Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(13) Einstichproben-T-Tests

T-Teststatistik

Einstichproben-T-Tests

Zweiseitige Einstichproben-T-Tests

Einseitige Einstichproben-T-Tests

Anwendungsbeispiel

Selbstkontrollfragen

T-Teststatistik

Definition (T-Teststatistik)

$X_1, ..., X_n \sim N(\mu, \sigma^2)$ 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

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

Bemerkungen

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

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

- Die T-Teststastitik 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

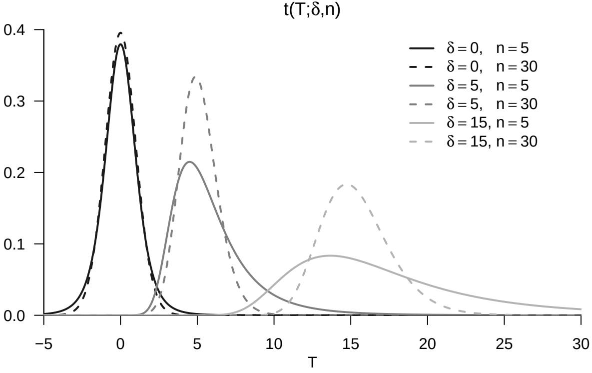
$$p : \mathbb{R} \to \mathbb{R}_{>0}, t \mapsto p(t) := \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n}{2}\right)(n\pi)^{\frac{1}{2}}} \times \int_0^\infty \tau^{\frac{n-1}{2}} \exp\left(-\frac{\tau}{2}\right) \exp\left(-\frac{1}{2}\left(t\left(\frac{\tau}{n}\right)^{\frac{1}{2}} - \delta\right)^2\right)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

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



T-Teststatistik

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

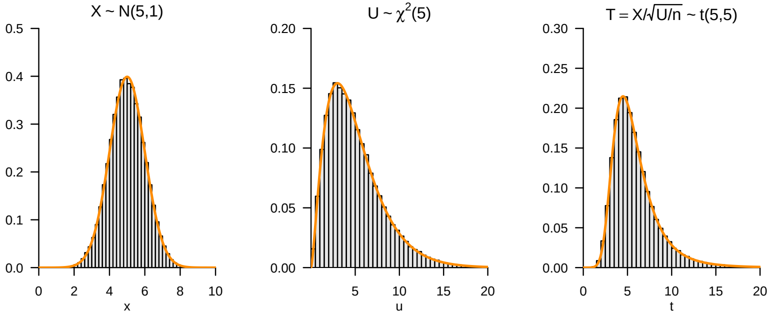
$$T := \frac{X}{\sqrt{U/n}} \tag{4}$$

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)

$X_1, ..., X_n \sim N(\mu, \sigma^2)$ 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

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

eine nichtzentrale t-Zufallsvariable mit Nichtzentralitätsparameter

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

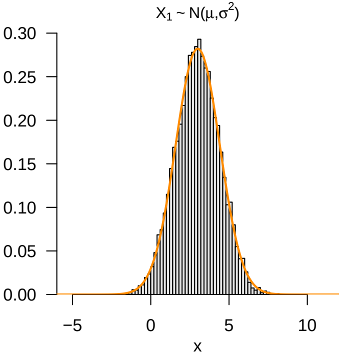
und Freiheitsgradparameter $n - 1$, es gilt also $T \sim t(d, n-1)$

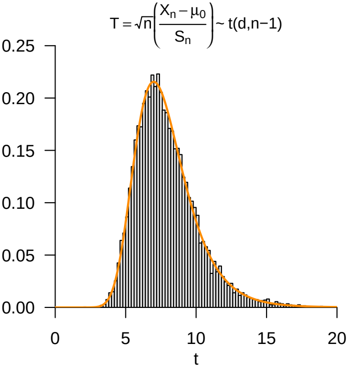
Bemerkung

- Wir verzichten auf einen Beweis.

T-Teststatistik

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





T-Teststatistik

Einstichproben-T-Tests

Zweiseitige Einstichproben-T-Tests

Einseitige Einstichproben-T-Tests

Anwendungsbeispiel

Selbstkontrollfragen

Einstichproben-T-Tests

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

Einstichproben-T-Tests

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-Teststatistik

Einstichproben-T-Tests

Zweiseitige Einstichproben-T-Tests

Einseitige Einstichproben-T-Tests

Anwendungsbeispiel

Selbstkontrollfragen

Zweiseitige Einstichproben-T-Tests

(1) Statistisches Modell in klassischer Form

$X_1, ..., X_n \sim N(\mu, \sigma^2)$ 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

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

wobei

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

bezeichnen.

