
μ_1	μ_2	β_j	veps_{1_j}	veps_{2_j}	X_{1j}	X_{2j}	X_j
10	12	-0.600	0.144	1.574	9.54	12.97	3.430
10	12	3.962	0.211	1.824	14.17	17.79	3.613
10	12	-2.874	0.651	1.102	7.78	10.23	2.451
10	12	-5.039	-0.148	-0.815	4.81	6.15	1.333
10	12	0.907	1.402	0.437	12.31	13.35	1.036
10	12	-0.207	-0.624	-0.600	9.17	11.19	2.024
10	12	2.197	-0.890	-0.489	11.31	13.71	2.400
10	12	-0.482	0.409	0.955	9.93	12.47	2.546
10	12	3.993	0.963	0.098	14.96	16.09	1.135
10	12	-1.217	-0.458	-1.439	8.32	9.34	1.018
10	12	-0.194	-1.259	-0.762	8.55	11.04	2.497
10	12	-0.589	0.363	-1.514	9.77	9.90	0.123

$\Rightarrow$ 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 $X_j, j = 1, \dots, n$

T-Tests

- T-Teststatistik
- Einstichproben-T-Tests
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - Unabhängige Stichproben
 - Abhängige Stichproben
- **Selbstkontrollfragen**

Selbstkontrollfragen

1. Definieren Sie die T-Teststatistik.
2. Geben Sie das Theorem zur nichtzentralen T-Transformation wieder.
3. Geben Sie das Theorem zur Verteilung der T-Teststatistik wieder.
4. Erläutern Sie das Anwendungsszenario eines Einstichproben-T-Tests.
5. Erläutern Sie mögliche Hypothesenszenarien eines Einstichproben-T-Tests.
6. Geben Sie das statistische Modell eines Einstichproben-T-Tests in klassischer Form an.
7. Geben Sie das statistische Modell eines Einstichproben-T-Tests in generativer Form an.
8. Erläutern Sie die Bedeutung des eines statistischen Modells in generativer Form.
9. Definieren Sie den zweiseitigen Einstichproben-T-Test.
10. Skizzieren Sie qualitativ die Testgütefunktion eines zweiseitigen Einstichproben-T-Test
11. Geben Sie die Definition des kritischen Werts für einen 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 Einstichproben-zweiseitigen T-Tests ab?
15. Skizzieren Sie qualitativ die Powerfunktion des zweiseitigen Einstichproben-T-Tests bei fester Stichprobengröße.

Selbstkontrollfragen

16. Skizzieren Sie qualitativ die Powerfunktion des zweiseitigen Einstichproben-T-Tests bei festem Nichtzentralitätsparameter.
17. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests bei unabhängigen Stichproben.
18. Nennen Sie drei Modellszenarien für Zweistichproben-T-Tests bei unabhängigen Stichproben.
19. Geben Sie das statistische Modell eines Zweistichproben-T-Tests bei unabhängigen Stichproben in klassischer Form an.
20. Geben Sie das statistische Modell eines Zweistichproben-T-Tests bei unabhängigen Stichproben in generativer Form an.
21. Definieren Sie die gepoolte Stichprobenstandardabweichung.
22. Definieren Sie die Zweistichproben-T-Teststatistik.
23. Definieren Sie den kritischen Wert für einen zweiseitigen Level- α_0 -Zweistichproben-T-Tests bei unabhängigen Stichproben.
24. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level- α_0 -Zweistichproben-T-Tests bei unabhängigen Stichproben.
25. Erläutern Sie das Szenario eines Zweistichproben-T-Tests bei unabhängigen Stichproben mit bekanntem Varianzverhältnis.
26. Definieren Sie das Behrens-Fisher Problem.
27. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests bei abhängigen Stichproben.
28. Geben Sie das statistische Modell eines Zweistichproben-T-Tests bei abhängigen Stichproben in generativer Form an.
29. Geben Sie das Theorem zur Verteilung der Differenzvariablen im Zweistichproben-T-Tests bei abhängigen Stichproben wieder.
30. 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.

Lehmann, E. L. 1986. *Testing Statistical Hypotheses*. Wiley Series in Probability and Statistics.

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>.