CHARLOTTE FRESENIUS HOCHSCHULE
UNIVERSITY OF PSYCHOLOGY

Statistik I

Einheit 8: t-Test

19.12.2024 | Prof. Dr. Stephan Goerigk

t-Test

Lernziele:

Sie lernen:

- Den t-Test - ein Signifikanztest zum Vergleich von 2 Mittelwerten in 3 Varianten
  - Ein-Stichproben t-Test
  - Unabhängiger t-Test (Zwei-Stichproben t-Test)
  - Abhängiger t-Test
- Die Voraussetzungen, die für die Durchführung eines t-Tests gegeben sein müssen
- Die Entscheidungsregel, basierend auf dem kritischen t-Wert (unter Annahme Signifikanzniveau $\alpha = .05$)
- Die Abgrenzung zwischen einem einseitigen und einem zweiseitigen t-Test

t-Test

Ein-Stichproben t-Test

- Hypothesen über μ einer normalverteilten Variable, wobei σ^2 unbekannt
- Mögliche Hypothesen:
  - H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$
  - H_0: $\mu \leq \mu_0$; H_1: $\mu > \mu_0$
  - H_0: $\mu \geq \mu_0$; H_1: $\mu < \mu_0$
- Prüft anhand des Mittelwerts einer Stichprobe ob der Erwartungswert in der entsprechenden Population gleich einem vorgegebenen Wert ist (dem unter H_0 erwarteten μ_0).
- Vergleich eines Stichprobenmittelwertes mit einem hypothetischen Populationsparameter μ_0.

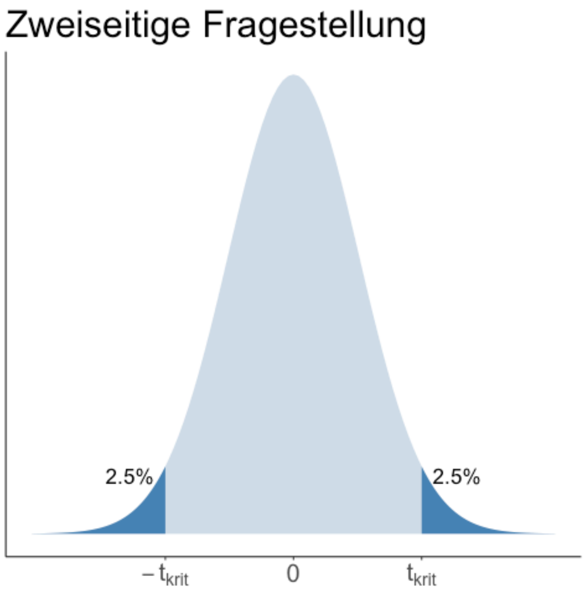
t-Test

Entwicklung eines Entscheidungskriteriums

- Zur Erinnerung: Warhscheinlichkeit, die H_0 abzulehnen, obwohl sie in Wirklichkeit gilt, heißt α-Fehler oder Fehler 1. Art
- t-Wert ist signifikant, wenn seine Auftretenswahrscheinlichkeit kleiner ist als das gewählte α
- Für die Signifikanzprüfung kann der t-Wert (t_{emp}) auch mit dem kritischen t-Wert (t_{krit}) verglichen werden (in t-Tabelle nachsehen)
- Die Wahl des Signikanzniveaus ist von inhaltlichen Überlegungen abhängig und wird oft als $\alpha = .05$ gewählt.

t-Test

Vergleich von ein- und zweiseitigen Fragestellungen



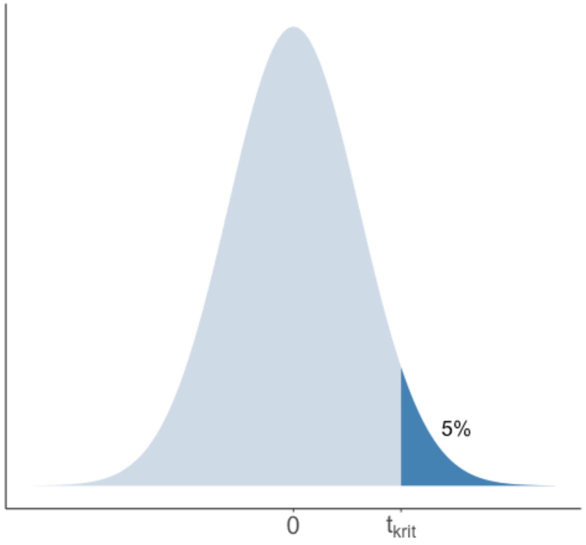
- Signifikanzniveau $\alpha = .05$ muss auf beide Seiten aufgeteilt werden
- Damit $\alpha = .05$ erreicht wird, darf t_{krit} nur 2.5% der Fläche abschneiden
- Auftretenswahrscheinlichkeit von t_{emp} muss kleiner als 2.5%
- Ist Betrag von t_{emp} größer als t_{krit} so ist Test signifikant