Zweiseitige Einstichproben-T-Tests

(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 $\varepsilon_i, i = 1, ..., n$) 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

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

Mit den WDF Transformationstheorem bei linear-affinen Abbildungen folgt dann

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

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

Zweiseitige Einstichproben-T-Tests

(3) Testhypothesen, Teststatistik, Test

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

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

respektive. Weiterhin betrachten wir die T-Teststatistik

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

und definieren den zweiseitigen kritischen Wert-basierten Test

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

(4) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

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

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

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

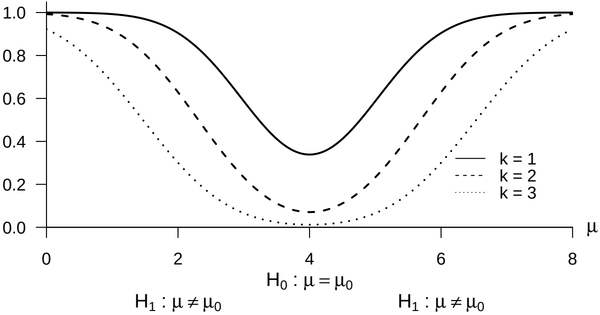
$$d_\mu := \sqrt{n}\left(\frac{\mu - \mu_0}{\sigma}\right) \tag{14}$$

und Freiheitsgradparameter $n-1$ bezeichnet.

(4) Analyse der Testgütefunktion

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

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



(4) Analyse der Testgütefunktion

Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testszenario ist definiert als

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

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

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

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

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

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

$$A =]-\infty, -k] \cup]k, \infty[. \tag{18}$$

Zweiseitige Einstichproben-T-Tests

(4) Analyse der Testgütefunktion

Beweis (fortgeführt)

Mit diesem Ablehungsbereich ergibt sich dann

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

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

$\square$

(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

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

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 $q_\phi(\mu) \le \alpha_0$ für alle $\mu \in \{\mu_0\}$, also hier $q_\phi(\mu_0) \le \alpha_0$, 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

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

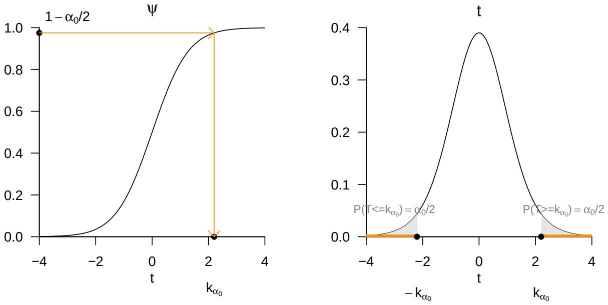
wobei $\psi(\cdot; d, n-1)$ und $\psi(\cdot; n-1)$ die KVF der nichtzentralen t-Verteilung mit Nichtzentralitätsparameter d und Freiheitsgradparameter $n-1$ sowie der t-Verteilung mit Freiheitsgradparameter $n-1$, respektive, bezeichnen. Sei nun also $k := k_{\alpha_0}$. Dann gilt

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

wobei die zweite Gleichung mit der Symmetrie der t-Verteilung folgt. Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \le \alpha_0$ 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 $k_{\alpha_0} := \psi^{-1}(1-\alpha_0/2; n-1)$ mit $n = 12$, $\alpha_0 := 0.05$ und Ablehnungsbereich



Zweiseitige Einstichproben-T-Tests

(5) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz $x_1, ..., x_n$ eine Realisation von $X_1, ..., X_n \sim N(\mu, \sigma^2)$ 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

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

- 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.

Zweiseitige Einstichproben-T-Tests

(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,

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

- 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

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

- 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.

Bestimmung des p-Wertes

- Es bleibt zu zeigen, dass gilt

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

- Dies aber folgt aus

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

(7) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

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

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.

