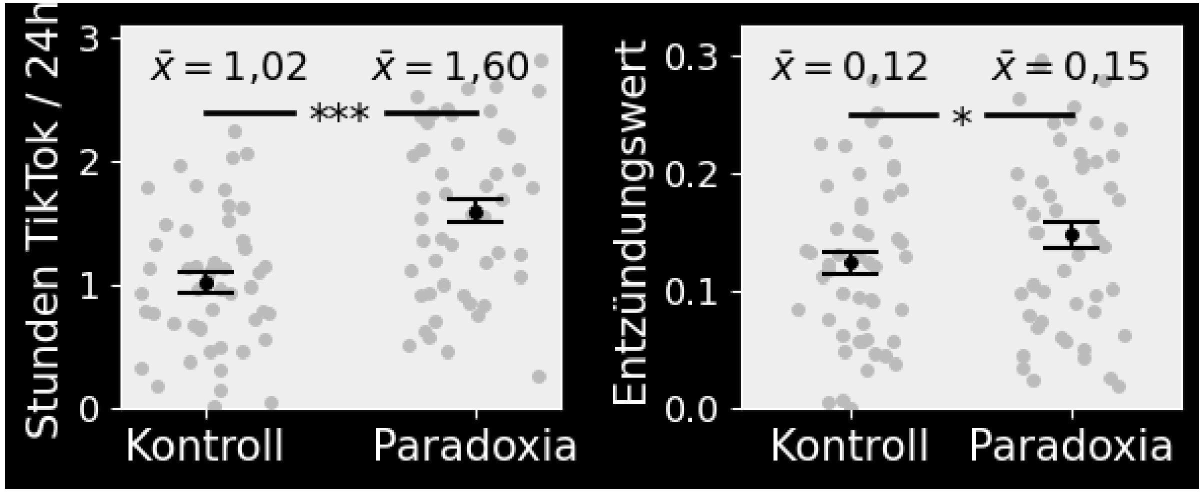
M24 Statistik 1: Wintersemester 23/24

Vorlesung 11: t-Test

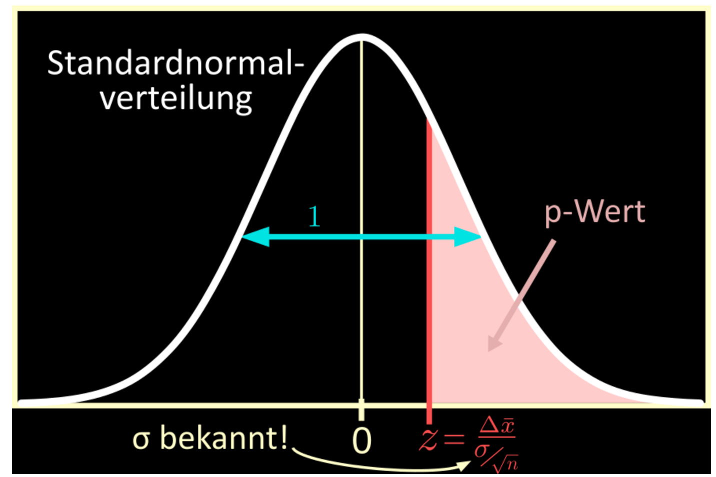
Prof. Matthias Guggenmos

Health and Medical University Potsdam

Erinnerung: den z-Test haben Sie unter der Voraussetzung durchgeführt, dass die Streuung σ in der Population bekannt ist (bzw. die Streuung*en* bei zwei Populationen).



Wie schon angesprochen, kommt dieser Fall in der Praxis allerdings nahezu nie vor und gilt auch für den vorliegenden Fall nicht. What shall you do?



Der Forschungsprozess

Auf Basis einer neuen/aktualisierten Theorie ergeben sich häufig neue Fragen und Hypothesen

Wundern, Raten, Fragen
(best case: bereits eine Theorie)

↓

Hypothese

↓

Operationalisierung

↓

Experiment / Studie

↓

Daten

↓

Auswertung / Ergebnis

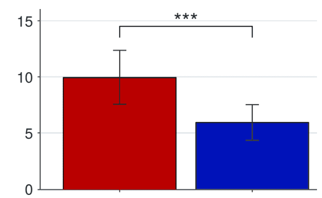
↓

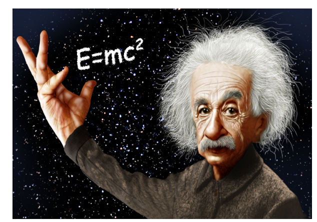
Schlussfolgerung / Interpretation

↓

Aufstellen oder Update einer Theorie

Player	Minutes	Points	Rebounds
A	41	20	6
B	30	29	7
C	22	7	7
D	26	3	3
E	20	19	8





Problemstellung

Ist die Standardabweichung σ einer Population bekannt, so können wir für die Prüfgröße $\frac{\Delta\bar{x}}{\sigma/\sqrt{n}}$ eine Normalverteilung annehmen. Der Grund ist, dass $\Delta\bar{x}$ auf Basis des zentralen Grenzwertsatzes als normalverteilt angenommen werden kann. Ist nun die Standardabweichung σ eine bekannte feste Größe, wird bei der Berechnung der Prüfgröße de facto eine normalverteilte Variable ($\Delta\bar{x}$) durch eine Konstante ($\sigma/\sqrt{n}$) geteilt, so dass auch die Prüfgröße normalverteilt ist. Aus diesem Grund können wir beim z-Test eine **Normalverteilung für die Prüfgröße z annehmen**.

Ist σ dagegen *nicht bekannt*, muss die Streuung auf Basis der Stichprobe als $\hat{\sigma}$ geschätzt werden. Die Schätzung von $\hat{\sigma}$ ist mit Unsicherheit behaftet und daher selbst eine **Zufallsvariable**. Wir teilen also eine normalverteilte Zufallsvariable $\Delta\bar{x}$ durch eine wie-auch-immer-verteilte (*) zweite Zufallsvariable $\hat{\sigma}$. **In diesem Fall können wir nicht mehr davon ausgehen, dass die Prüfgröße normalverteilt ist!**

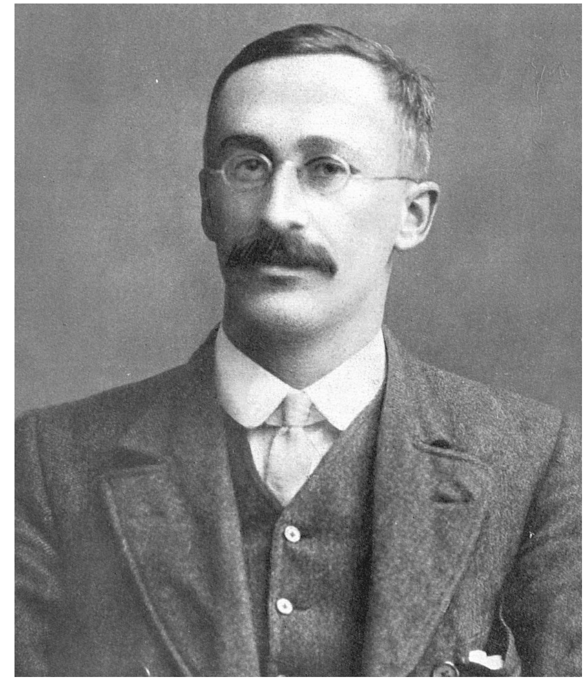
(* die wie-auch-immer-Verteilung ist bekannt: $\hat{\sigma}$ folgt der Chi-Verteilung bzw. χ-Verteilung)

Die Gretchen-Frage ist daher: welcher Verteilung folgt die Prüfstatistik

$$\frac{\Delta\bar{x}}{\hat{\sigma}/\sqrt{n}} \qquad ?$$

(mit Betonung auf dem ^ über $\hat{\sigma}$)

Die t-Verteilung

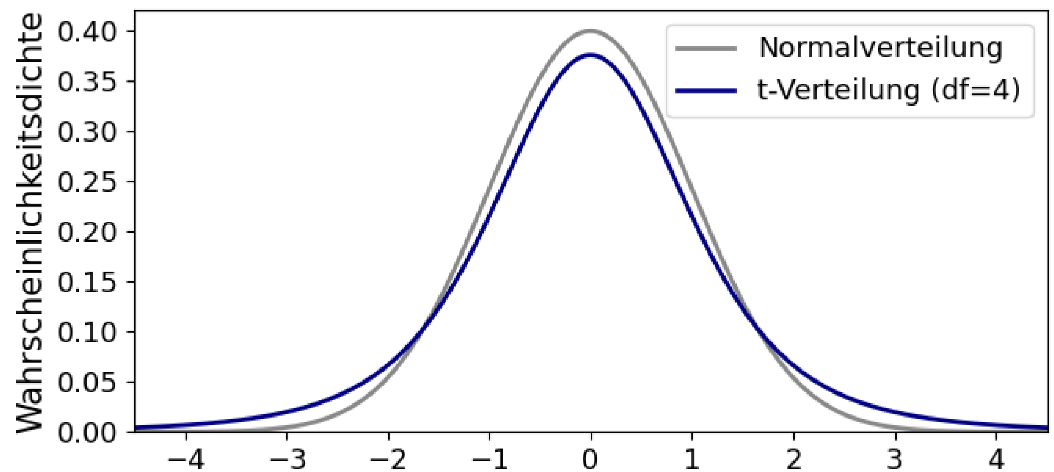


William Sealy Gosset (1876-1937)

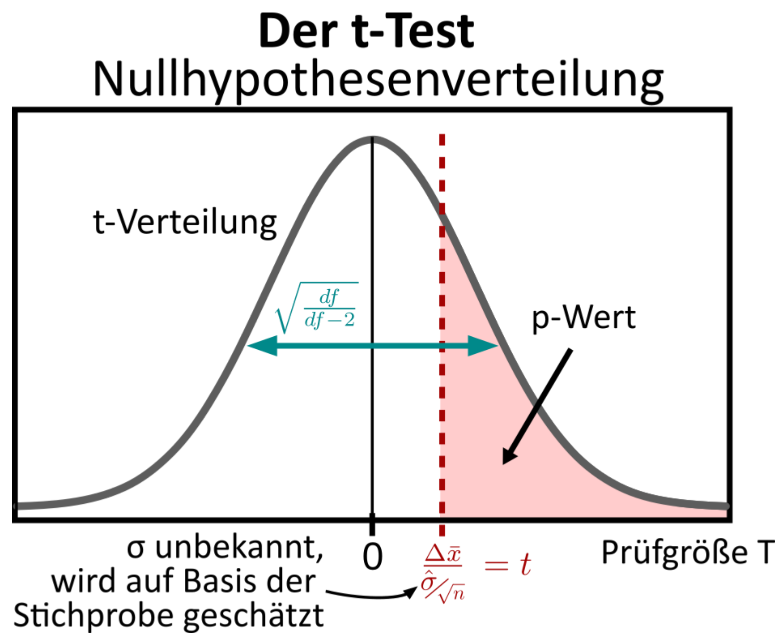
- An dieser Stelle kommt der englische Statistiker **William Sealy Gosset** in Spiel.
- Er wies 1908 nach, dass das Verhältnis einer normalverteilten Variable (z.B. $\Delta\bar{x}$) und einer Chi-verteilten Variable einer wohl-definierten Verteilung folgt, die als **t-Verteilung** bezeichnet wird.
- Die Formel der Verteilung ist vergleichsweise kompliziert, für die Praxis entscheidend ist jedoch, dass sie **durch einen einzigen Parameter definiert wird**: die Zahl der Freiheitsgrade (df):

$$t\text{-Verteilung:} \qquad f_t(x|\mathrm{df})$$

- Die **Zahl der Freiheitsgrade** hängt eng mit der Stichprobengröße n zusammen.
- Im Bild rechts ist eine t-Verteilung mit 4 Freiheitsgraden im Vergleich zur Normalverteilung aufgetragen.
- Erste Erkenntnis: die t-Verteilung hat etwas dickere Flanken (*fat tails*)!



t-Wert



- Wir kennen nun also die Form der Nullverteilung, wenn die Streuung als $\hat{\sigma}$ auf Basis der Stichprobe geschätzt werden muss: **t-Verteilung**
- In Korrespondenz mit dem Namen der Verteilung wird die Prüfgröße auf Basis von $\hat{\sigma}$ als **t-Wert** bezeichnet:

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

- Das Prinzip der p-Wert-Bestimmung ist exakt wie beim z-Test: es wird die Fläche unter der t-Verteilung relativ zum Prüfgrößenwert t bestimmt:
  - Gerichtete Hypothese $\Delta\bar{x} > 0$: Fläche rechts von t
  - Gerichtete Hypothese $\Delta\bar{x} < 0$: Fläche links von t
  - Ungerichtete Hypothese $\Delta\bar{x} \neq 0$:
    - Falls t < 0: Fläche links von t **PLUS** Fläche rechts von $-t$
    - Falls t > 0: Fläche links von $-t$ **PLUS** Fläche rechts von t
    - Oder allgemein: Fläche links von $-|t|$ **PLUS** Fläche rechts von $|t|$
- Merke: der *t-Wert* ist zur *t-Verteilung* wie der *z-Wert* zur *Standardnormalverteilung*!