t-Test

Vergleich von ein- und zweiseitigen Fragestellungen

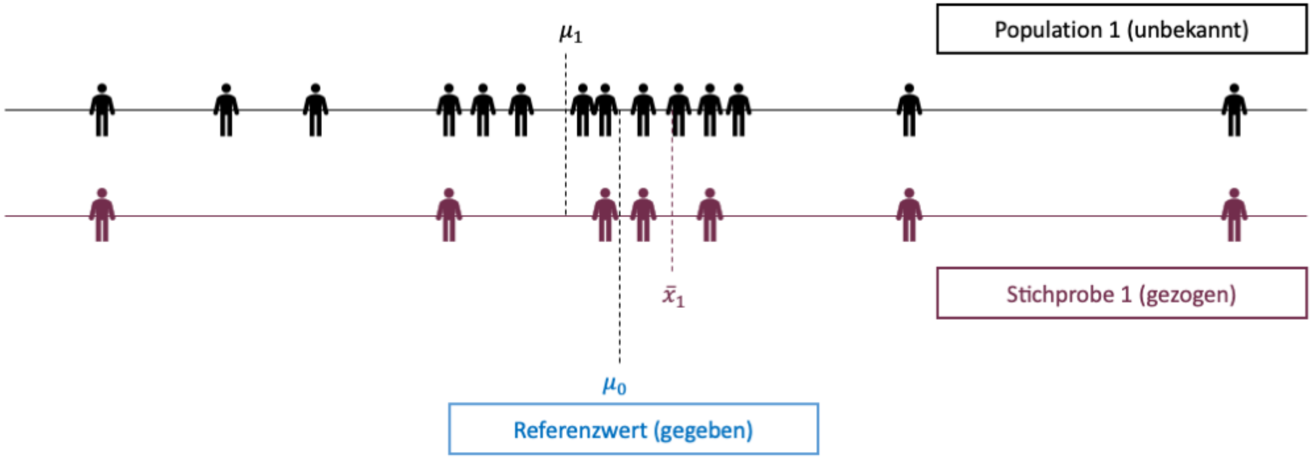
- Mittelwertsdifferenz muss in vorhergesagte Richtung auftreten
- Gesamte 5% liegen auf vorhergesagter Seite der Verteilung
- Folge: Gleiche empirische Mittelwertsdifferenz wird bei einseitigen Hypothesen leichter signifikant (Betrag von t_{krit} ist kleiner, bzw. Ablehnungsbereich ist größer).

Einseitige Fragestellung



t-Test

Ein-Stichproben t-Test



t-Test

Ein-Stichproben t-Test

Beispiel:

- Stichprobe von $n = 36$ Schülern mit Unterricht in Geometrie absolvieren Raumvorstellungstest normiert in Population auf $\mu = 100$
- Es soll die Hypothese geprüft werden, dass Schülern mit Geometrieunterricht im Schnitt besser sind.
- $n = 36$, Testpunkte normalverteilt, $\bar{x}_{Geo} = 101.32$, $s_{Geo} = 4.15$
- H_0: $\mu_{Geo} \leq \mu$; H_1: $\mu_{Geo} > \mu$; $\alpha = 0.05$

$$t = \frac{101.32 - 100}{\frac{4.15}{\sqrt{36}}} = 1.91$$

- $df = 35$; $t_{30;0.95} = 1.70$, $t_{40;0.95} = 1.68$
- Richtung stimmt: unter H_1 positiver t-Wert erwartet
- $1.91 > 1.69 \rightarrow H_0$ verwerfen
- Interpretation: Schüler mit Geometrieunterricht zeigen überdurchschnittliche Testleistungen.

t-Test

Unabhängige vs. abhängige Stichproben

Abhängige Stichproben:

- Elemente der zwei Stichproben können einander paarweise zugeordnet werden

Beispiele:

- die gleichen Personen wurden 2 Mal befragt (Messwiederholungen)
- es handelt sich um Paare (Geschwisterpaare, Ehepaare,..)
- Personen wurden aufgrund eines oder mehrerer Variablen parallelisiert wurden (z.B. aufgrund eines Vortests werden je zwei Personen mit gleicher Punktezahl zu Paaren zusammengefasst).

