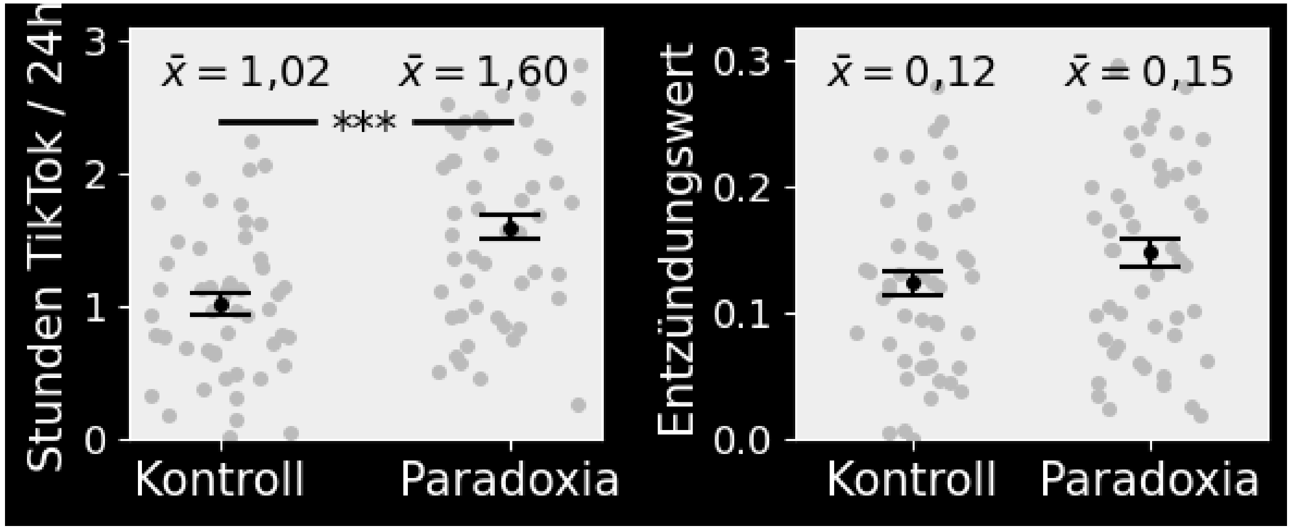
M24 Statistik 1: Sommersemester 2024

Vorlesung 12: Konfidenzintervalle

Prof. Matthias Guggenmos

Health and Medical University Potsdam



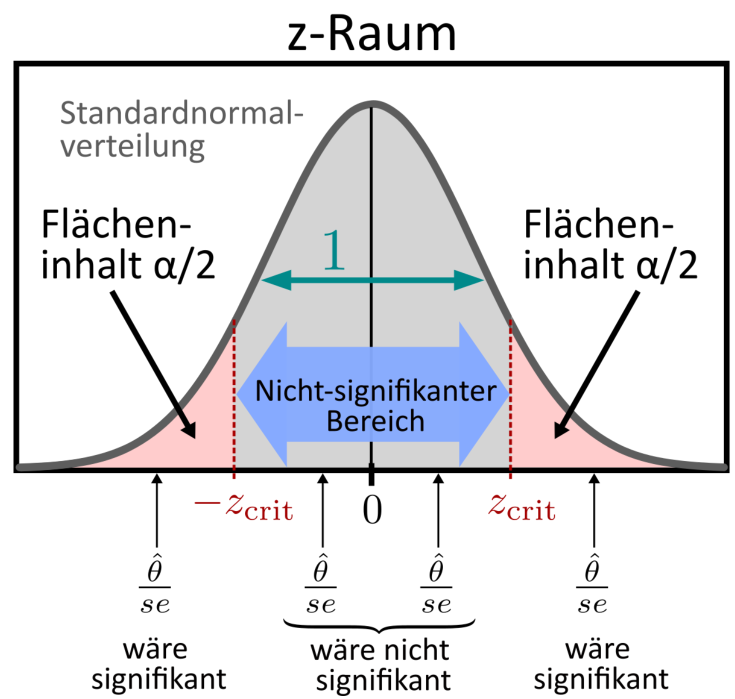
Sie haben nun die valide Signifikanz Ihrer Ergebnisse mithilfe des t-Tests bestimmt. Der Switch vom z- zum t-Test hat zur absurden Situation geführt, dass trotz sehr ähnlicher p-Werte, ersterer den Entzündungseffekt als signifikant wertet, letzterer aber nicht.

