M24 Statistik 1: Wintersemester 2024 / 2025

Seminar 09: t-Test

MSc Albert Anoschin & Prof. Matthias Guggenmos

Health and Medical University Potsdam

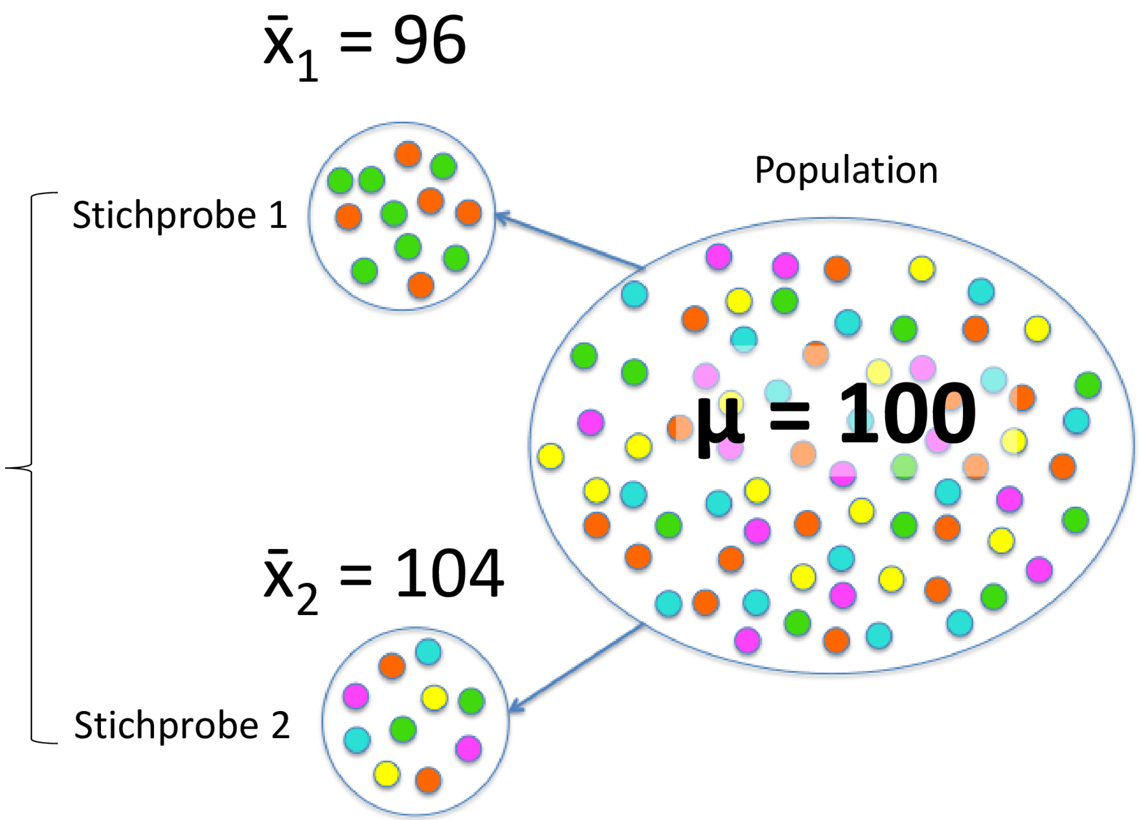
Wiederholungsfragen

1. Was ist das Ziel der Nullhypothesentestung (Signifikanztestung)?
2. Was ist das Signifikanzniveau und was sagt $\alpha = .05$ aus?
3. Was sagt der p-Wert aus?
4. Was versteht man unter einer empirischen und einer kritischen Prüfgröße?
5. Was versteht man unter einem Ablehnungsbereich?
6. Wovon ist die Größe des Ablehnungsbereichs abhängig?
7. Was zeigt die z-Wert-Tabelle?

Weshalb benötigen wir Signifikanztests?

Genauso wie die Messwerte einzelner Personen um den Mittelwert der Stichprobe streuen, streuen die Mittelwerte verschiedener Stichproben um den wahren Populationsmittelwert.

Diese Streuung der Stichprobenmittelwerte um den wahren Populationsmittelwert lässt sich mit der *Stichprobenverteilung* modellieren. Der Streuungsparameter der Stichprobenverteilung ist der *Standardfehler*.

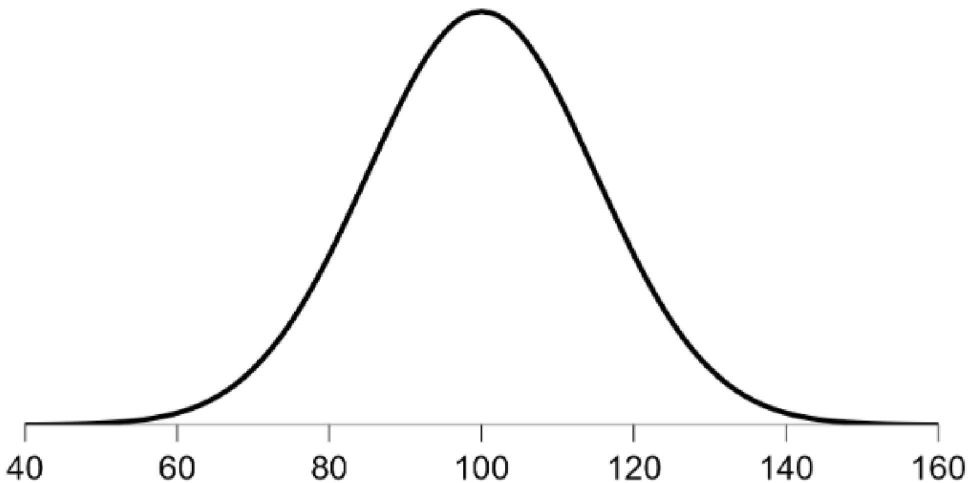


Hypothesentestung

Verteilung des IQ in der Population

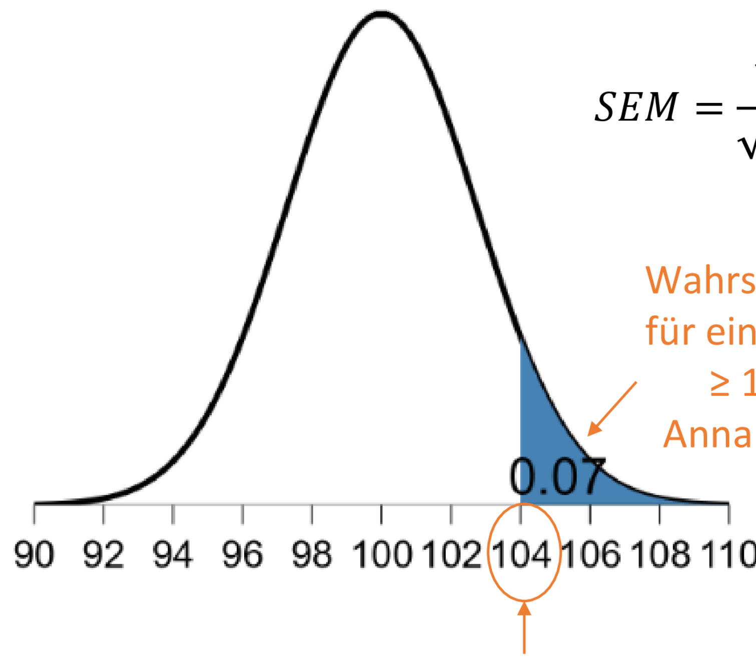
$$\mu = 100$$

$$\sigma = 15$$



Theoretische Stichprobenverteilung der Mittelwerte unter Annahme der H_0

$$SEM = \frac{15}{\sqrt{30}} = 2.73$$



Wahrscheinlichkeit für einen Wert von ≥ 104 unter Annahme der H0

Empirischer Mittelwert einer Stichprobe

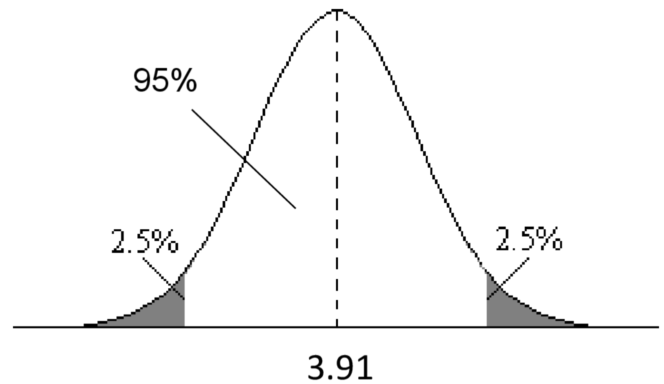
Prinzip der Einstichproben-Signifikanztestung

- Die H_0 nimmt an, dass **kein Unterschied** in einem Merkmal zwischen der Stichprobe und der Referenzpopulation besteht
- *Beispiel H0: Gewissenhaftigkeit (Merkmal) der Psychologen* $\bar{x} = 3.35$ *stimmt mit Gewissenhaftigkeit der Normpopulation* $\mu = 3.91$ *überein.*
- Anders formuliert: Laut H_0 stammt die Stichprobe der Psychologen aus der Normpopulation (wurde aus ihr gezogen)
- Folglich sollte der Mittelwert der gezogenen Psychologen-Stichprobe **innerhalb eines bestimmten Streuungsbereichs** um den Populationsmittelwert (der Norm) liegen
- Der Standardfehler gibt ein 68%iges Konfidenzintervall an: Mit 68% Wahrscheinlichkeit sollte der Stichprobenmittelwert innerhalb ± 1 SEM liegen, wenn die Stichprobe aus der Normpopulation stammt (→ Nullhypothese).
- Beim Signifikanztest wird ein 95%iges Konfidenzintervall angelegt: Liegt der Stichprobenmittelwert außerhalb dieses Streuungsbereichs, nimmt man an, dass die Stichprobe aus einer **anderen** Population als der Normpopulation stammt (mit 5% Irrtumswahrscheinlichkeit).

Erwartete Streuung der Stichprobenmittelwerte unter der $\mathbf{H_0}$

=

Stichprobenverteilung unter der $\mathbf{H_0}$

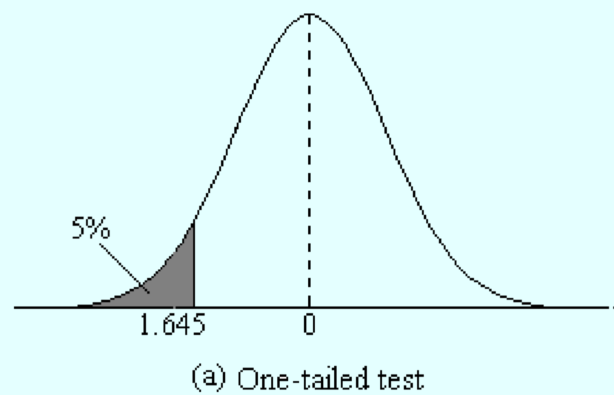


Übung z-Test

Sie möchten prüfen, ob Psychologen ($n = 17$) **weniger** gewissenhaft sind als eine Normpopulation. In der Normpopulation (altersgematchte Population mit hohem Bildungsniveau) ergab sich ein **Mittelwert** $\mu = \mathbf{3.91}$ und eine **Standardabweichung** $\sigma = \mathbf{0.83}$. Sie führen einen z-Test durch.

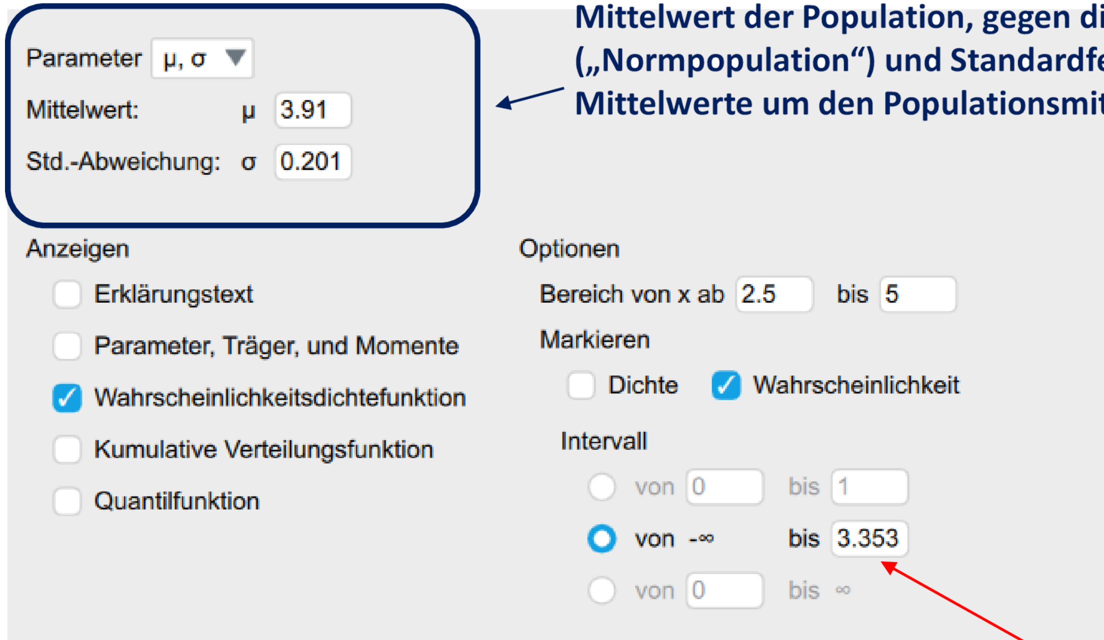
1. **Mittelwert** der Gewissenhaftigkeit von Psychologen:
   $\bar{x} = 3.353$
2. **Standardfehler** der Stichprobenverteilung unter der H_0:
   $SEM = \frac{0.83}{\sqrt{17}} = 0.201$