Unabhängige Stichproben:

- Es besteht keine Beziehung zwischen den Elementen der Stichproben.
- Werte in der einen Stichprobe erlauben keine Vorhersage über Werte in der anderen Stichprobe (unkorreliert).

Beispiele:

- Zufällige Zuteilung von Personen in Versuchsgruppe (VG) und Kontrollgruppe (KG) in einem Experiment
- Zufallsstichproben aus zwei unterschiedlichen Populationen
- Frauen und Männer (wobei es sich nicht um Paare handeln darf).

t-Test

t-Test für unabhängige Stichproben mit gleichen Varianzen

- Unterscheiden sich die Mittelwerte zweier unabhängiger Stichproben aus normalverteilten Grundgesamtheiten signifikant voneinander?
- Wichtigster Wert für t-Test (Effekt von Interesse): **Mittelwertsdifferenz** $\bar{x}_1 - \bar{x}_2$
- Die dichotome Gruppenvariable ist beim t-Test die UV, die numerische Variable, deren Mittelwerte berechnet werden, die AV

Ungerichtete Hypothese:

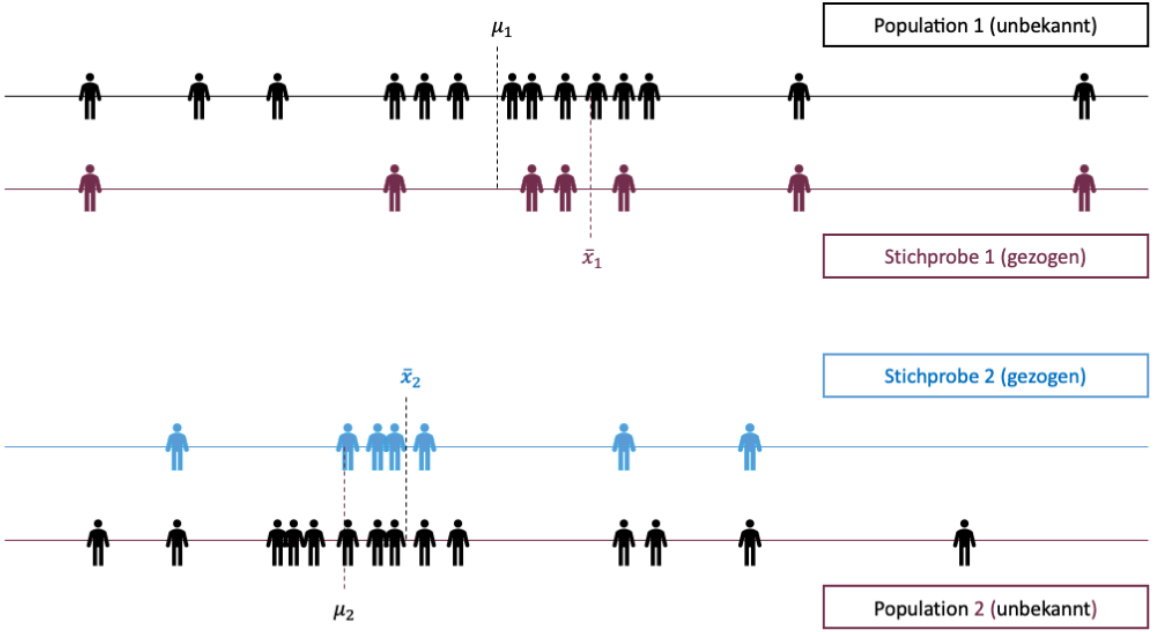
- H_0: $\mu_1 = \mu_2$ bzw. $\mu_1 - \mu_2 = 0$ und $\sigma_1 = \sigma_2 = \sigma$
- H_1: $\mu_1 \neq \mu_2$ bzw. $\mu_1 - \mu_2 \neq 0$ und $\sigma_1 = \sigma_2 = \sigma$

Gerichtete Hypothese z.B.:

