

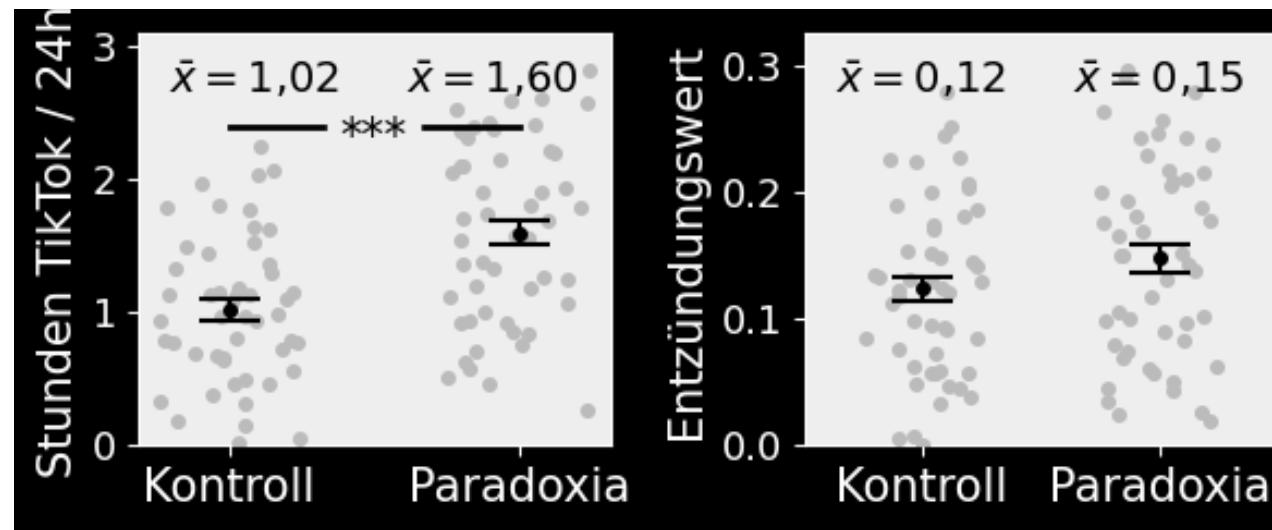
M24 Statistik 1: Wintersemester 23/24

Vorlesung 12: Konfidenzintervalle

Prof. Matthias Guggenmos

Health and Medical University Potsdam



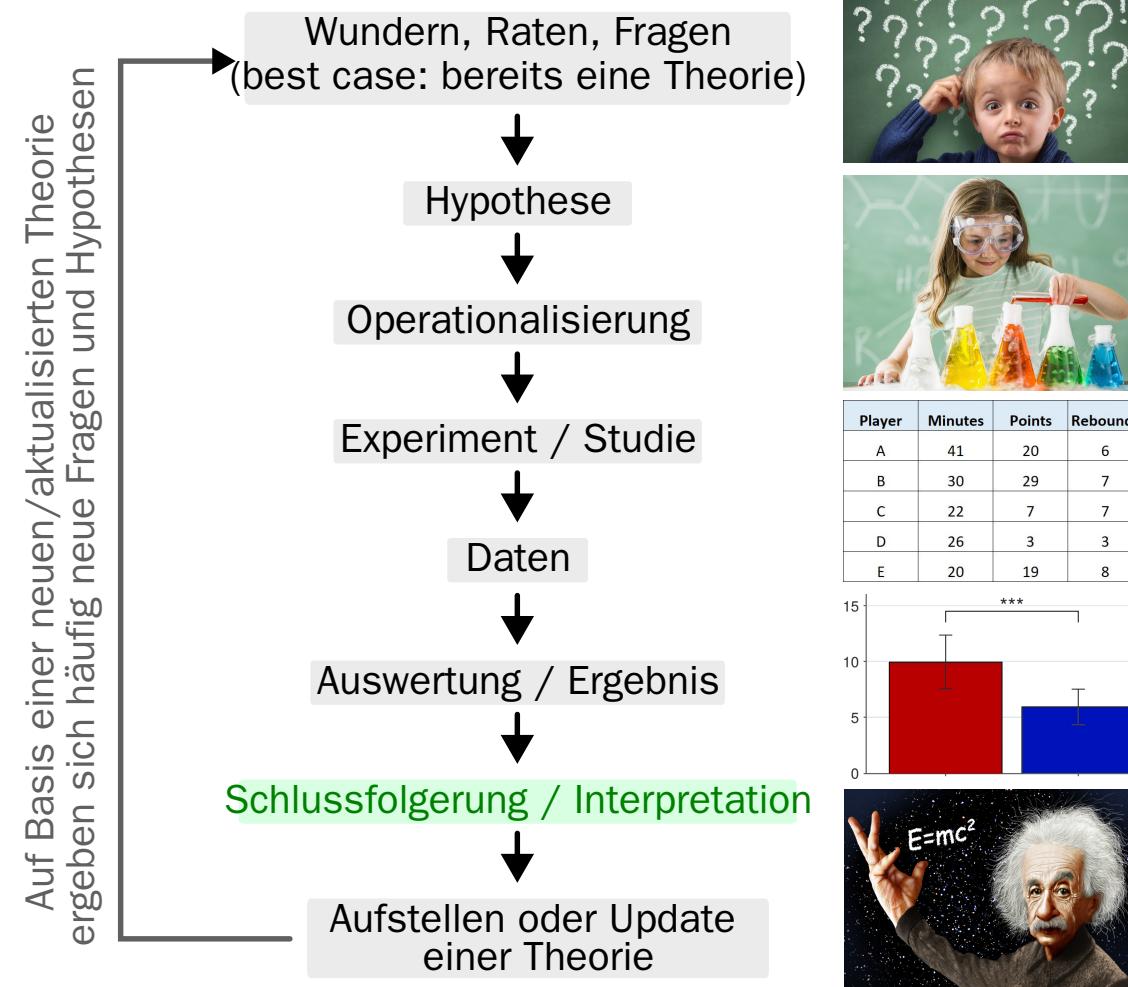


Sie haben nun die valide Signifikanz Ihrer Ergebnisse mithilfe des t-Tests bestimmt. Der Switch vom z- zum t-Test hat zur absurdum Situation geführt, dass trotz sehr ähnlicher p-Werte, ersterer den Entzündungseffekt als signifikant wertet, letzterer aber nicht.

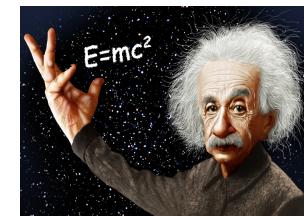
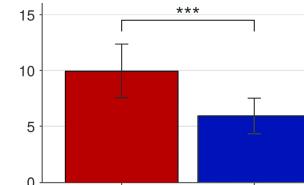
Um die Beurteilung der Signifikanz des Ergebnisses nicht einer Nachkommastelle zu überlassen, streben Sie für die Kommunikation der Ergebnisse gegenüber staatlichen Gremien und der Öffentlichkeit ein anderes Maß an.

Statt die Signifikanz mit Nein/Ja zu beurteilen, wollen Sie eine **graduelle Aussage** darüber treffen, wie vertrauenswürdig die jeweiligen Mittelwertschätzungen Ihrer Messungen sind. Diese Überlegung führt Sie zum **Konfidenzintervall**.

Der Forschungsprozess

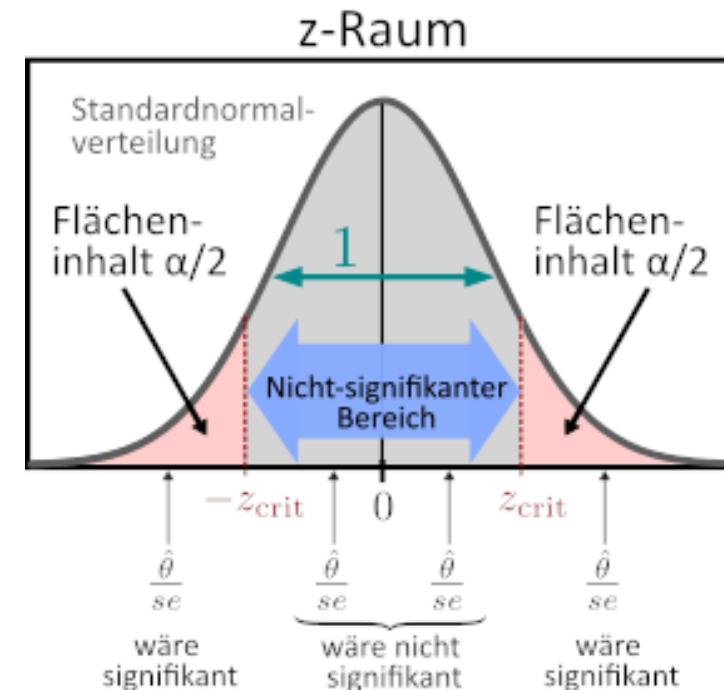


Player	Minutes	Points	Rebounds
A	41	20	6
B	30	29	7
C	22	7	7
D	26	3	3
E	20	19	8

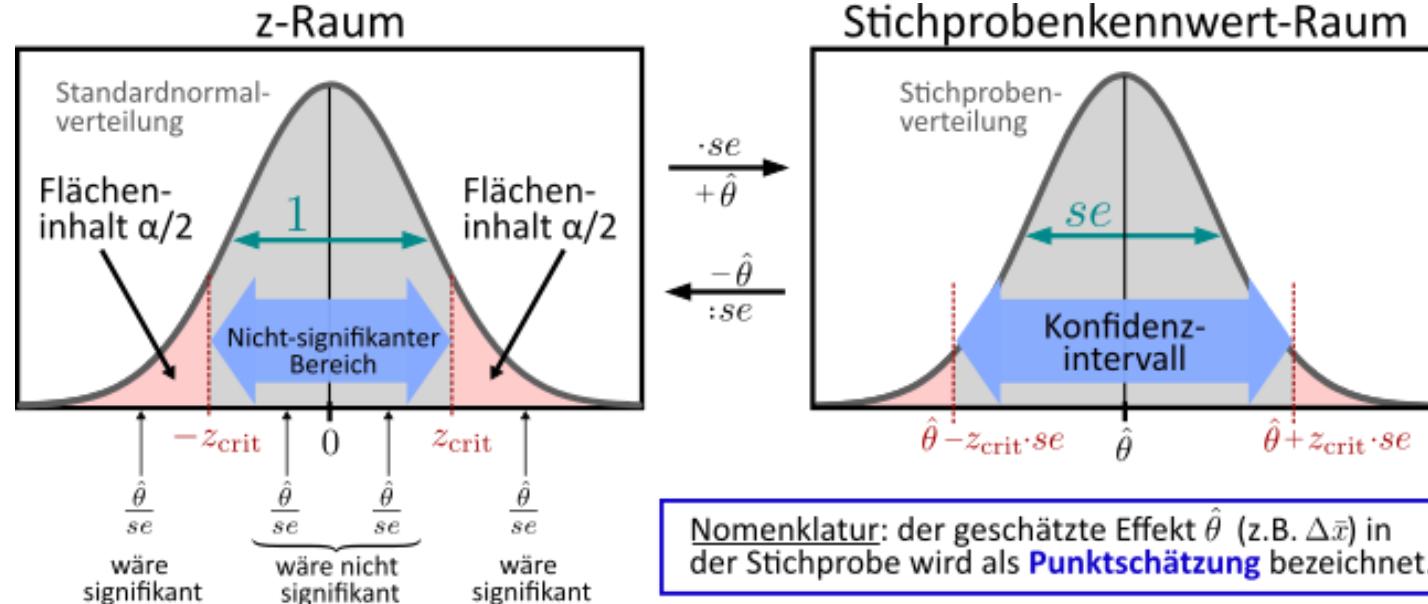


Rückblick

- Wir hatten das Konzept der Signifikanztestung anhand des **z-Tests** eingeführt.
- Anstatt die Signifikanz eines Effektes $\hat{\theta}$ (z.B. $\Delta\bar{x}, r$, usw.) unter der Stichprobenverteilung zu testen, standardisieren wir den Effekt zu $z = \frac{\hat{\theta} - \theta_0}{se}$ und testen ihn unter der Standardnormalverteilung (Hinweis: in den meisten Fällen ist $\theta_0 = 0$, weshalb wir $\theta_0 = 0$ ab jetzt weglassen).
- Für ein Signifikanzniveau α lassen sich die kritischen z-Werte $\pm z_{\text{crit}}$ bestimmen. Weist die empirische Prüfgröße $z = \frac{\hat{\theta}}{se}$ einen extremeren Wert als $\pm z_{\text{crit}}$ auf, ist der Effekt signifikant.
- Zugleich definiert das Signifikanzniveau ein Intervall $[-z_{\text{crit}}; z_{\text{crit}}]$, in dem der Effekt nicht signifikant wäre.
 - Dieses Intervall hängt nicht von den Daten $(\hat{\theta}, se)$ ab, sondern ausschließlich vom gewählten Signifikanzniveau α .
 - Das Intervall definiert also einen Bereich, in dem die Daten *unter der Nullhypothese* nicht als unwahrscheinlich gewertet werden (wobei das Kriterium wahrscheinlich/unwahrscheinlich durch das Signifikanzniveau α bestimmt wird).