H_0: **Psychologen ≥ Population**
H_1: **Psychologen < Population**

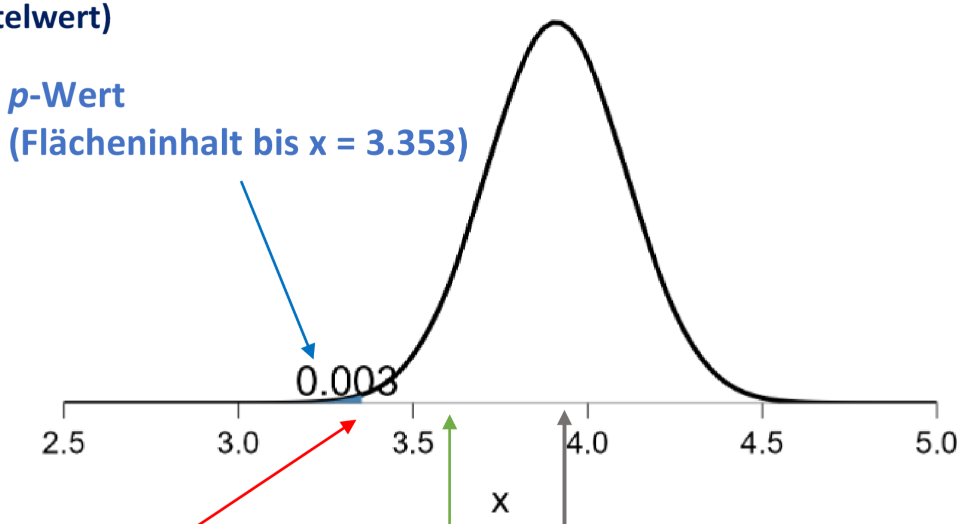


Signifikanztest ohne z-Transformation

Stichprobenverteilung unter der H0 ohne z-Transformation: Mittelwert der Population, gegen die getestet wird („Normpopulation“) und Standardfehler (erwartete Streuung der Mittelwerte um den Populationsmittelwert)



p-Wert (Flächeninhalt bis x = 3.353)



Prüfgröße: Mittelwert der Stichprobe
$\bar{x}$ = 3.353

Lage des kritischen Werts

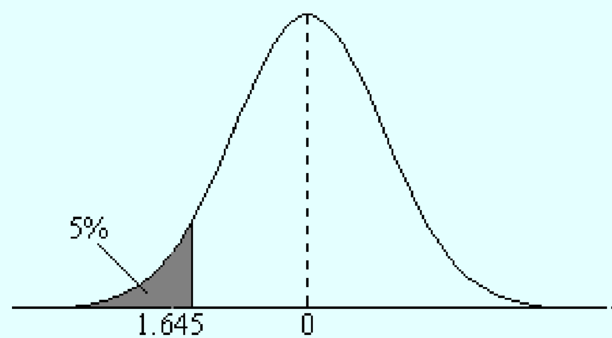
Mittelwert der Population
μ = 3.91

Übung z-Test

Sie möchten prüfen, ob Psychologen $(n = 17)$ **weniger** gewissenhaft sind als eine Normpopulation. In der Normpopulation (altersgematchte Population mit hohem Bildungsniveau) ergab sich ein **Mittelwert** $\mu = \mathbf{3.91}$ und eine **Standardabweichung** $\sigma = \mathbf{0.83}$. Sie führen einen z-Test durch.

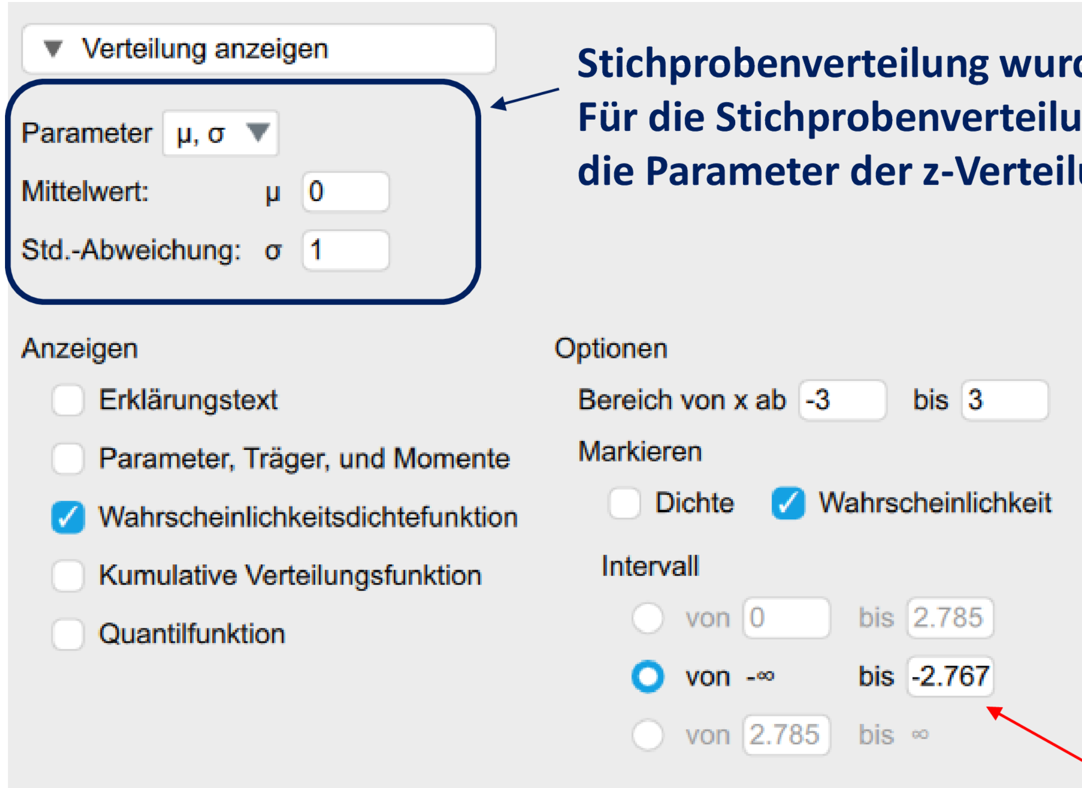
H_0: **Psychologen ≥ Population**
H_1: **Psychologen < Population**

1. **Mittelwert** der Gewissenhaftigkeit von Psychologen:
   $\bar{x} = 3.353$
2. **Standardfehler** der Stichprobenverteilung unter der H_0:
   $SEM = \frac{0.83}{\sqrt{17}} = 0.201$
3. **Empirische Prüfgröße** $z = (3.353 - 3.91)/0.201 = -2.77$
4. Lage des Ablehnungsbereichs: rechtsseitig, $z \leq -1.645$
5. Effektstärke $d = ?$

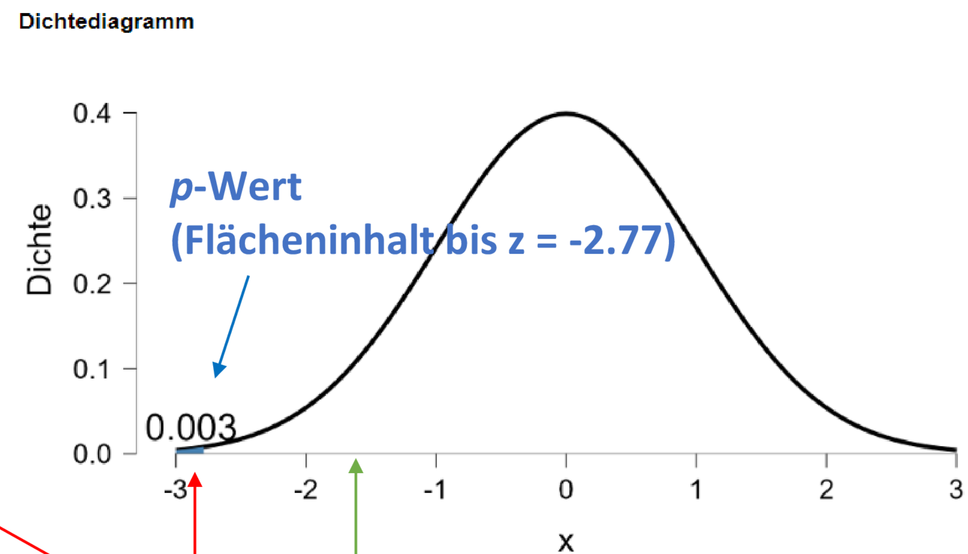


(a) One-tailed test

z-Test (linksseitig)



Stichprobenverteilung wurde in den z-Raum transformiert:
Für die Stichprobenverteilung unter der H0 werden hier daher die Parameter der z-Verteilung eingetragen (Standardnormalverteilung)



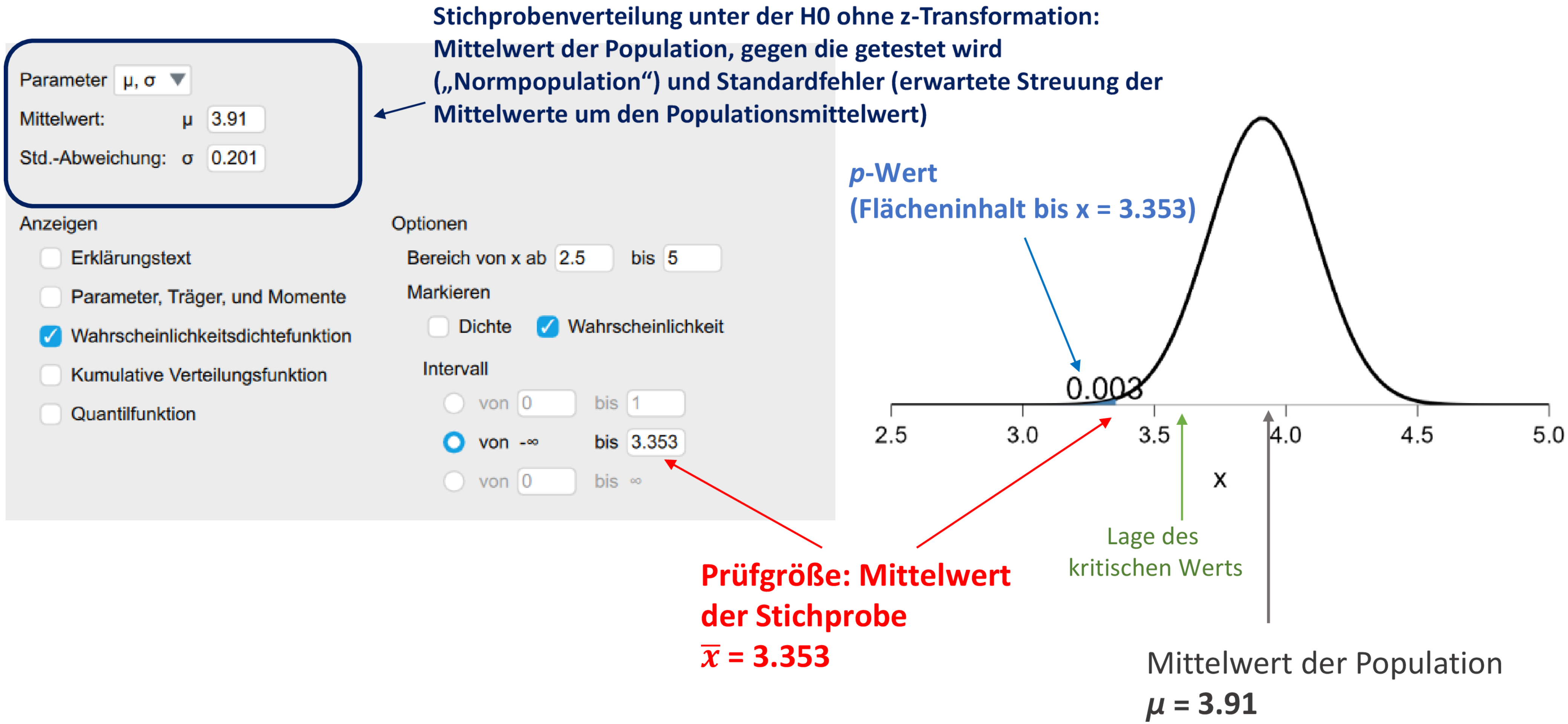
***p*-Wert**
(Flächeninhalt bis z = -2.77)