- H_0: $\mu_1 \leq \mu_2$ bzw. $\mu_1 - \mu_2 \leq 0$ und $\sigma_1 = \sigma_2 = \sigma$
- H_1: $\mu_1 > \mu_2$ bzw. $\mu_1 - \mu_2 > 0$ und $\sigma_1 = \sigma_2 = \sigma$

t-Test

t-Test für unabhängige Stichproben mit gleichen Varianzen



t-Test

t-Test für unabhängige Stichproben mit gleichen Varianzen

Voraussetzungen:

- Unabhängige Stichproben
- Metrische AV
- Normalverteilung in beiden Populationen
- Homogene Varianzen

Folgen verletzter Voraussetzungen:

- Sind Voraussetzungen für t-Test erfüllt, ist er der mächtigste Test zum Vergleich zweier unabhängiger Stichproben.
- Ist eine (oder mehrere) der Voraussetzungen nicht erfüllt, liegt keine t-Verteilung vor, das reale α entspricht nicht dem vorgegebenen α und es kommt zu Fehlentscheidungen.

t-Test

t-Test für unabhängige Stichproben mit gleichen Varianzen

Prüfung der Voraussetzungen:

Unabhängige Stichproben

- Kein formeller Test $\rightarrow$ Beurteilung anhand der Kenntnis Studiendesigns (z.B. Liegt eine Messwiederholung vor?)

Metrische AV

- Kein formeller Test $\rightarrow$ Beurteilung anhand der Kenntnis des Messinstruments

Normalverteilung in beiden Populationen

- Prüfung der NV anhand der Verteilung in der Stichprobe (Rückschluss auf Population)
- Graphische Prüfung: Histogramm oder QQ-Plot
- Prüfung mit Signifikanztest (Wenn signifikant $\rightarrow$ NV-Annahme verletzt):
  - Shapiro-Wilk Test (empfohlen für $3 \leq n \leq 5.000)$
  - Kolgomorow-Smirnov Test
  - berechnen wir nicht händisch $\rightarrow$ in Klausur wird angegeben, ob NV in Population angenommen werden kann
- Viele Statistiker nehmen mittlerweile an, dass Tests ab balancierten Gruppengrößen von $n_1 = n_2 \geq 30$ robust sind

Homogene Varianzen

- Prüfung mittels Levene-Test (F-Test) $\rightarrow$ siehe nächste Folie

t-Test

t-Test für unabhängige Stichproben mit gleichen Varianzen

Vergleich zweier unabhängiger Varianzen: F-Test

- Unterscheiden sich die Varianzen zweier unabhängiger Stichproben aus normalverteilten Grundgesamtheiten?
- H_0: $\sigma_1^2 = \sigma_2^2$
- H_1: $\sigma_1^2 \neq \sigma_2^2$

Teststatistik:

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}; df_1 = n_1 - 1, df_2 = n_2 - 1$$

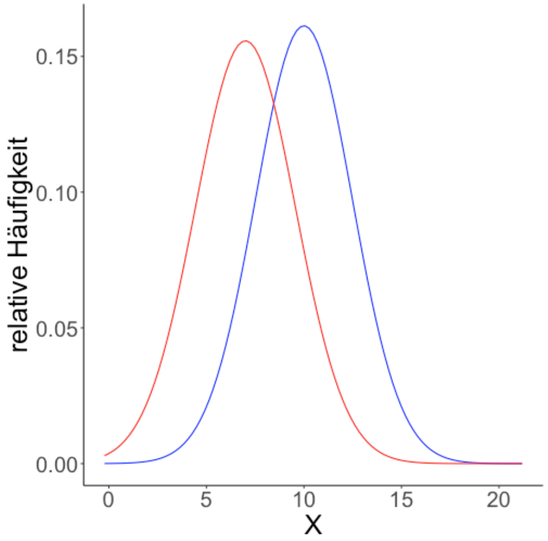
- Voraussetzungen: NV in jeder Population
- ACHTUNG: Größere Varianz muss im Zähler stehen!
- H_0: $\sigma_1^2 = \sigma_2^2$ wird verworfen, wenn $F > F(df_1, df_2, 1 - \alpha/2)$ (kritischer Wert)

t-Test

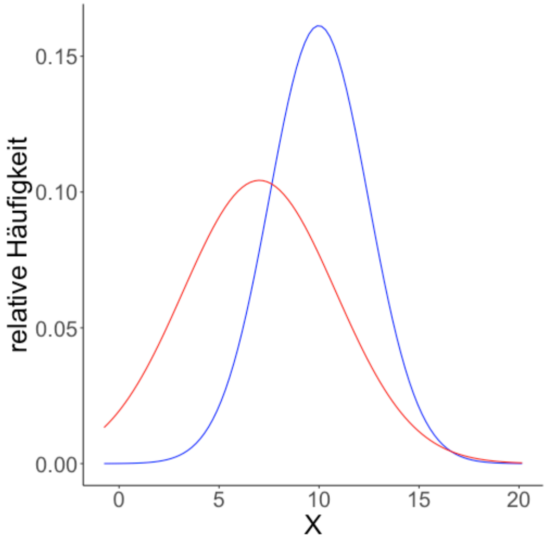
t-Test für unabhängige Stichproben mit gleichen Varianzen