Standardfehler bei Mittelwertdifferenzen: Cheat sheet

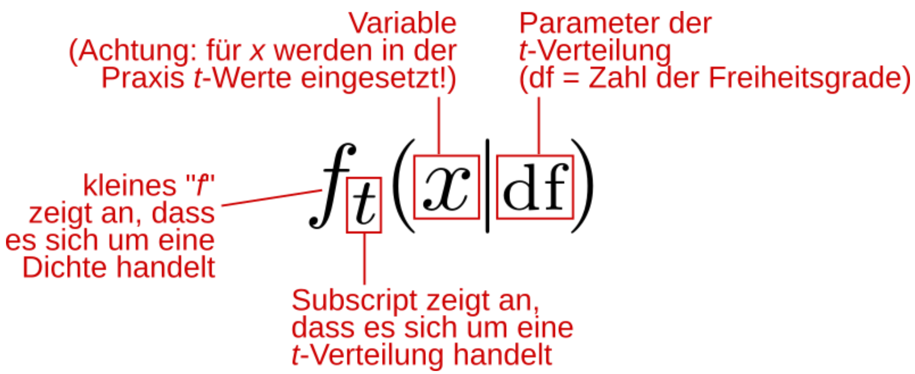
Das folgende Cheat sheet gibt Ihnen eine Übersicht über die Berechnung des Standardfehlers $\hat{se}$ im Nenner der Prüfgröße $t = \frac{\Delta\bar{x}}{\hat{se}}$:

Fall	Berechnung des Standardfehlers $\hat{se}$
Differenz des Mittelwertes <u>einer Gruppe</u> und einem Referenzwert μ_0 (→ Einstichproben-t-Test)	$\hat{se} = \frac{\hat{\sigma}}{\sqrt{n}}$
Mittelwertdifferenz zweier Bedingungen A und B in <u>einer Gruppe</u> (→ Differenzen-t-Test)	$\hat{se} = \frac{\hat{\sigma}_\Delta}{\sqrt{n}}$ mit $\hat{\sigma}_\Delta = \sqrt{\hat{\sigma}_A^2 + \hat{\sigma}_B^2 - 2\,\hat{Cov}(X_A, X_B)}$
Mittelwertdifferenz <u>zweier unabhängiger Gruppen</u> A und B (ähnliche Varianzen) (→ Klassischer Zweistichproben-t-Test)	$\hat{se} = \hat{\sigma}_{\text{pooled}}\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$ mit $\hat{\sigma}_{\text{pooled}} = \sqrt{\frac{(n_A-1)\hat{\sigma}_A^2+(n_B-1)\hat{\sigma}_B^2}{n_A+n_B-2}}$
Mittelwertdifferenz <u>zweier unabhängiger Gruppen</u> A und B (unterschiedliche Varianzen) (→ Welch's t-Test)	$\hat{se} = \sqrt{\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_B}}$

Dieses Cheat Sheet gilt für den Fall, dass die Populationsvarianzen σ_A^2 bzw. σ_B^2 **nicht bekannt** sind. In diesem Sinne ist es das in der Praxis relevante Cheat Sheet, da die Populationsvarianzen nahezu nie bekannt sind.

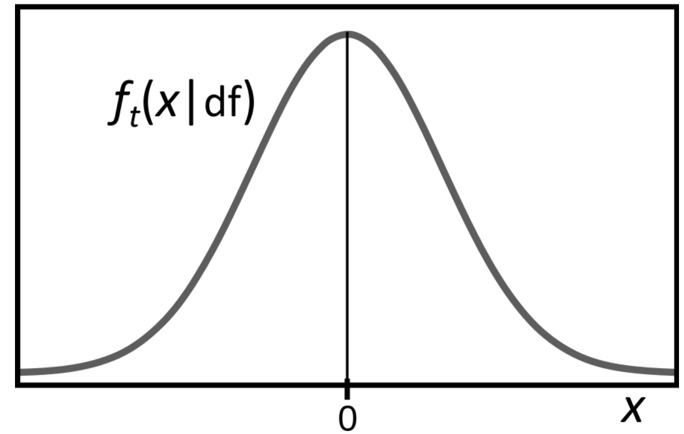
Funktionen der t-Verteilung

Wie die Normalverteilung wird auch die t-Verteilung durch eine **Wahrscheinlichkeitsdichte** definiert:



t-Verteilung

Wahrscheinlichkeitsdichte

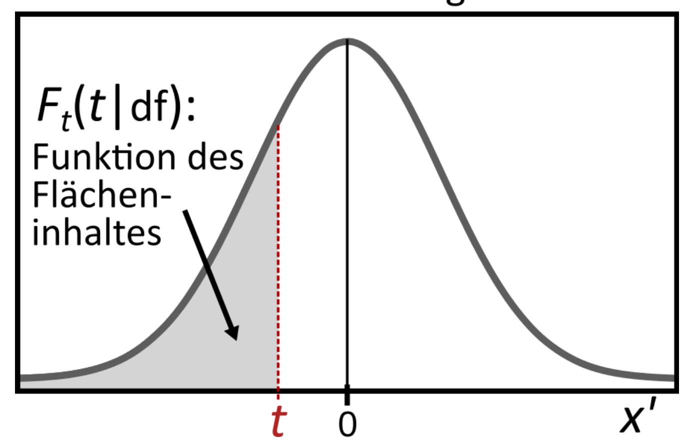


Die zugehörige **kumulative Verteilungsfunktion** wird der Konvention entsprechend mit einem großen F denotiert:

Kumulative Verteilungsfunktion:

$$F_t(x|\mathrm{df}) = \int_{-\infty}^{x} f_t(x'|\mathrm{df})\,dx'$$

Kumulative Verteilungsfunktion



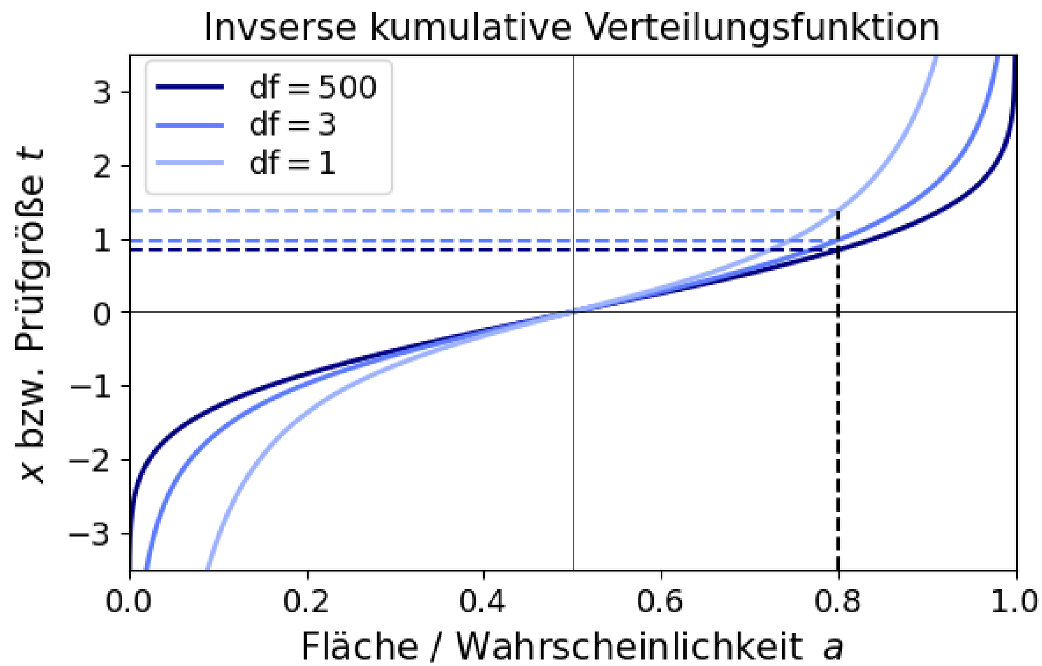
Die kumulative Verteilungsfunktion ordnet jedem t-Wert den Flächeinhalt unter der t-Verteilung im Bereich $[-\infty; t]$ zu.

Funktionen der t-Verteilung

Zuweilen stellt sich auch die umgekehrte Frage: wie lautet der t-Wert für einen bestimmten Flächeninhalt (z.B. einen bestimmten p-Wert)?

Diesen Zusammenhang stellt die **inverse kumulative Verteilungsfunktion** her:

$$t = F_t^{-1}(a|\mathrm{df})$$



wobei a die Fläche (area) bezeichnet. Da die t-Verteilung eine Wahrscheinlichkeitsdichte ist, kann die Fläche a nur Werte zwischen 0 und 1 annehmen.

Die Funktion wird auch **Quantilfunktion** genannt, da sie z.B. für $a = 0{,}8$ den Wert t zurück gibt, der 80% der t-Verteilung (von $-\infty$ gerechnet) umfasst.

Nota Bene

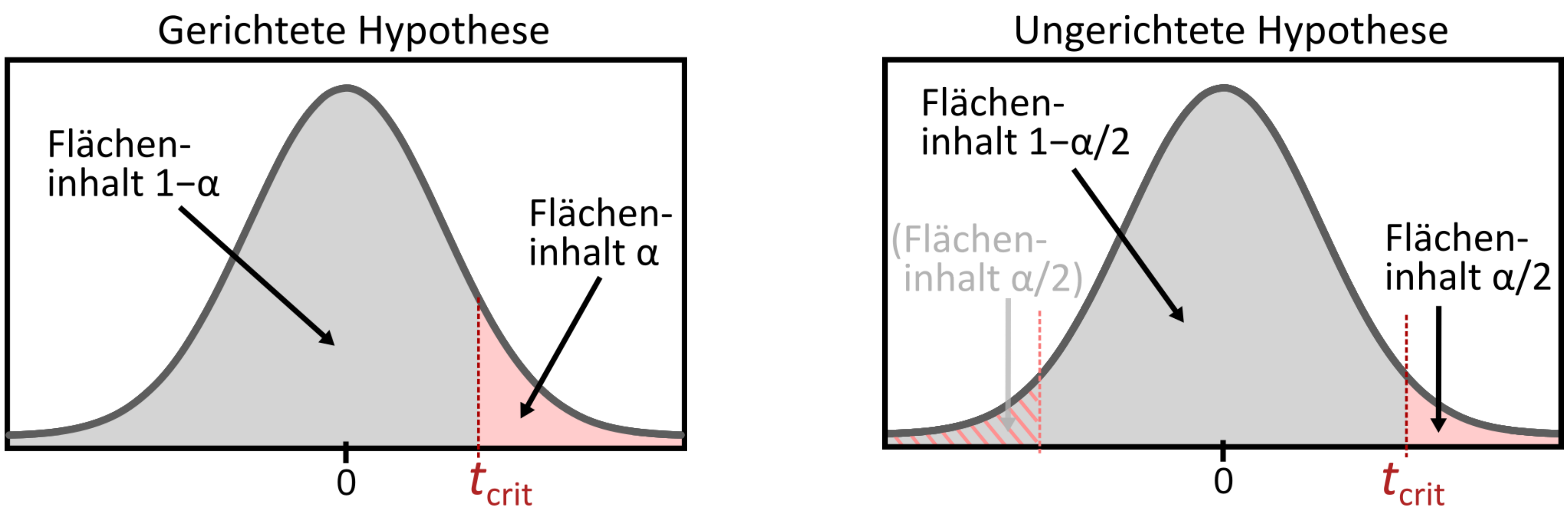
Sämtliche Funktionen (Wahrscheinlichkeitsdichte, kumulative Verteilungsfunktion, inverse kumulative Verteilungsfunktion) sind zu kompliziert zur manuellen Berechnung und werden mit dem Computer ausgewertet. Zusätzlich bietet die t-Tabelle die Möglichkeit, kritische t-Werte für ausgewählte Signifikanzniveaus nachzuschlagen.

Funktionen der t-Verteilung

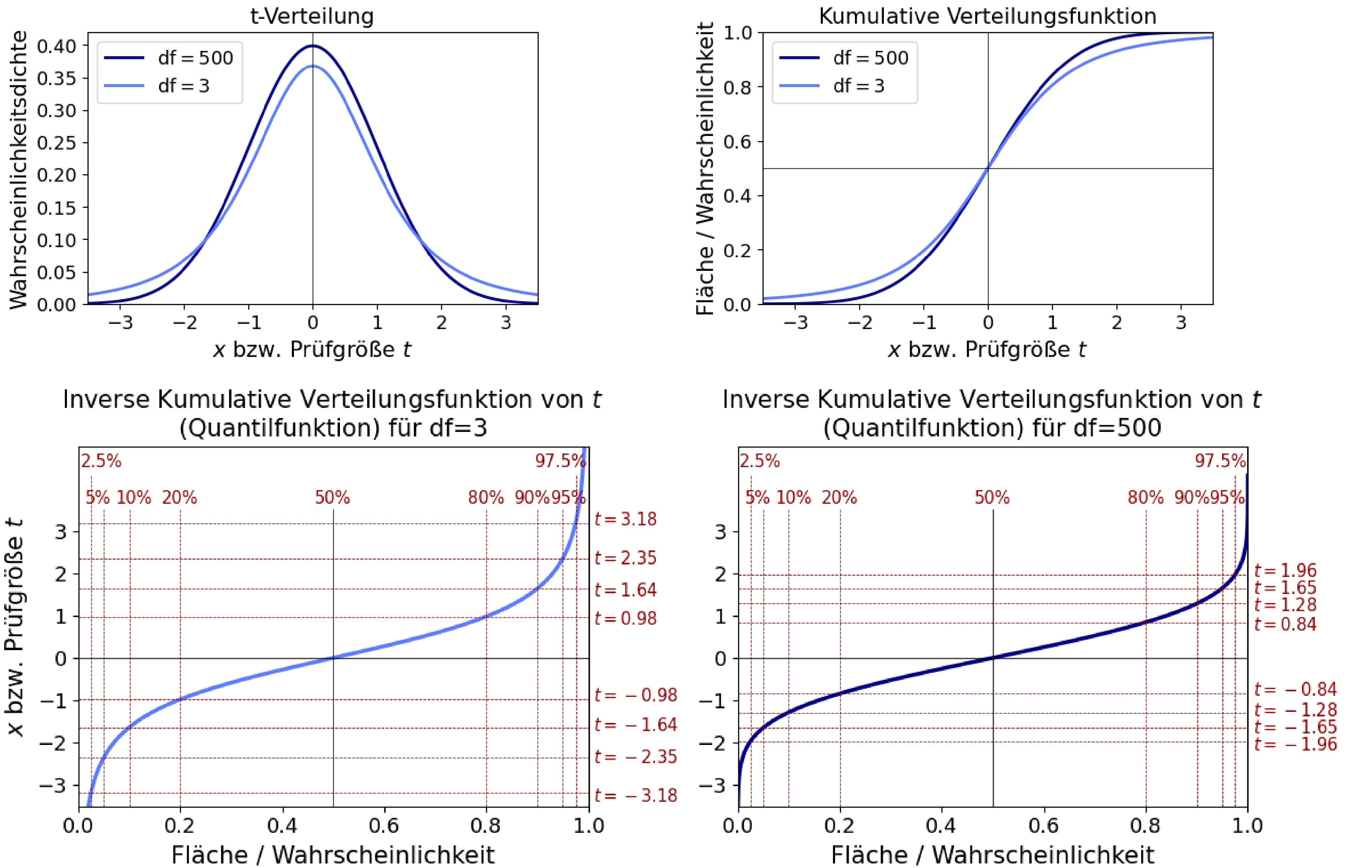
Mit der inversen kumulativen Verteilungsfunktion können **kritische t-Werte** für gegebene Signifikanzniveaus α berechnet werden, wie sie von der **t-Tabelle** bereitgestellt werden.