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

empirischer z-Wert (Prüfgröße)

kritischer z-Wert = - 1.645

Signifikanztest ohne z-Transformation

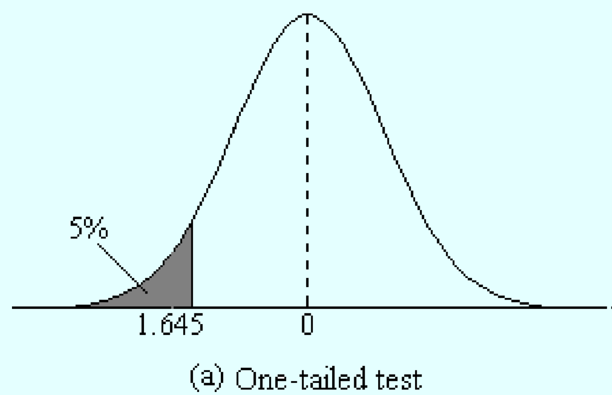


Übung z-Test

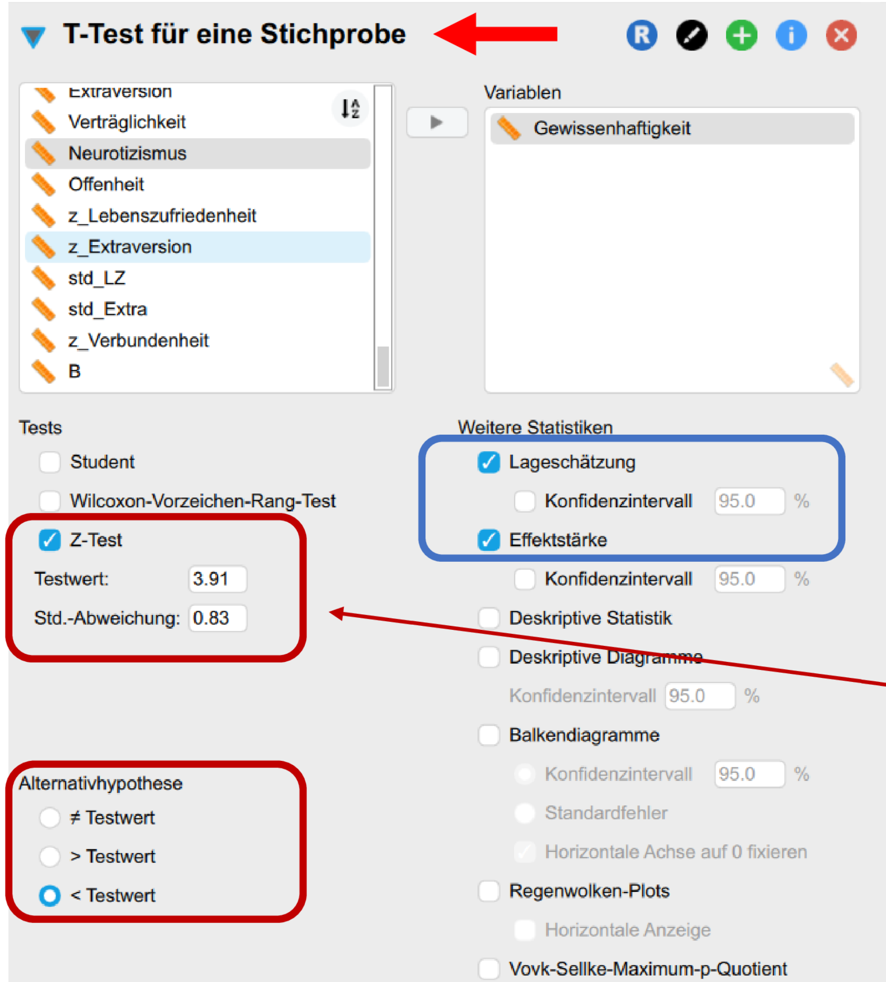
Sie möchten prüfen, ob Psychologen ($n = 17$) **weniger** gewissenhaft sind als eine Normpopulation. In der Normpopulation (altersgematchte Population mit hohem Bildungsniveau) ergab sich ein **Mittelwert** $\mu = \mathbf{3.91}$ und eine **Standardabweichung** $\sigma = \mathbf{0.83}$. Sie führen einen z-Test durch.

H_0: **Psychologen ≥ Population**
H_1: **Psychologen < Population**

1. **Mittelwert** der Gewissenhaftigkeit von Psychologen: $\bar{x} = 3.353$
2. **Standardfehler** der Stichprobenverteilung unter der H_0: $SEM = \frac{0.83}{\sqrt{17}} = 0.201$
3. Empirische Prüfgröße** $z = (3.353 – 3.91)/0.201 = -2.77$
4. Lage des Ablehnungsbereichs: rechtsseitig, $z \leq -1.645$
5. Effektstärke $d = (3.353 – 3.91)/0.83 = -0.67$



z-Test mit JASP (im Menü: t-Test für eine Stichprobe)



T-Test für eine Stichprobe

	Z	p	Mittelwertsdifferenz	Cohens d	Std.-Fehler Cohens d
Gewissenhaftigkeit	−2.767	0.003	−0.557	−0.671	0.268

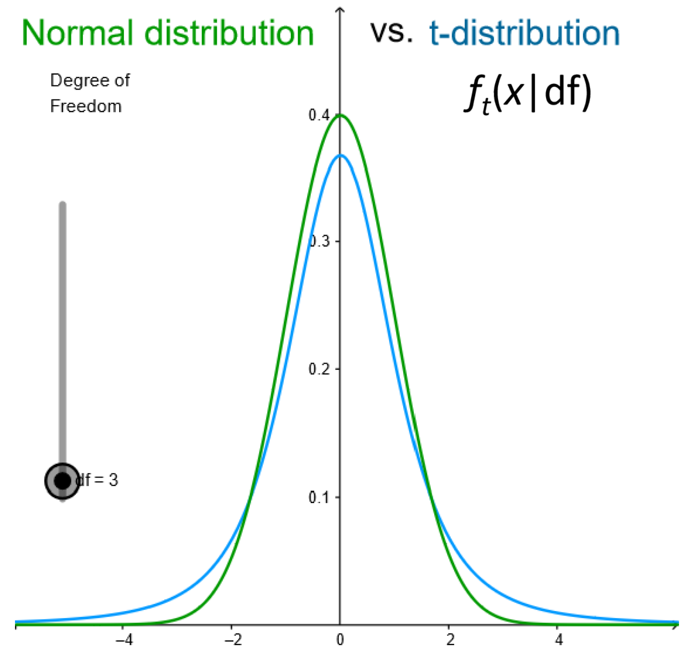
Hinweis. Beim Z-Test wird die Effektstärke mit Cohens *d* angegeben (basierend auf der eingegebenen Standardabweichung der Grundgesamtheit).
Hinweis. Beim Z-Test gibt die Alternativhypothese an, dass der Mittelwert kleiner als 3.91 ist.
Hinweis. Für die Z-Test ist die Schätzung der Lagedifferenz durch die Stichprobenmittelwertdifferenz *d* gegeben.
Hinweis. Z-Test.

Kennwerte der Populationsverteilung (Achtung: nicht Standardfehler! Die Stichprobenverteilung wird basierend auf der Populationsstandardabweichung und der Stichprobengröße berechnet)

Signifikanztestung in der Praxis: Einführung in den t-Test

Signifikanztestung in der Praxis: t-Test

- In der Praxis sind die Populationsparameter μ und σ unbekannt, sodass z-Test nicht durchgeführt werden kann!
- Anstelle der Standardnormalverteilung (z-Verteilung) erfolgt die Signifikanztestung mittels der Student'schen t-Verteilung.
- Die t-Verteilung ist durch einen einzigen Parameter definiert: den Freiheitsgraden df (abhängig von n).
- Anstelle des empirischen z-Werts wird ein empirischer t-Wert als Prüfgröße berechnet und mit einem kritischen t-Wert verglichen, um die Signifikanz festzustellen.
- Die Berechnung des t-Werts folgt dem selben Prinzip, allerdings diesmal mit geschätzter Populationsstreuung $\hat{\sigma}$

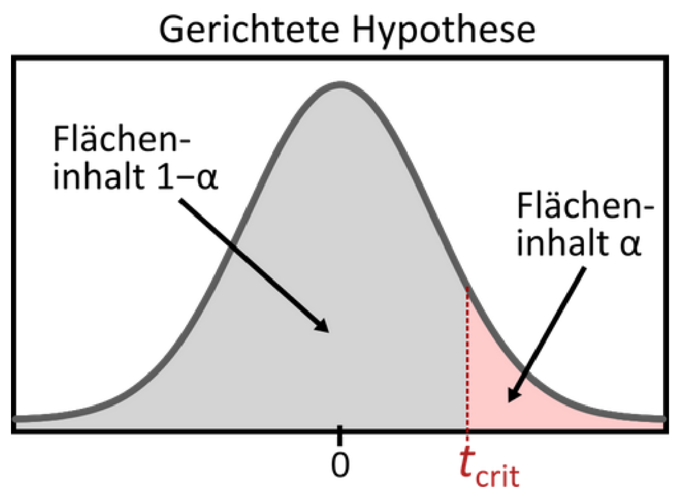


$$t = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}} \qquad z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Signifikanztestung in der Praxis: t-Test

$$t = \frac{\bar{x} - \mu}{\hat{\sigma} / \sqrt{n}}$$

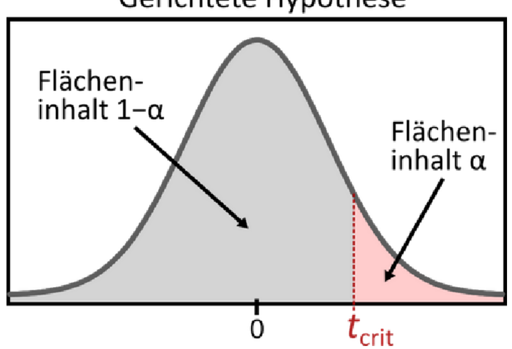
- Das statistische Verfahren **bei unbekannten Populationsparametern** wird folglich ***t-Test*** genannt.
- **Er dient der Prüfung von:**
  - Unterschied eines Stichprobenmittelwerts von einem erwarteten Wert (Einstichproben-t-Test, vgl. z-Test zuvor).
  - Mittelwertunterschieden zwischen zwei abhängigen Stichproben (vgl. Messwiederholung / within-subjects design).
  - Mittelwertunterschieden zwischen zwei unabhängigen Stichproben (vgl. between-subjects design)
- Wie beim z-Test ist eine einseitige und zweiseitige Testung möglich.



Selbes Prinzip!

Die t-Tabelle

- Die Freiheitsgrade df sind abhängig von der Stichprobengröße n.
  - Bei einer Stichprobe: $\mathrm{df} = n - 1$
  - Bei Vergleich zweier Gruppen: $\mathrm{df} = n_A + n_B - 2$
- **Somit gibt es für jede Stichprobengröße eine eigene t-Verteilung mit eigenen kritischen t-Werten!**



t-Verteilung

In der Tabelle finden sich die kritischen *t*-Werte. Das Signifikanzniveau wird durch die Fläche angegeben. Beim einseitigen Testen auf dem 5%-Niveau beträgt die relevante Fläche 0,95; beim zweiseitigen Testen entsprechend 0,975. Der empirische *t*-Wert muss gleich groß oder größer sein als der kritische *t*-Wert aus der Tabelle, um auf dem entsprechenden Niveau signifikant zu sein.

	Fläche						
df	0,8	0,85	0,9	0,95	0,975	0,99	0,995
1	1,377	1,964	3,078	6,314	12,706	31,821	63,657
2	1,001	1,386	1,886	2,92	4,303	6,965	9,925
3	0,978	1,25	1,638	2,353	3,182	4,541	5,841
4	0,941	1,19	1,533	2,132	2,776	3,747	4,604
5	0,92	1,156	1,476	2,015	2,571	3,365	4,032
6	0,906	1,134	1,44	1,943	2,447	3,143	3,707
7	0,896	1,119	1,415	1,895	2,305	2,998	3,5
8	0,889	1,108	1,397	1,86	2,306	2,896	3,355
9	0,883	1,1	1,383	1,833	2,262	2,821	3,25
10	0,879	1,093	1,372	1,813	2,228	2,764	3,169
11	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,873	1,083	1,356	1,782	2,179	2,681	3,055

Beispiel: t-Test für abhängige Stichproben

- Sie prüfen in einer Stichprobe von 4 Personen, ob die Teilnahme an einem Statistikseminar die Punktzahl in einem Statistik-Testat verbessert. In Ihrem Datensatz liegen für jede Person die Punktzahlen zu Zeitpunkt 1 (vor Teilnahme) und zu Zeitpunkt 2 (nach Teilnahme) vor.
- Die Mittelwertdifferenz beträgt 5 Punkte.
- Berechnen Sie die Freiheitsgrade $\mathrm{df} = n - 1$.
- Lesen Sie den kritischen t-Wert in der t-Tabelle ab.
- **Ihr empirischer t-Wert beträgt 1.497.**
- Ist das Ergebnis signifikant? Was ist Ihre Schlussfolgerung hinsichtlich des Effekts eines Statistikseminars?

	ID	Stats_Pre	Stats_Post
1	1	47	51
2	2	52	61
3	3	59	70
4	4	81	77

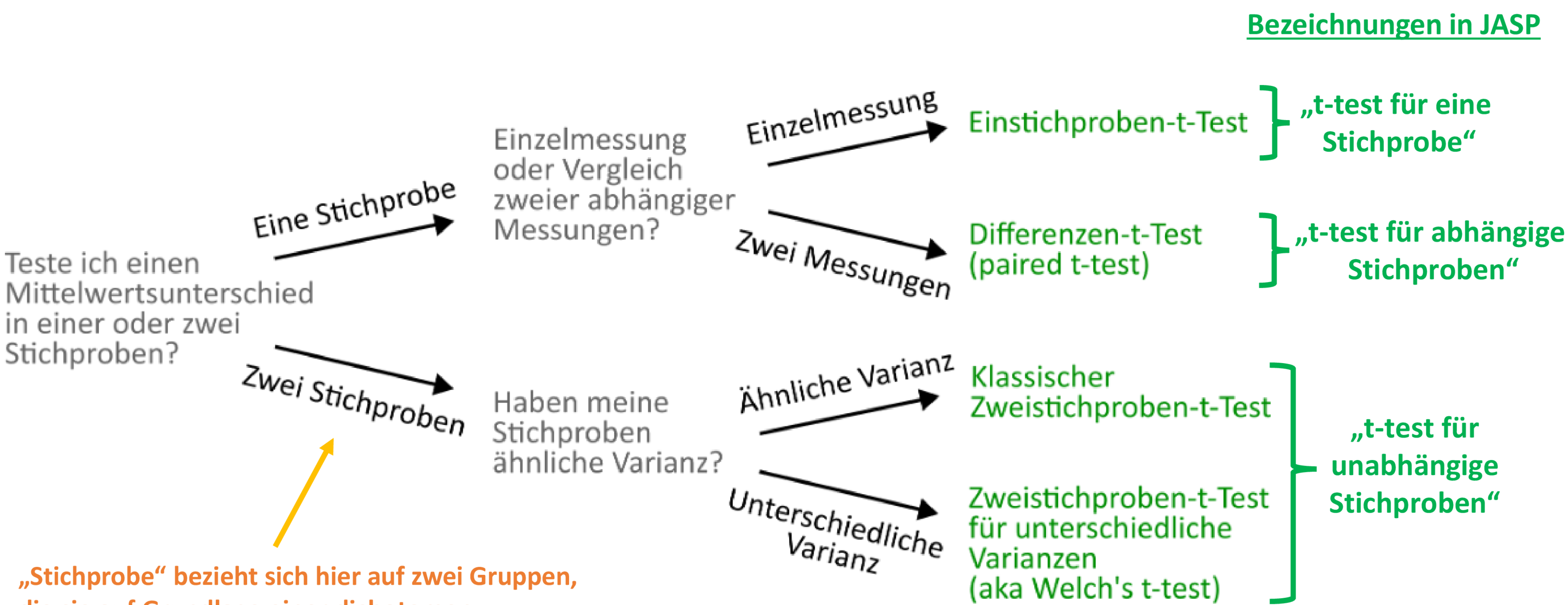
	Stats_Pre	Stats_Post
Valid	4	4
Missing	0	0
Mean	59.750	64.750
Std. Deviation	14.997	11.266

t-Verteilung

In der Tabelle finden sich die kritischen *t*-Werte. Das Signifikanzniveau wird durch die Fläche angegeben. Beim einseitigen Testen auf dem 5%-Niveau beträgt die relevante Fläche 0,95; beim zweiseitigen Testen entsprechend 0,975. Der empirische *t*-Wert muss gleich groß oder größer sein als der kritische *t*-Wert aus der Tabelle, um auf dem entsprechenden Niveau signifikant zu sein.