Vergleich zweier unabhängiger Varianzen: F-Test

Beispiel: Gruppenvergleich mit homogenen Varianzen



Beispiel: Gruppenvergleich mit inhomogenen Varianzen



t-Test

t-Test für unabhängige Stichproben mit gleichen Varianzen

Vergleich zweier unabhängiger Varianzen: F-Test

F-Test - Beispiel 1

- Im Rahmen eines Experimentes hört die Versuchsgruppe während der Bearbeitung eines Gedächtnistests Hintergrundmusik, die Kontrollgruppe bearbeitet den Test ohne Hintergrundmusik.
- Frage, die Levene-Test beantwortet: Streuen die Testleistungen der beiden Bedingungen unterschiedlich stark?

Hypothesen des F-Tests:

- H_0: $\sigma^2_{VG} = \sigma^2_{KG}$
- H_1: $\sigma^2_{VG} \neq \sigma^2_{KG}$

t-Test

t-Test für unabhängige Stichproben mit gleichen Varianzen

Vergleich zweier unabhängiger Varianzen: F-Test

F-Test - Beispiel 1

Gegeben:

- $s_{VG}^2 = 8.5$
- $s_{KG}^2 = 4.7$
- $n_{VG} = n_{KG} = 50$
- NV der Daten in VG und KG kann angenommen werden.

$$F = \frac{8.5}{4.7} = 1.81; df_1 = 49, df_2 = 49$$

- Nachprüfen in F-Tabelle: $F_{50,50;0.975} = 1.75 < 1.81 \rightarrow H_0$ wird verworfen

- Levene-Test ist signifikant $\rightarrow$ Die Varianzen sind unterschiedlich (nicht homogen).

t-Test

Prüfgröße t, unabhängige Stichproben, homogene Varianzen

- t ist der Wert, welcher auf der t-Verteilung liegt und uns eine Wahrscheinlichkeitseinschätzung für die Mittelwertsdifferenz erlaubt:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)\cdot\hat{\sigma}_1^2+(n_2-1)\cdot\hat{\sigma}_2^2}{(n_1-1)+(n_2-1)}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}; df = n_1 + n_2 - 2$$

Verbal:

$$t = \frac{\text{Mittelwert Gruppe 1 - Mittelwert Gruppe 2}}{\text{geschätzter Standardfehler der Mittelwertsdifferenz}}$$

- Der Effekt (Mittelwertsdifferenz) wird gewichtet mit der Stichprobengröße und der Streuung innerhalb der Gruppen
- Zweiseitige H_0 wird verworfen, wenn $|t| > t(df; 1 - \alpha/2)$ (kritischer Wert)
- Einseitige H_0 wird verworfen, wenn Abweichung in die erwartete Richtung und $|t| > t(df; 1 - \alpha)$

t-Test

t-Test für unabhängige Stichproben mit gleichen Varianzen

Standardfehler der Mittelwertsdifferenz

- Standardfehler der Mittelwertsdifferenz bei $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\sigma^2 \cdot (\frac{1}{n_1} + \frac{1}{n_2})}$$

- Varianzschätzung innerhalb (der Stichproben)

$$\hat{\sigma}^2 = \frac{(n_1 - 1) \cdot \hat{\sigma_1^2} + (n_2 - 1) \cdot \hat{\sigma_2^2}}{(n_1 - 1) + (n_2 - 1)}$$

Schätzung des Standardfehlers der Mittelwertsdifferenz

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{(n_1 - 1) \cdot \hat{\sigma_1^2} + (n_2 - 1) \cdot \hat{\sigma_2^2}}{(n_1 - 1) + (n_2 - 1)} \cdot (\frac{1}{n_1} + \frac{1}{n_2})}$$

t-Test

t-Test für unabhängige Stichproben mit gleichen Varianzen

Standardfehler der Mittelwertsdifferenz

Warum Varianzschätzung innerhalb?