Bei gerichteter Hypothese ist der kritische t-Wert bei einem Signifikanzniveau von α gleich $F_t^{-1}(1-\alpha|\mathrm{df})$, bei ungerichteter Hypothese gleich $F_t^{-1}(1-\frac{\alpha}{2}|\mathrm{df})$. Für die Angabe eines kritischen t-Wertes wird üblicherweise als Subscript die kritische Fläche und die Zahl der Freiheitsgrade df angegeben:

$$\text{Gerichtet:} \quad t_{\text{crit}} = t_{(1-\alpha,\,\mathrm{df})} \qquad \text{Ungerichtet:} \quad t_{\text{crit}} = t_{(1-\frac{\alpha}{2},\,\mathrm{df})}$$

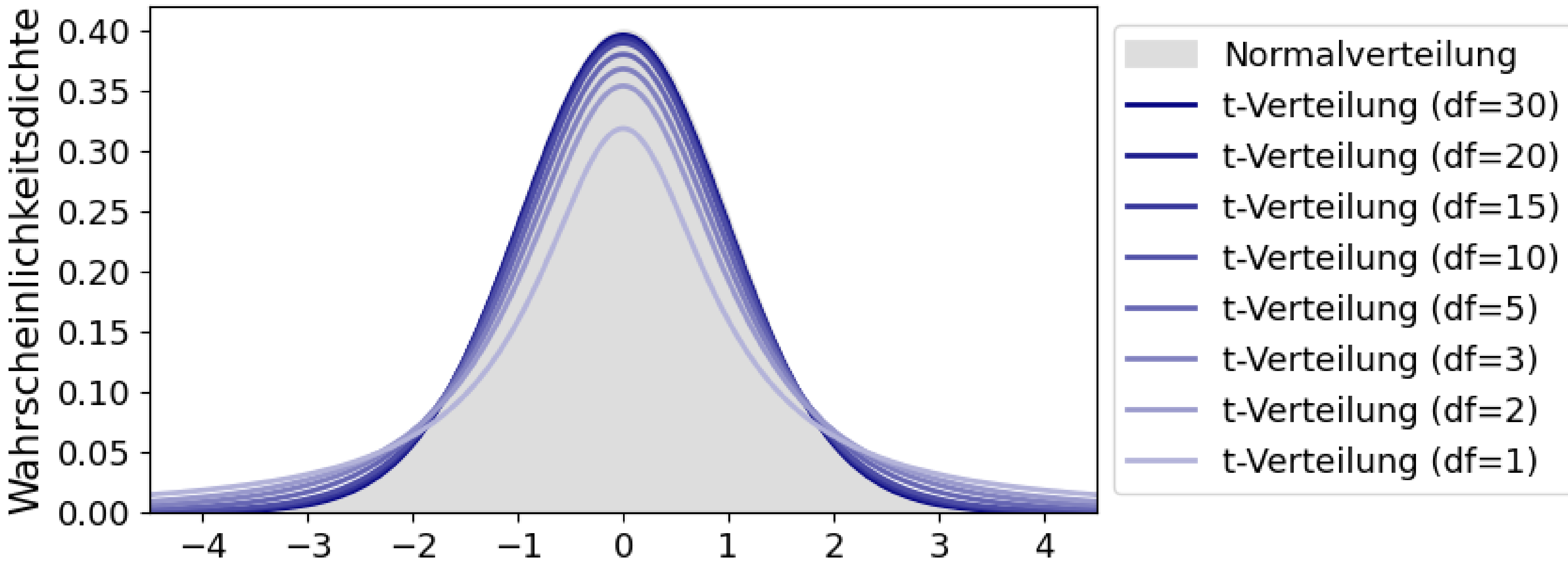


Funktionen der t-Verteilung



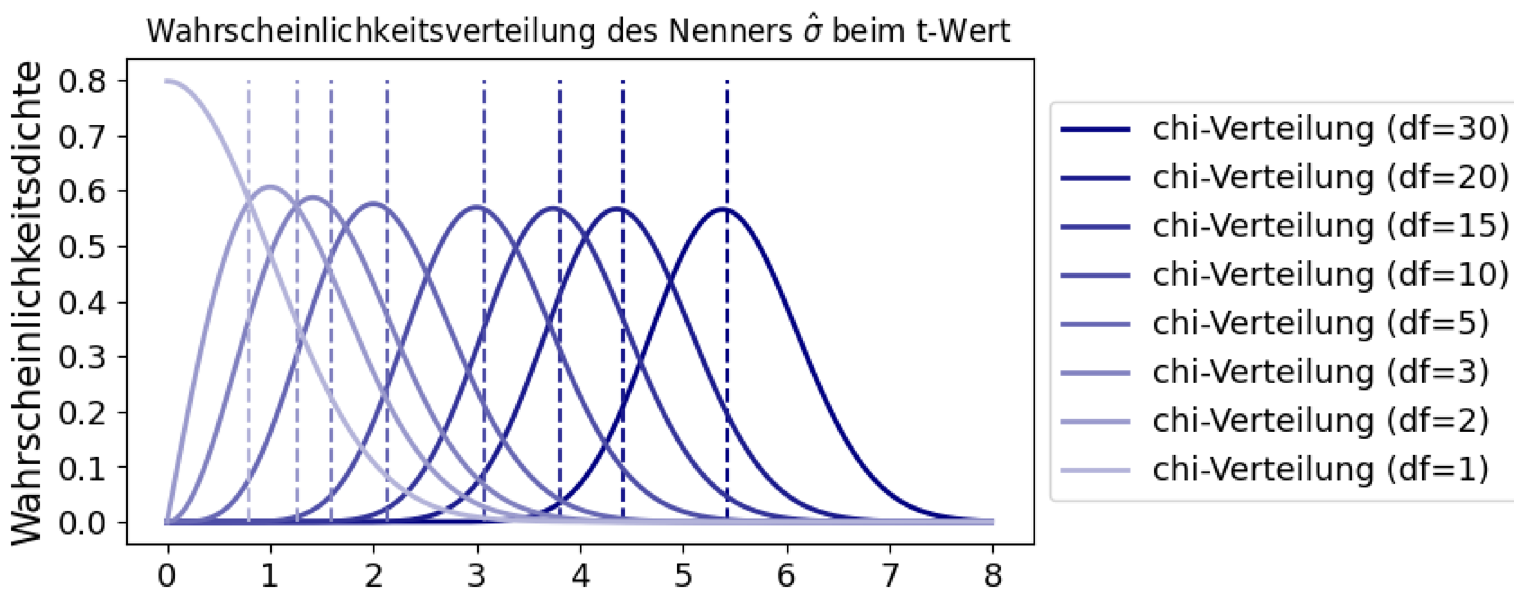
t-Verteilung versus Normalverteilung

- Der einzige Parameter der t-Verteilung, die Zahl der Freiheitsgrade df, bestimmt die Form der Verteilung.
- Die Zahl der Freiheitsgrade df hängt eng mit der Stichprobengröße n zusammen (z.B. $\mathrm{df} = n - 1$ bei einem Mittelwertsunterschied abhängiger Messungen).
- Je größer df/Stichprobengröße, desto ähnlicher wird die t-Verteilung der Normalverteilung!



Intuition: verdickte Flanken der t-Verteilung

- Warum sind die Flanken der Verteilung der Testgröße $t = \frac{\Delta\bar{x}}{\hat{\sigma}/\sqrt{n}}$ stärker ausgeprägt, wenn $\hat{\sigma}$ auf Basis der Stichprobe geschätzt werden muss?
- Der Grund liegt in der (chi-)Verteilung der Zufallsvariable $\hat{\sigma}$ im Nenner:
  - Bei kleinen Stichprobengrößen (kleines df), gibt es eine Assymmetrie der Verteilung hin zu Werten kleiner dem Mittelwert (welcher die korrekte Schätzung von σ repräsentiert — in der Abbildung gestrichelte Linien)
  - D.h. wir *teilen* $\Delta\bar{x}$ *überproportional häufig durch zu kleine Werte*, wodurch die *Teststatistik* $t = \frac{\Delta\bar{x}}{\hat{\sigma}/\sqrt{n}}$ *überproportional häufig zu extreme (negative oder positive) Werte* liefert.
  - Dies führt zu den stärkeren Flanken der t-Verteilung!



KEIN KLAUSUR STOFF

Standardisierte versus unstandardisierte t-Verteilung

Vor der Einführung des z-Tests hatten wir zunächst die **unstandardisierte Normalverteilung** kennengelernt, die durch zwei Parameter definiert ist: Mittelwert μ und Standardabweichung σ:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Für die Nullhypothese gilt idR $\mu = 0$:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$$

Verwenden wir statt des eigentlichen Kennwertes die **standardisierte Prüfgröße** $\frac{Kennwert}{se}$, vereinfacht sich die Nullhypothesen-Verteilung zur **Standardnormalverteilung**:

$$f(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Die Standardnormalverteilung hat keinen Parameter mehr (nur noch die *Variable* x).

Die **t-Verteilung** hatten wir dagegen direkt auf Basis der standardisierten Prüfgröße $\frac{Kennwert}{\hat{se}}$ eingeführt. Die t-Verteilung ist aus diesem Grund bereits eine **standardisierte Verteilung**, die nicht mehr vom Mittelwert oder Streuung abhängt. Im Gegensatz zur Standardnormalverteilung hat sie aber noch einen Parameter: die Zahl der Freiheitsgrade df.

Unstandardisierte t-Verteilung

KEIN KLAUSUR STOFF

Die Formel für die klassische (standardisierte) t-Verteilung lautet:

$$f_t(x|\mathrm{df}) = \frac{\Gamma\left(\frac{\mathrm{df}+1}{2}\right)}{\sqrt{\mathrm{df}\,\pi}\;\Gamma\left(\frac{\mathrm{df}}{2}\right)}\left(1+\frac{1}{\mathrm{df}}x^2\right)^{-\frac{\mathrm{df}+1}{2}} \qquad (\Gamma \text{ ist die Gamma-Funktion})$$

Wie die Standardnormalverteilung hat die t-Verteilung Mittelwert 0 (daher kann sie als Nullhypothesenverteilung fungieren). Ihre Streuung ist jedoch nicht exakt 1, sondern hängt von der Zahl der Freiheitsgrade ab: $\sigma^2 = \frac{\mathrm{df}}{\mathrm{df}-2}$.

Tatsächlich gibt es auch zur t-Verteilung ein unstandardisiertes Pendant, die **nicht-standardisierte t-Verteilung**, die von Mittelwert μ und Streuung σ abhängt:

$$f_t(x|\mu,\sigma,\mathrm{df}) = \frac{\Gamma\left(\frac{\mathrm{df}+1}{2}\right)}{\sigma\sqrt{\mathrm{df}\,\pi}\;\Gamma\left(\frac{\mathrm{df}}{2}\right)}\left(1+\frac{1}{\mathrm{df}}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\mathrm{df}+1}{2}}$$

Ähnlich wie beim z-Test hat es sich in der Praxis aber durchgesetzt immer mit standardisierten Prüfgrößen (wie z oder t) zu arbeiten. Daher finden die unstandardisierte t- und Normalverteilung im Kontext der Hypothesentestung selten eine Anwendung.

Der **Vorteil standardisierter Prüfgrößen** ist, dass diese vergleichbar zwischen Studien sind. Ein bestimmter z- oder t-Wert hat eine Aussagekraft, ohne den Standardfehler einer Studie zu kennen.

t-Test

Zur Durchführung des t-Tests müssen wir uns über drei Faktoren Gedanken machen:

1. Welche Art von t-Test benötige ich?

Siehe Entscheidungsbaum auf der nächsten Folie.