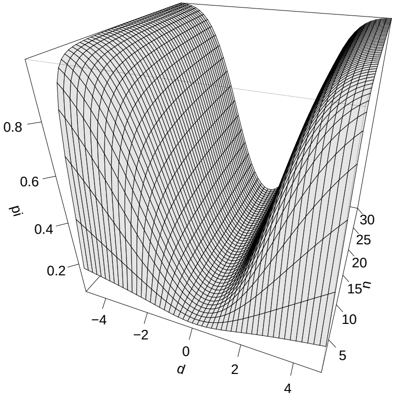
Es ergibt sich die bivariate reellwertige Funktion

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

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.

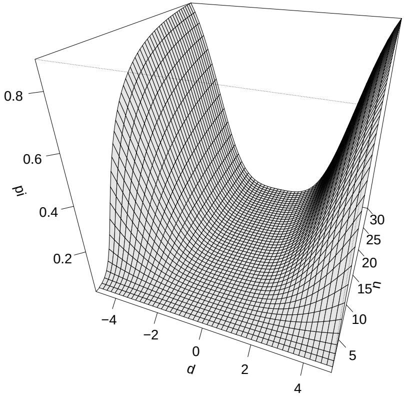
(7) Analyse der Powerfunktion

Powerfunktion für $\alpha_0 = 0.05$



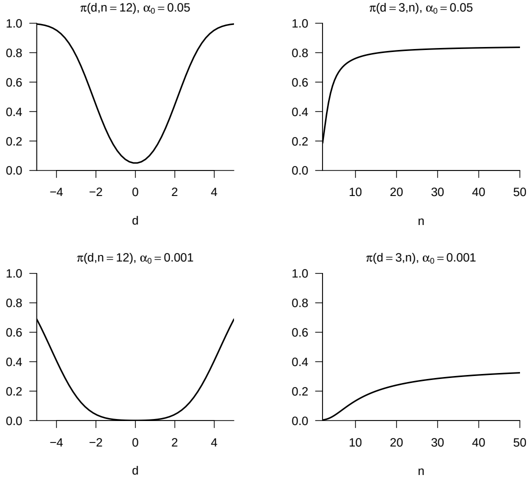
(7) Analyse der Powerfunktion

Powerfunktion für $\alpha_0 = 0.001$



(7) Analyse der Powerfunktion

Powerfunktionen für $\mu_0 = 0$



Zweiseitige Einstichproben-T-Tests

(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.

$\Rightarrow$ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

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

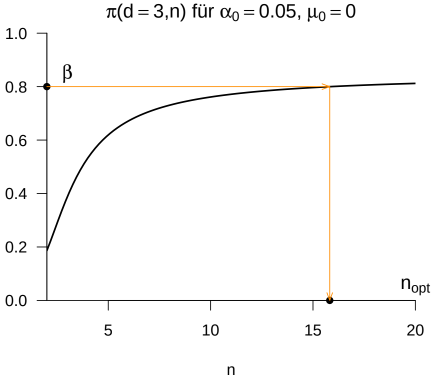
$\Rightarrow$ Wenn man d schon kennen würde, würde man den Test nicht durchführen.

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



T-Teststatistik

Einstichproben-T-Tests

Zweiseitige Einstichproben-T-Tests

Einseitige Einstichproben-T-Tests

Anwendungsbeispiel

Selbstkontrollfragen

(1) Statistisches Modell in klassischer Form

$X_1, ..., X_n \sim N(\mu, \sigma^2)$ 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

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

wobei

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

bezeichnen.

(3) Testhypothesen, Teststatistik, Test

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

$$H_0 : \mu \leq \mu_0 \Leftrightarrow \Theta_0 := \,] - \infty, \mu_0] \tag{31}$$

und die zusammengesetzte Alternativhypothese

$$H_1 : \mu > \mu_0 \Leftrightarrow \Theta_1 := \,]\mu_0, \infty[. \tag{32}$$

Weiterhin betrachten wir die T-Teststatistik

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

und definieren den einseitigen kritischen Wert-basierten Test

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

(4) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

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

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

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

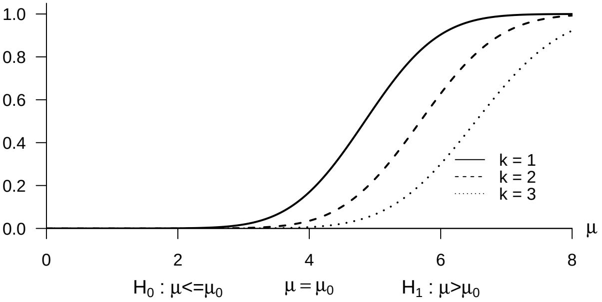
$$d_\mu := \sqrt{n}\left(\frac{\mu - \mu_0}{\sigma}\right) \tag{36}$$

und Freiheitsgradparameter $n-1$ bezeichnet.