Konfidenzintervall



- Projizieren wir den nicht-signifikanten Bereich der Standardnormalverteilung in den Raum der Stichprobenverteilung ($\cdot se, +\hat{\theta}$), so wird er **Konfidenzintervall** genannt und umfasst folgenden Wertebereich:

$$\text{Konfidenzintervall: } CI = \hat{\theta} \pm z_{\text{crit}} \cdot se$$

- Wir stellen zwei Punkte fest:

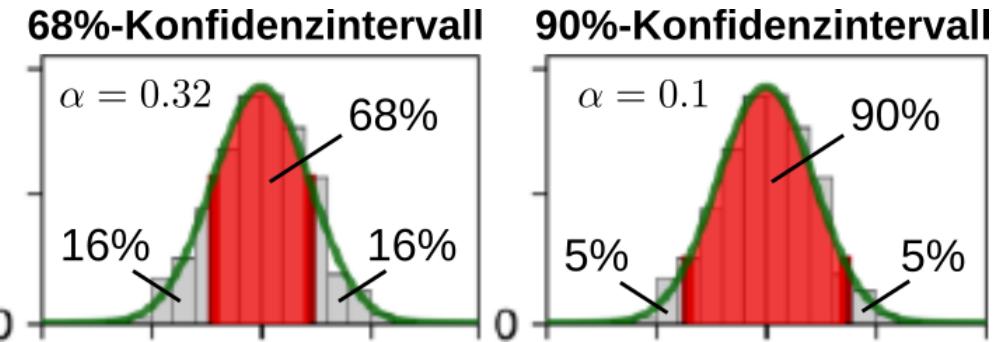
- Die Größe des Intervalls hängt jetzt von den Daten ab ($\hat{\theta}, se$).
- Das Konfidenzintervall umfasst $100 \cdot (1-\alpha)\%$ der Stichprobenverteilung (also zB 95% bei $\alpha = 0.05$).

Konfidenzniveau

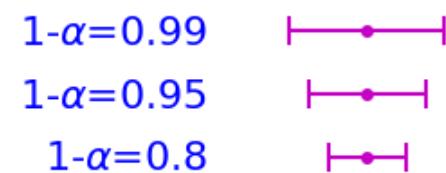
- Konfidenzintervalle können für verschiedene **Konfidenzniveaus** (auch “Überdeckungswahrscheinlichkeiten”) angegeben werden:

$$\text{Konfidenzniveau: } 1 - \alpha$$

- Das gängste Konfidenzniveau verwendet $\alpha = 0,05$ und wird dementsprechend als **95%-Konfidenzintervall** bezeichnet.
 - $\alpha = 0,01 \rightarrow 99\%-Kofidenzintervall$
 - $\alpha = 0,1 \rightarrow 90\%-Kofidenzintervall$
 - usw.
- Grundsätzlich gilt: je höher das zugrundegelegte Konfidenzniveau, desto größer auch das Konfidenzintervall.



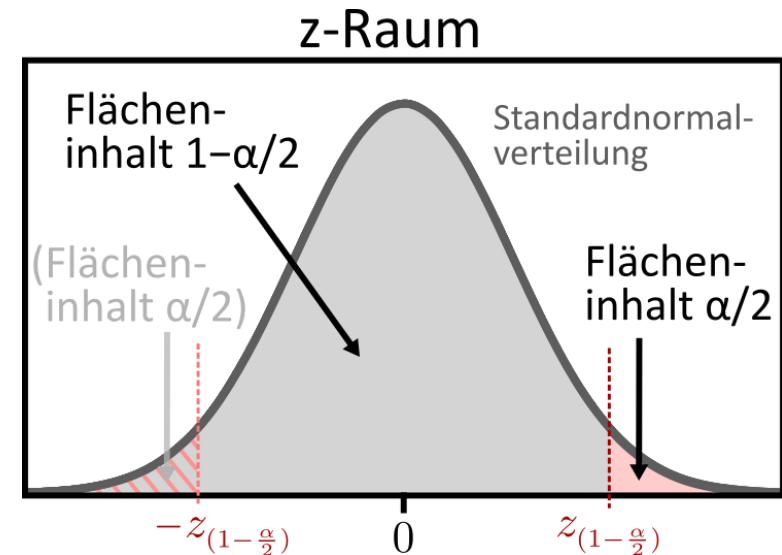
exercise had a moderate effect on college students' depressive mood (standardized mean difference = -0.63); the 95% CI was [-0.80, -0.46], which result was statistically significant. Finally, contrary to the con-
Beispiel für die Verwendung eines 95%-Konfidenzintervalls in einer
Publikation¹



Konfidenzniveau und kritischer z-Wert

- Jedes Konfidenzniveau $1 - \alpha$ ist mit einem kritischen z-Wert z_{crit} verbunden.
- Da Konfidenzintervalle immer zweiseitig sind, verteilt sich die Irrtumswahrscheinlichkeit α auf beide Flanken der Standardnormalverteilung. Deshalb hat sich auch folgende Nomenklatur für kritische z-Werte eingebürgert:

$$z_{\text{crit}} = z_{(1-\frac{\alpha}{2})}$$



- Zu jedem **Konfidenzniveau $1 - \alpha$** lässt sich also ein **kritischer z-Wert $z_{(1-\frac{\alpha}{2})}$** zuordnen.
- Damit haben wir folgende aktualisierte Form des Konfidenzintervalls:

Konfidenzintervall: $CI = \hat{\theta} \pm z_{(1-\frac{\alpha}{2})} \cdot se$

- Der kritische z-Wert wird wie bisher auf Basis der Verteilungsfunktion Φ der Standardnormalverteilung berechnet (\rightarrow z-Tabelle!):

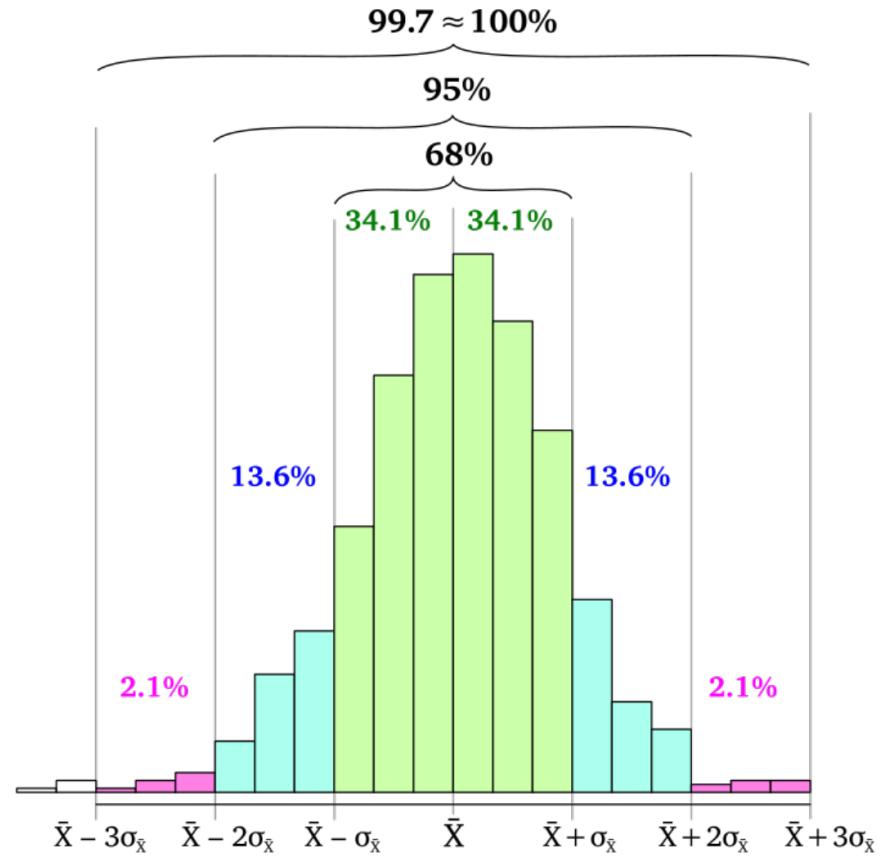
$$z_{(1-\frac{\alpha}{2})} = \Phi \left(1 - \frac{\alpha}{2} \right)$$

Sie kennen bereits ein Konfidenzintervall

Konfidenzintervall.. umfasst einen gewissen Prozentsatz der Stichprobenverteilung – does that ring a bell?

Auch der **Standardfehler**, als Streuungsmaß der Stichprobenverteilung, umfasst einen gewissen Bereich dieser Verteilung. Um genau zu sein entspricht der Standardfehler ja einfach der Standardabweichung der Stichprobenverteilung und umfasst damit gemäß der **68-95-99.7-Regel** ungefähr **68%** der **Stichprobenverteilung**.

In gewisser Weise ist der **Standardfehler** also ein recht **laxes Konfidenzintervall** mit einem Konfidenzniveau von lediglich 68%. In seiner Funktion als Konfidenzintervall findet der Standardfehler aufgrund des niedrigen Konfidenzniveaus jedoch kaum Verwendung.



Definition über die Signifikanz

Was genau hat das Konfidenzintervall mit dem Signifikanzniveau α zu tun, das wir bisher nur im Kontext der Nullhypothese betrachtet haben? Dies führt uns zu einer präziseren Definition von Konfidenzintervallen:

Definition Ein **Konfidenzintervall** definiert einen Bereich möglicher Populationsparameter θ , die nicht signifikant verschieden von der Punktschätzung $\hat{\theta}$ sind (auf Basis des Signifikanzniveaus α).

Einfacher gesprochen enthält das Konfidenzintervall also mögliche Populationsparameter θ , auf Basis derer die Punktschätzung $\hat{\theta}$ nicht unwahrscheinlich erscheinen würde.



Auch wenn das Konfidenzintervalls um die Punktschätzung herum konstruiert wird, kann ein Signifikanztest nie auf Basis einer Verteilung um die Punktschätzung durchgeführt werden. Wie bei der Nullhypothese basieren auch die (hypothetischen) Signifikanztests in der Definition des Konfidenzintervalls auf Verteilungen um fixe angenommene Werte der Population. Berechnet wird *immer* die Wahrscheinlichkeit der Stichprobendaten gegeben einer Annahme über den Populationsparameter θ (bei der Nullhypothese $\theta = 0$).

Intuition zu Konfidenzintervallen

Was bedeutet es, ob ein Konfidenzintervall größer oder klein ist? Auf Basis unserer ersten Definition können wir feststellen:

- Ist das **Konfidenzintervall** bei einem gegebenen α **verhältnismäßig klein**, so gäbe es auch nur einen kleinen Bereich möglicher Populationsparameter die kompatibel (im Sinn von “nicht signifikant verschieden”) mit unserer Punktschätzung sind.
 - Wir können in diesem Fall davon ausgehen, dass unsere **Punktschätzung recht präzise** ist.
- Ist das **Konfidenzintervall** (bei gleichem α) **verhältnismäßig groß**, so ist der Bereich kompatibler Populationsparameter größer.
 - In diesem Fall ist unsere **Punktschätzung mit mehr Unsicherheit behaftet**.



Dürfte man den Namen von Konfidenzintervall neu wählen, wäre **Kompatibilitätsintervall** ein sinnvoller(er) Begriff, da das Intervall im obigen Sinne eben genau **jene Populationsparameter einschließt, die kompatibel mit dem Stichprobenkennwert sind**.

Intuition zu Konfidenzintervallen

Die bisherigen Erkenntnisse führen uns zu einer Charakterisierung von Konfidenzintervallen, die **für die Praxis** ausreichend und zentral ist:

Das **Konfidenzintervall** bemisst die **Vertrauenswürdigkeit** einer **Punktschätzung** $\hat{\theta}$ (\bar{x} , $\Delta\bar{x}$, r , usw.) aus. **Je kleiner** das Intervall, **desto präziser** ist der Stichprobenschätzwert $\hat{\theta}$ in Bezug auf den wahren Parameters θ .