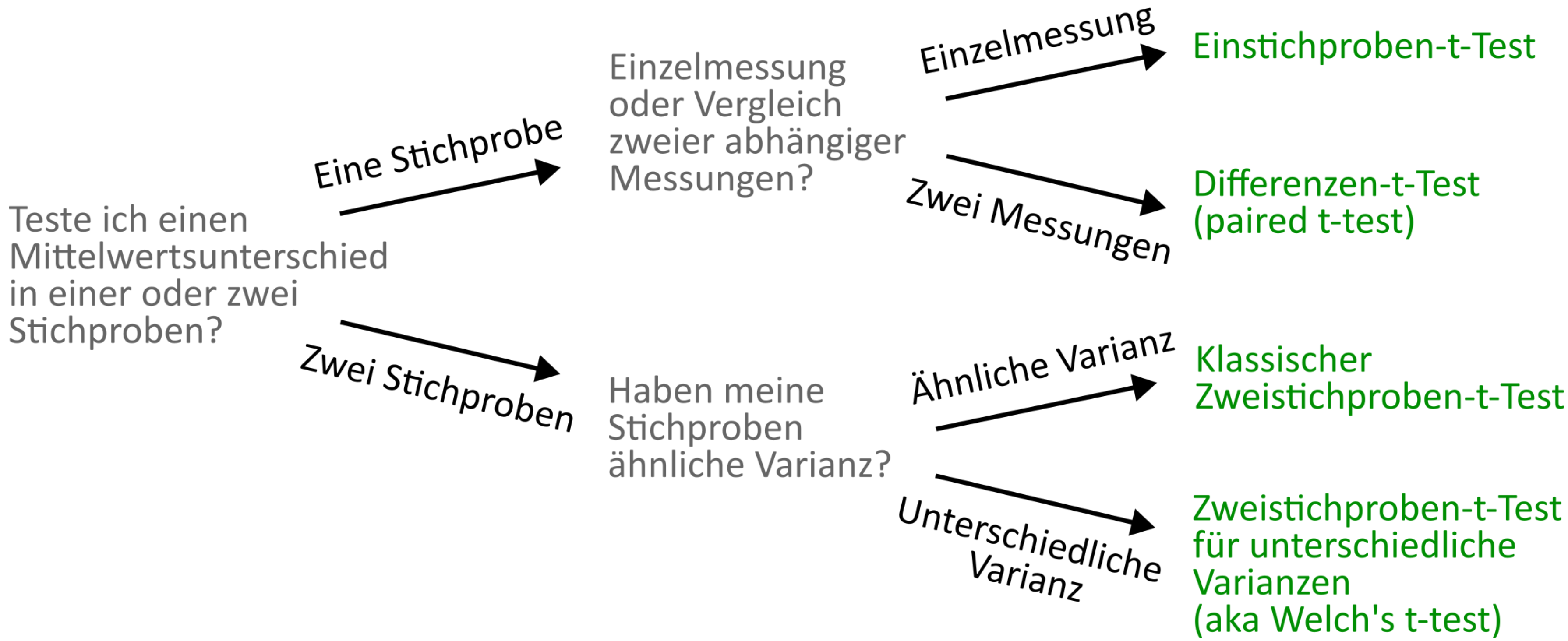
2. Einseitige vs. zweiseitige Testung

Ist mein Test einseitig (gerichtete Hypothese) oder zweiseitig (ungerichtete Hypothese)? Welchen Wert bestimme ich als Signifikanzniveau α? $\Rightarrow$ siehe Vorlesung 10!

3. Zahl der Freiheitsgrade

Die Zahl der Freiheitsgrade ist als Parameter für die t-Verteilung notwendig, und damit auch notwendig um Flächen (wie p-Werte) unter der Verteilung zu berechnen. De facto übernehmen das heute Computerprogramme, jedoch ist es nach wie vor Usus die Zahl der Freiheitsgrade eine statistischen Testes in wissenschaftlichen Veröffentlichungen anzugeben.

Familie der t-Tests



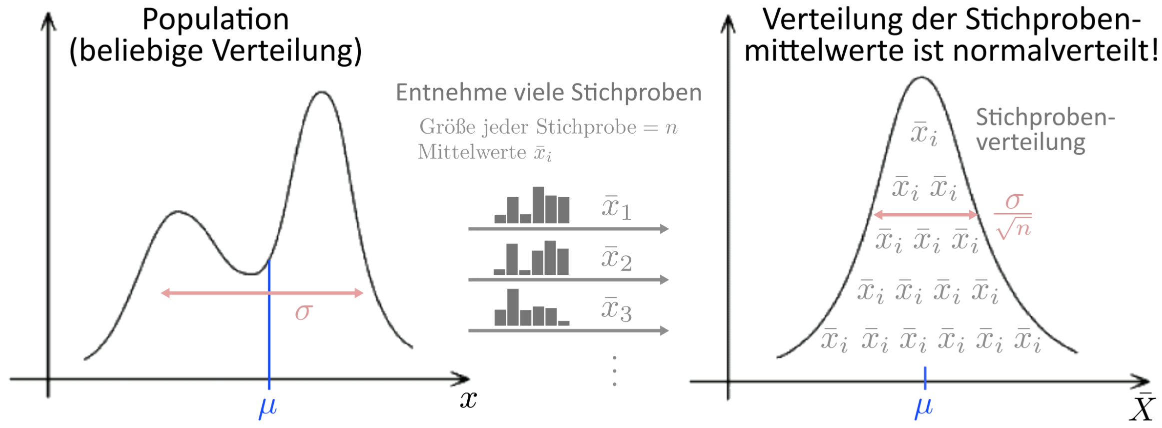
t-Test: Voraussetzungen

- Mindestens **intervallskalierte Daten**.
- **Hinreichende Normalverteilung** in der betrachteten Variable in der Population (Zweistichproben-t-Test: innerhalb beider Gruppen!)
  - Allerdings zeigen Simulationsstudien, dass der t-Test sehr „robust“ gegen Verletzungen der Normalverteilungsannahme ist.
  - Problematisch sind stark schiefe Verteilungen (rechts- oder linksschief) bei kleinen Stichprobengrößen.
  - Kann von keiner Normalverteilung ausgegangen werden: nicht-parametrische Testverfahren (z.B. Mann-Whitney-U-Test).
- **Ähnliche Varianzen** in beiden Gruppen (sonst Welch’s t-test).

Mr. T

But wait..

.. hatten wir nicht festgestellt, dass die Stichprobenverteilung gemäß dem zentralen Grenzwertsatz bei *beliebigen* Populationsverteilungen der Variable normalverteilt ist? (vgl. Vorlesung 08)



Warum sollten die Populationsverteilungen also plötzlich normalverteilt sein müssen?

Zwei Gründe:

- Damit $t = \text{Stichprobenkennwert} / \hat{se}$ einer t-Verteilung folgt, müssen der Stichprobenkennwert und der Standardfehler $\hat{se}$ unabhängig sein (dürfen nicht korrelieren). Dies ist (theoretisch) nur erfüllt, wenn die gemessenen Variablen eine Normalverteilung folgen — tatsächlich lässt sich genau mit dieser Bedingung die Normalverteilung ableiten.
- Der zentrale Grenzwertsatz gilt streng genommen nur für $n \to \infty$, was natürlich in empirischen Studien nie gegeben ist. Ist die Stichprobengröße begrenzt, ist die Stichprobenverteilung nur normalverteilt, wenn auch die gemessene Variable in der Population normalverteilt ist (Satz von Cramer).

Aber wie erwähnt: trotz dieser theoretischen Fallstricke ist der t-Test in der Praxis auch bei nicht-normalverteilten Populationsvariablen recht robust.

KEIN KLAUSUR STOFF

Zahl der Freiheitsgrade

- Die **Zahl der Freiheitsgrade** gibt die **Zahl der frei variierbaren Werte** bei der Berechnung eines statistischen Kennwertes an.
- Vereinfacht[1] gesprochen (wir werden eine Ausnahme kennenlernen) kann die Zahl der Freiheitsgrade berechnet werden, als die Stichprobengröße (Anzahl Datenpunkte insgesamt) *minus* die Anzahl der Parameter, die für die Berechnung des statistischen Kennwertes geschätzt werden müssen.

Beispiel

Beispiel Varianz: Angenommen, wir wollen für einer Stichprobe aus vier Werten die Varianz bestimmen:

$$\sigma^2 = \frac{1}{n}\sum_{1}^{4}(x_i - \bar{x})$$

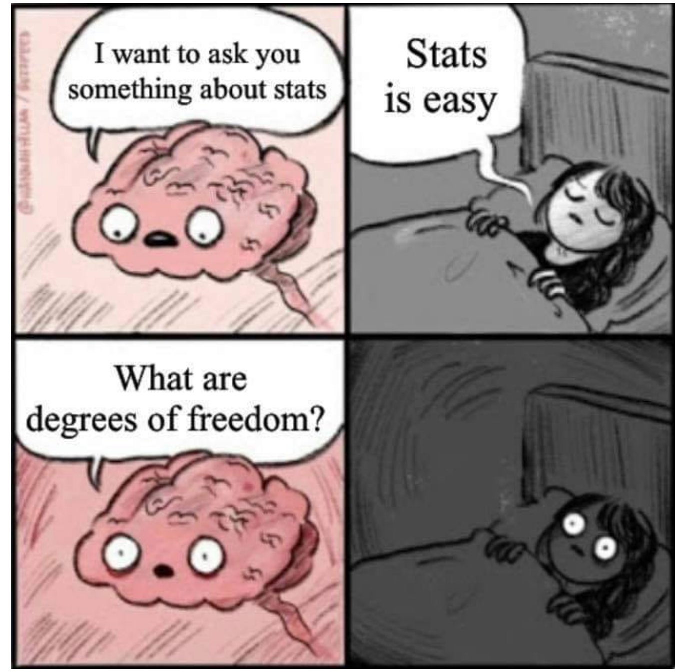
Aus der Formel der Varianz erkennen wir, dass wir zur Berechnung der Varianz **einen Parameter** aus den Daten bestimmen müssen, nämlich der Mittelwert $\bar{x}$. Die Zahl der Freiheitsgrade ergibt sich damit plain & simple nach obiger Definition als **Stichprobengröße n minus 1**.

Für die Intuition entscheidend ist folgende Frage bzw. Gedankenexperiment: wie viele der $n = 4$ Werte kann ich frei variieren unter der Bedingung, dass der Mittelwert gleich $\bar{x}$ beträgt? — dies ist die Zahl der Freiheitsgrade! Wie man sich leicht überlegen kann, könnten wir die ersten $n - 1 = 3$ Werte völlig frei bestimmen und hätten mit dem vierten Wert die Möglichkeit den Mittelwert $\bar{x} = 2$ zu erreichen. Bei (2, 4, 1) wäre der vierte Wert 1. Bei (-10000, 0, 0) wäre der vierte Wert 10008. Den vierten Wert allerdings ergibt sich wie gerade gesehen automatisch — wir können ihn nicht mehr frei bestimmen. Diese Logik gilt nicht nur für den Mittelwertsparameter, sondern für jeden Parameter der "auf dem Weg" zur Berechnung einer statistischen Größe geschätzt werden muss.

Zahl der Freiheitsgrade

KEIN KLAUSUR STOFF

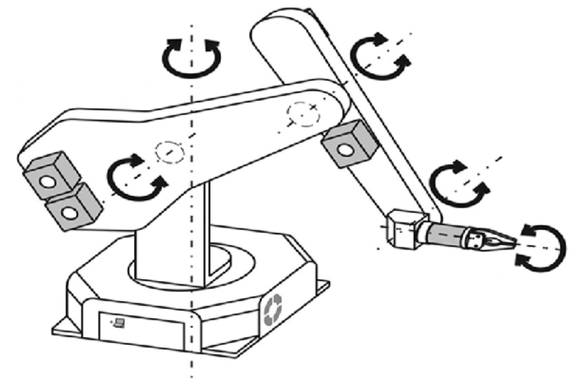
- Generell gilt: **wir wollen (idealerweise) eine möglichst große Zahl von Freiheitsgraden** für die Bestimmung unseres statistischen Kennwertes, denn alle Freiheitsgerade die nicht "unterwegs verloren gehen" (bei der Berechnung von Zwischenparametern), fließen in die Schätzung des Kennwertes ein und präzisieren damit dessen Schätzung.
- Das Konzept der Freiheitsgrade wird **besonders wichtig bei komplexeren Modellen**, wie sie in Statistik 2 behandelt werden (multiple Regression). Dort gehen schnell eine große Zahl von Freiheitsgraden "flöten" und die Zahl der Freiheitsgrade muss daher bereits bei der Wahl eines Modells berücksichtig werden.



Bildnachweis[2]

- Der **Begriff "Freiheitsgrad" ist wenig intuitiv**, weil die Datenpunkte nur im Gedankenexperiment "frei" gewählt werden können — in Realität sind sie natürlich gegeben.
- Ein intuitiverer Ausdruck wäre **effektive Stichprobengröße**: wie viele Datenpunkte aus meiner Stichprobe kann ich effektiv zur Schätzung meines statistischen Zielkennwerts nutzbar machen?