	Fläche						
df	0,8	0,85	0,9	0,95	0,975	0,99	0,995
1	1,377	1,964	3,078	6,314	12,706	31,821	63,657
2	1,001	1,386	1,886	2,92	4,303	6,965	9,925
3	0,978	1,25	1,638	2,353	3,182	4,541	5,841
4	0,941	1,19	1,533	2,132	2,776	3,747	4,604
5	0,92	1,156	1,476	2,015	2,571	3,365	4,032
6	0,906	1,134	1,44	1,943	2,447	3,143	3,707
7	0,896	1,119	1,415	1,895	2,305	2,998	3,5
8	0,889	1,108	1,397	1,86	2,306	2,896	3,355
9	0,883	1,1	1,383	1,833	2,262	2,821	3,25
10	0,879	1,093	1,372	1,813	2,228	2,764	3,169
11	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,873	1,083	1,356	1,782	2,179	2,681	3,055

Entscheidungsbaum t-Test



„Stichprobe“ bezieht sich hier auf zwei Gruppen, die sie auf Grundlage einer dichotomen Variable in ihrem Datensatz bilden können.

Signifikanztestung von Zusammenhängen

- Korrelationen und Regressionskoeffizienten werden ebenfalls mit dem t-Test auf Signifikanz geprüft (in JASP allerdings mit dem Modul Regression → Korrelation).

Teste ich einen Zusammenhang (Korrelation/Regression) in einer Stichprobe? → Rechne *r* oder b_1 in einen t-Wert um → Einstichproben-t-Test

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

$$\left(\text{Reminder: } r = \frac{\hat{\sigma}_X}{\hat{\sigma}_Y} b_1\right)$$

df = n – 2

Gruppenarbeit mit „Semesterdatensatz“

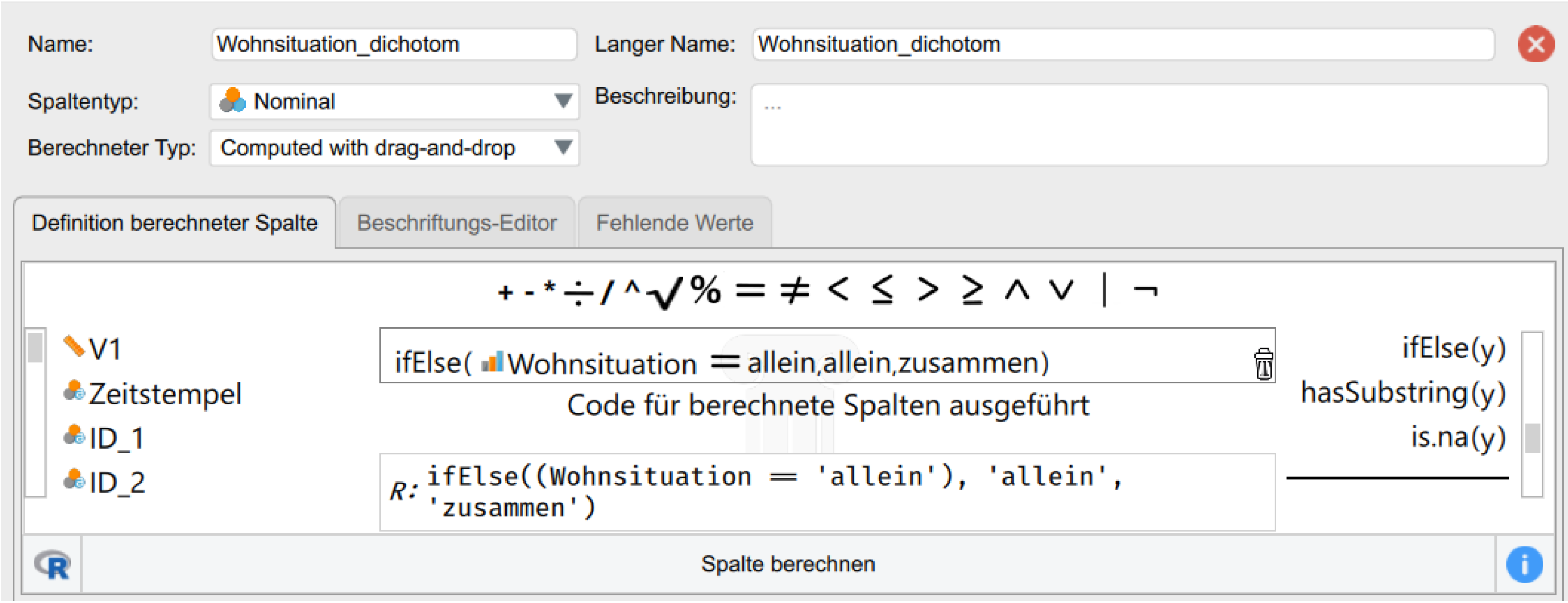
1. Bilden Sie Gruppen á 3 Personen.
2. Schauen Sie im TMD_Codebook nach, welche Variablen wir erfasst haben. Sie können auch diesen Link nutzen, um sich die Umfrage erneut anzuschauen: https://t1p.de/fkytt
3. Formulieren Sie eine **Unterschiedshypothese** (H_1 und H_0), die Sie mittels t-Test prüfen können (Tipp: Sie benötigen eine metrische Variable und eine dichotome Variable)
4. Führen Sie einen Signifikanztest in JASP durch. Nutzen Sie den Datensatz *TMD_Data_BFI.jasp*.
5. Berichten Sie das Ergebnis mit allen notwendigen statistischen Kennwerten (siehe Formulierungshilfe nächste Folie)
6. Stellen Sie Ihre Hypothesen und Ergebnisse in einer Powerpoint-Präsentation zusammen.

Dichotomisierung von Variablen

- Mit dem t-Test lassen sich maximal zwei Gruppen/Ausprägungen/Zeitpunkte miteinander vergleichen.
- Wenn Gruppenvariablen mit mehr als zwei Ausprägungen vorliegen, müssen diese dichotomisiert werden (z.B. hat „Wohnsituation" in unserem Datensatz 3 Ausprägungen).
- Metrische Variablen können ebenfalls dichotomisiert werden (z.B. Gruppe 1: Alter > 20 Jahre, Gruppe 2: Alter < 20 Jahre).
- Für die Dichotomisierung muss in JASP eine neue Variable berechnet werden. Nutzen Sie dafür den „ifElse" Befehl:
  - ifElse(test,then,else).
  - Beispiel: ifElse(Alter<20,1,0): Wenn Alter < 20 ist, nimmt die neu berechnete Gruppenvariable den Wert 1 an, ansonsten nimmt sie den Wert 0 an.

Dichotomisierung von Variablen

Berechnung der neuen Variable „Wohsituation_dichotom“ mit 2 Ausprägungen (allein, zusammen)



Bestimmung des empirischen t-Werts

- Die empirische Prüfgröße t wird auf Grundlage des Mittelwertunterschieds berechnet, der am Standardfehler standardisiert wird.
- Je nach durchgeführtem Vergleich (Einzelmessung, zwei abhängige Gruppen, zwei unabhängige Gruppen) unterscheidet sich die Berechnung des Standardfehlers!
- Bei dem Vergleich zweier unabhängiger Gruppen muss zunächst geprüft werden, ob die Varianzen der untersuchten Variable in beiden Gruppen ähnlich oder unähnlich sind. Die Berechnung des Standardfehlers und der Freiheitsgrade unterscheidet sich in Abhängigkeit der Varianzhomogenität.
- Siehe Formelsammlung für detaillierte Darstellung.

$$t = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}} = \frac{\bar{x} - \mu}{\hat{se}}$$

Varianzen gelten als ähnlich, wenn:

$$0.5 < \frac{\hat{\sigma}_A}{\hat{\sigma}_B} < 2$$

Übung 1: t-Test in JASP

Personen, die verträglich sind, zeichnen sich durch Altruismus und Hilfsbereitschaft aus. Hohe Werte bei diesem Persönlichkeitsmerkmal sind charakterisiert durch Adjektive wie mitfühlend, nett, warm, vertrauensvoll, hilfsbereit, kooperativ und nachsichtig. Menschen mit niedrigen Werten werden als streitbar, egozentrisch, gegensätzlich und misstrauisch gegenüber den Absichten anderer beschrieben. Sie verhalten sich eher wettbewerbsorientiert als kooperativ.

Sie erwarten einen Unterschied in der Verträglichkeit zwischen Menschen, die ländlich leben und Menschen, die in der Stadt leben. Formulieren Sie eine gerichtete Unterschiedshypothese. Prüfen Sie in unserem Datensatz (TMD_Data_BFI.jasp), ob sie tatsächlich einen signifikanten Unterschied in die erwartete Richtung finden.

1. Prüfen Sie zunächst anhand der Daumenregel, ob die Varianzen in beiden Gruppen (städtisch vs. ländlich) ähnlich sind.
2. Rechnen Sie in JASP einen t-Test. Wählen Sie das korrekte Verfahren.
3. Lassen Sie sich eine passende Abbildung der Ergebnisse ausgeben.
4. Berichten Sie das Ergebnis in einem ausformulierten Satz und nennen Sie alle notwendigen Kennwerte inkl. der Effektstärke.

Formulierungshilfe für Ergebnisse (Beispiele)

- Studierende in Erfurt berichteten eine signifikant höhere Bildschirmzeit als Studierende in Potsdam, **t(df) = t-Wert, p-Wert, Effektstärke d**.
- Wir fanden einen signifikanten Unterschied im Sinnerleben zwischen Frauen und Männern, $t(231) = -3.16, p = 0.002, d = \ldots$.
- Übereinstimmend mit unserer Hypothese waren Psychologiestudierende signifikant gewissenhafter als Nicht-Studierende, $t(\mathrm{df}) = \ldots, p < .001, d = \ldots$.

Nicht signifikante Ergebnisse:

- Wir fanden keinen signifikanten Unterschied im Optimismus zwischen […]
- Für Korrelationen: Wir fanden keine Evidenz für einen Zusammenhang zwischen Optimismus und Gewissenhaftigkeit, $r = 0.02, p = 0.752$.

Lösung Übung 2

Hypothesen

- H_1: Auf dem Land lebende Personen sind verträglicher als in der Stadt lebende Personen (ländlich > städtisch).
- H_0: Auf dem Land lebende Personen sind gleich verträglich oder weniger verträglich als in der Stadt lebende Personen (ländlich ≤ städtisch).

Deskriptive Statistik

	Verträglichkeit	
	ländlich	städtisch
Gültig	9	8
Fehlend	0	0
Mittelwert	3.056	3.875
Standardabweichung	1.184	0.582
Minimum	1.500	3.000
Maximum	4.500	5.000

Varianzhomogenität:

- Varianzen sind unähnlich: $1.184/0.582 = 2.03 > 2$
- Es muss ein Welch‘s t-Test durchgeführt werden.

T-Test für unabhängige Stichproben

	Test	Statistik	df	p	Cohens d	Std.-Fehler Cohens d
Verträglichkeit	Student	−1.771	15.000	0.952	−0.861	0.527
	Welch	−1.840	11.936	0.955	−0.878	0.528

Hinweis. Bei allen Tests gibt die Alternativhypothese an, dass Gruppe *ländlich* größer ist als Gruppe *städtisch* .

Lösung Übung 2

T-Test für unabhängige Stichproben

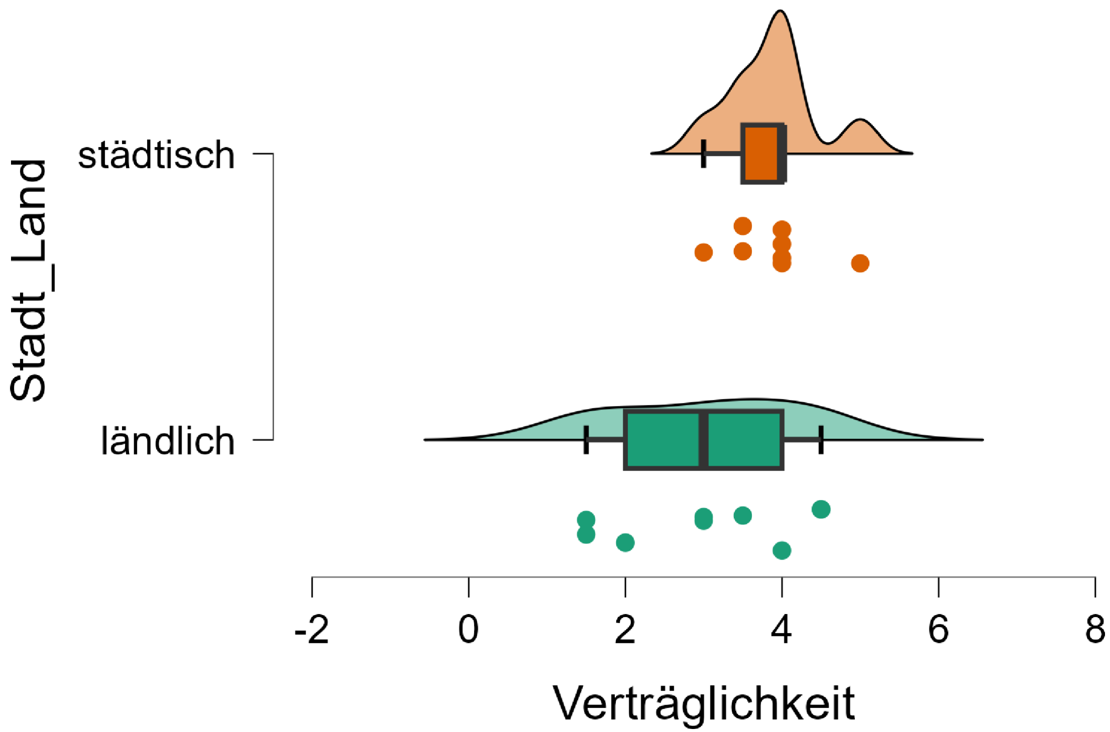
	Test	Statistik	df	p	Cohens d	Std.-Fehler Cohens d
Verträglichkeit	Student	−1.771	15.000	0.952	−0.861	0.527
	Welch	−1.840	11.936	0.955	−0.878	0.528

Hinweis. Bei allen Tests gibt die Alternativhypothese an, dass Gruppe *ländlich* größer ist als Gruppe *städtisch* .