- Unter H_1 beide Stichproben normalverteilt mit gleichem σ, aber unterschiedlichen Mittelwerten
- Würde man die Stichproben zu einer einzigen zusammenfassen und die Varianz berechnen, entstünde eine zweigipfelige Verteilung und man erhielte eine größere Varianz
- Man spricht bei Varianzschätzung innerhalb auch von gepoolten Varianzen

t-Test

t-Test für unabhängige Stichproben mit gleichen Varianzen

Beispiel 1:

- Es soll überprüft werden, ob zusätzliches autogenes Training einen positiven Effekt bei der Behandlung von Depressionen hat.
- Klassischer Zwei-Gruppen-Versuchsplan: eine Gruppe von Patientinnen erhält nur die konventionelle Therapie (KG), eine zweite Gruppe erhält zusätzlich autogenes Training (VG).
- Operationalisierung des Effekts der Therapien: Scoredifferenz (vorher minus nachher) in einem Depressionsfragebogen.
- H_0: $\mu_{VG} \leq \mu_{KG}, \sigma^2_{VG} = \sigma^2_{KG}$
- H_1: $\mu_{VG} > \mu_{KG}, \sigma^2_{VG} = \sigma^2_{KG}$

Gegeben:

- VG: $\bar{x}_{VG} = 7.22, s^2_{VG} = 6.12; n_{VG} = 53$
- KG: $\bar{x}_{KG} = 4.91, s^2_{KG} = 6.54; n_{KG} = 51$
- NV gegeben, $\alpha = .05$

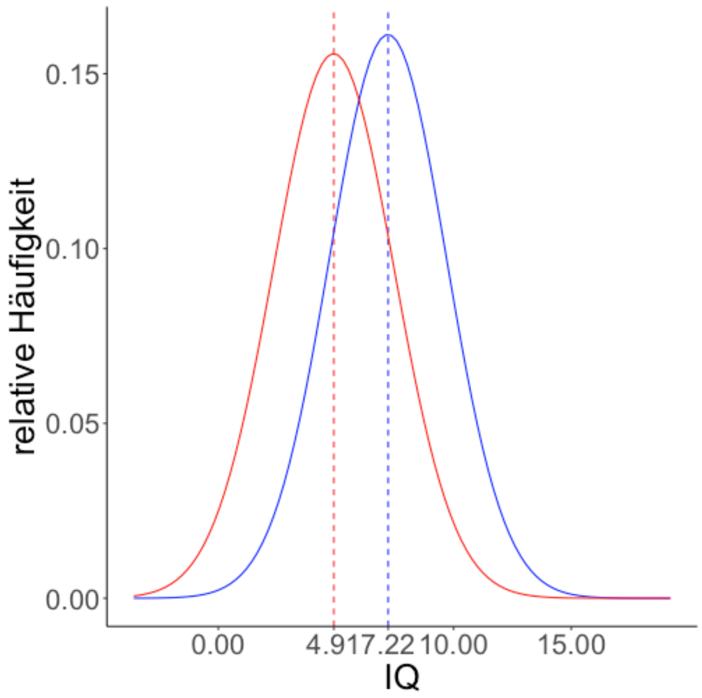
t-Test

t-Test für unabhängige Stichproben mit gleichen Varianzen

Beispiel 1:

Gegeben:

- VG: $\bar{x}_{VG} = 7.22, s^2_{VG} = 6.12; n_{VG} = 53$
- KG: $\bar{x}_{VG} = 4.91, s^2_{VG} = 6.54; n_{VG} = 51$
- NV gegeben, $\alpha = .05$



t-Test

t-Test für unabhängige Stichproben mit gleichen Varianzen

Beispiel 1:

Homogenität der Varianzen: H_0: $\sigma_1^2 = \sigma^2$, $H_1 : \sigma_1^2 \neq \sigma^2$

$$F = \frac{6.54}{6.12} = 1.07; df_1 = 50, df_2 = 52$$

- Prüfung in F-Tabelle: $F_{50,50;0.975} = 1.75 > 1.07 \rightarrow H_0$ wird beibehalten, Varianzen sind homogen.

$$t = \frac{7.22 - 4.91}{\sqrt{\frac{52 \cdot 6.12 + 50 \cdot 6.54}{52+50}(\frac{1}{53} + \frac{1}{51})}} = 4.64; df = 53 + 51 - 2 = 102$$