Zahl der Freiheitsgrade beim t-Test



- Für die Zahl der Freiheitsgrade beim t-Test gilt, dass die Zahl der Freiheitsgrade ausschließlich auf Basis der Streuung $\hat{\sigma}$ im Nenner bestimmt wird.
- Es reicht in fast allen Fällen also aus, zu zählen, wie viele Mittelwerte für die Berechnung von $\hat{\sigma}$ im Nenner des t-Wertes für die verschiedenen Tests verwendet werden.
  - Die glorreiche Ausnahme ist der Zweistichprobentest mit unterschiedlichen Varianzen, bei der die Freiheitsgrade nur näherungsweise (und immer noch kompliziert!) mit der sog. Welch–Satterthwaite-Gleichung bestimmt werden können.
- Eine berechtigte Frage ist, warum beim t-Test die Streuung $\hat{\sigma}$ nicht auch als Freiheitsgrad abgezogen werden muss (schließlich wird dieser Parameter bei der Berechnung des t-Wertes verwendet).
- Der Grund ist subtil und hängt damit zusammen, dass wir die t-Verteilung *unter der Voraussetzung verwenden, dass der Standardfehler geschätzt werden musste*. Der Verlust dieses Freiheitsgrades ist also in der Verteilung selbst bereits berücksichtigt und muss post-hoc nicht mehr als zusätzlicher Freiheitsgrad abgezogen werden.[3]

Zahl der Freiheitsgrade beim t-Test

Test	Frage	Zahl der Freiheitsgrade
Einstichprobentest mit Einzelmessung	Ist $\bar{x}$ größer als ein Referenzwert μ_0?	$\text{df} = n - 1$, da für die Berechnung von t genau ein Zwischenparameter bestimmt werden muss (der Mittelwert $\bar{x}$)
Einstichprobentest mit zwei abhängigen Messungen	Unterscheiden sich die Mittelwerte $\bar{x}_A$ und $\bar{x}_B$ zwischen zwei Bedingungen A und B? In diesem Fall kann man auch fragen: ist der Mittelwert der Differenzvariable ($\Delta\bar{x} = \overline{X_A - X_B}$) verschieden von Null?	$\text{df} = n - 1$, da für die Berechnung von t genau ein Zwischenparameter bestimmt werden muss (der Mittelwert $\overline{X_A - X_B}$)
Zweistichprobentest (ähnliche Varianzen)	Unterscheiden sich die Mittelwerte $\bar{x}_A$ und $\bar{x}_B$ zwischen zwei Gruppen A und B? ($\Delta\bar{x} = \bar{x}_A - \bar{x}_B$)	$\text{df} = n_A + n_B - 2$, da für die Berechnung von t genau zwei Zwischenparameter bestimmt werden müssen (die Mittelwerte $\bar{x}_A$ und $\bar{x}_B$)
Zweistichprobentest (unterschiedliche Varianzen)	Unterscheiden sich die Mittelwerte $\bar{x}_A$ und $\bar{x}_B$ zwischen zwei Gruppen A und B? ($\Delta\bar{x} = \bar{x}_A - \bar{x}_B$)	$\text{df} = \frac{\left(\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_B}\right)^2}{\frac{\left(\hat{\sigma}_A^2/n_A\right)^2}{n_A-1} + \frac{\left(\hat{\sigma}_B^2/n_B\right)^2}{n_B-1}}$ Falls $n_A = n_B = n$: $\text{df} = (n-1)\left(1 + \frac{2}{\left(\frac{\hat{\sigma}_A}{\hat{\sigma}_B}\right)^2 + \left(\frac{\hat{\sigma}_B}{\hat{\sigma}_A}\right)^2}\right)$

Freiheitsgrade beim Zweistichproben-t-Test mit unterschiedlicher Varianz

Auch wenn die Formel für die Freiheitsgrade beim Zweistichproben-t-Test mit unterschiedlicher Varianz recht kompliziert aussieht, ist eine nähere Betrachtung durchaus aufschlussreich.

Für $n_A = n_B = n$ gilt:

$$\mathrm{df} = (n-1)\left(1 + \frac{2}{\left(\frac{\hat\sigma_A}{\hat\sigma_B}\right)^2 + \left(\frac{\hat\sigma_B}{\hat\sigma_A}\right)^2}\right)$$

Zwei Fälle sind interessant:

Fall 1: Sind die Varianzen gleich (entgegen der Annahme), d.h. $\hat\sigma_A = \hat\sigma_B$, so vereinfacht sich die Formel zu

$$\mathrm{df} = (n-1)(1 + \frac{2}{2}) = 2n - 2,$$

also exakt die Formel des klassichen Zweistichprobentests.

KEIN KLAUSUR STOFF

Freiheitsgrade beim Zweistichproben-t-Test mit unterschiedlicher Varianz

Fall 2: Sind die Varianzen extrem unterschiedlich, so wird entweder $\left(\frac{\hat{\sigma}_A}{\hat{\sigma}_B}\right)^2$ oder $\left(\frac{\hat{\sigma}_B}{\hat{\sigma}_A}\right)^2$ extrem groß und die Formel vereinfacht sich zu

$$\text{df} = (n-1)(1+\frac{2}{\infty}) = (n-1)(1+0) = n-1$$

also exakt die Formel des Einstichprobentests.

Für extrem ungleiche Varianzen reduzieren sich also die Freiheitsgrade — und damit die *effektive Stichprobengröße* — auf die Größe einer der beiden verglichenen Gruppen.

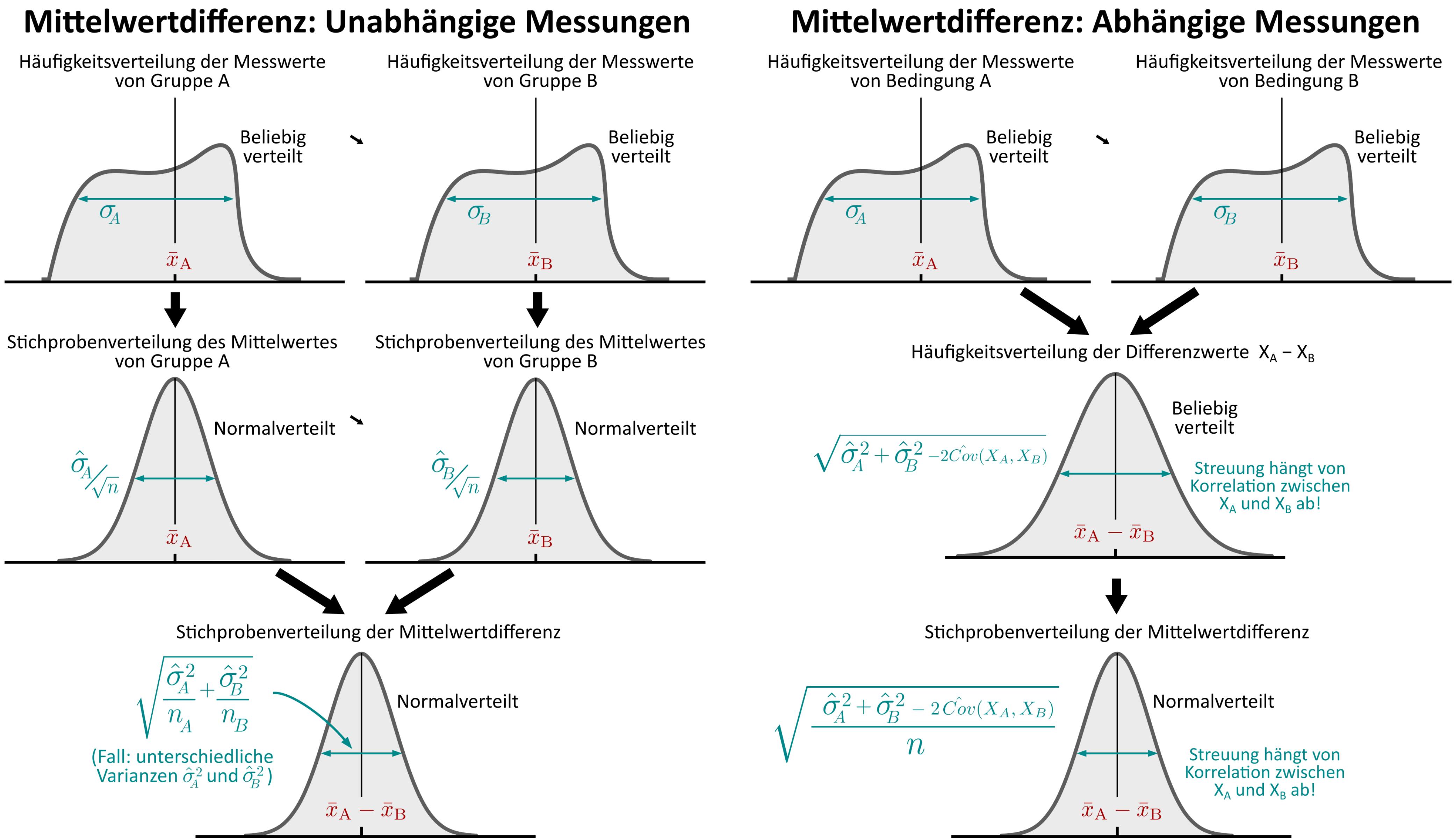
Das ist durchaus intuitiv: ist z.B. $\hat{\sigma}_A$ extrem viel kleiner als $\hat{\sigma}_B$, so spielt die Varianz der Gruppe A nahezu keine Rolle mehr. Die Versuchspersonen dieser Gruppe gehen also für die Berechnung des Standardfehlers “verloren”, was sich entsprechend auf die Freiheitsgrade auswirkt.

Im Grenzfall spielt die Varianz dieser Gruppe überhaupt keine Rolle mehr, sondern nur noch ihr Mittelwert. Die Gruppe hat damit die gleiche Funktion wie der Referenzwert μ_0 beim Einstichproben-t-Test.

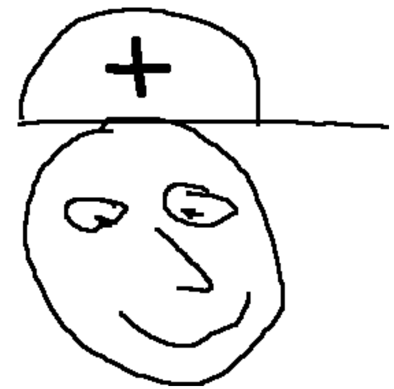
KEIN KLAUSUR STOFF

t-Test für Mittelwertunterschiede

Stichprobenverteilungen von Mittelwertdifferenzen (σ unbekannt)



Beispiel: t-Test für unabhängige Gruppen



Betrachten wir das Beispiel Med- versus Psych-Nasen für den Fall, dass wir die Standardabweichungen σ_{med} und σ_{psych} von Nasenlängen in der Med- und Psych-Population nicht kennen.

Es gilt immer noch:

$$\Delta\bar{x} = \bar{x}_{\text{med}} - \bar{x}_{\text{psych}} = 0{,}2cm$$

.. aber σ_{med} und σ_{psych} müssen jetzt aus unseren Messwerten geschätzt werden. Für das Beispiel nehmen wir an:

$$\hat{\sigma}_{\text{med}} = 0{,}5 \qquad\qquad \hat{\sigma}_{\text{psych}} = 0{,}25$$

Wir schlagen die Formel für den Standardfehler bei unabhängigen Messungen und ungleichen Varianzen nach und setzen ein:

$$\hat{se} = \sqrt{\frac{\hat{\sigma}^2_{\text{med}}}{n_{\text{med}}} + \frac{\hat{\sigma}^2_{\text{psych}}}{n_{\text{psych}}}} \overset{(Computer)}{=} 0{,}102$$

Erinnerung: es galt $n_{\text{med}} = n_{\text{psych}} = 30$.

Beispiel: t-Test für unabhängige Gruppen

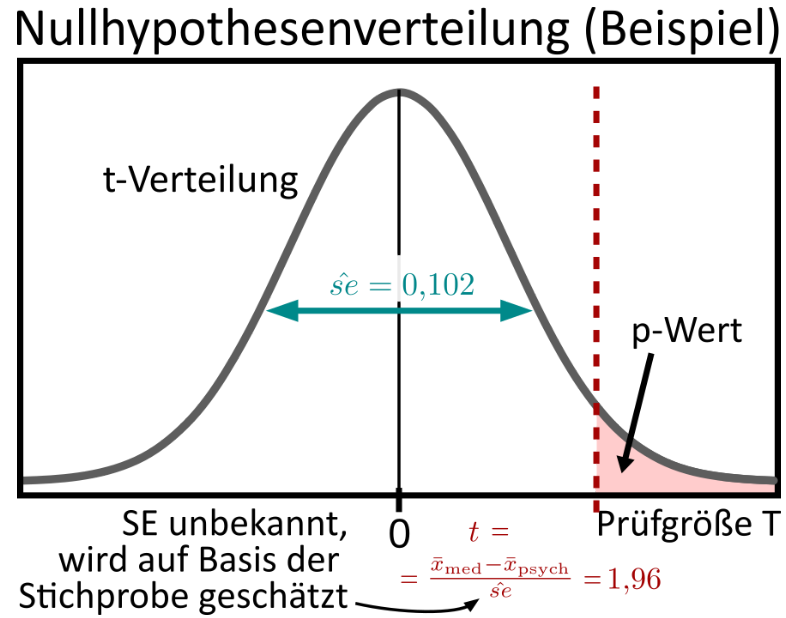
Mit dem Standardfehler bewaffnet, können wir nun den t-Wert berechnen:

$$t = \frac{\Delta\bar{x}}{\hat{se}} = 1{,}96$$

Es fehlen noch die Freiheitsgrade. Bei ungleichen Varianzen ist die Formel recht sperrig:

$$\mathrm{df} = (n-1)\left(1 + \frac{2}{\left(\frac{\hat{\sigma}_{\mathrm{med}}}{\hat{\sigma}_{\mathrm{psych}}}\right)^2 + \left(\frac{\hat{\sigma}_{\mathrm{psych}}}{\hat{\sigma}_{\mathrm{med}}}\right)^2}\right) =$$

$$= (30-1)\left(1 + \frac{2}{\left(\frac{0{,}5}{0{,}25}\right)^2 + \left(\frac{0{,}25}{0{,}5}\right)^2}\right) = 42{,}6$$



Aufgrund der gerichteten (positiven) Hypothese ist der p-Wert die Fläche *rechts des t-Wertes* unter der t-Verteilung:

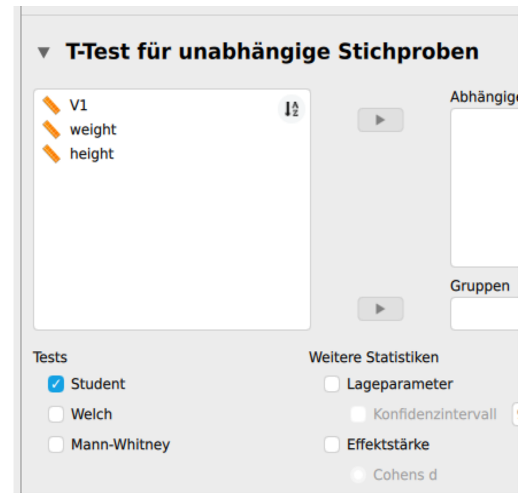
$$p = \int_t^\infty f_t(x|\mathrm{df})dx = \int_{1{,}96}^\infty f_t(x|42{,}6)dx = 1 - F_t(1{,}96|42{,}6) \overset{(Computer)}{=} 0{,}028$$

Der p-Wert ist dem p-Wert des z-Tests ($0{,}023$) sehr ähnlich. Dies ist wenig überraschend, da a) der Standardfehler einen ähnlichen Wert aufwies ($se = 2$ beim z-Test) und b) die Zahl er Freiheitsgrade so hoch ist, dass der Unterschied Normalverteilung vs. t-Verteilung marginal ist.

t-Test für unabhängige Stichproben

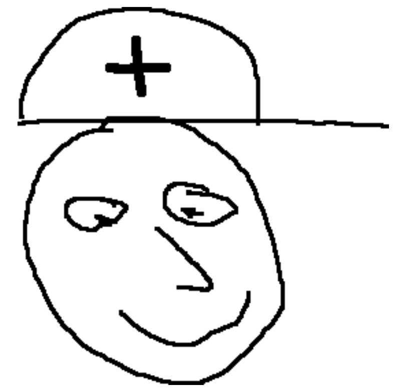
Klassischer Zweistichproben-t-Test (ähnliche Varianzen) versus Welch's t-Test (unterschiedliche Varianzen) — welcher der beiden Tests sollte nun verwendet werden?