Ergebnis

- Wir fanden keine Evidenz für unsere Hypothese, dass auf dem Land lebende Personen verträglicher sind als in der Stadt lebende Personen, $t(11.936) = -1.840, p = .955, d = -0.878$.
- Bei Testung in die andere Richtung findet sich Evidenz für eine größere Verträglichkeit bei in der Stadt lebenden Personen!

Lösung Übung 2



- Im Regenwolkenplot sind die ungleichen Varianzen der beiden Gruppen gut erkennbar.
- Entgegen der Hypothese zeigt sich, dass die durchschnittliche Verträglichkeit der in der Stadt lebenden Personen höher ist als die Verträglichkeit auf dem Land lebender Personen.

Prüfung auf Varianzhomogenität in JASP

Annahmeprüfungen

Prüfung auf Gleichheit der Varianzen (Levenes)

	F	df_1	df_2	p
Verträglichkeit	4.543	1	15	0.050

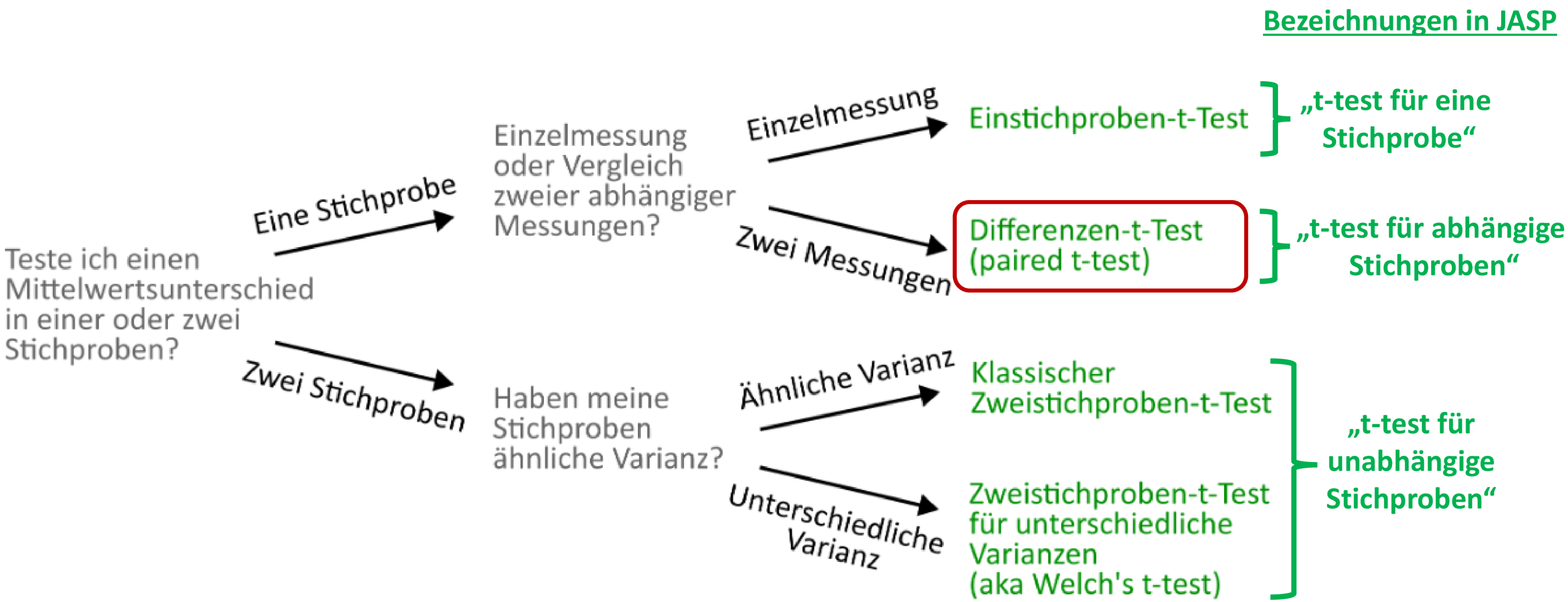
Annahmeprüfungen

Prüfung auf Gleichheit der Varianzen (Brown-Forsythe)

	F	df_1	df_2	p
Verträglichkeit	4.481	1	15	0.051

- Der Levene‘s Test und Brown-Forsythe-Test prüfen die Hypothese (H_1), dass die Varianzen in beiden Gruppen ungleich sind. Hier werden die Tests knapp nicht signifikant ($p \geq 0.05$).
- Die Verfahren stehen allerdings in Kritik, denn das Ziel der Nicht-Ablehnung der H_0 ist problematisch: **Die Nichtverwerfung von H_0 ist kein Beweis für die H_0.** Ein nicht-signifikanter Test zeigt lediglich an, dass es nicht genügend Evidenz gibt, um die H_0 zurückzuweisen. Das Fehlen von Evidenz ist nicht dasselbe wie Evidenz für das Fehlen eines Effekts.

Übung 2: Differenzen-t-Test (abhängige Stichproben)



Übung 2: Hausaufgabe

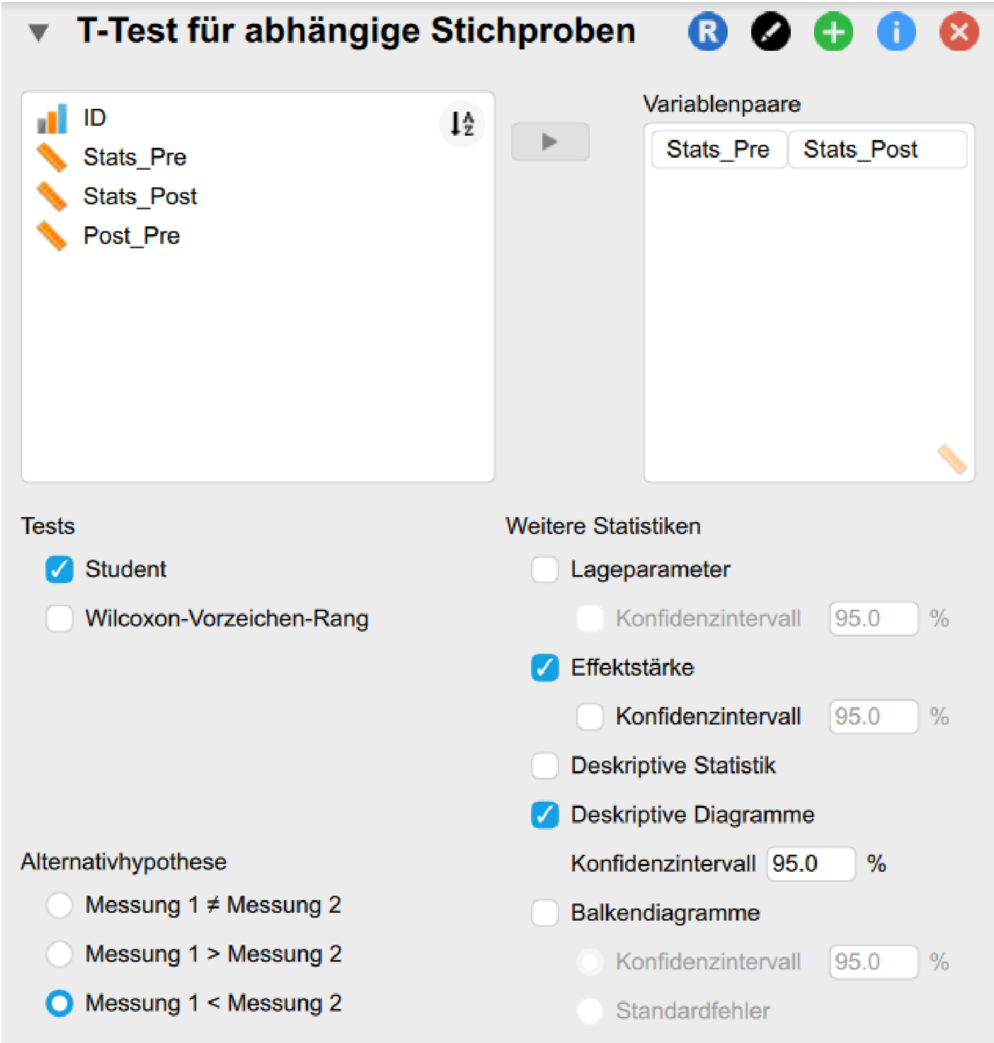
1. Laden Sie den Datensatz *Stats_Within.jasp* herunter.
2. Öffnen Sie die Datei in JASP.
3. Prüfen Sie, ob sich die Testat-Punktzahlen der 20 Personen nach Teilnahme am Statistik-Seminar signifikant verbessern. Führen Sie dazu einen t-Test für abhängige Stichproben durch.
4. Lassen Sie sich in JASP Abbildungen zur Veranschaulichung des Mittelwertunterschieds anzeigen.
5. Beschreiben Sie das Ergebnis in 1-2 Sätzen. Die Sätze sollten Ihre Hypothese, das Ergebnis des t-Tests und die relevanten statistischen Kennwerte enthalten.

Part 2

Übung 2: Hausaufgabe

1. Laden Sie den Datensatz *Stats_Within.jasp* herunter.
2. Öffnen Sie die Datei in JASP.
3. Prüfen Sie, ob sich die Testat-Punktzahlen der 20 Personen nach Teilnahme am Statistik-Seminar signifikant verbessern. Führen Sie dazu einen t-Test für abhängige Stichproben durch.
4. Lassen Sie sich in JASP Abbildungen zur Veranschaulichung des Mittelwertunterschieds anzeigen.
5. Beschreiben Sie das Ergebnis in 1-2 Sätzen. Die Sätze sollten Ihre Hypothese, das Ergebnis des t-Tests und die relevanten statistischen Kennwerte enthalten.

Lösung Übung 2



T-Test für abhängige Stichproben

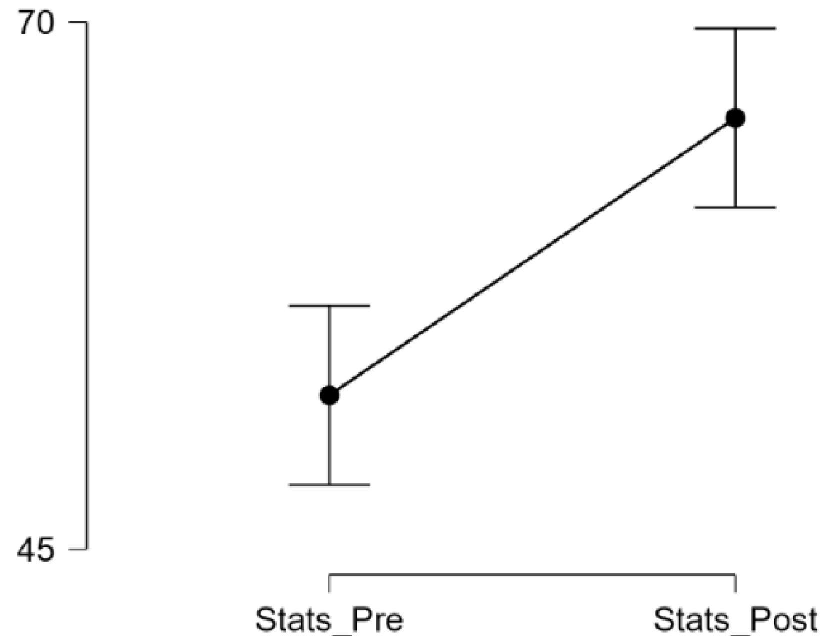
Messung 1		Messung 2	t	df	p	Cohens d	Std.-Fehler Cohens d
Stats_Pre	-	Stats_Post	−4.580	19	< .001	−1.024	0.159

Hinweis. Bei allen Tests gibt die Alternativhypothese an, dass Stats_Pre kleiner als Stats_Post ist.
Hinweis. Students T-Test.