- Prüfung in t-Tabelle: $t_{102;0.95} = 1.66 < 4.64 \rightarrow H_0$ verwerfen
- Interpretation: Ergebnis spricht dafür, dass autogenes Training zusätzlichen Effekt hat.

t-Test

t-Test für unabhängige Stichproben mit gleichen Varianzen

Bewertung des t-Werts:

- t-Wert schneidet geiwssen Prozentsatz der Fläche einer t-Verteilung ab
- Wahrscheinlichkeit des t-Werts gibt Wahrscheinlichkeit an, ob Nullhypothese zutrifft
- Ergibt t-Test (Nullhypothesentest) eine geringe Wahrscheinlichkeit, ist Ablehnung der Nullhypothese mögllich
- die zugrundeliegenden Populationen haben nicht den gleichen, sondern verschiedene Mittelwerte
- Entscheidungen eines t-Tests sind nie zu 100% sicher

t-Test

Welch's t-Test

- Was wenn Varianzen nicht homogen sind? $\rightarrow$ Annahme des unabhängigen t-Tests verletzt
- Prüfgröße t mit gepoolten Varianzen nicht anwendbar
- Näherungslösung: Prüfgröße t nach Welch

$$t_{Welch} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Es wird eine Korrektur der Freiheitsgrade (df) vorgenommen:

$$df = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{(\frac{s_1^2}{n_1})^2}{n_1 - 1} + \frac{(\frac{s_2^2}{n_2})^2}{n_2 - 1}}$$

- korrigierte df abrunden auf die nächste ganze Zahl
- kritischen Wert aus t-Tabelle ablesen

t-Test

Freiheitsgerade einer t-Verteilung

- Exakte Form der t-Verteilung ist abhängig vom Stichprobenumfang $\rightarrow$ deckt sich nicht exakt mit z-Verteilung
- Unterschied zwischen t-Verteilung und z-Verteilung $\rightarrow$ in t-Verteilung müssen 2 Schätzer eingehen
    - empirische Mittelwertsdifferenz
    - Standardfehler der Mittelwertsdifferenz

Strebt N gegen unendlich nähert sich t-Verteilung der z-Verteilung an.

t-Test

Abhängiger t-Test

- wird z.B. bei Messwiederholungen eingesetzt $(t_0 - t_1)$
- Ähnliches Prinzip wie unabhängiger t-Test
- Betrachtet nicht die Mittelwerte beider Zeitpunkte sondern die Differenz der Werte jeder einzelnen Versuchsperson

$\rightarrow$ es geht nur der Unterschied der Messwerte zwischen 1. und 2. Messung in die Auswertung mit ein

- allgemeine Unterschiede, die zwischen den Personen zu beiden Messzeitpunkten wirken gehen nicht mit ein
- Der relevante Effekt für den abhängigen t-Test ist also:

$$\bar{x}_d = \frac{\sum_{i=1}^{N} d_i}{N}$$

Hypothesen:

- H_0: $\mu_d \leq 0$ bzw. $\mu_d = 0$
- H_1: $\mu_d > 0$ bzw. $\mu_d \neq 0$

t-Test

Abhängiger t-Test

Berechnung der Teststatistik:

Da dieser Test die Verteilung der Mittelwerte von Differenzen betrachtet, ergibt sich eine andere Schätzung der Streuung:

$$t_{abhängig} = \frac{\bar{x}_d}{\hat{\sigma}_{\bar{x}_d}}$$

Berechnung des Standardfehlers der Differenzen:

$$\hat{\sigma}_{\bar{x}_d} = \frac{\hat{\sigma}_d}{\sqrt{N}}$$

Schätzung der Streuung der Differenzen:

$$\hat{\sigma}_d = \sqrt{\frac{\sum_{i=1}^{N}(d_i - \bar{x}_d)^2}{N-1}}$$

Berechnung der Freiheitsgrade:

- $df = N - 1$ (Anzahl der Messwertpaare - 1)

t-Test

Abhängiger t-Test

Beispiel 1:

- Übungseffekt bei Wiederholung einer motorischen Aufgabe (Tippen einer kurzen Sequenz).
- AV = Anzahl richtiger Sequenzen; UV = Zeitpunkt (dichotom): t_0 vs. t_1
- Ungerichtete Hypothese $\mu_d = 0$: Übung könnte Leistung verbessern vs. Ermüdung könnte Leistung verschlechtern.