- In der Psychologie sind ungleiche Varianzen die realistischere Annahme.
- Selbst bei randomisiert-kontrollierten Studien, bei denen Versuchspersonen den Gruppen aus einer identischen Ursprungs-Population zugewiesen werden, kann das Treatment selbst einen Einfluss auf die Varianz haben.
- Prinzipiell gibt es Testverfahren zur Überprüfung der Varianzgleichheit zwischen Gruppen, allerdings sind diese nicht unproblematisch
  - Ein Grund:auf Varianzgleichheit wird idR auf Basis eines nicht-signifikanten Ergebnis in einem solchen Test angenommen — "absence of evidence is not evidence of absence"[4]
- Herrscht Varianzgleichheit vor und ist die Stichprobengröße nicht extrem klein in einer Gruppe, kommen der Student'sche t-Test und Welch's t-Test zum gleichen Ergebnis.
- Vor diesem Hintergrund wird empfohlen[5], bereits ab moderaten Gruppengrößen (ca. $n \geq 8$ pro Gruppe) *immer* Welch's t-Test anzuwenden. In den meisten Statistikprogrammen ist dieser Test implementiert.



Option für Welch's t-test in JASP.

Beispiel: t-Test für abhängige Messungen



Sie führen ein Experiment *innerhalb der Med-Gruppe* (n=32) durch. In einer experimentellen Intervention stellen Sie den Med-Studierenden die Frage

> Studieren Sie Medizin, weil es der Wunsch Ihrer Eltern ist?

Sie messen die Nasenlängen dabei sowohl vor (“pre”), als auch nach (“post”) der Intervention. Ihre ungerichtete Hypothese ist, dass sich die Nasenlängen vor und nach der Intervention auf einem Signifikanzniveau $\alpha = 0{,}05$ unterscheiden.

Die Mittelwertdifferenz betrage $\Delta\bar{x} = \bar{x}_{\text{post}} - \bar{x}_{\text{pre}} = 0{,}1cm$.

Wir nehmen der Einfachheit halber an, dass die beiden Messungen X_{pre} und X_{post} unkorreliert sind, d.h. $\hat{Cov}(X_{\text{pre}}, X_{\text{post}}) = 0$. Die Standardabweichungen beider Messungen seien $\hat{\sigma}_{\text{pre}} = \hat{\sigma}_{\text{post}} = 0,5$. Die Standardabweichung $\hat{\sigma}_{\Delta}$ der Differenzvariable ΔX ist damit:

$$\hat{\sigma}_{\Delta} = \sqrt{\hat{\sigma}^2_{\text{pre}} + \hat{\sigma}^2_{\text{post}} - 2\,\hat{Cov}(X_{\text{pre}}, X_{\text{post}})} = \sqrt{(0,5)^2 + (0,5)^2 - 2 \cdot 0} = \frac{1}{\sqrt{2}}$$

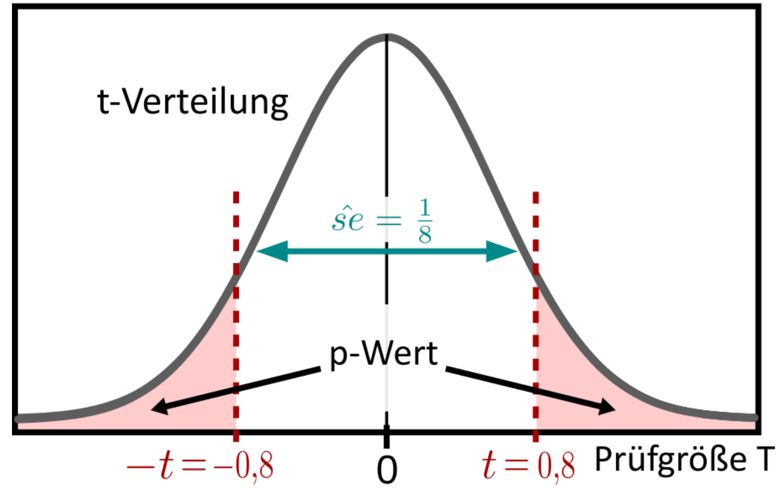
$$\text{Standardfehler:} \qquad \hat{se} = \frac{\hat{\sigma}_{\Delta}}{\sqrt{n}} = \frac{1}{\sqrt{2}\sqrt{32}} = \frac{1}{8}$$

Beispiel: t-Test für abhängige Messungen

Nun können wir den t-Wert berechnen:

$$t = \frac{\Delta\bar{x}}{\hat{se}} = \frac{0{,}1}{1/8} = 0{,}8$$

Aufgrund der ungerichteten Hypothese ist der p-Wert die Summe der Flächen unter der t-Verteilung links von $t = -0{,}8$ und rechts von $t = +0{,}8$. Bei $\mathrm{df} = n - 1 = 31$ Freiheitsgraden gilt:



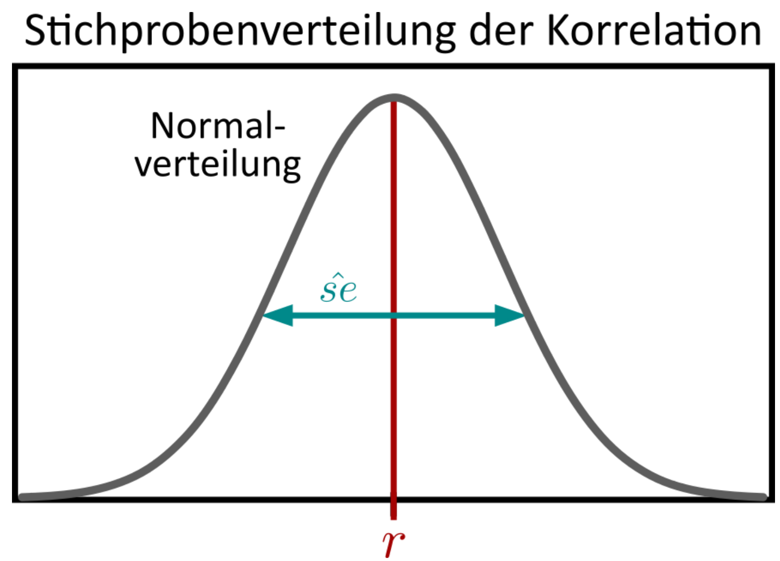
$$p = \int_{-\infty}^{-t} f_t(x|\mathrm{df})dx + \int_{t}^{\infty} f_t(x|\mathrm{df})dx \overset{(Symmetrie)}{=} 2 \cdot \int_{-\infty}^{-t} f_t(x|\mathrm{df})dx =$$

$$= 2 \cdot F_t(-t|\mathrm{df}) \overset{(einsetzen)}{=} 2 \cdot F_t(-0{,}8|31) \overset{(Computer)}{=} 0{,}43$$

Der p-Wert ist also deutlich größer als unser Signifikanzniveau $\alpha = 0{,}05$ und wir können die Nullhypothese $\Delta\bar{x} = 0$ nicht ablehnen. More research is needed!

t-Tests für Zusammenhänge

Stichprobenverteilung der Korrelation



- Analog zu Mittelwertsunterschieden kann auch die **Stichprobenverteilung der Korrelation** aufgestellt werden: als **Normalverteilung** mit dem **Mittelwert unserer Kennwertschätzung (hier r)** und einer Streuung, die dem **Standardfehler** ($\hat{se}$) des Kennwertes entspricht.
- Den Standardfehler der Korrelation haben wir bereits in Vorlesung 08 kennengelernt:

$$\hat{se}(r) = \sqrt{\frac{1-r^2}{n-2}}$$

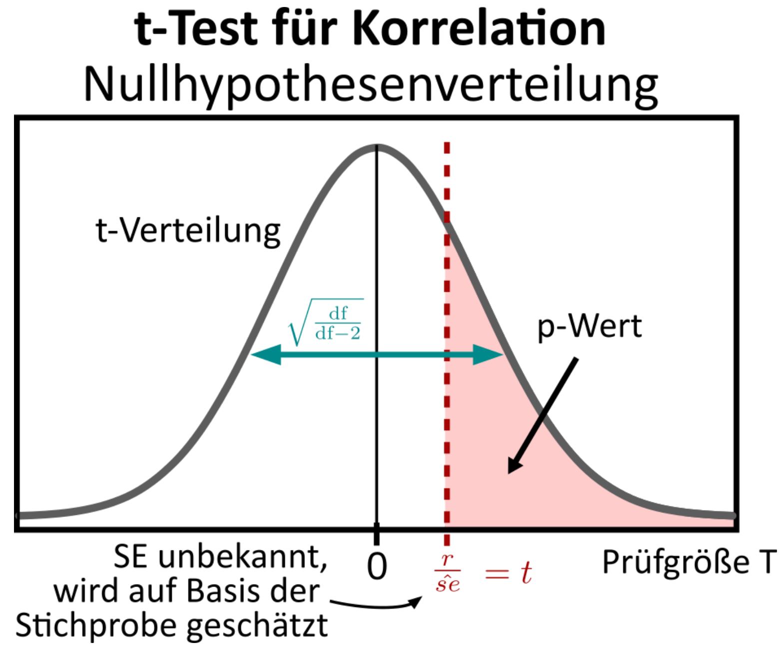
- Auch dieser Standardfehler ist eine Schätzung auf Basis der Stichprobe und entsprechend stellt die standardisierte Prüfgröße einen t-Wert dar:

$$t = \frac{r}{\hat{se}(r)} = r\sqrt{\frac{n-2}{1-r^2}}$$

(Präziser gesagt handelt es sich auch hier wieder um das Verhältnis einer normalverteilten Variable $(1-r^2)$ und einer chi-verteilten Variable ($\hat{se}(r)$) — ein solches Verhältnis führt zu einer t-verteilten Prüfgröße)

Nullhypothese und Hypothesentestung bei Korrelationen

- Die **Nullhypothesenverteilung** der Korrelation entspricht also der **t-Verteilung**.
- Die Nullhypothese entspricht der Annahme, dass der wahre Zusammenhang in der Population $\rho = 0$ ist.
- Auch hier können wir gerichtete und ungerichtete Hypothesen testen:
  - Gerichtete Hypothesen:
    - die Korrelation ist größer 0 ($\rho > 0$)
    - die Korrelation ist kleiner 0 ($\rho < 0$)
  - Ungerichtete Hypothese: die Korrelation ist ungleich 0 ($\rho \neq 0$)



- Die Berechnung der p-Werte erfolgt analog wie bei Mittelwertunterschieden.

Die Zahl der Freiheitsgrade der Korrelation ist $\mathrm{df} = n - 2$. Grund: es müssen **zwei** Mittelwerte bestimmt werden, der Mittelwert $\bar{x}$ der Variable X und der Mittelwert $\bar{y}$ der Variable Y.

Beispiel: Bestimmung des p-Wertes bei der Korrelation

- Nehmen wir an, die Korrelation von Größe und Gewicht in einer Stichprobe von $n = 12$ betrage $r = 0{,}4$.
- Wir wollen testen, ob die Korrelation auf einem Signifikanzniveau $\alpha = 0{,}05$ signifikant größer als 0 ist.
- Berechnung von der Prüfgröße t:

$$t = r\sqrt{\frac{n-2}{1-r^2}} = 0{,}4\sqrt{\frac{12-2}{1-0{,}4^2}} \overset{(Computer)}{=} 1{,}38$$

- Wir sehen in der Tabelle, dass der kritische t-Wert für $\mathrm{df} = n - 2 = 10$ Freiheitsgrade bei einem (einseitigen!) Signifikanzniveau von $\alpha = 0{,}05$ gleich $2{,}228$ beträgt.
- Die Nullhypothese wird also nicht abgelehnt, und der Effekt als nicht signifikant gewertet.