(4) Analyse der Testgütefunktion

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

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



(4) Analyse der Testgütefunktion

Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testszenario ist definiert als

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

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

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

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

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

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

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

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

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

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 $q_\phi(\mu) \le \alpha_0$ 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

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

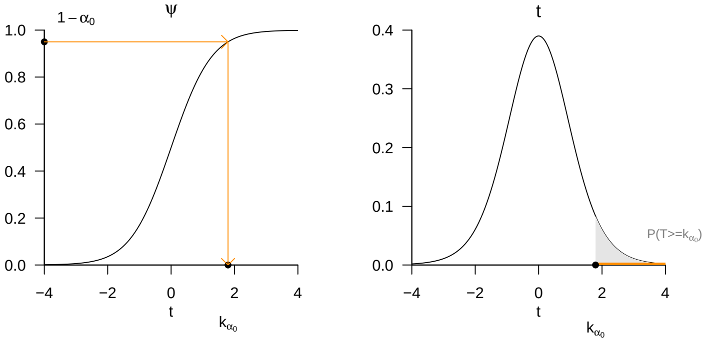
wobei $\psi(\cdot; d, n-1)$ und $\psi(\cdot; n-1)$ die KVF der nichtzentralen t-Verteilung mit Nichtzentralitätsparameter d und Freiheitsgradparameter $n-1$ sowie der t-Verteilung mit Freiheitsgradparameter $n-1$, respektive, bezeichnen. Sei nun also $k := k_{\alpha_0}$. Dann gilt

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

Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \le \alpha_0$ 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



Einseitige Einstichproben-T-Tests

(5) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz $x_1, ..., x_n$ eine Realisation von $X_1, ..., X_n \sim N(\mu, \sigma^2)$ mit unbekannten Parametern μ und $\sigma^2 > 0$ ist.
- Man möchte entscheiden ob für ein $\mu_0 \in \mathbb{R}$ eher $H_0 : \mu \le \mu_0$ 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

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

- 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 $t \ge \psi^{-1}(1 - \alpha_0; n-1)$ abgelehnt werden. Für diese α_0 gilt, wie unten gezeigt,
$$\alpha_0 \ge \mathbb{P}(T \ge t). \tag{45}$$
- Das kleinste $\alpha_0 \in [0,1]$ mit $\alpha_0 \ge \mathbb{P}(T \ge t)$ ist dann $\alpha_0 = \mathbb{P}(T \ge t)$, also folgt
$$\text{p-Wert} = \mathbb{P}(T \ge t) = 1 - \psi(t; n-1). \tag{46}$$
- 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

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

- Dies aber folgt aus

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

(7) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

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

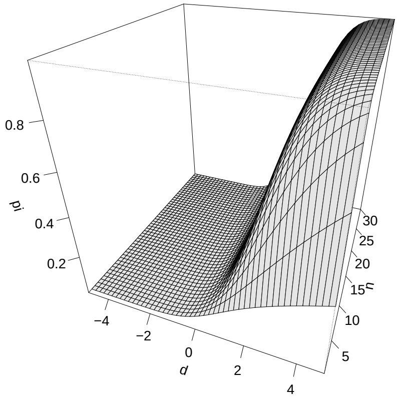
bei kontrolliertem Testumfang, also für $k_{\alpha_0} := \psi^{-1}(1-\alpha_0; n-1)$ mit festem α_0 als Funktion des Nichtzentralitätsparameters und des Stichprobenumfangs. Namentlich hängt hier k_{α_0} auch von n ab.