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.

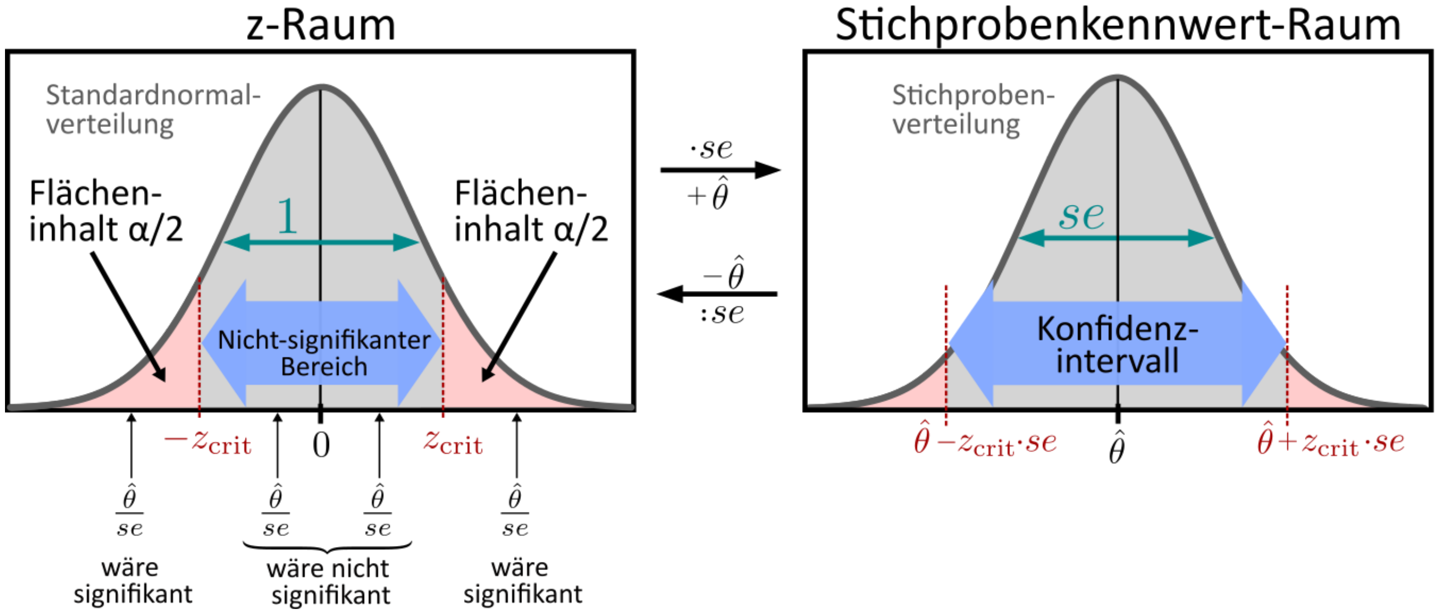
Statt die Signifikanz mit Nein/Ja zu beurteilen, wollen Sie eine **graduelle Aussage** darüber treffen, wie vertrauenswürdig die jeweiligen Mittelwertschätzungen Ihrer Messungen sind. Diese Überlegung führt Sie zum **Konfidenzintervall**.

Rückblick

- Wir hatten das Konzept der Signifikanztestung anhand des **z-Tests** eingeführt, der auf der standardisierten Prüfgröße $z = \frac{\hat{\theta}}{se}$ basiert.
- Für ein gegebenes Signifikanzniveau α lässt sich ein kritischer z-Wert $\pm z_{\text{crit}}$ bestimmen: gilt $z < -z_{\text{crit}}$ oder $z > +z_{\text{crit}}$, so wird ein Effekt als signifikant gewertet (zweiseitiger Test).
- Zugleich definiert das Signifikanzniveau ein Intervall $[-z_{\text{crit}}; z_{\text{crit}}]$, in dem der Effekt <u>nicht</u> signifikant wäre.
  - Dieses Intervall — in der Abbildung bezeichnet als “Nicht-signifikanter Bereich” — hängt nicht von den Daten $(\hat{\theta}, se)$ ab, sondern ausschließlich vom gewählten Signifikanzniveau α.
  - Im nicht-signifikanten Bereich werden Stichprobeneffekte als *mit der Nullhypothese kompatibel* gewertet, bzw. ein Effekt in diesem Intervall würde die *Nullhypothese nicht ablehnen*.