Tabelle der t-Verteilung

	Fläche				
df	0,85	0,9	0,95	0,975	0,99
1	1,964	3,078	6,314	.2,706	31,821
2	1,386	1,886	2,92	4,303	6,965
3	1,25	1,638	2,353	3,182	4,541
4	1,19	1,533	2,132	2,776	3,747
5	1,156	1,476	2,015	2,571	3,365
6	1,134	1,44	1,943	2,447	3,143
7	1,119	1,415	1,895	2,305	2,998
8	1,108	1,397	1,86	2,306	2,896
9	1,1	1,383	1,833	2,262	2,821
10	1,093	1,372	1,813	2,228	2,764
30	1,055	1,31	1,697	2,042	2,459
40	1,05	1,303	1,684	2,021	2,423
60	1,046	1,296	1,071	1,997	2,39
120	1,041	1,289	1,658	1,98	2,358
∞	1,039	1,282	1,645	1,96	2,326

Beispiel: Bestimmung des p-Wertes bei der Korrelation

- Statistische Programme geben den p-Wert in der Regel automatisch mit an und es ist somit ersichtlich, ob es sich um eine statistisch signifikante Korrelation handelt.
- **Standardmäßig** handelt es sich dabei immer um einen **Test für eine ungerichtete Hypothese**.
- War die Hypothese dagegen gerichtet, gilt:
  - Halbiere den p-Wert der ungerichteten Hypothese, falls das Vorzeichen von r in Richtung der Hypothese ist.
  - Verdoppele den p-Wert der ungerichteten Hypothese, falls das Vorzeichen von r in gegensätzlicher Richtung der Hypothese ist.
  - Im Beispiel wäre also der Korrelationskoeffizient $2 \times 0{,}198 = 0{,}396$, falls wir einen negativen Zusammenhang vorhergesagt hätten, und $0{,}198/2 = 0{,}099$, falls wir einen positiven Zusammenhang vorhergesagt hätten.

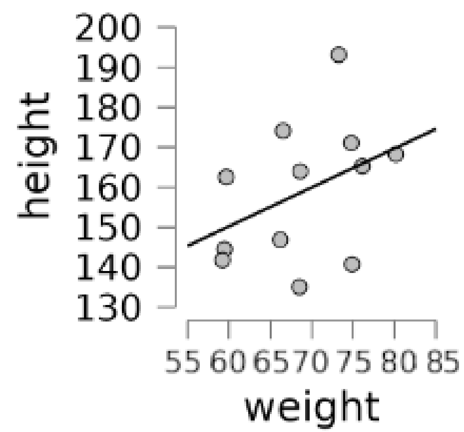
Korrelation

```
jaspRegression::Correlation(
    version = "0.17.2",
    scatterPlot = TRUE,
    variables = list("height", "weight"))
```

Pearsons Korrelationen

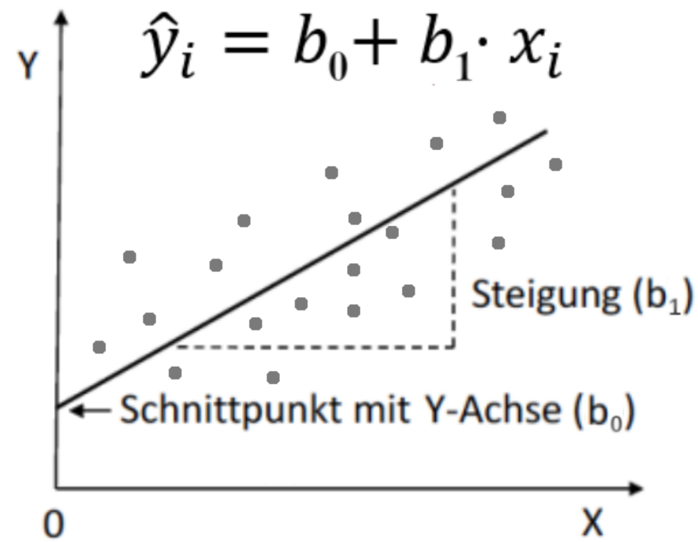
Variable		height	weight
1. height	Pearsons r	—	
	p-Wert	—	
2. weight	Pearsons r	0.400	—
	p-Wert	0.198	—

Streudiagramm



Auswahl in JASP: Regression → Klassisch → Korrelation. Zusätzliche Selektion des Streudiagramms.

Inferenzstatistik für Regressionskoeffizienten



- Die Hypothesentestung für den Regressionskoeffizienten der einfachen Regression (d.h. 1 UV) unterscheidet sich nicht von der Korrelation.
- Auch für den Regressionskoeffizienten können wir einen Standardfehler definieren:

$$\hat{se}(b_1) = \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} \sqrt{\frac{1 - r^2}{n - 2}}$$

- .. und einen darauf basierenden t-Wert:

$$t = \frac{b_1}{\hat{se}(b_1)}$$

- Die Zahl der Freiheitsgrade ist wie bei der Korrelation $\mathrm{df} = n - 2$, da auch hier für die Berechnung des Standardfehlers die beiden Mittelwerte $\bar{x}$ und $\bar{y}$ bestimmt werden müssen.
- Der Rest ist bekannt.

Inferenzstatistik für Regressionskoeffizienten

- Während die Regressionssteigung abhängt davon, welche Variable als UV (bzw. x) und welche als AV (bzw. y) definiert wird, sind die t-Werte (und damit auch die p-Werte) unabhängig von der Rollenverteilung der Variablen.
- Dies ergibt sich direkt aus dem Zusammenhang von r und b_1 (vgl. Vorlesung 06):

$$b_1 = \frac{\hat{\sigma}_y}{\hat{\sigma}_x} r$$

- Damit ist die Prüfgröße t:

$$t = \frac{b_1}{SE(b_1)} = \frac{\frac{\hat{\sigma}_Y}{\hat{\sigma}_X} r}{\frac{\hat{\sigma}_Y}{\hat{\sigma}_X}\sqrt{\frac{1-r^2}{n-2}}} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

- Der t-Wert bei der einfachen Regression ist also identisch zum t-Wert der Korrelation und insbesondere nur noch abhängigg von r und nicht mehr b_1
- Da für die Korrelation r die Rollenverteilung der beiden Zusammenhangsvariablen unerheblich ist, folgt, dass der p-Wert bei der Regression ebenso wenig von der Zuordnung der Variablen als UV und AV abhängt.

KEIN KLAUSUR STOFF

Testung des y-Achsenabschnitts bei der Regression

- Wir haben bislang den Achsenabschnitt b_0 aus der Diskussion außen vor gelassen.
- Im Kontext der Regression wird der Achsenabschnitt b_0 idR nicht getestet.
- Grund: die Frage, ob Variable Y bei $X = 0$ einen y-Achsenabschnitt aufweist, der signifikant von Null verschieden ist, ist im Kontext der Regression selten interessant — schließlich ist es ja der ganze Zweck der Regression die systematische Verändeurng von Y in X zu analysieren und dabei gerade nicht nur einen bestimmten Wert von X zu betrachten.
- Nichts desto trotz ist auch die Schätzung von b_0 mit Unsicherheit verbunden, die durch folgenden Standardfehler definiert ist:

$$\hat{se}(b_0) = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}}$$

KEIN KLAUSUR STOFF

Testung des y-Achsenabschnitts bei der Regression

- Ein Spezialfall ist das lineare Modell ohne Steigung (bzw. ohne x!):

$$\hat{y} = b_0$$

- In diesem Modell kommt dem Regressionskoeffizienten b_0 und dessen Unsicherheit eine interessantere Bedeutung zu: die Frage ob b_0 signifikant verschieden von 0 ist, ist hier gleichbedeutend mit der Frage ob die Zufallsvariable Y **signifikant verschieden von Null ist**.
- Mit anderen Worten: dieses Modell ist nichts anderes als ein Einstichproben-t-Test!
- Dies beinhaltet auch den Vergleich zweier abhängiger Messungen A und B, wenn wir zuvor Y als Differenzvariable definieren: $Y = Y_A - Y_B$
- .. dann testet

$$\hat{y} = b_0$$

- die Frage, ob die Mittelwertsdifferenz von A und B signifikant verschieden von Null ist.

KEIN KLAUSUR STOFF

“Common statistical tests are linear models”

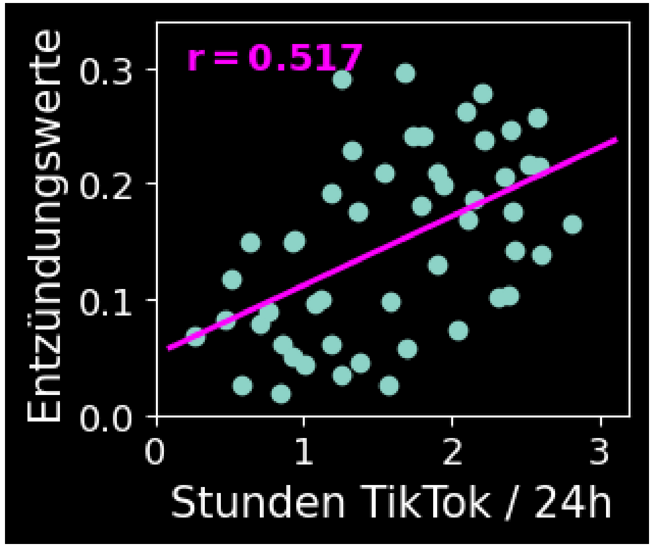
Die Parallele von bekannten statistischen Tests und (generalisierter) linearer Regression lässt sich auf alle Tests erweitern, die wir in Statistik 1 und Statistik 2 kennenlernen.

Hier eine Auswahl der Tests aus Statistik 1:

Test	Lineares Modell	Spezifierung	Nullhypothese
Pearson-Korrelation	$\hat{y} = \beta_0 + \beta_1 x$	-	$\mathcal{H}_0 : \beta_1 = 0$
Spearman-Korrelation	$\text{Rang}(\hat{y}) = \beta_0 + \beta_1 \text{Rang}(x)$	-	$\mathcal{H}_0 : \beta_1 = 0$
Einstichproben-t-Test	$\hat{y} = \beta_0$	-	$\mathcal{H}_0 : \beta_0 = 0$
Differenzen-t-Test	$\hat{y_A} - \hat{y_B} = \beta_0$	-	$\mathcal{H}_0 : \beta_0 = 0$
Zweistichproben-t-Test	$\hat{y} = \beta_0 + \beta_1 x$	$x = 0$ für alle Datenpunkte von Gruppe A und $x = 1$ für alle Datenpunkte von Gruppe B → Regression auf binäre x-Variable!	$\mathcal{H}_0 : \beta_1 = 0$

Die Analogie von bekannten statistischen Tests und linearem Modell ist lange bekannt, ging aber 2021 durch Beiträge von Jonas Lindeløv viral[6] [7] [8].

KEIN KLAUSUR STOFF

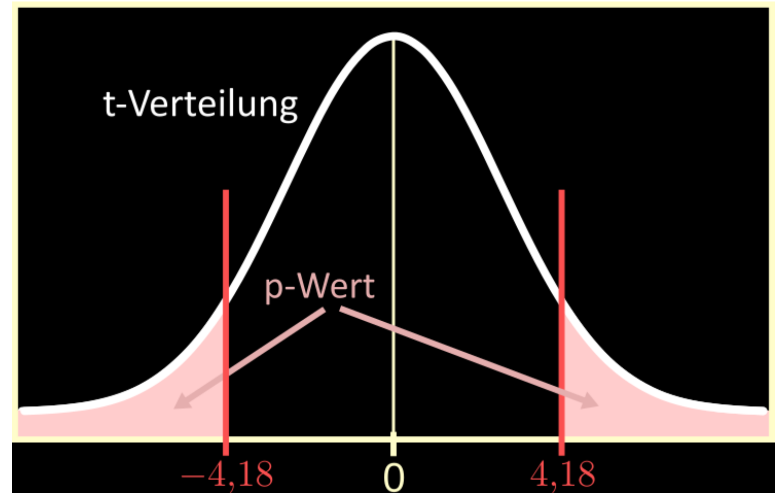


Stichwort Signifikanz von Zusammenhängen: die Signifikanz des unerwarteten Zusammenhangs von TikTok-Onlinezeit und Entzündungswerten haben Sie bislang nicht getestet.

Basierend auf dem Korrelationskoeffizienten $r = 0{,}517$ ergibt sich die Prüfgröße t direkt:

$$t = \frac{r}{\hat{se}(r)} = r\sqrt{\frac{n-2}{1-r^2}} = 0{,}517\sqrt{\frac{50-2}{1-0{,}517^2}} = 4{,}18$$

Der Zusammenhang war unerwartet und es gab dementsprechend keine gerichtete Hypothese. Der p-Wert ist in diesem Fall also der doppelte Wert der Fläche rechts des (positiven) t-Werts von $4{,}18$, oder analog, der doppelte Wert links des negierten t-Werts $-4{,}18$.



$$p = 2\int_{-\infty}^{-4{,}18} f_t(x|\mathrm{df})dx = 2F_t(-4{,}18|48) = 0{,}0001$$

(Die Zahl der Freiheitsgrade beim t-Test der Korrelation ist $n-2$)

Der Zusammenhang von TikTok-Onlinezeit und Entzündungswerten ist also deutlich signifikant.

Trotz des hochsignifikanten Effektes bleiben Sie skeptisch — Ihr größtes Fragezeichen: was ist Ursache, was ist Wirkung? Folgen erhöhte Entzündungswerte tatsächlich auf TikTok-Konsum (einschlägiger Kanäle)? Oder ist es einfach so, dass Erkrankte (mit erhöhten Entzündungswerten) Rat und Solidarität auf TikTok suchen?