Konfidenzniveau

Es ist verführerisch anzunehmen, dass z.B. ein 95%-Konfidenzniveau bedeutet, dass der wahre Populationsparameter zu 95% innerhalb des Konfidenzintervalls liegt. **Dem ist nicht so.**

Wäre dem so, hätten wir auf magische Weise im Framework der Nullhypotesentestung die Möglichkeit eingeführt, die Wahrscheinlichkeit einer Hypothese zu bemessen (die Wahrscheinlichkeit einer Hypothese ist immer gleich der Wahrscheinlichkeit eines Parameterwertes oder -wertebereiches). **Dies bleibt im Nullhypotesenframework unmöglich.**

Der Populationskennwert ist ein fixer Wert und keine Zufallsvariable. Entweder liegt er in einem gegebenen Konfidenzintervall ($p=1$) oder nicht ($p=0$).

Die Crux liegt im Detail der Formulierung “in einem gegebenen Konfidenzintervall”. **Nach der Erhebung der Stichprobe** (also für ein gegebenes Konfidenzintervall) ist es wie erwähnt im Kontext der Nullhypotesentestung sinnlos eine Wahrscheinlichkeit anzugeben, dass der Populationskennwert in diesem Intervall liegt.

Anders ist die Situation **vor Erhebung einer Stichprobe** (d.h. bevor das Konfidenzintervall gegeben ist) – in diesem Fall können wir folgende Feststellung treffen:

Das Konfidenzniveau gibt die Wahrscheinlichkeit an, mit der das Konfidenzintervall einer zu erhebenden Stichprobe den wahren Populationsparameter einschließt.

Konfidenzniveau

Diese Feststellung ist zugegebenermaßen etwas sehr theoretisch – schließlich interessieren wir uns in der Praxis für die Interpretation des Konfidenzintervalls genau dann, wenn wir Daten erhoben haben und das Intervall berechnet haben.

Alternativen wir wenden daher das gleiche Prinzip einfach auf zukünftige hypothetische Stichprobenerhebungen nach der aktuellen Stichprobe an. Wir gehen dabei von der **Wahrscheinlichkeit** einer einzelnen (noch vor uns liegenden) Stichprobenerhebung zu einer **Frequenz** bei zukünftigen Stichprobenerhebungen über. Dies führt zur **Textbuchdefinition** des **Konfidenzniveaus**:

Definition

Das **Konfidenzniveau** gibt die relative Häufigkeit an, mit der Konfidenzintervalle zukünftiger Stichproben den wahren Populationsparameter einschließen.

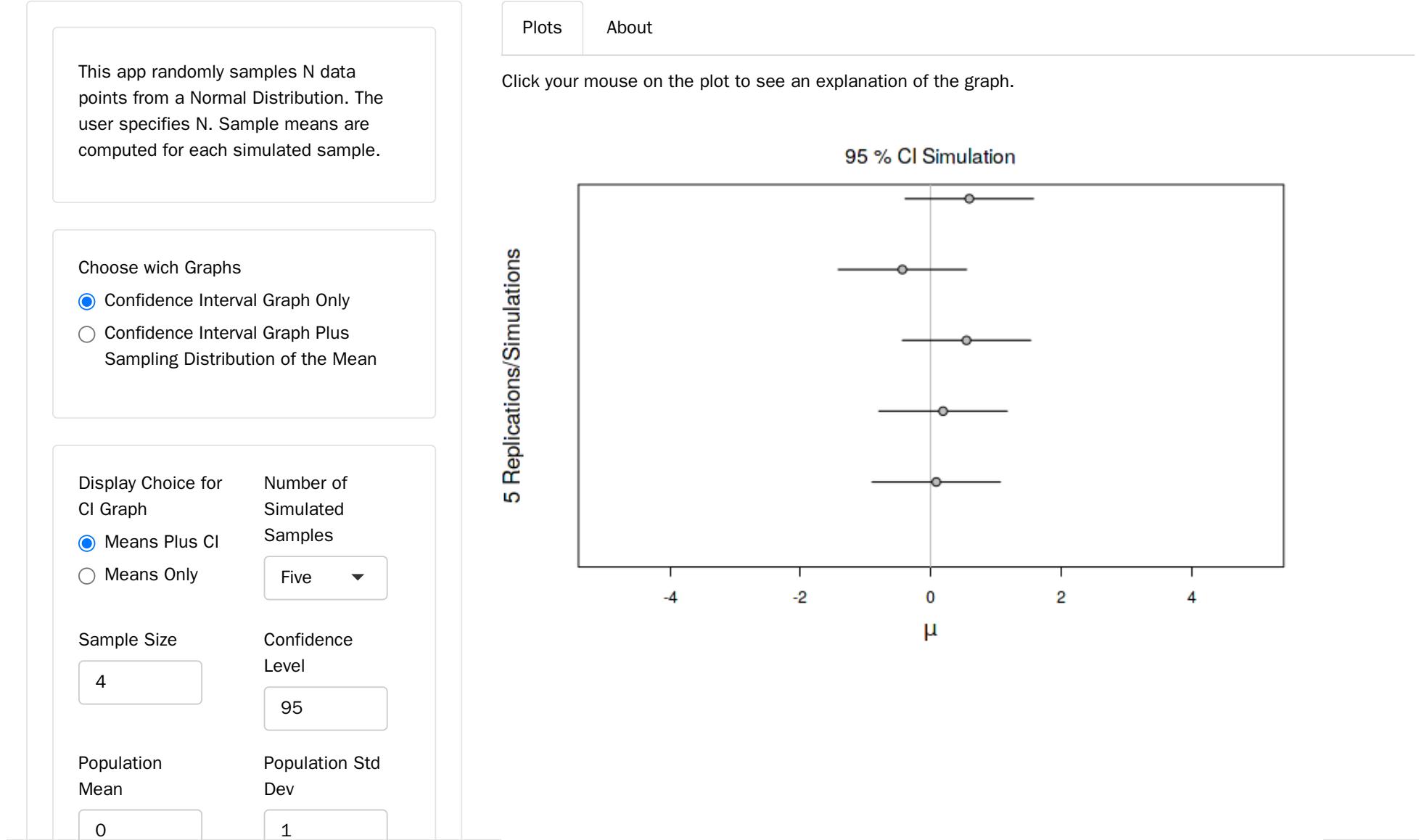
Bei einem 95%-Konfidenzintervall erwarten wir also, dass 95% der hypothetischen (zukünftigen) Stichproben ein Konfidenzintervall aufweisen, das den Populationsparameter einschließt.

Erscheint Ihnen die Definition des Konfidenzniveaus ähnlich unintuitiv wie die des Konfidenzintervalls selbst? Agreed. Es gilt die *Erste Regel von Konfidenzintervallen*:

Konfidenzintervalle sind **nützlich** aber **nicht intuitiv**.

<https://shiny.rit.albany.edu/stat/confidence/>

Confidence Interval for a Sample Mean: A simulation



Konfidenzintervall: what it is and what it isn't

Korrekte Aussagen zu einem 95%-Konfidenzintervall

Schließt 95% der Fläche unter der Stichprobenverteilung ein.



Auf lange Sicht werden 95% der erhobenen Stichproben ein Konfidenzintervall aufweisen, das den Populationsparameter einschließt.



Definiert einen Bereich möglicher Populationsparameter θ , die nicht signifikant verschieden von der Punktschätzung $\hat{\theta}$ sind (bei einem 5%-Signifikanzniveau).



Entspricht einem "Kompatibilitätsbereich" — Populationswerte innerhalb des Intervalls sind kompatibel mit dem Stichprobenkennwert.



Falsche Aussagen zu einem 95%-Konfidenzintervall

Enthält den Populationsparameter mit einer Wahrscheinlichkeit von 95%.



Enthält 95% der Datenpunkte der vorliegenden Stichprobe.



Schließt erwartbar 95% der zukünftigen Stichprobeneffekte ein.



Konfidenzintervall: ein Beispiel

Als Beispiel nehmen wir eine **Studie** an, in der Sie die **durchschnittliche Nasenlänge von Psychologiestudierenden** untersuchen. Als Referenzwert ist Ihnen der Bundesnasenlängendurchschnitt von $\mu_0 = 5\text{cm}$ bekannt – diese Referenz beachten wir aber zunächst nicht.

Sie erheben eine Stichprobe mit n Psychologiestudierenden und berechnen das Konfidenzintervall Ihrer Mittelwertsschätzung \bar{x} mit einem Konfidenzniveau von $1 - \alpha$.

Das Konfidenzintervall lautet auf diese Studie angepasst:

$$CI = \bar{x} \pm z_{(1-\frac{\alpha}{2})} \cdot \frac{\sigma}{\sqrt{n}}$$



Konfidenzintervall: ein Beispiel

Nehmen wir für eine erste Berechnung eines Konfidenzintervalls an, dass Sie eine Stichprobe von $n = 10$ untersuchen und Nasenlängen in der Psych-Population eine Streuung von $\sigma = 0,5$ aufweisen. Das Konfidenzniveau $1 - \alpha$ solle 95% sein (also $\alpha = 0,05$).

In einem ersten Schritt müssen wir dafür den kritischen z-Wert $z_{(1-\frac{\alpha}{2})}$ berechnen:

$$z_{(1-\frac{\alpha}{2})} = \Phi\left(1 - \frac{\alpha}{2}\right) = \Phi\left(1 - \frac{0,05}{2}\right) = \Phi(0,975) \stackrel{\substack{(z\text{-Tabelle)} \\ (\text{o. Computer})}}{=} 1,96$$

- Aufgrund der Beliebtheit des 95%-Konfidenzniveaus der assoziierte z-Wert **1,96** ein sehr wichtiger Wert, den zu kennen es sich lohnt.
- Der erhaltene Mittelwert in Ihrer Studie sei $\bar{x} = 4,5\text{cm}$.
- Wir setzen alle Werte ein, um das Konfidenzintervall für diese Mittelwertschätzung zu erhalten:

$$CI_{95} = 4,5 \pm 1,96 \cdot \frac{0,5}{\sqrt{10}} = 4,5 \pm 0,31 \quad \stackrel{(\text{Computer})}{\longrightarrow} \quad CI_{95} = [4,19; 4,81]$$

⇒ Der Mittelwert $\bar{x} = 4,5\text{cm}$ ist mit einem 95%-Konfidenzintervall $[4,19\text{cm}; 4,81\text{cm}]$ verbunden.



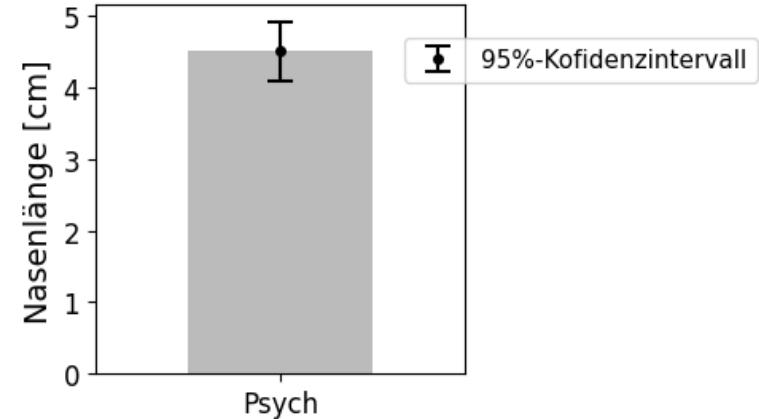
Konfidenzintervall: ein Beispiel

- Dieses Konfidenzintervall könnten wir nun zum Beispiel in einem Balkendiagramm als **Fehlerbalken** abtragen.
- Im Gegensatz zu Standardfehler/abweichung ist das Konfidenzintervall (wie der Name sagt) bereits ein Intervall, d.h. es kann ohne \pm -Rechnung direkt abgetragen werden.
 - Im Bild rechts sehen wir, dass die obere Kante des Fehlerbalkens der oberen Grenze des Konfidenzintervalls entspricht und die untere Kante der unteren Grenze.
- Im Fließtext wird es zum Beispiel so angegeben:

Die Messung der Nasenlängen ergab einen Mittelwert von $4,5\text{cm}$ ($CI_{95} = [4,19\text{cm}; 4,81\text{cm}]$).

- Oder ausführlicher (wenn Sie auf der Suche nach Worten für Ihre Bachelorarbeit sind):

Die Messung der Nasenlängen ergab einen Mittelwert von $4,5\text{cm}$. Das 95%-Konfidenzintervall der Mittelwertschätzung zeigte eine untere Grenze von $4,19\text{cm}$ und eine obere Grenze von $4,81\text{cm}$ an.



Einflussfaktoren auf das Konfidenzintervall

Für das Verständnis von Konfidenzintervallen ist es gewinnbringend zu untersuchen, wie das Konfidenzintervall von den relevanten Variablen einer Studie abhängt.