Es ergibt sich die bivariate reellwertige Funktion

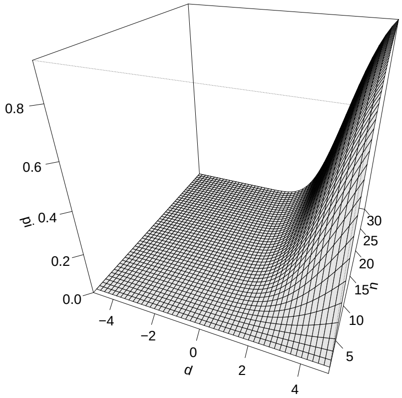
$$\pi : \mathbb{R} \times \mathbb{N} \to [0,1], (d,n) \mapsto \pi(\mu, n) := 1 - \psi(k_{\alpha_0}; d, n-1) \tag{50}$$

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

Powerfunktion für $\alpha_0 = 0.05$

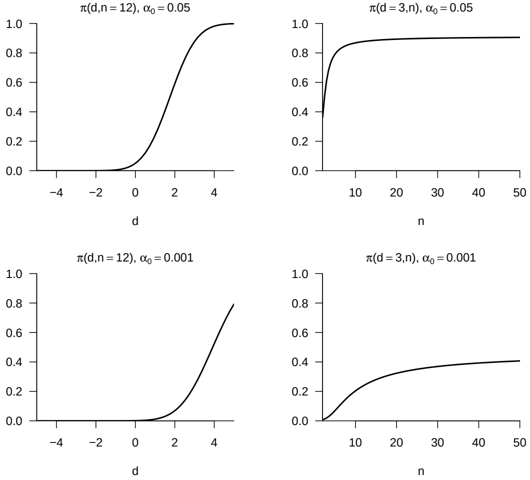


Powerfunktion für $\alpha_0 = 0.001$



Einseitige Einstichproben-T-Tests

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.

$\Rightarrow$ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

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

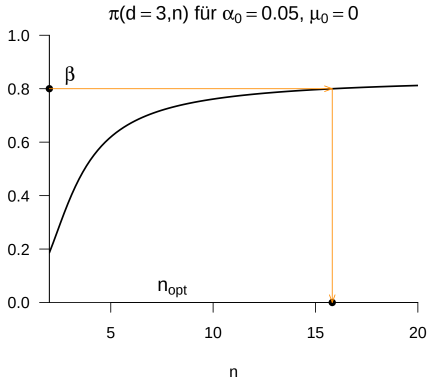
$\Rightarrow$ Wenn man d schon kennen würde, würde man den Test nicht durchführen.

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



T-Teststatistik

Einstichproben-T-Tests

Zweiseitige Einstichproben-T-Tests

Einseitige Einstichproben-T-Tests

Anwendungsbeispiel

Selbstkontrollfragen

T-Teststatistik

Einstichproben-T-Tests

Zweiseitige Einstichproben-T-Tests

Einseitige Einstichproben-T-Tests

Anwendungsbeispiel

Selbstkontrollfragen

Selbstkontrollfragen

1. Definieren Sie die T-Teststatistik.
2. Erläutern Sie die T-Teststatistik.
3. Geben Sie das Theorem zur nichtzentralen T-Transformation wieder.
4. Geben Sie das Theorem zur Verteilung der T-Teststatistik wieder.
5. Erläutern Sie das Anwendungsszenario eines Einstichproben-T-Tests.
6. Erläutern Sie mögliche Hypothesenszenarien eines Einstichproben-T-Tests.
7. Geben Sie das statistische Modell eines Einstichproben-T-Tests in klassischer Form an.
8. Geben Sie das statistische Modell eines Einstichproben-T-Tests in generativer Form an.
9. Erläutern Sie die Bedeutung des eines statistischen Modells in generativer Form.
10. Definieren Sie den zweiseitigen Einstichproben-T-Test.
11. Skizzieren Sie qualitativ die Testgütefunktion eines zweiseitigen Einstichproben-T-Test
12. Geben Sie die Definition des kritischen Werts für einen zweiseitigen Level-α_0-Einstichproben-T-Test wieder.
13. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level-α_0-Einstichproben-T-Tests.
14. Geben Sie die Definition des p-Wertes Werts für einen zweiseitigen Einstichproben-T-Test wieder.
15. Von welchen Werten hängt die Powerfunktion eines Einstichproben-zweiseitigen T-Tests ab?
16. Skizzieren Sie die Powerfunktion des zweiseitigen Einstichproben-T-Tests bei fester Stichprobengröße.
17. Skizzieren Sie die Powerfunktion des zweiseitigen Einstichproben-T-Tests bei festem Nichtzentralitätsparameter.

References

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