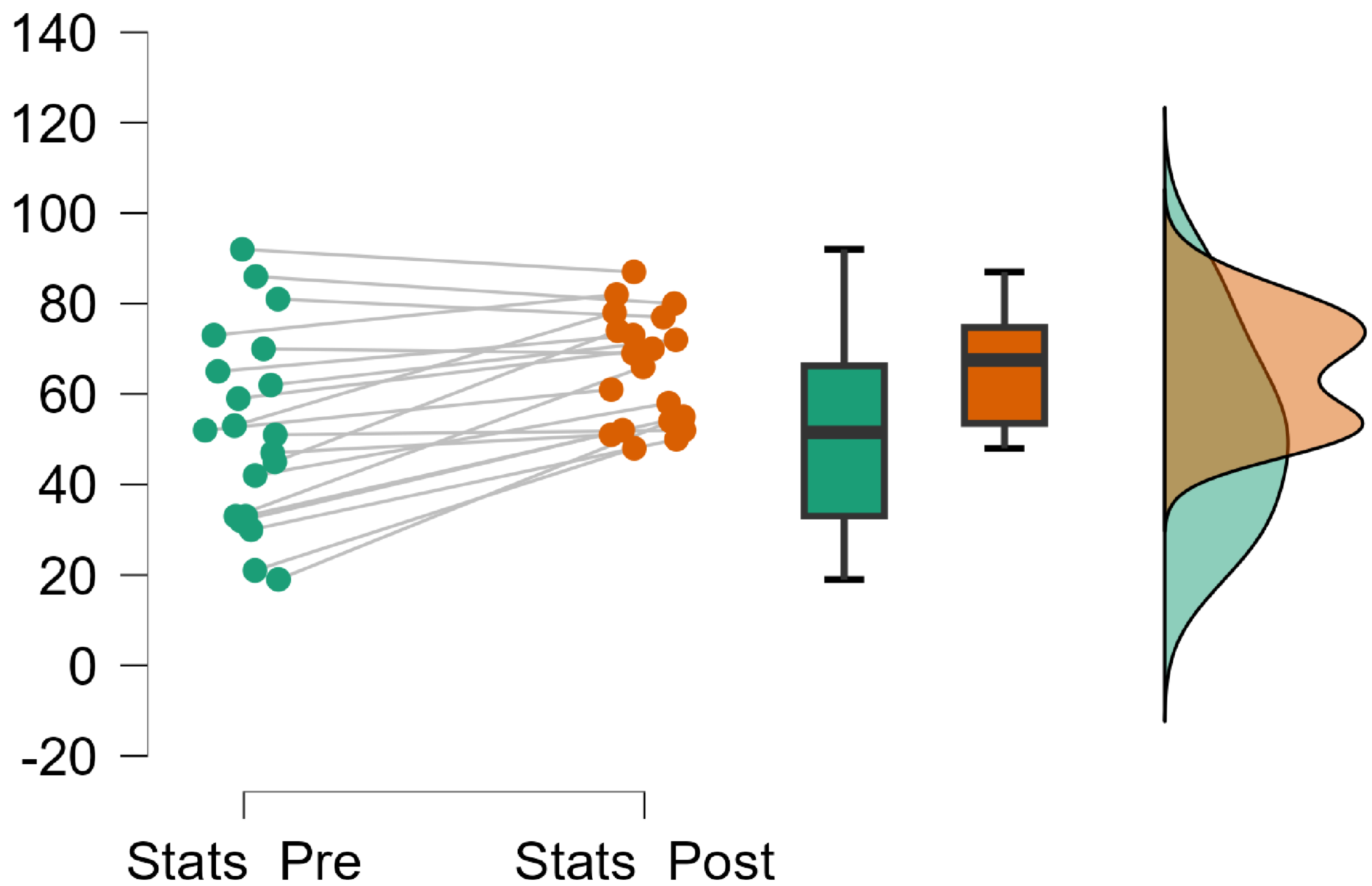
Deskriptive Statistiken

Deskriptive Diagramme

Stats_Pre - Stats_Post



Lösung Übung 2



t-Test für abhängige Stichproben

Bei abhängigen Stichproben gilt: Zur Berechnung des SEM und des **empirischen t-Werts** muss **zunächst die Standardabweichung $\hat{\sigma}_\Delta$ der Differenzwerte bestimmt werden.**

$$t = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}} = \frac{\bar{x} - \mu}{s\hat{e}}$$

Mittelwertdifferenz: abhängige Stichproben

$$\hat{se} = \frac{\hat{\sigma}_\Delta}{\sqrt{n}} \text{ mit } \hat{\sigma}_\Delta = \sqrt{\frac{1}{n-1} \sum \left(\Delta x_i - \Delta\bar{x}\right)^2}$$

Berechnung des t-Werts für abhängige Vergleiche

- Berechnen Sie den Standardfehler und den empirischen t-Wert für das vorangegangene Beispiel!
- Nutzen Sie die Standardabweichung $\hat{\sigma}_\Delta \approx 6.68$.

$$\hat{se} = SEM = \frac{6.68}{\sqrt{4}} = 3.34$$

$$t = \frac{5}{3.34} = 1.497$$

Standardfehler

Mittelwertdifferenz: abhängige Stichproben

$$\hat{se} = \frac{\hat{\sigma}_\Delta}{\sqrt{n}} \text{ mit } \hat{\sigma}_\Delta = \sqrt{\frac{1}{n-1}\sum(\Delta x_i - \Delta\bar{x})^2}$$

t-Test

Mittelwertdifferenz: abhängige Stichproben

$$t = \frac{\bar{x}_A - \bar{x}_B}{\hat{se}} \quad \text{mit} \quad \text{df}=n-1$$

t-Test für unabhängige Stichproben

- Wie zuvor muss zur Berechnung des empirischen t-Werts zunächst der **Standardfehler (SEM)** bestimmt werden.
- Der Berechnungsweg unterscheidet sich je nachdem, ob die beiden Gruppen ähnliche oder unähnliche Varianzen aufweisen.
- Bei **ähnlichen Varianzen** wird die **gepoolte Standardabweichung** zur Berechnung des SEM herangezogen.
- Bei unähnlichen Varianzen findet eine Korrektur der Freiheitsgrade df statt (Welch-Test).

Varianzen gelten als ähnlich, wenn:

$$0.5 < \frac{\hat{\sigma}_A}{\hat{\sigma}_B} < 2$$

Ähnliche Varianzen

Standardfehler

Mittelwertdifferenz: unabhängige Stichproben

Varianzen in A und B ähnlich $(0{,}5 < \frac{\hat{\sigma}_A}{\hat{\sigma}_B} < 2)$:

$$\boxed{\hat{se} = \hat{\sigma}_{\text{pooled}} \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \quad \text{mit}$$

t-Test

Mittelwertdifferenz: unabhängige Stichproben

$$t = \frac{\bar{x}_A - \bar{x}_B}{\hat{se}} \quad \text{mit} \quad \boxed{\begin{array}{c}\text{df} = n_A + n_B - 2\\ \text{(ähnliche Varianzen)}\end{array}}$$

Unähnliche Varianzen

Varianzen in A und B unähnlich:

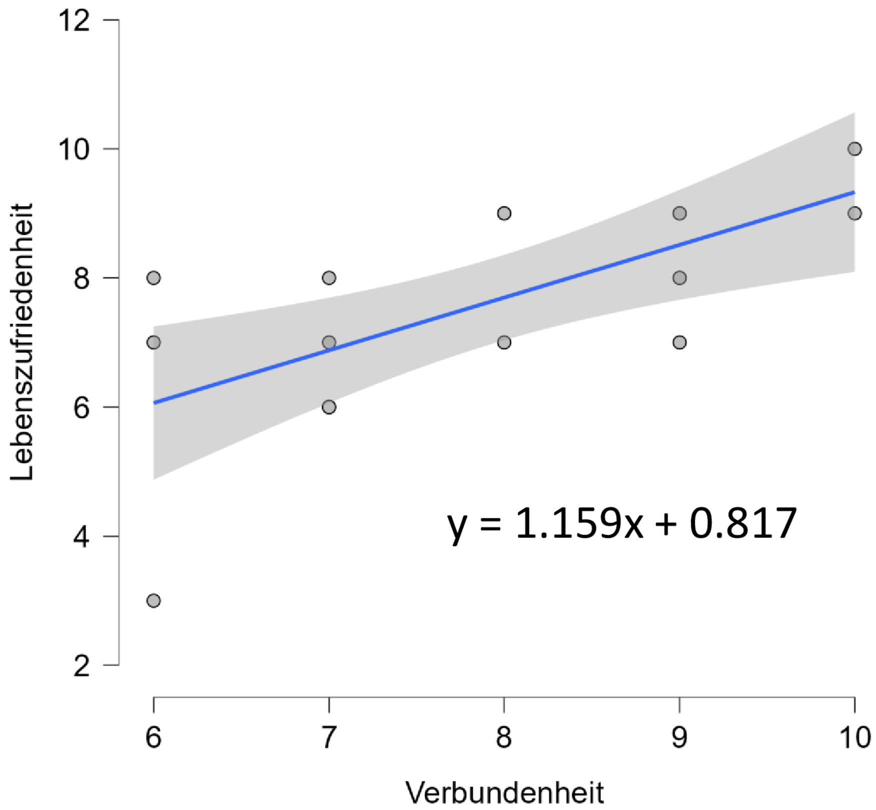
$$\hat{se} = \sqrt{\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_B}}$$

Standardfehler

$$\text{df} = \frac{\left(\hat{\sigma}_A^2/n_A + \hat{\sigma}_B^2/n_B\right)^2}{\frac{\left(\hat{\sigma}_A^2/n_A\right)^2}{n_A - 1} + \frac{\left(\hat{\sigma}_B^2/n_B\right)^2}{n_B - 1}}$$

df für Welch t-Test

Übung 3: t-Test für Regressionskoeffizienten



- Berechnen Sie einen t-Test um festzustellen, ob es einen **signifikant positiven** Zusammenhang zwischen den Variablen Verbundenheit und Lebenszufriedenheit in Ihrer Stichprobe von $\mathbf{n = 17}$ gibt. Der standardisierte Regressionskoeffizient ist $\beta = \mathbf{0.664}$.
- Bestimmen Sie zunächst r, den empirischen t-Wert und die Freiheitsgrade
- Nutzen Sie das JASP-Modul *Verteilungen* → *Skalierte verschobene Student's t* um den *p*-Wert zu bestimmen. Tragen Sie die dazu die Freiheitsgrade ein und lassen Sie sich die Wahrscheinlichkeit für das korrekte Intervall anzeigen.
- Berechnen Sie eine lineare Regression mit JASP. Stimmt das Ergebnis überein?

Signifikanztestung von Zusammenhängen

- Korrelationen und Regressionskoeffizienten werden ebenfalls mit dem t-Test auf Signifikanz geprüft (in JASP allerdings mit dem Modul Regression → Korrelation).

Teste ich einen Zusammenhang (Korrelation/Regression) in einer Stichprobe? → Rechne *r* oder b_1 in einen t-Wert um → Einstichproben-t-Test

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

$$\left(\text{Reminder: } r = \frac{\hat{\sigma}_X}{\hat{\sigma}_Y} b_1\right)$$

df = n – 2

Übung 3: t-Test für Regressionskoeffizienten

- $n = 17$
- Der standardisierte Regressionskoeffizient ist $\beta = 0.664.$

$$\beta = r \quad \text{(bei univariater Regression)}$$

$$\text{df} = n - 1 = 15$$
$$t_{\text{crit}} = 1.753$$

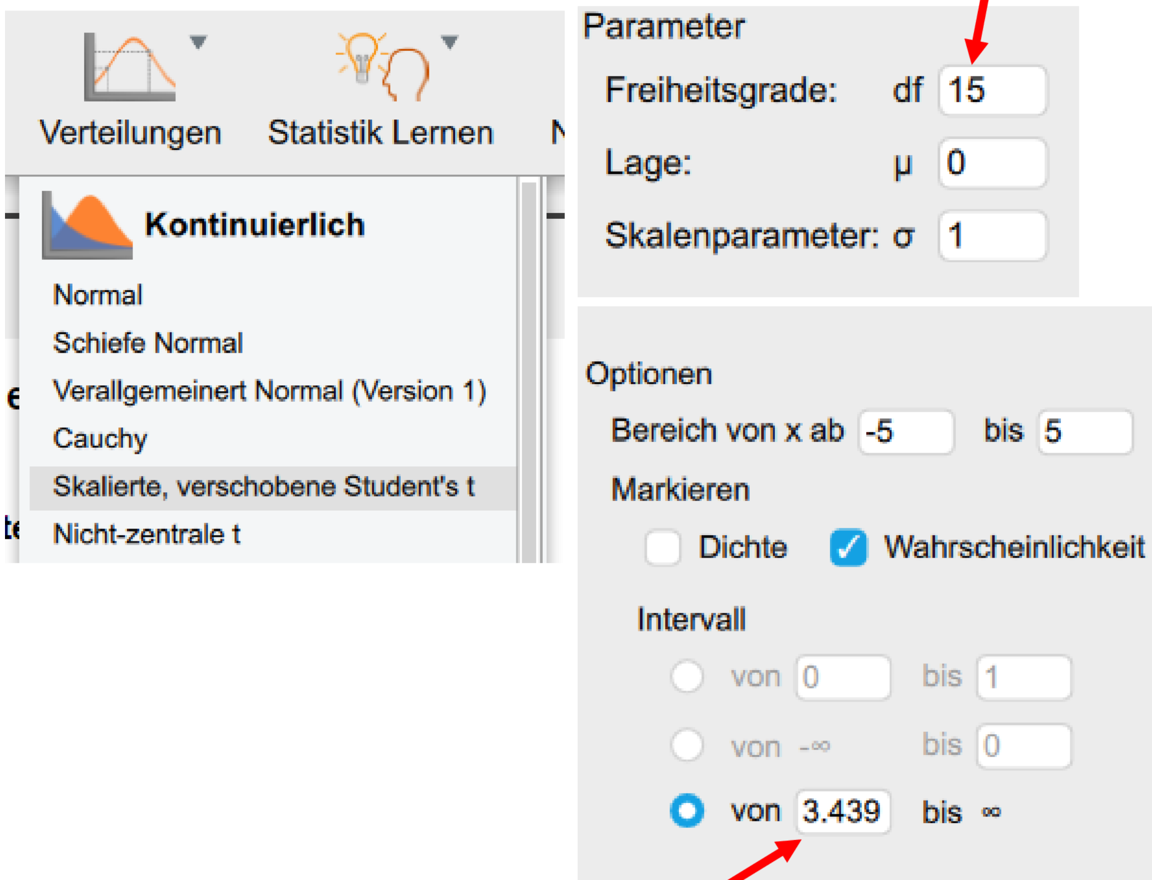
Rechne r oder b_1 in einen t-Wert um:

$$t = r\sqrt{\frac{n-2}{1-r^2}} = 3.439$$

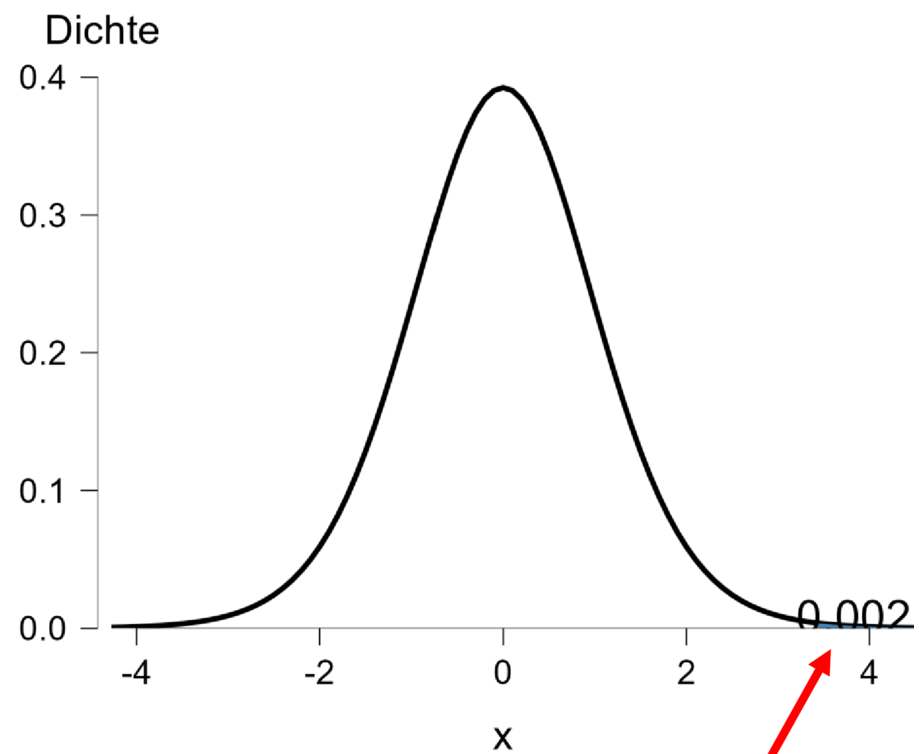
				Fläche
df	0,8	0,85	0,9	0,95
1	1,377	1,964	3,078	6,314
2	1,001	1,386	1,886	2,92
3	0,978	1,25	1,638	2,353
4	0,941	1,19	1,533	2,132
5	0,92	1,156	1,476	2,015
6	0,906	1,134	1,44	1,943
7	0,896	1,119	1,415	1,895
8	0,889	1,108	1,397	1,86
9	0,883	1,1	1,383	1,833
10	0,879	1,093	1,372	1,813
11	0,876	1,088	1,363	1,796
12	0,873	1,083	1,356	1,782
13	0,87	1,079	1,35	1,771
14	0,868	1,076	1,345	1,761
15	0,866	1,074	1,341	1,753

Übung 3: t-Test für Regressionskoeffizienten

Freiheitsgrade



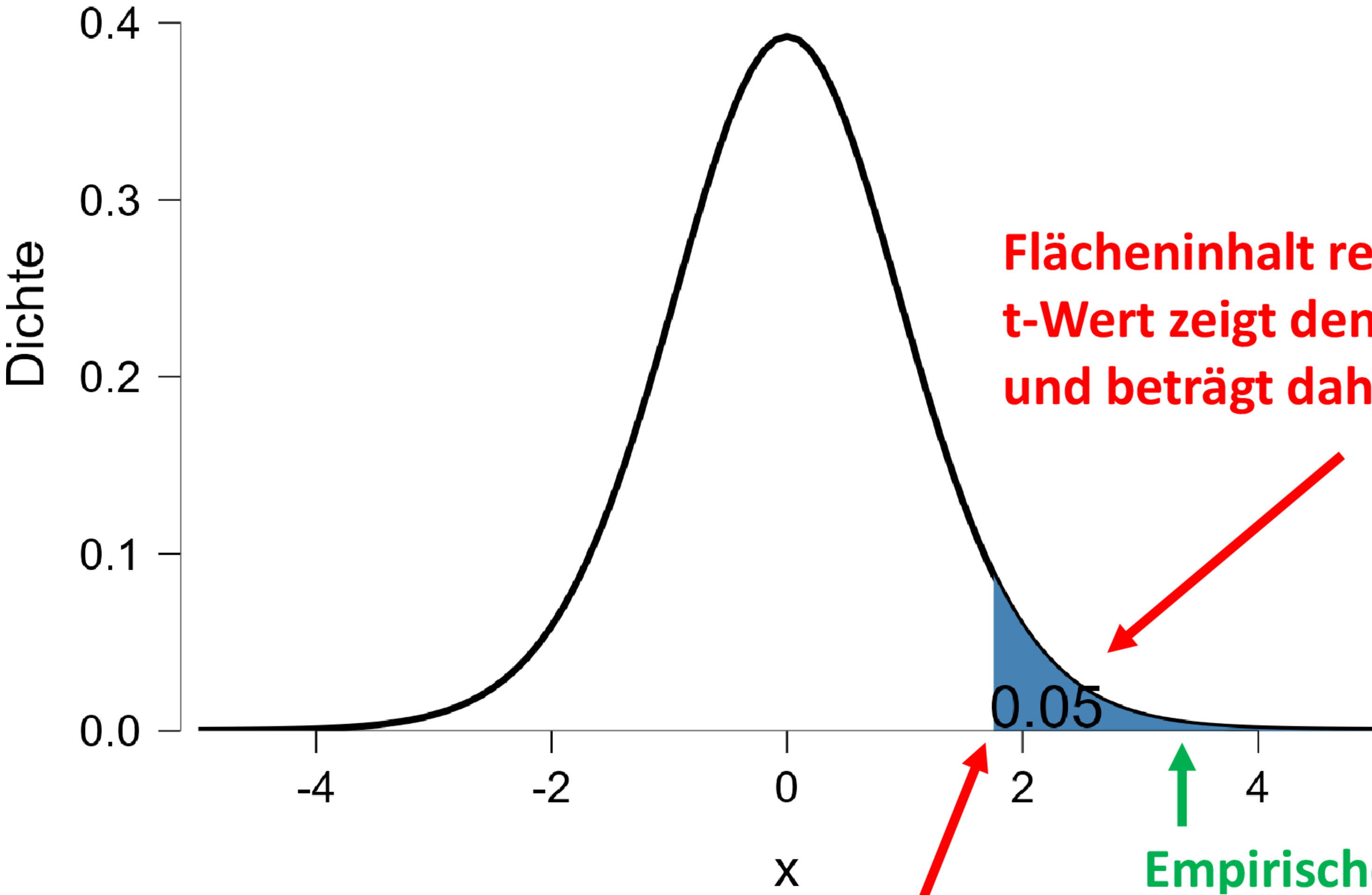
t-Verteilung unter H0 mit df = 15



empirischer t-Wert: t = 3.439

p-Wert (rechtsseitiger Test):
***p* = 0.002**

Übung 3: t-Test für Regressionskoeffizienten



Flächeninhalt rechtsseitig vom kritischen t-Wert zeigt den Ablehnungsbereich und beträgt daher $p = 0.05$

Lage des kritischen t-Werts für df = 15: $t_{crit} = 1.753$

Empirischer t-Wert von t = 3.439 liegt etwa hier (Fläche rechtsseitig davon ist $p = 0.002 < 0.05$)

Übung 3: Auszug aus der Regressionstabelle

Koeffizienten ▼

				Beta	empirischer t-Wert	p-Wert (zweiseitig)
Modell		Unstandardisiert	Standardfehler	Standardisiert	t	p
H_0	(Konstante)	7.647	0.402		19.039	< .001
H_1	(Konstante)	1.159	1.914		0.605	0.554
	Verbundenheit	0.817	0.238	0.664	3.435	0.004

- Regressionskoeffizienten werden in JASP immer zweiseitig getestet.
- Die Richtung des Zusammenhangs stimmt mit Ihrer gerichteten Hypothese überein (positiver Zusammenhang) → Der zweiseitige p-Wert von .004 kann halbiert werden, um den einseitigen p-Wert zu erhalten.

JASP Outputs verstehen

- Welche Hypothesen wurden geprüft?
- Was sind AV und UV?
- Welches Testverfahren wurde durchgeführt?
- Wie groß ist die Stichprobe?
- Was ist das Ergebnis? Ist es signifikant?

	t	df	p	Cohens d	Std.-Fehler Cohens d
Verträglichkeit	3.626	15	0.001	1.930	0.662

Hinweis. Bei allen Tests gibt die Alternativhypothese an, dass Gruppe *wohnen in Gemeinschaft* größer ist als Gruppe *wohnen allein* .

Deskriptive Statistiken

Gruppen, Deskriptiv

	Gruppe	N	Mittelwert	SD	Standardfehler	Variationskoeffizient
Verträglichkeit	wohnen in Gemeinschaft	12	3.875	0.644	0.186	0.166
	wohnen allein	5	2.400	1.025	0.458	0.427

JASP Outputs verstehen

- Welche Hypothesen wurden geprüft?
- Was sind AV und UV?
- Welches Testverfahren wurde durchgeführt?
- Wie groß ist die Stichprobe?
- Was ist das Ergebnis? Ist es signifikant?

Variable		Screentime	Sozialkontakt	Lebenszufriedenheit	Neurotizismus
1. Screentime	Pearsons r	—			
	p-Wert	—			
2. Sozialkontakt	Pearsons r	−0.150	—		
	p-Wert	0.566	—		
3. Lebenszufriedenheit	Pearsons r	−0.616	0.271	—	
	p-Wert	0.008	0.292	—	
4. Neurotizismus	Pearsons r	0.208	−0.711	−0.380	—
	p-Wert	0.423	0.001	0.132	—

JASP Outputs verstehen

- Welche Hypothesen wurden geprüft?
- Was sind AV und UV?
- Welches Testverfahren wurde durchgeführt?
- Wie groß ist die Stichprobe?
- Was ist das Ergebnis? Ist es signifikant?

	t	df	p	Mittelwertsdifferenz	Std.-Fehler Differenz
Lebenszufriedenheit	1.207	15	0.246	1.000	0.829
Einsamkeit	−2.388	15	0.031	−2.530	1.060
Extraversion	3.158	15	0.006	1.265	0.401
Verträglichkeit	1.081	15	0.297	0.553	0.512
Gewissenhaftigkeit	−0.218	15	0.830	−0.098	0.451
Neurotizismus	−2.413	15	0.029	−1.000	0.414

Gruppen, Deskriptiv ▼

	Gruppe	N	Mittelwert
Lebenszufriedenheit	optimstisch	11	8.000
	pessimistisch	6	7.000
Einsamkeit	optimstisch	11	2.636
	pessimistisch	6	5.167
Extraversion	optimstisch	11	3.682
	pessimistisch	6	2.417
Verträglichkeit	optimstisch	11	3.636
	pessimistisch	6	3.083
Gewissenhaftigkeit	optimstisch	11	3.318
	pessimistisch	6	3.417
Neurotizismus	optimstisch	11	3.000
	pessimistisch	6	4.000

JASP Outputs verstehen

- Welche Hypothesen wurden geprüft?
- Was sind AV und UV?
- Welches Testverfahren wurde durchgeführt?
- Wie groß ist die Stichprobe?
- Was ist das Ergebnis? Ist es signifikant?

Modell-Zusammenfassung - Sinn im Leben

Modell	R	R^2	Korrigiertes R^2	RMSE
H_0	0.000	0.000	0.000	1.139
H_1	0.537	0.289	0.241	0.992

ANOVA

Modell		Quadratsumme	df	Mittlere Quadratsumme	F	p
H_1	Regression	5.997	1	5.997	6.094	0.026
	Residuum	14.761	15	0.984		
	Gesamt	20.757	16			

Hinweis. Das konstante Modell wurde ausgelassen, da keine bedeutsame Information angezeigt werden kann.

Koeffizienten

Modell		Unstandardisiert	Standardfehler	Standardisiert	t	p
H_0	(Konstante)	5.397	0.276		19.537	< .001
H_1	(Konstante)	6.235	0.416		14.987	< .001
	Screentime	−0.195	0.079	−0.537	−2.469	0.026

Gerichtete Signifikanztestung

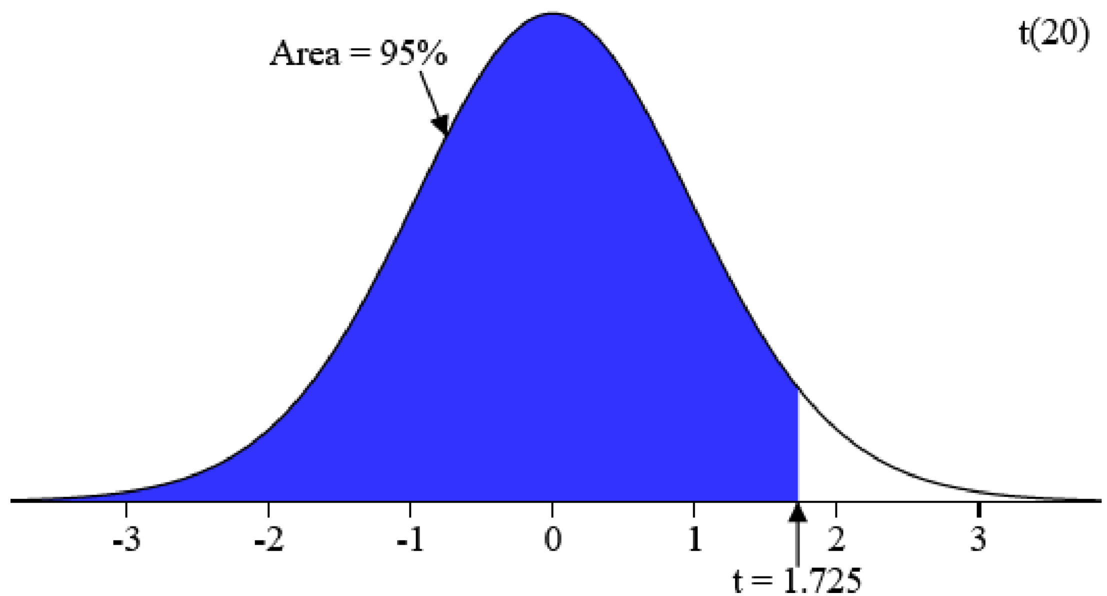
Haben Psychologiestudierende, die ein Achtsamkeitstraining absolviert haben (Gruppe 1), einen höheren durchschnittlichen Konzentrationswert als jene, die das Training nicht absolviert haben (Gruppe 2)?