Aus der Formel

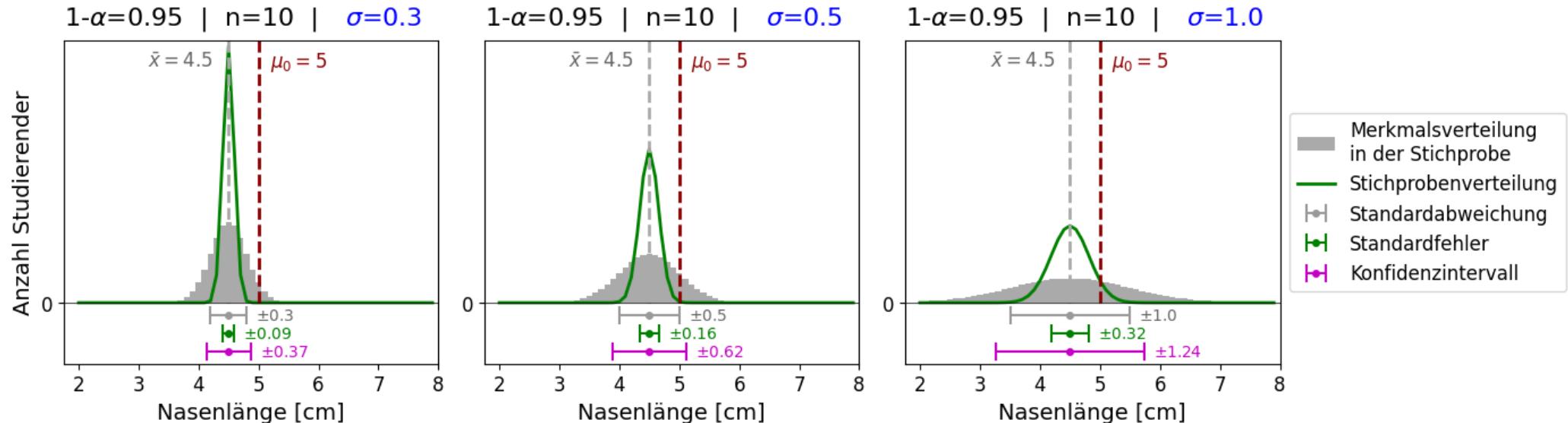
$$CI = \hat{\theta} \pm z_{(1-\frac{\alpha}{2})} \cdot se \quad \text{bzw.} \quad CI = \hat{\theta} \pm z_{(1-\frac{\alpha}{2})} \cdot \frac{\sigma}{\sqrt{n}}$$

ergeben sich direkt die **Einflussfaktoren auf das Konfidenzintervall**:

- Die **Streuung σ** in der Population
- Die **Stichprobengröße n**
- Das **Konfidenzniveau $1 - \alpha$** bzw. der damit verbundene kritische z-Wert $z_{(1-\frac{\alpha}{2})}$

Einflussfaktoren auf das Konfidenzintervall

Einflussfaktor 1: je größer die Streuung σ des Merkmals in der Population, desto größer ist das Konfidentintervall.

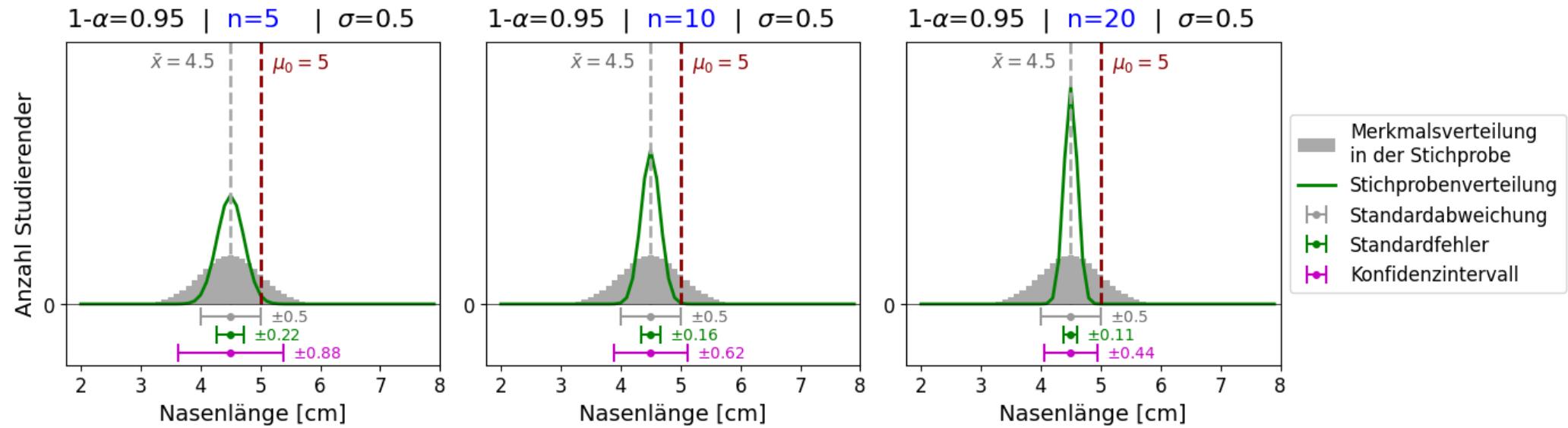


Interpretation: Je größer die Streuung in der Population, desto größer auch der Standardfehler, desto unsicherer ist man bezüglich der Schätzung des Mittelwertes, desto größer das Konfidenzintervall.

$$CI = \bar{x} \pm z_{(1-\frac{\alpha}{2})} \cdot \frac{\sigma}{\sqrt{n}}$$

Einflussfaktoren auf das Konfidenzintervall

Einflussfaktor 2: je größer die Stichprobe n , desto *kleiner* ist das Konfidenzintervall.

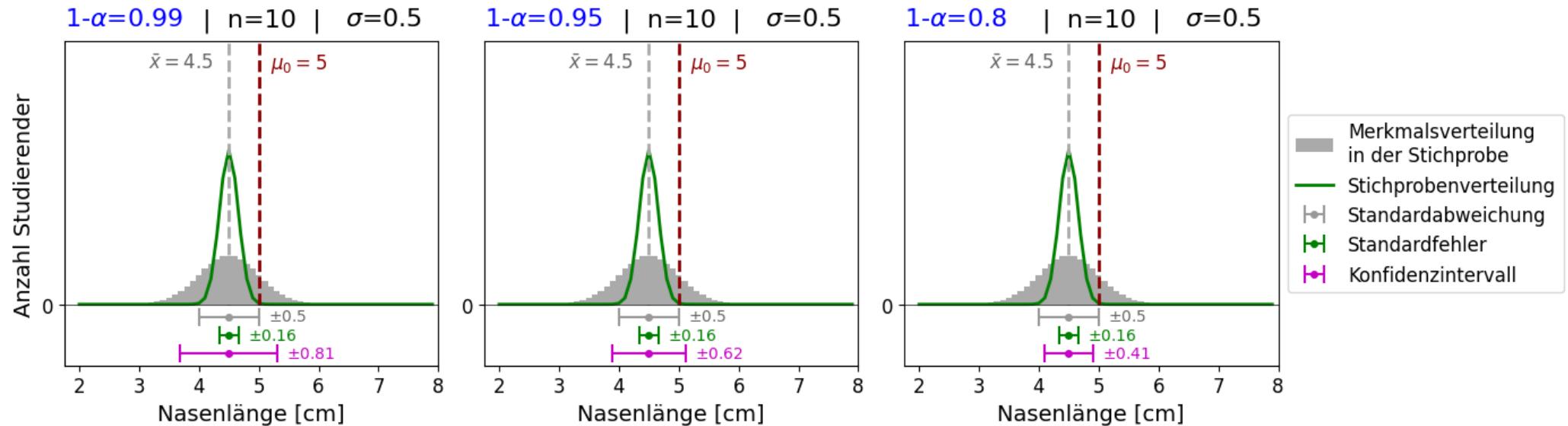


Interpretation: Je größer die Stichprobe, desto kleiner ist der Standardfehler, desto sicherer ist man bezüglich der Schätzung des Mittelwertes, desto kleiner das Konfidenzintervall.

$$CI = \bar{x} \pm z_{(1-\frac{\alpha}{2})} \cdot \frac{\sigma}{\sqrt{n}}$$

Einflussfaktoren auf das Konfidenzintervall

Einflussfaktor 3: je höher das Konfidenzniveau $1 - \alpha$, desto größer ist das Konfidenzintervall.



Interpretation:

Je größer das Konfidenzintervall $1 - \alpha$, desto kleiner das Signifikanzniveau α , desto größer ist das Intervall potentieller Populationsparameter θ , die nicht signifikant verschieden vom Stichprobenkennwert $\hat{\theta}$ sind.

$$CI = \Delta\bar{x} \pm z_{(1-\frac{\alpha}{2})} \cdot \frac{\sigma}{\sqrt{n}}$$

Alternativ, für Pragmatiker: je kleiner α , desto größer $z_{(1-\frac{\alpha}{2})}$, desto größer das Konfidenzintervall.

Konfidenzintervalle in der Praxis

Anwendungsfälle für Konfidenzintervalle

In der Praxis kommen Konfidenzintervallen zwei (subtil) unterschiedliche Funktionen zu.

Anwendung als Signifikanzmaß

Bei Effekten wie z.B. Mittelwertunterschieden ist Hauptinteresse idR, ob der Effekt von Null verschieden ist. In diesem Fall ist häufig die relevante Frage an das Konfidenzintervall, ob der Wert Null im Intervall enthalten ist – in diesem Fall wäre der Effekt nämlich signifikant auf Basis des α -Wertes des verwendeten Konfidenzniveaus.

“Der Gruppenunterschied betrug $\Delta\bar{x} = 0,2\text{cm}; CI_{95} = [-0,1\text{cm}; 0,4\text{cm}]$ ” \rightarrow impliziert einen nicht-signifikanten Effekt auf einem Signifikanzniveau $\alpha = 0,05$.

“Der Gruppenunterschied betrug $\Delta\bar{x} = 0,2\text{cm}; CI_{95} = [0,1\text{cm}; 0,3\text{cm}]$ ” \rightarrow impliziert einen signifikanten Effekt auf einem Signifikanzniveau $\alpha = 0,05$.

Anwendung als Streuungsmaß

Insbesondere im Fall von reinen Rohmaßen wie einem Mittelwert \bar{x} oder Modellparametern ist häufig nicht interessant, ob diese von Null verschieden sind; stattdessen interessiert uns die Genaugkeit unserer Parameterschätzung (“Wie sehr kann ich meinem Mittelwert \bar{x} vertrauen? Wie sicher ist die Schätzung meines Modellparameters?”).

In beiden Anwendungsfällen ist jedoch die Berechnung von Konfidenzintervallen exakt gleich – der Unterschied liegt lediglich im Auge des Betrachters bzw. im Ziel der Analyse.

Mittelwerte und Mittelwertunterschiede

- Für die Einführung der Konfidenzintervalle hatten wir implizit wieder **angenommen, dass die Populationsstreuung σ bekannt ist**, und daher z-Werte $z_{(1-\frac{\alpha}{2})}$ bestimmt werden können.
- Wir uns mittlerweile bekannt ist, ist dies **in der Praxis nahezu nie gegeben**.
- Bei Mittelwerten \bar{x} und Mittelwertunterschieden $\Delta\bar{x}$ ist dies ein Problem, denn die Annahme einer Normalverteilung ist nicht mehr zulässig (bzw. nur annähernd bei größeren Stichproben).
- Die Abhilfe für diese Problematik ist allerdings denkbar einfach — wir **ersetzen die kritischen z-Werte einfach durch kritische t-Werte**.

Konfidenzintervalle für Mittelwerte:

$$CI = \bar{x} \pm t_{(1-\frac{\alpha}{2}, \text{ df})} \cdot \frac{\hat{\sigma}}{\sqrt{n}}$$

(Beachte auch den Übergang von σ zu $\hat{\sigma}$, da wir nun die Streuung auf Basis der Stichprobe schätzen)

- Die Berechnung der Streuung σ ist beim einfach Mittelwert denkbar einfach, da es nur eine Stichprobe und nur eine Messung gibt.