Konfidenzintervall



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

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

- Das Konfidenzintervall umfasst 100·(1-α)% der Stichprobenverteilung (also z.B. 95% bei $\alpha = 0.05$).

Definition

Die Projektion des nicht-signifikanten Bereiches in den Stichprobenkennwert-Raum führt uns zu einer Definition von Konfidenzintervallen:

Defin ition 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, die als statistisch kompatibel mit der Punktschätzung $\hat{\theta}$ gewertet werden.

Nota Bene

Im Kontext von Konfidenzintervallen wird der Stichprobenkennwert $\hat{\theta}$ häufig als **Punktschätzung** bezeichnet. Damit wird betont, dass es sich 1) um einen Einzelwert und keinen Wertebereich handelt ("Punkt…"), und 2) dass dieser Wert nur eine Schätzung des wahren Wertes darstellt ("…schätzung").

Durch die Projektion in den Stichprobenkennwert-Raum ($\cdot se$, $+\hat{\theta}$) vergleichen wir konzeptuell gesehen nicht mehr $\hat{\theta}$ mit der Null (wie bei der Signifikanztestung), sondern mögliche Populationsparameter θ mit der Punktschätzung $\hat{\theta}$. Letztlich sind Signifikanztestung und Konfidenzintervalle aber lediglich zwei Betrachtungsweisen desselben Sachverhaltes.

Intution zu Konfidenzintervallen

Was bedeutet es, ob ein Konfidenzintervall größer oder klein ist? Auf Basis unserer 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**.
- Zusammenfassend bemisst das Konfidenzintervall also die **Vertrauenswürdigkeit einer Punktschätzung** $\hat{\theta}$ ($\bar{x}, \Delta\bar{x}, r$, usw.). **Je kleiner** das Intervall, **desto präziser** ist unsere Schätzung $\hat{\theta}$ des Populationsparameters θ.

Nota Bene

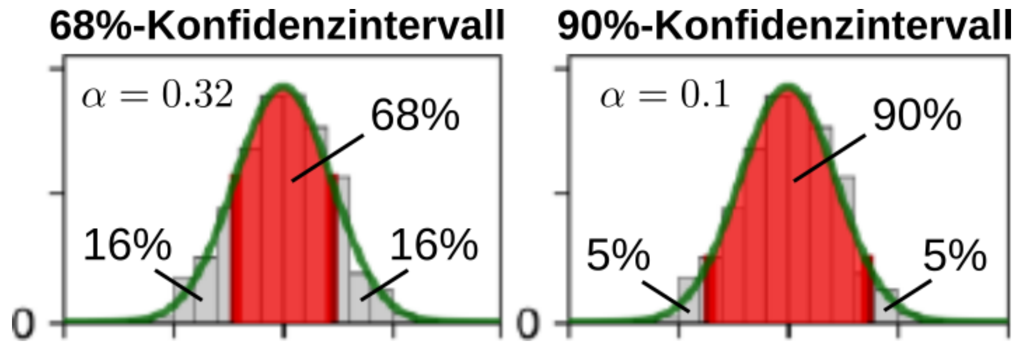
Dürfte man den Namen von Konfidenzintervall neu wählen, wäre **Kompatibilitäsintervall** ein sinnvoller(er) Begriff, da das Intervall im obigen Sinne eben genau **jene Populationsparameter einschließt, die kompatibel mit dem Stichprobenkennwert sind**.

Konfidenzniveau

Konfidenzniveau

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

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

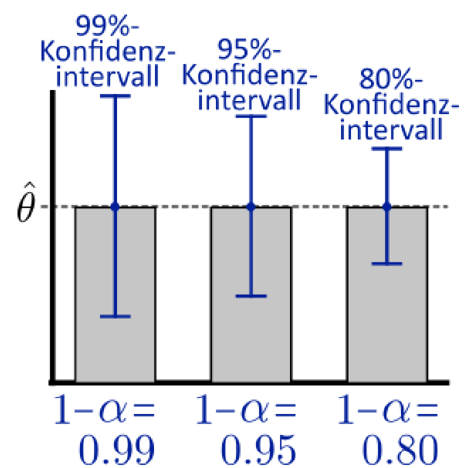


- Das gängiste Konfidenzniveau verwendet $\alpha = 0.05$ und wird dementsprechend als **95%-Konfidenzintervall** bezeichnet.
  - $\alpha = 0.01$ → 99%-Kofidenzintervall
  - $\alpha = 0.1$ → 90%-Kofidenzintervall
  - usw.