Um die Gruppen*unterschiede* nun ebenfalls mithilfe des t-Tests zu überprüfen, listen Sie nochmals alle relevanten Werte auf, die Sie zum Teil bereits beim z-Test bestimmt hatten:

	Fallzahl n	$\Delta\bar{x}$	**Standardfehler** $\hat{se}$	**Freiheitsgrade** df
TikTok	2×50	$0{,}577$	$0{,}123$	$93{,}7$
Entzündung	2×50	$0{,}0243$	$0{,}0147$	$96{,}2$

- Neu hinzugekommen sind die Freiheitsgrade, die Sie in der Tabelle mit der Formel für ungleiche Varianzen bestimmt haben:

$$\mathrm{df} = (n-1)\left(1 + \frac{2}{\left(\frac{\hat{\sigma}_{\text{control}}}{\hat{\sigma}_{\text{paradoxia}}}\right)^2 + \left(\frac{\hat{\sigma}_{\text{paradoxia}}}{\hat{\sigma}_{\text{control}}}\right)^2}\right)$$

- Zur Erinnerung:
  - TikTok: $\hat{\sigma}_{\text{control}} = 0{,}55$; $\hat{\sigma}_{\text{paradoxia}} = 0{,}68$
  - Entzündung: $\hat{\sigma}_{\text{control}} = 0{,}068$; $\hat{\sigma}_{\text{paradoxia}} = 0{,}078$

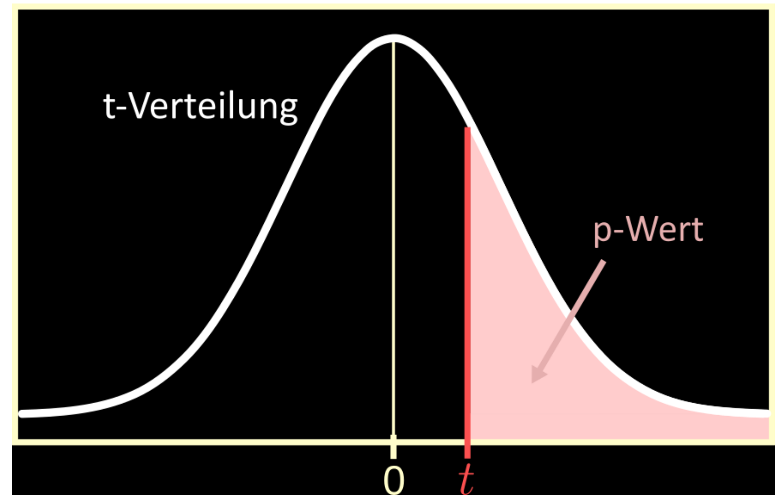
Mit diesen Information erhalten Sie folgende t-Werte (* in diesem Fall identisch mit den z-Werten!):

$$\text{TikTok:} \qquad t = \frac{\Delta \bar{x}}{\hat{se}} = \frac{0{,}577}{0{,}123} = 4{,}69$$

$$\text{Entzündung:} \qquad t = \frac{\Delta \bar{x}}{\hat{se}} = \frac{0{,}0243}{0{,}0147} = 1{,}65$$

Statt der Standardnormalverteilung integrieren Sie nun einfach die entsprechende Fläche unter der t-Verteilung:

$$p = \int_t^\infty f_t(x|\mathrm{df})dx = 1 - F_t(t|\mathrm{df})$$



(*) Warum identisch? Weil wir für die z-Werte — in Abwesenheit anderweitig bekannter Werte — ebenfalls die Stichprobenstreuung verwendet hatten.

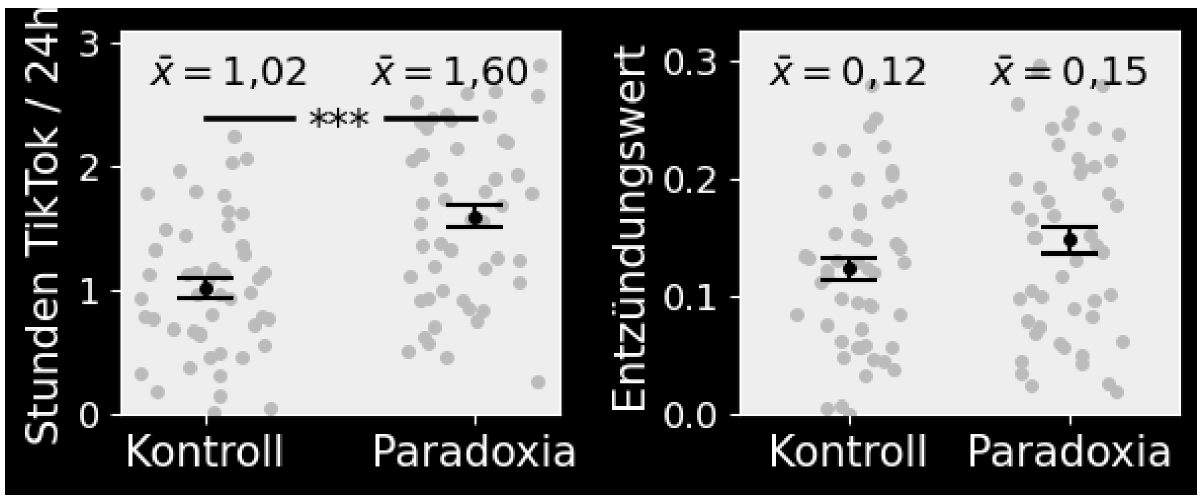
Die große Frage ist: würden weiterhin beide Effekte signifikant bleiben? Antwort:

$$\text{TikTok:} \qquad p = 1 - F_t(4{,}83|93{,}7) \overset{\text{(Computer)}}{=} 0{,}000003$$

$$\text{Entzündung:} \qquad p = 1 - F_t(1{,}65|96{,}2) \overset{\text{(Computer)}}{=} 0{,}051$$

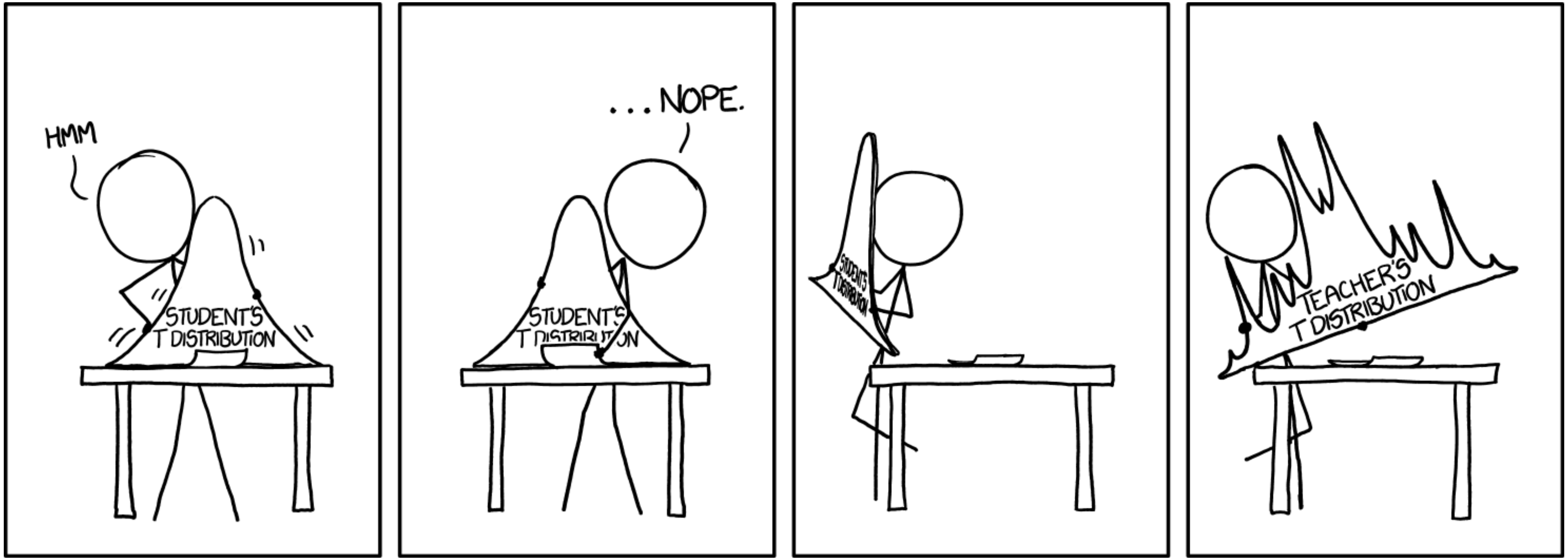
Kurios — das Pendel schlägt exakt auf der anderen Seite des Signifikanzniveaus aus! Der Entzündungseffekt ist nicht mehr signifikant. Der TikTok-Effekt bleibt stabil.

Bei beiden Effekten zeigt sich die leicht konservativere Natur der t-Verteilung. Die stärkeren Flanken im Vergleich zur Normalverteilung führen zu geringfügig höheren p-Werten. Notgedrungen müssen Sie Ihre zentrale Grafik anpassen und ein Signifikanzsternchen entfernen:



Vor dem Hintergrund dieser neuen Ergebnisse, stellen sich eine Reihe von Fragen:

- Ist die Interpretation der Signifikanz fundamental verschieden zwischen dem z- und dem t-Test?
- Ist die Entzündungshypothese eindeutig widerlegt?
- Können Sie aus dem hochsignifikanten Effekt der TikTok-Hypothese und dem nicht-signifikanten Effekt der Entzündungshypothese schlussfolgern, dass der TikTok-Effekt signifikant stärker ist?
- Können Sie schlusfolgern, dass Paradoxia eindeutig durch den TikTok-Effekt verursacht wird?

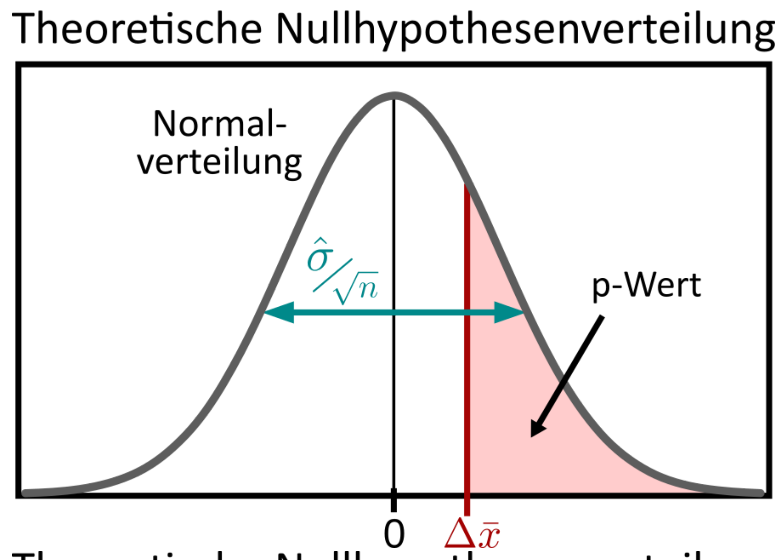

HMM
STUDENT'S
T DISTRIBUTION
...NOPE.
STUDENT'S
TEACHER'S
T DISTRIBUTION

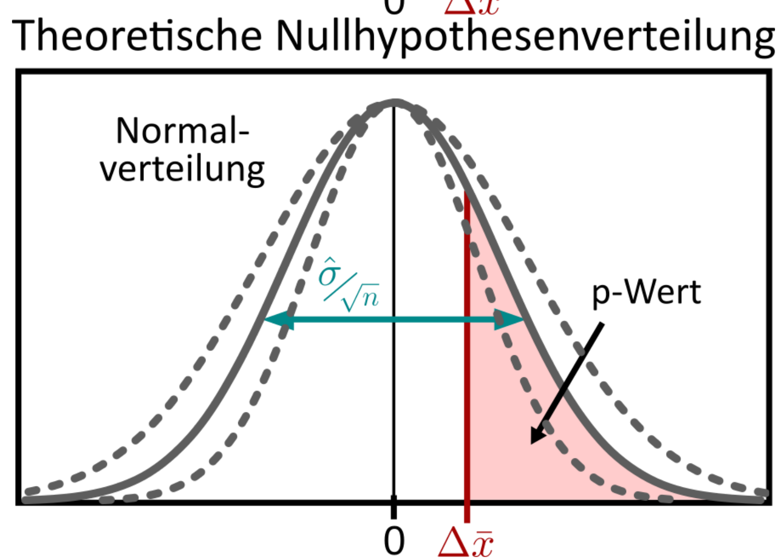
xkcd#1347

Appendix

Herleitung der t-Verteilung

- Um ein Gefühl dafür zu bekommen, wie die t-Verteilung hergeleitet werden kann, ist es sinnvoll nochmals das Testen des **unstandardisierten Mittelwertunterschiedes** $\Delta\bar{x}$ zu betrachten.
- Da die Streubreite $\hat{\sigma}/\sqrt{n}$ der Stichprobenverteilung nicht genau bekannt ist, sondern mit Unsicherheit behaftet ist (siehe gestrichelte Linien im unteren Bild), ist der p-Wert nicht eindeutig bestimmt.
- Es ist dennoch möglich, den (erwarteten) p-Wert zu berechnen, in dem man nicht nur die Fläche auf Basis der eigentlichen Schätzung $\hat{\sigma}$ berechnet, sondern *für alle möglichen* Werte von σ.
- Die so erhalteten Flächen mittelt man gewichtet an der Wahrscheinlichkeit jedes möglichen Wertes von σ — dadurch wird die Unsicherheit von $\hat{\sigma}$ “marginalisiert”.





- Betrachtet man diese Art der Integration mit Marginalisierung nicht für einen konkreten Fall, sondern für die theoretische Verteilung, ergibt sich die t-Verteilung.

KEIN KLAUSUR STOFF

Fußnoten

1. https://stats.stackexchange.com/a/17148/62140
2. https://www.facebook.com/SaraswathiAnalytics/photos/a.101349091331739/428667868599858
3. https://stats.stackexchange.com/questions/226483/does-the-determination-of-the-mean-and-sd-imply-the-loss-of-one-or-two-degrees-o
4. Altman DG, Bland JM (1995) Absence of evidence is not evidence of absence. BMJ 311:485.
5. Delacre M, Lakens D, Leys C (2017) Why Psychologists Should by Default Use Welch’s t-test Instead of Student’s t-test. International Review of Social Psychology 30:92.
6. https://stats.stackexchange.com/questions/303269/common-statistical-tests-as-linear-models
7. https://lindeloev.github.io/tests-as-linear/
8. https://twitter.com/jonaslindeloev/status/1110907133833502721