Mittelwerte und Mittelwertunterschiede

Konfidenzintervalle für Mittelwertunterschiede:

Die Formel für das Konfidenzintervall von Mittelwertunterschieden $\Delta\bar{x}$ lautet analog:

$$CI = \Delta\bar{x} \pm t_{(1-\frac{\alpha}{2})} \cdot \hat{se}$$

bzw.

$$CI = \Delta\bar{x} \pm t_{(1-\frac{\alpha}{2})} \cdot \frac{\hat{\sigma}}{\sqrt{n}}$$

- Für die Berechnung der Populationsstreuung $\hat{\sigma}$ bzw. des Standardfehlers \hat{se} gelten alle Formeln (inkl. Regeln zu Freiheitsgraden der t-Verteilung) die wir im Kontext des t-Tests kennengelernt haben.
- Auch hier hängt der Standardfehler davon ab, ob es sich um Messungen in verschiedenen Gruppen handelt, ob die Varianzen ähnlich sind, usw.

Mittelwerte und Mittelwertunterschiede

Mittelwertunterschiede: Beispiel aus der Praxis

Bei der Darstellung von Mittelwertunterschieden kann es sinnvoll sein, zusätzlich zum Konfidenzintervall des Unterschiedes auch die Konfidenzintervalle der einzelnen Mittelwerte zu zeigen.

Konfidenzintervall je Mittelwert

Konfidenzintervall für den Unterschied

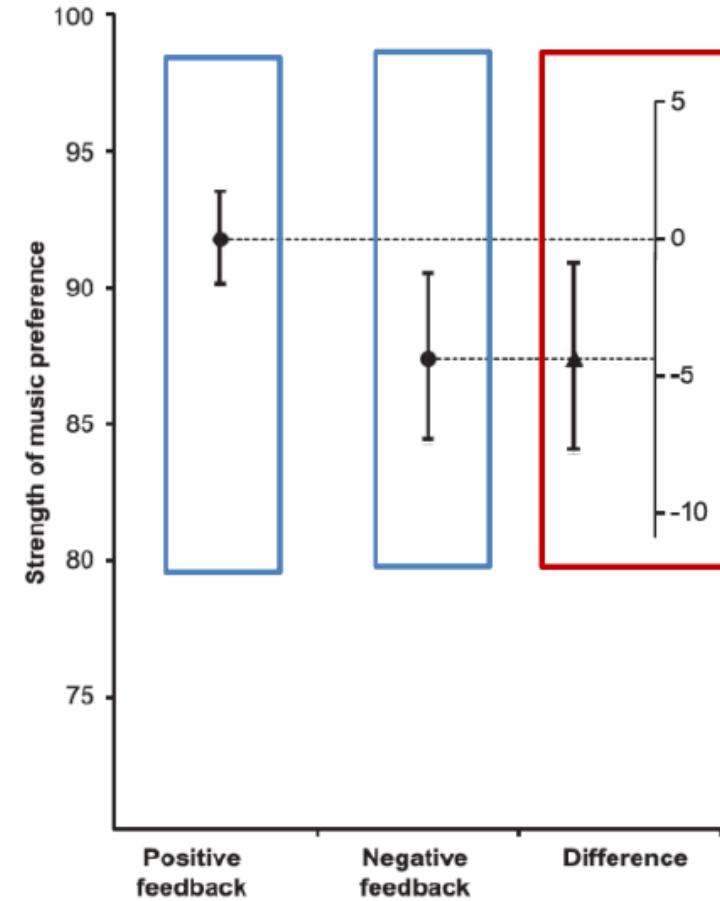
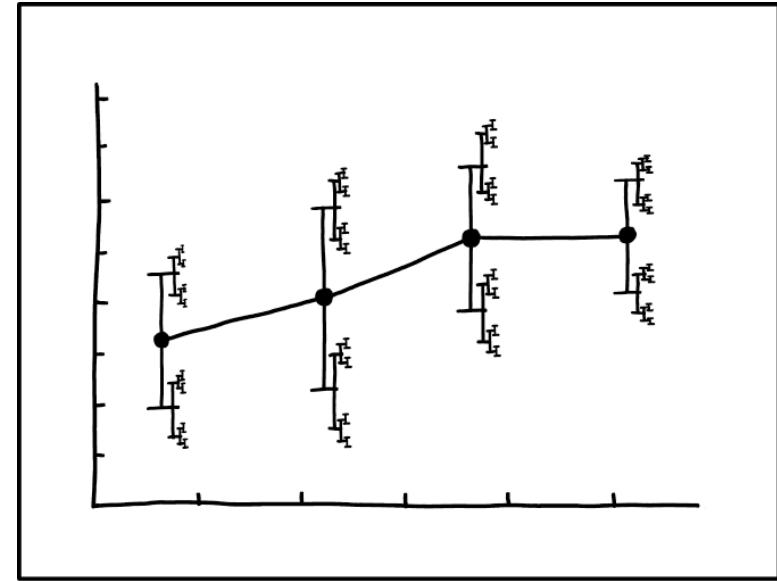


Figure 1. Means and 95% confidence intervals (CIs) for the positive and negative social feedback conditions in Session 3, together with the mean difference between these conditions and its 95% CI on a floating axis (see Cumming, 2013, for a detailed explanation of CIs of effects).

Konfidenzintervalle für Streuungsmaße

- Auch für **Streuungsmaße** wie die Varianz oder Standardabweichung können Konfidenzintervalle berechnet werden – allerdings tritt dieser Fall wesentlich seltener auf.
- Grund: Streuungsmaße sind **selten Gegenstand wissenschaftlicher Fragestellungen**.
- Ein wichtige Ausnahme, bei der die Streuung von Interesse ist, ist die Varianz von Mittelwerten im Rahmen einer **Varianzanalyse** (→ Statistik 2!).



I DON'T KNOW HOW TO PROPAGATE
ERROR CORRECTLY, SO I JUST PUT
ERROR BARS ON ALL MY ERROR BARS.

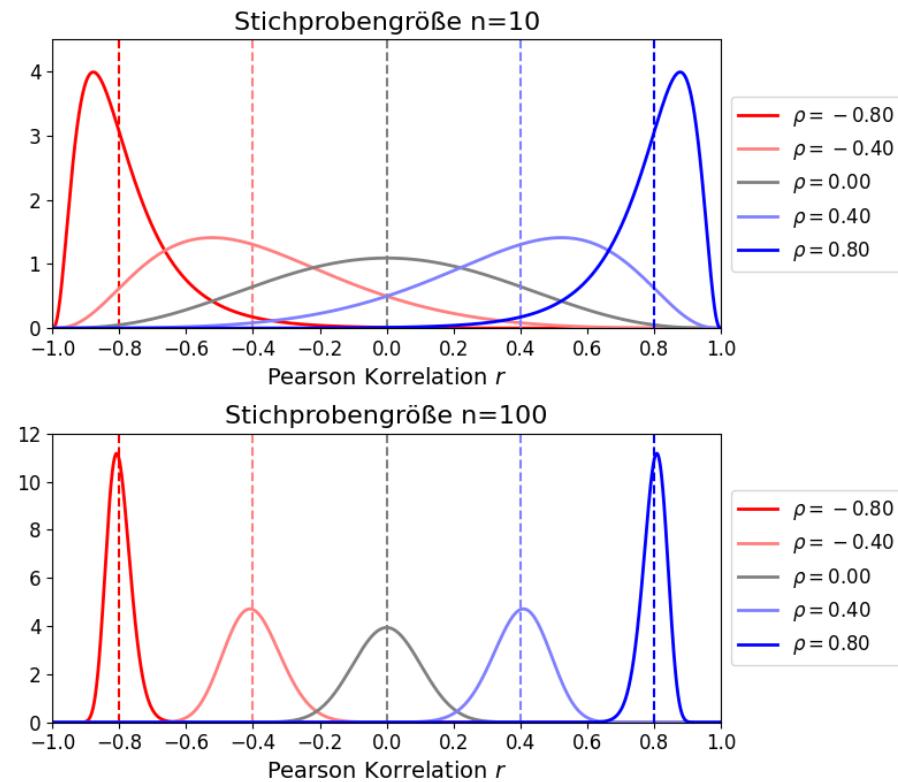
xkcd#2110²

Konfidenzintervalle für die Korrelation

- Ein häufiger Anwendungsfall für Konfidenzintervalle sind Zusammenhangsmaße wie der Pearson-Korrelationskoeffizient r .
- Den dafür notwendigen Standardfehler haben wir bereits kennengelernt:

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

- Wie aber sieht die Stichprobenverteilung des Korrelationskoeffizienten aus?
- Antwort: kompliziert³. Grund: der Korrelationskoeffizient ist auf die Grenzen -1 und $+1$ beschränkt, was insbesondere für Populationskorrelationen ρ nahe ± 1 zu einer stark asymmetrischen Stichprobenverteilung führt.
- Die Annahme einer Normalverteilung ist hier mit Sicherheit nicht sinnvoll. Was also tun?

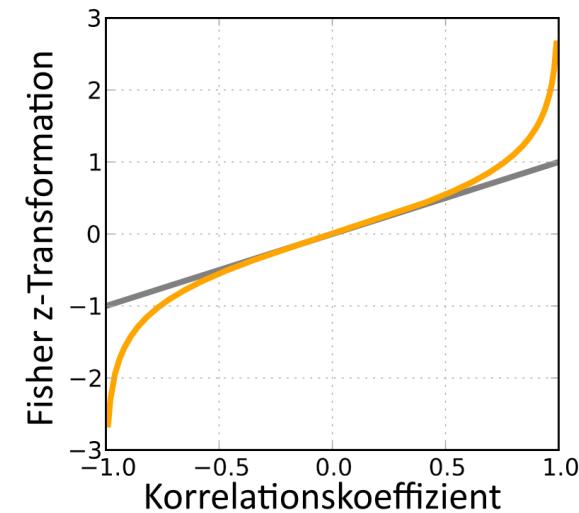


Konfidenzintervalle für die Korrelation

- Die Lösung für diese Problematik stammt von Ronald Fisher: die Korrelationskoeffizienten werden **z-transformiert** (nicht zu verwechseln mit dem z-Wert!)

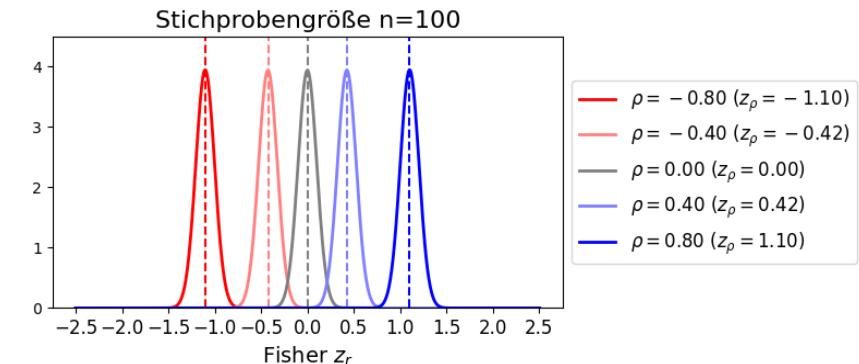
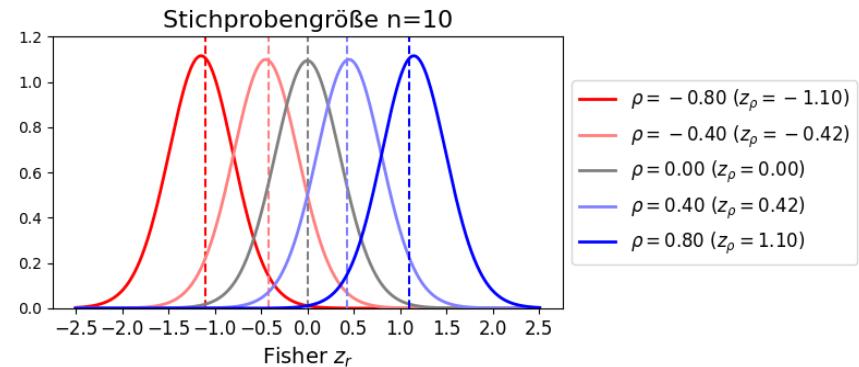
$$\text{Fisher z-Transformation: } z_r = \operatorname{artanh}(r)$$

(Hinweis: artanh ist der *Areatangens hyperbolicus*)



- Nach der z-Transformation ist die Stichprobenverteilung ungefähr normalverteilt.
- Der Standardfehler ist jetzt nicht mehr von Korrelation abhängig, sondern für jede Populationskorrelation gleich:

$$\hat{se} = \frac{1}{\sqrt{n-3}}$$



Konfidenzintervalle für die Korrelation

Die gesamte Prozedur für Konfidenzintervallen von Korrelationskoeffizienten ist wie folgt:

1. Fisher z-Transformation

Transformation des Korrelationskoeffizienten zu Fisher z_r :

$$z_r = \operatorname{artanh}(r)$$

2. Berechnung des Konfidenzintervalls

Aufgrund der asymptotischen Normalverteilung kann das Konfidenzintervall für z_r mithilfe der z-Werte der Standardnormalverteilung (z.B. aus der z-Tabelle) angegeben werden:

$$CI_z = z_r \pm z_{(1-\frac{\alpha}{2})} \cdot \hat{se}$$

bzw. nach Einsetzen des Standardfehlers:

$$CI_z = z_r \pm z_{(1-\frac{\alpha}{2})} \cdot \frac{1}{\sqrt{n-3}}$$

3. Rücktransformation in Korrelation

Sind z_r^u und z_r^o die untere und obere Grenze des Konfidenzintervalls von z_r aus Schritt 2, so lässt sich dieses mithilfe des *Tangens Hyperbolicus* in eine Korrelation zurücktransformieren:

$$CI = [\tanh(z_r^u); \tanh(z_r^o)]$$

Konfidenzintervalle für die Korrelation: Beispiel

- Gehen wir zurück zu unserem Beispiel aus der letzten Vorlesung, bei dem Sie in einer Stichprobe $n = 12$ eine Korrelation zwischen Größe und Gewicht von $r = 0,4$ gemessen hatten.
- Ein Signifikanztest hatte gezeigt, dass die Korrelation von Null nicht signifikant verschieden war (Signifikanzniveau $\alpha = 0,05$). Der Test war sogar einseitig und wir erwarten daher, dass das 95%-Konfidenzintervall den Wert Null erst recht einschließen muss.