Alternativhypothese (H_1): $\mu_1 > \mu_2$

	Fläche						
df	0,8	0,85	0,9	0,95	0,975	0,99	0,995
1	1,377	1,964	3,078	6,314	12,706	31,821	63,657
2	1,001	1,386	1,886	2,92	4,303	6,965	9,925

Wie lautet die Nullhypothese?

Nullhypothese (H_0): $\mu_1 \leq \mu_2$



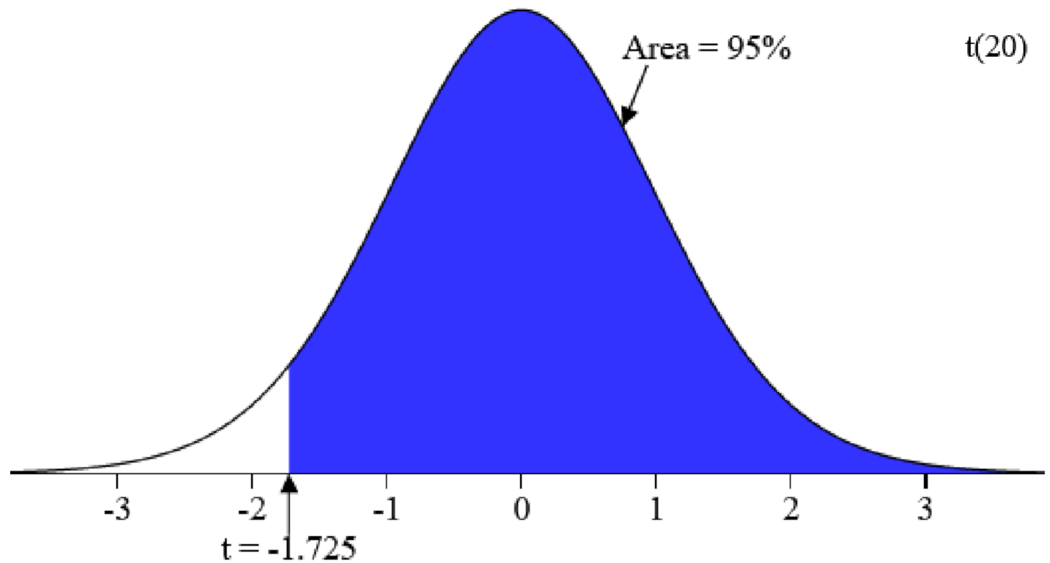
Gerichtete Signifikanztestung

*Haben Psychologiestudierende, die ein Achtsamkeitstraining absolviert haben (**Gruppe 2**), einen höheren durchschnittlichen Konzentrationswert als jene, die das Training nicht absolviert haben (**Gruppe 1**)?*

Nullhypothese (H_0): $\mu_1 \geq \mu_2$

Alternativhypothese (H_1): $\mu_1 < \mu_2$

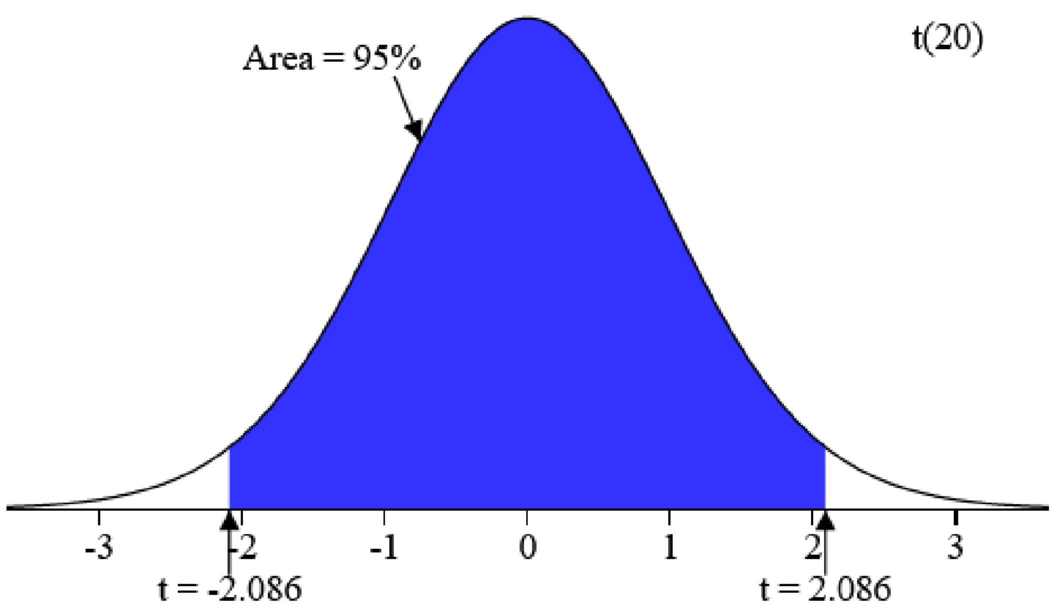
	Fläche						
df	0,8	0,85	0,9	$\boxed{0,95}$	0,975	0,99	0,995
1	1,377	1,964	3,078	6,314	12,706	31,821	63,657
2	1,001	1,386	1,886	2,92	4,303	6,965	9,925



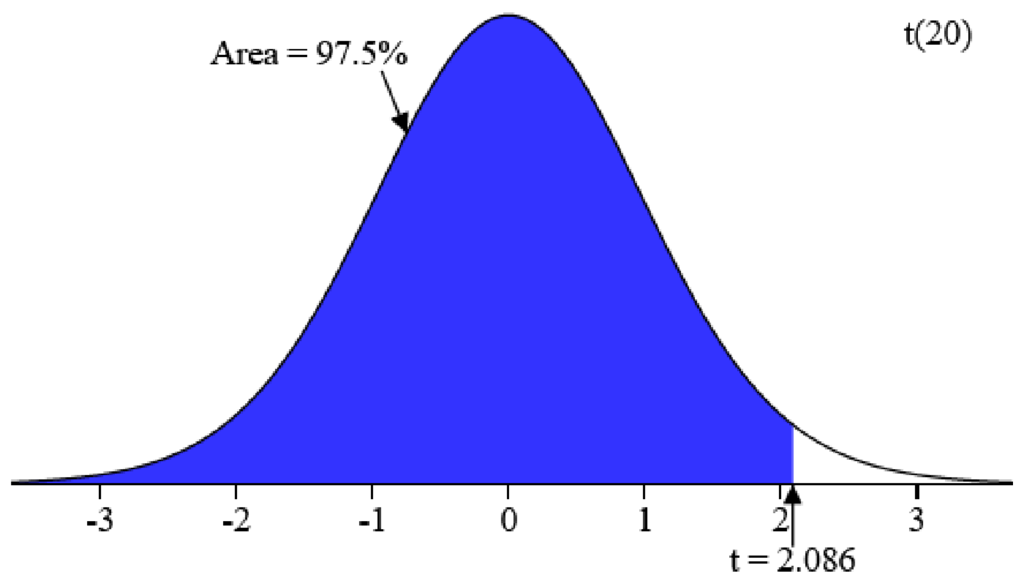
Ungerichtete Signifikanztestung

- Die t-Tabelle enthält kritische t-Werte für Flächeninhalte links von den gezeigten kritischen t-Wert.
- Bei zweiseitiger Testung müssen daher t-Werte in der Spalte 0.975 abgelesen werden. Anschließend wird der abgelesene kritische t-Wert mit dem Betrag des empirischen t-Werts verglichen.

Was Sie testen



Was die t-Tabelle zeigt



	Fläche						
df	0,8	0,85	0,9	0,95	$\boxed{0{,}975}$	0,99	0,995
1	1,377	1,964	3,078	6,314	12,706	31,821	63,657
2	1,001	1,386	1,886	2,92	4,303	6,965	9,925

Aufgabe 1: t-Test (manuell)

Fünf Personen nahmen an einer Medikamentenstudie teil. Es wurden zu zwei Zeitpunkten Reaktionszeiten (in ms) in einer Aufmerksamkeitsaufgabe gemessen. T1 enthält Reaktionszeiten vor der Medikamenteneinnahme, T2 nach der Medikamenteneinnahme. Lösen Sie die folgenden Aufgaben. Runden Sie auf zwei Nachkommastellen (beim t-Wert auf drei Stellen).

1. Bestimmen Sie die Differenzen in der Reaktionszeit zwischen den beiden Messzeitpunkten.
2. Bestimmen Sie die durchschnittliche Reaktionszeit zu T1 und T2, und die durchschnittliche Reaktionszeitdifferenz.
3. Bestimmen Sie die Streuung der Reaktionszeiten und der Reaktionszeitdifferenzen (Schätzung der Standardabweichung)
4. Prüfen Sie, ob die Varianz der Reaktionszeiten zu beiden Messzeitpunkten ähnlich ist.
5. Bestimmen Sie den Standardfehler der Mittelwertdifferenz.
6. Berechnen Sie den t-Wert.
7. Prüfen Sie mittels t-Tabelle, ob sich die Reaktionszeiten zwischen den beiden Messzeitpunkten signifikant unterscheiden (zweiseitiger Test).
8. Berechnen Sie die Effektstärke *d*.

Datei auf Teams: Tabelle_Übungen_S24.xlsx

Person	T1	T2	Differenz
1	530	620	
2	90	250	
3	230	440	
4	180	230	
5	320	760	
Mittelwert			
Standardabweichung			
Varianzen gleich?			
Standardfehler			
t-Wert			
kritischer t-Wert			
Signifikant? (ja/nein)			
Cohen's d			

Aufgabe 2: t-Test (manuell)

Die vorherige Studie wird wiederholt. Diesmal wird eine Experimentalgruppe (EG), welche das Medikament erhielt mit einer Kontrollgruppe (KG), welche ein Placebo erhielt, verglichen. Lösen Sie die folgenden Aufgaben. Runden Sie auf zwei Dezimalstellen (beim t-Test: 3 Dezimalstellen). Testen Sie die Hypothese, dass die Einnahme des Medikaments zu einer verlangsamten Reaktion führt.

1. Berechnen Sie die Mittelwerte und Standardabweichungen der Reaktionszeiten in beiden Gruppen.
2. Prüfen Sie, ob die Varianzen in beiden Gruppen ähnlich sind.
3. Bestimmen Sie die gepoolte Standardabweichung und den Standardfehler.
4. Berechnen Sie den t-Wert für einen Gruppenvergleich.
5. Bestimmen Sie die Freiheitsgrade und lesen Sie den kritischen t-Wert aus der Tabelle in der Formelsammlung ab.
6. Prüfen Sie auf Signifikanz.
7. Berechnen Sie die Effektstärke *d*.

Datei auf Teams: Tabelle_Übungen_S24.xlsx

Messung	EG	KG
1	560	130
2	990	420
3	890	110
4	1200	90
5	780	510
Mittelwert		
Standardabweichung		
Varianzen gleich?		
gepoolte Standardabweichung		
Standardfehler		
t-Wert		
kritischer t-Wert		
Signifikant? (ja/nein)		
Cohen's d		

Aufgabe 3: Kovarianz, Korrelation, Regression

Bei 6 Personen wurde Depression (auf einer Skala von 0 bis 6) und Achtsamkeit (auf einer Skala von 1 bis 7) gemessen. Nutzen Sie die Formelsammlung zur Bearbeitung der folgenden Aufgaben. Runden Sie auf zwei Nachkommastellen.

1. Berechnen Sie die mittlere Ausprägung von Depression und Achtsamkeit über alle Personen hinweg.
2. Berechnen Sie die Populationsschätzer der Varianz und Standardabweichung für beide Variablen
3. Berechnen Sie für jede Person den z-Wert auf den beiden Variablen.
4. Berechnen Sie die Kovarianz und die Korrelation.
5. Bestimmen Sie den t-Wert (3 Nachkommastellen) für die Korrelation.
6. Vergleichen Sie den empirischen t-Wert mit dem kritischen t-Wert (s. t-Tabelle) und entscheiden Sie, ob die Korrelation signifikant ist (zweiseitiger Test).
7. Sie wollen die Ausprägung von Depression mit der Achtsamkeit vorhersagen. Bestimmen Sie X und Y. Berechnen Sie den Achsenabschnitt und die Steigung und stellen Sie die Regressionsgleichung auf.
8. Berechnen Sie den standardisierten Steigungskoeffizienten Beta und vergleichen Sie ihn mit dem Korrelationskoeffizienten.

Datei auf Teams: Tabelle_Übungen_S24.xlsx

Person	Depression	Achtsamkeit	Depression z-Wert	Achtsamkeit z-Wert
1	3	5		
2	2	6		
3	4	4		
4	1	7		
5	5	3		
6	0	5		
Mittelwert				
Varianz				
Standardabweichung				
Kovarianz				
Korrelation				
t-Wert der Korrelation				
kritischer t-Wert				
Signifikant? (ja/nein)				
Achsenabschnitt				
Steigung				
Beta				
Regressionsgleichung				