Gegeben:

- $N = 36$ Teilnehmer:innen
- $\bar{x}_d = 0.722$ und $\hat{\sigma}_d = 4.186$
- $\alpha = .05$

t-Test

Abhängiger t-Test

- $N = 36$ Teilnehmer:innen
- $\bar{x}_d = 0.722$ und $\hat{\sigma}_d = 4.186$
- $\alpha = .05$

$$t_{abhängig} = \frac{\bar{x}_d}{\hat{\sigma}_{\bar{x}_d}} = \frac{\bar{x}_d}{\frac{\hat{\sigma}_d}{\sqrt{N}}} = \frac{0.722}{\frac{4.186}{\sqrt{36}}} = \frac{0.722}{0.698} = 1.035$$

- kritischer Wert (nächster df Wert in Tabelle ist 30) von $t_{30,0.975} = 2.042$

- t-Wert ist nicht größer als kritischer Wert.

- Interpretation: Es besteht kein signifikanter Unterschied zwischen den beiden Messzeitpunkten.

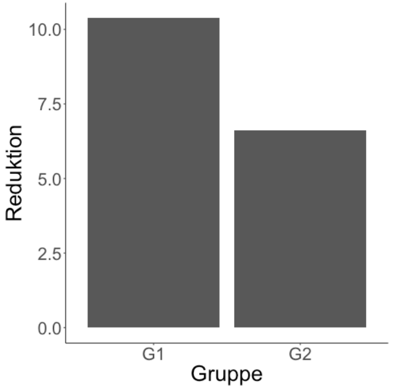
t-Test

Berechnung in R

Unabhängiger t-Test:

Beispiel: Vergleich der Reduktion von Symptomatik zwischen 2 Gruppen $(N = 16)$

Gruppe	Mittelwert	SD
G1	10.38	2.26
G2	6.62	2.13




```r
t.test(Reduktion ~ Gruppe, data = df, var.equal = T)
```

```
##
##     Two Sample t-test
##
## data:  Reduktion by Gruppe
## t = 3.4093, df = 14, p-value = 0.004233
## alternative hypothesis: true difference in means between
## 95 percent confidence interval:
##  1.390909 6.109091
## sample estimates:
## mean in group G1 mean in group G2
##           10.375            6.625
```

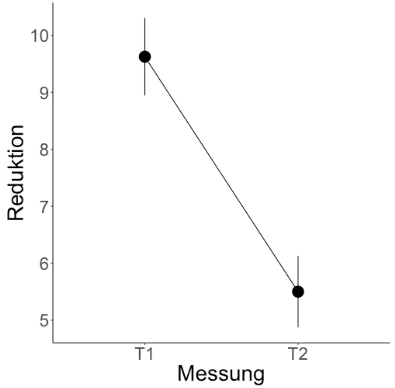
t-Test

Berechnung in R

Abhängiger t-Test:

Beispiel: Prüfung, ob Symptomreduktion zwischen 2 Zeitpunkten signifikant ist $(N = 8)$

Messung	Mittelwert	SD
T1	9.62	1.92
T2	5.50	1.77




```
t.test(Reduktion ~ Messung, data = df, paired = T)
```

```
##
##     Paired t-test
##
## data:  Reduktion by Messung
## t = 5.0834, df = 7, p-value = 0.001426
## alternative hypothesis: true mean difference is not equal
## 95 percent confidence interval:
##  2.20618 6.04382
## sample estimates:
## mean difference
##           4.125
```

Take-aways

- t-Test ist eine Auswertungsverfahren für den **Vergleich von 2 Mittelwerten**
- Vergleich von Mittelwerten 2er Gruppen erfolgt mittels **unabhängigem t-Test**, von Mittelwerten 2er Zeitpunkte mit **abhängigem t-Test** und Vergleich von Mittelwert mit vorgegebenem Referenzwert mit **Ein-Stichproben t-Test**
- **Voraussetzungen** für unabhängige t-Test umfassen unabhängige Daten, Intervallskalenniveau, Normalverteilung und Varianzhomogenität
- Vorgehen: Berechnung von t_{emp} und Vergleich mit t_{krit}, welcher aus **t-Tabelle** abgelesen wird
- Vorsicht: t-Test kann zu **Fehlentscheidungen** führen (s.h. α-Fehler und β-Fehler)
- Prüfung, ob Effekt (Mittelwertunterschiede/Mittelwertsdifferenzen), die in Stichprobe gemessen wurden auf Population **generalisierbar** sind.