1. Fisher z-Transformation

$$z_r = \text{artanh}(r) = \text{artanh}(0,4) = 0,424$$

2. Berechnung des Konfidenzintervalls

Standardfehler: $\hat{se} = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{12-3}} = \frac{1}{3}$

Kritischer z-Wert (z.B. aus z-Tabelle): $z_{(1-\frac{\alpha}{2})} = z_{(1-\frac{0,05}{2})} = z_{(0,975)} = 1,96$

Konfidenzintervall:

$$CI_z = z_r \pm z_{(1-\frac{\alpha}{2})} \cdot \hat{se} = 0,424 \pm 1,96 \cdot \frac{1}{3} = [-0,230; 1,077]$$

Konfidenzintervalle für die Korrelation: Beispiel

3. Rücktransformation in Korrelation

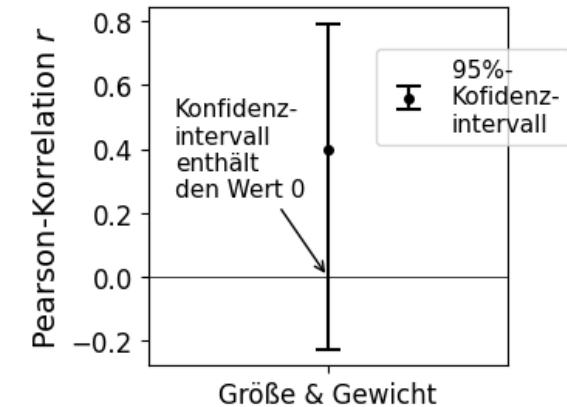
Mithilfe des *Tangens Hyperbolicus* transformieren wir das Konfidenzintervall CI_z zurück in Korrelationswerte:

$$CI = [\tanh(z_r^u); \tanh(z_r^o)] = [\tanh(-0,230); \tanh(1,077)] = [-0,226; 0,792]$$

Das Konfidenzintervall für unsere Korrelation $r = 0,4$ lautet also $[-0,226; 0,792]$. Das ist ein sehr großer Bereich, der sowohl negative als auch positive Werte umfasst.

Damit ist insbesondere auch der Wert 0 im Konfidenzintervall enthalten und wir können direkt ablesen, dass ein Signifikanztest mit $\alpha = 0,05$ nicht signifikant wäre.

Warum? Weil das Konfidenzintervall jene Korrelationsbereiche umfasst, in denen die Populationskorrelation ρ nicht signifikant verschieden von unserem Stichprobenwert r ist. Insbesondere ist also auch der Wert $\rho = 0$ nicht signifikant verschieden von unserem Wert $r = 0,4$ und damit die Nullhypothese **nicht abgelehnt**.



Konfidenzintervalle für Regressionskoeffizienten

- Im Gegensatz zu Korrelationskoeffizienten sind Regressionskoeffizienten b_i nicht auf das Intervall $[0; 1]$ beschränkt.
- In diesem Fall gilt die gleiche Logik wie bei Mittelwert(unterschied)en: ist der Standardfehler bekannt kann der z-Wert verwendet werden, *im allgemeinen Fall* jedoch sollte der t-Wert verwendet werden:

$$CI = b_i \pm t_{(1-\frac{\alpha}{2}, n-2)} \cdot \hat{se}$$

(Wir erinnern uns: für Korrelation und Steigung müssen 2 Mittelwerte berechnet werden (die von x und y), daher wird die t-Verteilung mit $df = n - 2$ Freiheitsgraden angewandt.)

Im Fall der Steigung (b_1) gilt:

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

Konfidenzintervalle für Regressionssteigung: graphische Darstellung

- Moderne Statistikprogramme können die Unsicherheit die mit der Steigung der einfachen Regression verbunden ist, auch graphisch zum Ausdruck bringen.
- Um das illustrieren, verwenden wir erneut das Größe-Gewicht-Beispiel.
- Wir nehmen als Standardabweichungen an $\hat{\sigma}_{\text{Größe}} = 16,5$ und $\hat{\sigma}_{\text{Gewicht}} = 6,8$. Damit ergibt sich die Steigung b_1 und der zugehörige Standardfehler:

$$b_1 = \frac{\hat{\sigma}_{\text{Gewicht}}}{\hat{\sigma}_{\text{Größe}}} \cdot r = \frac{6,8}{16,5} \cdot 0,4 \approx 0,16$$

$$\hat{s}e(b_1) = \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} \sqrt{\frac{1 - r^2}{n - 2}} = \frac{6,8}{16,5} \sqrt{\frac{1 - 0,4^2}{12 - 2}} \approx 0,12$$

- Der kritische t-Wert lautet:

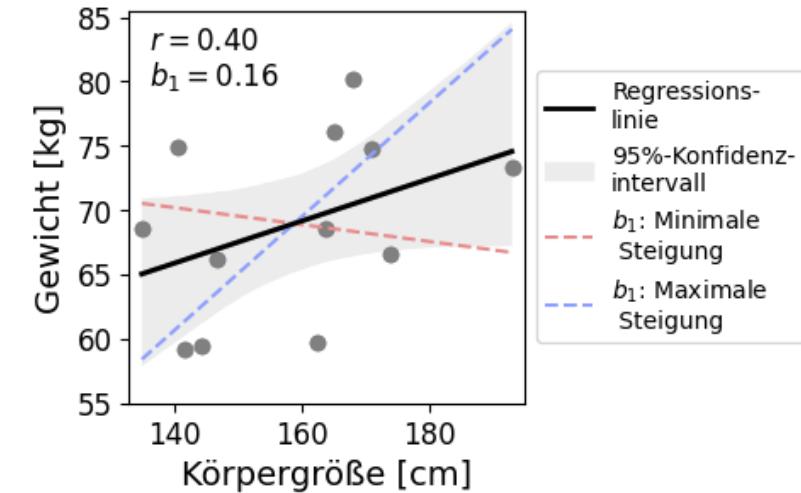
$$t_{(1 - \frac{\alpha}{2}, n-2)} = t_{(0.975, 10)} \approx 2,23$$

- Damit ergibt sich das Konfidenzintervall zu

$$CI = b_1 \pm t_{(1 - \frac{\alpha}{2}, n-2)} \cdot \hat{s}e = 0,16 \pm 2,23 \cdot 0,12 = 0,16 \pm 0,27 = [-0,11; 0,43]$$

Konfidenzintervalle für Regressionssteigung: graphische Darstellung

- Die maximale Steigung innerhalb des 95%-Konfidenzintervalls beträgt also $-0,11$ und die minimale Steigung $0,43$.
- In Regressionsplots wird der Bereich *aller mit dem Konfidenzintervall kompatiblen Steigungen* häufig als farblich hinterlegte Fläche angedeutet (hellgrau in der Abbildung rechts).



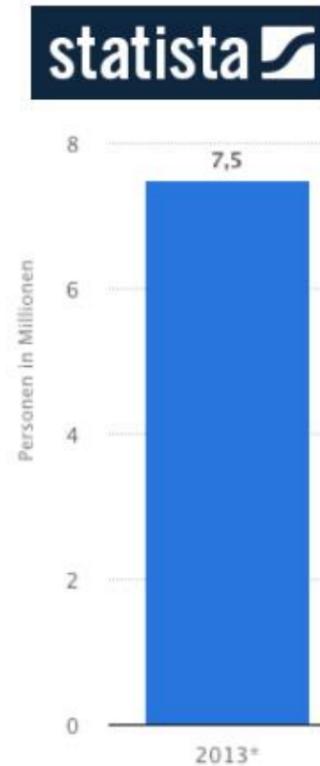
Beispiel für einen Regressionsplot mit 95%-Konfidenzintervall (hellgrau). Die minimale (rot) und maximale (blau) mit dem Intervall kompatible Regressionsgerade wird idR nicht eingetragen und wird hier nur zur Veranschaulichung gezeigt.

Konfidenzintervalle für Anteile / relative Häufigkeiten

Beispiel: Wie viele Menschen in Deutschland ernähren sich vegetarisch? Statista liefert einen Anteil von 7,5 % (Grundlage 25.677 Befragte). Das RKI liefert einen Anteil von 4 % (Grundlage 6.933 Befragte).



Für die Auswertung nutzten die RKI-Forscher Daten von 6933 Menschen, deren Essverhalten zwischen 2008 und 2011 analysiert worden war. Dabei wurde jeweils



2013: 25.677 Befragte, Hochrechnung auf 70,33 Mio. Personen

Können wir für die beiden Angaben ein Konfidenzintervall angeben?

Konfidenzintervalle für Anteile / relative Häufigkeiten

- **Relative Häufigkeiten (Anteile)** setzen eine Teilmenge k , die ein bestimmtes Kriterium erfüllt, ins Verhältnis zur Gesamtmenge n :

$$\hat{p} = \frac{k}{n}$$

- Das $\hat{\cdot}$ bringt hier zum Ausdruck, dass wir in der Regel nicht die gesamte Population N untersuchen, sondern eine Stichprobe der Größe n und daher der gemessene Anteil \hat{p} nur eine Schätzung für den wahren Anteil p in der Population darstellt.
- Im Vegetarier-Beispiel wäre interessant, mit welcher Unsicherheit die Prozentschätzungen behaftet sind → Konfidenzintervall!
- Die Stichprobenverteilung von relativen Häufigkeiten p ist eine **Binomialverteilung**. Ihr Standardfehler ist bekannt:

$$\hat{s}_e(\hat{p}) = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \quad (\text{Achtung: } p \text{ steht hier für den Anteil ("proportion") und nicht den p-Wert!})$$

- Der Standardfehler wäre ein erstes mögliches Maß für die Unsicherheit unserer Schätzung (zur Erinnerung: der Standardfehler ist das 68%-Konfidenzintervall)

Konfidenzintervalle für Anteile / relative Häufigkeiten

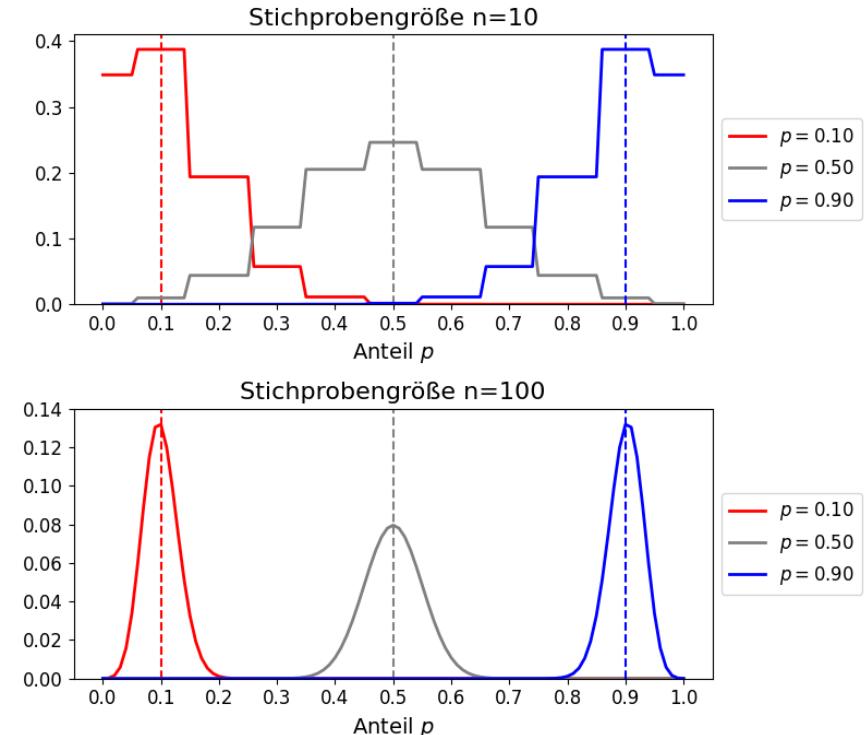
- Ähnlich wie der Korrelationskoeffizient sind Anteile auf einen Bereich eingeschränkt (0 bis 1 bzw. 0 bis 100%) und weisen daher eine **asymmetrische Stichprobenverteilung** auf.
- Sind die Stichprobengrößen n nicht zu klein und die Anteile nicht zu extrem (zu nah bei 0 oder 1), kann die Binomialverteilung jedoch durch die **Normalverteilung approximiert** werden.
- In diesem Fall können z-Werte für die Konstruktion des Konfidenzintervalls verwendet werden^{4 5}:

$$CI(\hat{p}) = \hat{p} \pm z_{(1-\frac{\alpha}{2})} \cdot \hat{s}e(\hat{p})$$

- Im Computerzeitalter sollten allerdings für die Konfidenzintervalle von Anteilen präzisere Schätzungen wie das **Wilson-Intervall**⁶ verwendet werden.



Beachte: bei präziseren Schätzmethoden wie dem Wilson-Intervall ist der wahrscheinlichste Populationswert (und damit der Ankerpunkt für das Konfidenzintervall) nicht mehr \hat{p} sondern ein etwas kleinerer (bei $\hat{p} < 0.5$) bzw. größerer (bei $\hat{p} > 0.5$) Wert. Zudem handelt es sich um asymmetrischer Konfidenzintervalle.



Stichprobenverteilung für Anteile (Binomialverteilung) für Stichproben der Größe $n = 10$ und $n = 100$.

Konfidenzintervalle für Anteile / relative Häufigkeiten

- In unserem Vegetarierbeispiel sind die Fallzahlen sehr groß ($n = 6933$ und $n = 25.677$), so dass eine Näherung mit der Normalverteilung angemessen ist. Wir können diese daher zur Einschätzung der (Un)Sicherheit der Prozentangaben nutzen.
- Für die Standardfehler erhalten wir:

$$\hat{se}_{\text{RKI}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = \sqrt{\frac{0,04(1 - 0,04)}{6933}} = 0,0024$$

$$\hat{se}_{\text{Statista}} = \sqrt{\frac{0,075(1 - 0,075)}{25677}} = 0,0016$$

- Den kritischen z-Wert für ein 95%-Intervall kennen wir mittlerweile aus dem FF: 1,96
- Damit ergibt sich für die Konfidenzintervalle:

$$CI_{\text{RKI}} = \hat{p} \pm z_{(1-\frac{\alpha}{2})} \cdot \hat{se}(\hat{p}) = 0,04 \pm 1,96 \cdot 0,0024 = [0,035; 0,045]$$

$$CI_{\text{Statista}} = 0,075 \pm 1,96 \cdot 0,0016 = [0,072; 0,078]$$

- Beide Schätzungen sind also recht präzise – und widersprechen sich!

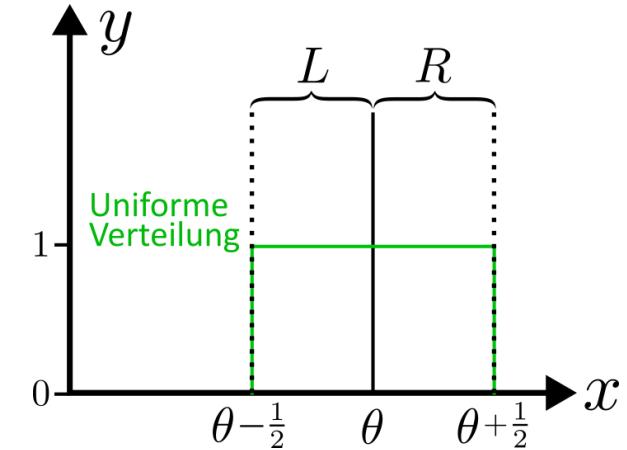
Konfidenzintervall-Beispiel von Welch



- Auf den Statistiker Bernard Welch (“Welch’s t-Test”) geht ein berühmtes Beispiel zurück, das eindrucksvoll zeigt, warum die Interpretation des Konfidenzniveaus als “*Wahrscheinlichkeit, dass der Populationsparameter im Intervall liegt*” nur vor der Beobachtung der Daten zulässig ist.
- Es sei bekannt, dass eine Zufallsvariable X einer uniformen Verteilung folgt, die im Intervall $[\theta - \frac{1}{2}; \theta + \frac{1}{2}]$ angesiedelt ist.
- Der genaue “Shift” θ der Verteilung auf der x-Achse ist unbekannt und soll durch Messen von X bestimmt werden.
- Tatsächlich werden nur zwei Messwerte von X erhoben, x_1 & x_2 .

Vor der Messung:

- Bei exakt zwei durchzuführenden Messungen x_1 und x_2 wissen wir, dass das Intervall $[\min(x_1, x_2); \max(x_1, x_2)]$ ein 50%-Konfidenzintervall für θ darstellt.
- Warum? Der Populationsparameter θ befindet sich *genau dann* in diesem Intervall, wenn x_1 und x_2 auf unterschiedlichen Seiten von θ liegen. Diese Wahrscheinlichkeit ist genau 50%.
 - Warum? Die Wahrscheinlichkeit, dass x_1 links von θ (also im Intervall L) liegt ist $\frac{1}{2}$, ebenso die Wahrscheinlichkeit, dass x_2 rechts von θ (also im Intervall R) liegt. Die Gesamtwahrscheinlichkeit für diesen Fall ist also $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Der umgekehrte Fall, dass x_1 rechts von θ und x_2 links von θ liegt hat ebenfalls die Wahrscheinlichkeit $\frac{1}{4}$. Die kombinierte Wahrscheinlichkeit beider Fälle ist daher $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ oder 50%.



Konfidenzintervall-Beispiel von Welch



Nach der Messung:

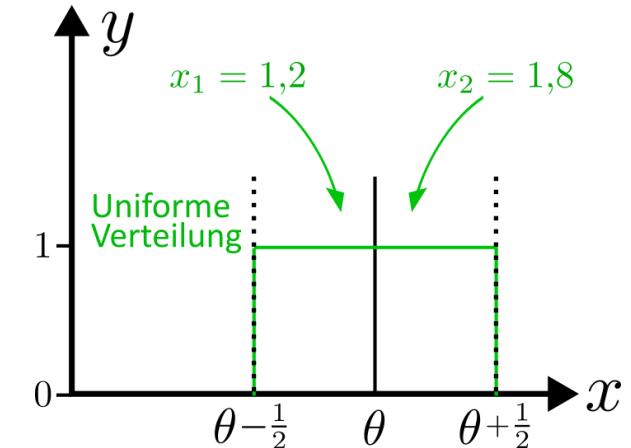
Nehmen wir an, dass wir nun zwei Messwerte x_1 und x_2 erhoben haben und außerdem, dass diese weiter als 0,5 auseinanderliegen ($|x_1 - x_2| > 0,5$). Um es konkret zu machen, sagen wir $x_1 = 1,2$ und $x_2 = 1,8$.

In dieser Situation wissen wir nun mit **100% Sicherheit**, dass θ in unserem Intervall

$$CI = [\min(x_1, x_2); \max(x_1, x_2)] = [1,2; 1,8]$$

liegt. Warum? θ kann nicht kleiner als 1,3 sein, denn sonst wäre es nicht möglich gewesen, einen Messwert 1,8 zu erhalten (beachte, dass der Bereich rechts und links von θ nur 0,5 umfasst!). Ebenso kann θ nicht größer als 1,7 sein, denn sonst wäre es nicht möglich gewesen, einen Messwert von 1,2 zu erhalten.

⇒ die Wahrscheinlichkeit, dass θ im Konfidenzintervall $[1,2; 1,8]$ liegt, ist 100%.



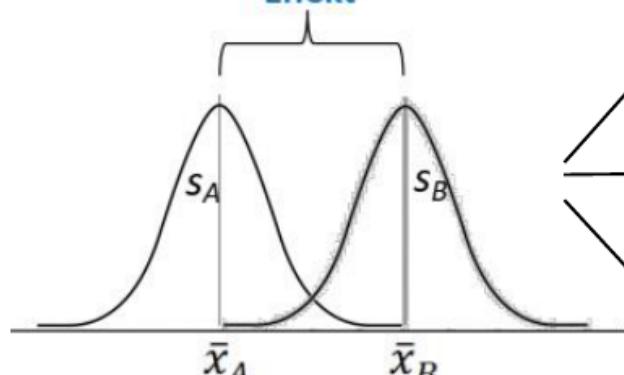
Bottom line: das **Konfidenzniveau** gibt nur vor der Messung die Wahrscheinlichkeit an, dass der Populationsparameter θ im Konfidenzintervall liegen wird. Nach der Messung – und damit in der Praxis – gilt dies nicht mehr.

Inferenzstatistik: Zusammenfassung

Wird durch die Standardisierung anhand der Streuungen zur

Effektgröße

Effekt

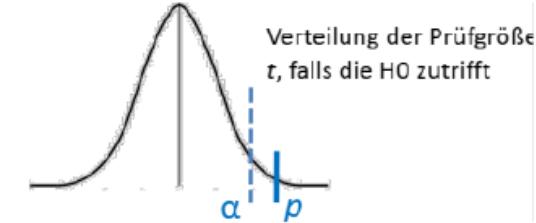
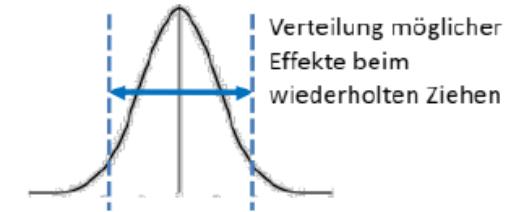
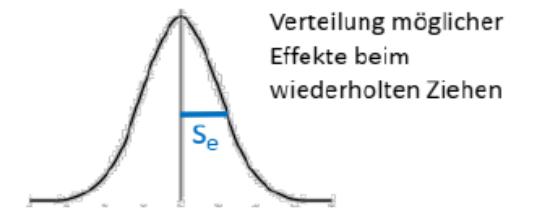


Zwei unabhängige Stichproben

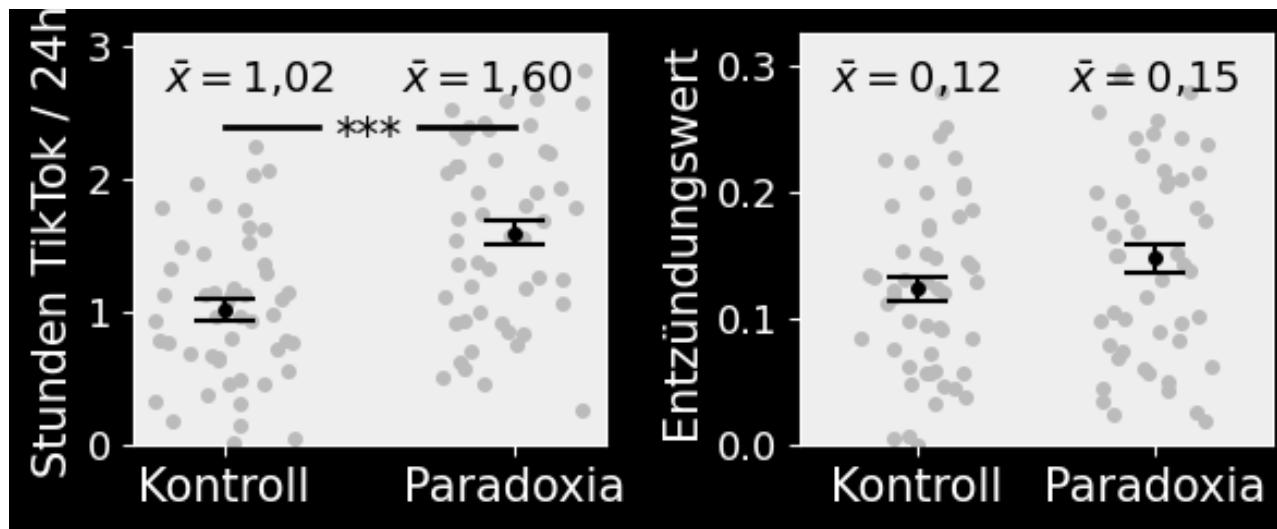
$$(N_{\text{gesamt}} = n_a + n_b)$$

Drei mögliche inferenzstatistische Aussagen:

- 1) Schätzung des *Standardfehlers* des Effekts anhand dessen Stichprobenverteilung
- 2) Angabe eines *Konfidenzintervalls* für den Effekt anhand dessen Stichprobenverteilung
- 3) Berechnung der Prüfgröße t und Prüfung auf Signifikanz mithilfe der t -Verteilung ($p < \alpha?$)



27



Zurück zur Ergebnispräsentation Ihrer Paradoxa-Taskforce. Sie wollen für die beiden zentralen Gruppenvergleiche 95%-Konfidenzintervalle berechnen:

$$CI = \Delta\bar{x} \pm t_{(1-\frac{\alpha}{2}, df)} \cdot \hat{se} \quad \text{mit } \alpha = 0,05$$

Sie holen nochmals die bereits berechneten Größen von den t-Tests aus der Schublade:

	Fallzahl N	$\Delta\bar{x}$	Standardfehler \hat{se}	Freiheitsgrade df	$t_{(0.975, df)}$
TikTok	2×50	0,577	0,123	93,7	1,986
Entzündung	2×50	0,0243	0,0147	96,2	1,985

Neu hinzugekommen ist hier lediglich der kritische t-Wert $t_{(1-\frac{\alpha}{2}, df)}$ für $\alpha = 0,05$.

Damit lauten die Konfidenzintervalle:

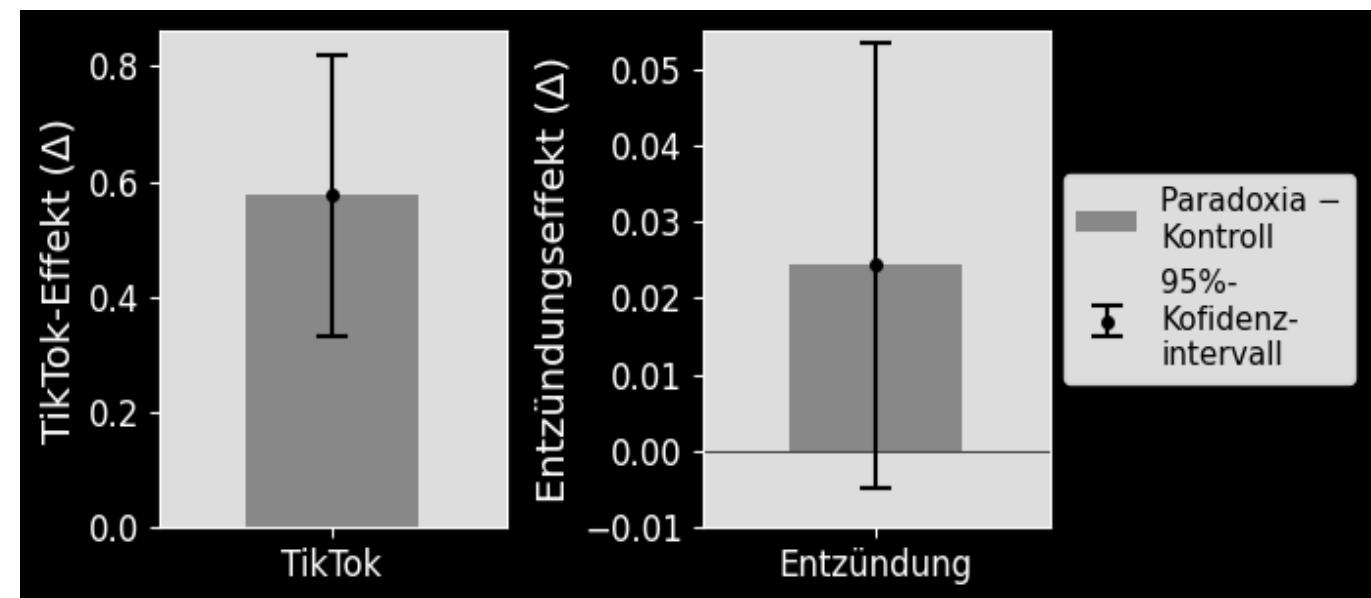
$$\begin{aligned}\text{TikTok: } CI &= \Delta\bar{x} \pm t_{(1-\frac{\alpha}{2}, df)} \cdot \hat{s}e = \\ &= 0,577 \pm 1,986 \cdot 0,123 = [0,333; 0,821]\end{aligned}$$

$$\begin{aligned}\text{Entzündung: } CI &= \Delta\bar{x} \pm t_{(1-\frac{\alpha}{2}, df)} \cdot \hat{s}e = \\ &= 0,0243 \pm 1,985 \cdot 0,0147 = [-0,0049; 0,0535]\end{aligned}$$



Das Ergebnis der Konfidenzintervalle bestätigt den Signifikanztest: der **TikTok-Effekt schließt die Null (deutlich) nicht ein** und ist daher signifikant; der **Entzündungs-Effekt schließt die Null (knapp) ein** und ist daher **nicht signifikant**.

Sie veranschaulichen die Ergebnisse:



Bei den **Konfidenzintervallen der Korrelation** zwischen TikTok-Onlinezeit und Entzündungswerten erwarten Sie aufgrund der Stärke und Signifikanz der Korrelation, dass die Null klar außerhalb des 95%-Konfidenzintervalls liegt.

Im ersten Schritt z-transformieren Sie den Korrelationskoeffizienten $r = 0,517$:

$$z_r = \text{artanh}(r) = \text{artanh}(0,517) = 0,572$$

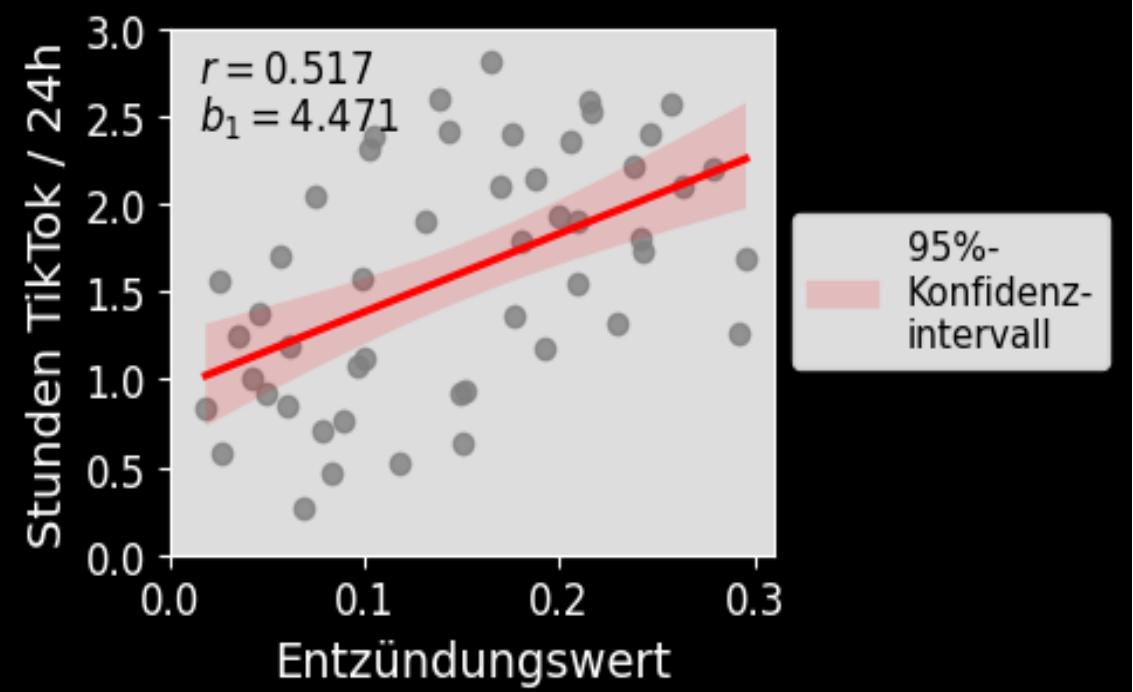
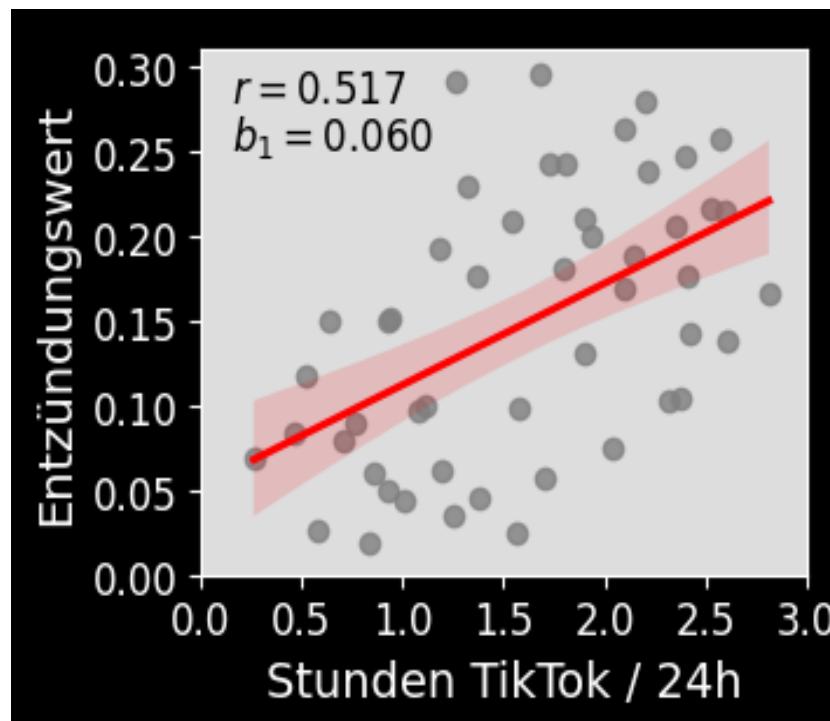
Fisher z-Werte sind in guter Näherung normalverteilt, entsprechend verwenden Sie die Prüfgröße $z_{(1-\frac{\alpha}{2})}$:

$$\begin{aligned} CI_z &= z_r \pm z_{(1-\frac{\alpha}{2})} \cdot \frac{1}{\sqrt{N-3}} = 0,572 \pm z_{(1-\frac{0,05}{2})} \cdot \frac{1}{\sqrt{50-3}} = \\ &= 0,572 \pm 1,96 \cdot \frac{1}{\sqrt{47}} = [0,286; 0,858] \end{aligned}$$



Wie erwartet liegt der Korrelationskoeffizient $r = 0,517$ mit einem 95%-Konfidenzintervall von $[0,286; 0,858]$ klar im positiven Bereich.

Auch für diese Zusammenhangsanalyse erstellen Sie eine verfeinerte Abbildung, in der Sie zusätzlich zur Regressionssteigung auch deren Konfidenzintervall zeigen (dessen Berechnung Sie aber Ihrem Statistikprogramm überlassen):



Fußnoten

1.

Huang X, Wang Y, Zhang H (2023) Effects of physical exercise intervention on depressive and anxious moods of college students: A meta-analysis of randomized controlled trials. Asian Journal of Sport and Exercise Psychology Available at: <https://www.sciencedirect.com/science/article/pii/S2667239123000011> [Accessed September 14, 2023].

2. <https://xkcd.com/2110/>

3.

Die exakte Wahrscheinlichkeitsdichte der Stichprobenverteilung findet sich z.B. auf Wikipedia:

https://en.wikipedia.org/wiki/Pearson_correlation_coefficient#Using_the_exact_distribution

4.

Zu beachten ist allerdings, dass für die Verwendung des z-Wertes nicht die Verteilung von p relevant ist, sondern die Verteilung der Prüfgröße $\frac{\hat{p}}{\hat{s}e}$. Hier handelt es sich dann wieder um

das Verhältnis einer (näherungsweise) normalverteilten und einer chi-verteilten Größe, also in Summe einer t-verteilten Prüfgröße. Da es sich aber *ohnehin* nur um eine Näherung handelt, sind die Differenzen zwischen Normal- vs. t-Verteilung für die Prüfgröße vernachlässigbar und der Einfachheit halber wird die Normalverteilung verwendet. Siehe auch

[Stackexchange](#).

5. <https://www.econometrics.blog/post/the-wilson-confidence-interval-for-a-proportion/>

6. https://de.wikipedia.org/wiki/Konfidenzintervall_f%C3%BCr_die_Erfolgswahrscheinlichkeit_der_Binomialverteilung#Wilson-Intervall