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 Publikation[1]

- Grundsätzlich gilt: **je höher das zugrundegelegte Konfidenzniveau, desto größer das zugehörige Konfidenzintervall eines Effektes $\hat{\theta}$.**

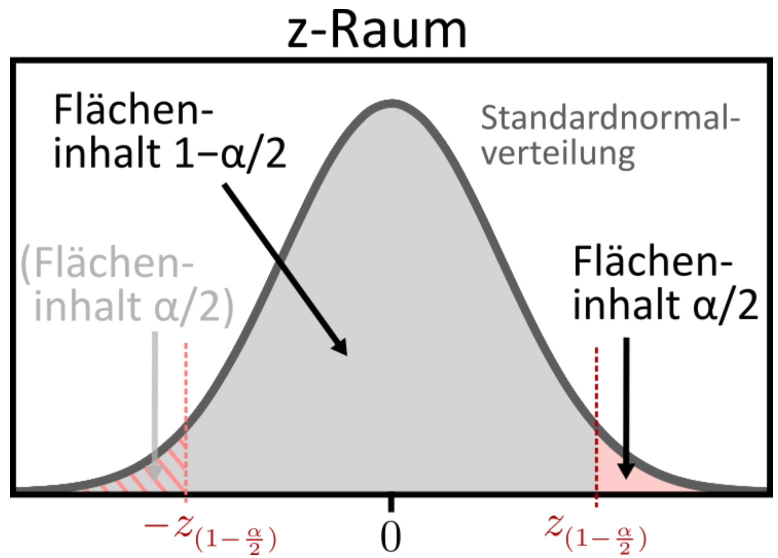


- Bonus: eine präzise (aber wenig intuitive) Interpretation des Konfidenzniveaus und ein zugehöriges Beispiel von Bernard Welch.

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. Folgende Nomenklatur für kritische z-Werte hat sich deshalb 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 Formel des Konfidenzintervalls:

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

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

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

Populationsstreuung nicht bekannt

- Wie bereits festgestellt ist die zentrale Annahme bei z-Werten — Populationsstreuungen gegeben — in der Praxis nahezu nie zutreffend.
- Die Abhilfe ist hier prinzipiell analog zur Signifikanztestung: sind der **Standardfehler se oder die Merkmalsstreuungen σ nicht bekannt**, so **ersetzen wir den kritischen z-Wert $z_{(1-\frac{\alpha}{2})}$ durch den kritischen t-Wert $t_{(1-\frac{\alpha}{2},\ \mathrm{df})}$**:

$$\text{Konfidenzintervall:} \qquad CI = \hat{\theta} \pm t_{(1-\frac{\alpha}{2},\ \mathrm{df})} \cdot \hat{se}$$

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


Nota Bene

Eine wichtige Ausnahme sind **Konfidenzintervalle zu Korrelationskoeffizienten**: hier hat es sich eingebürgert auch dann einen z-Test zu verwenden, wenn die Populationsstreuungen unbekannt sind. Da Korrelationskoeffizienten aber ohnehin ein Spezialfall sind und eine spezielle Transformation benötigen, behandeln wir diese noch gesondert.

Konfidenzintervall: ein erstes Beispiel

Beispiel: Berechnung des Konfidenzintervalls

Als Beispiel nehmen wir eine **Studie** an, in der Sie die **durchschnittliche Nasenlänge $\bar{x}$ von Psychologiestudierenden** untersuchen. (Als Referenzwert ist Ihnen der Bundesnasenlängendurchschnitt von $\mu_0 = 5cm$ bekannt — dieser Wert ist aber zunächst nicht relevant)

Sie erheben eine Stichprobe mit $n = 10$ Psychologiestudierenden und berechnen das Konfidenzintervall Ihrer Mittelwertsschätzung $\bar{x}$ mit einem Konfidenzniveau von $1 - \alpha$.

Der Effekt $\hat{\theta}$ ist in diesem Fall ein Mittelwert $\bar{x}$. Da bei einer Einzelmessung außerdem $\hat{se} = \frac{\hat{\sigma}}{\sqrt{n}}$ gilt, lautet das Konfidenzintervall in diesem Beispiel:

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

Außerdem gilt bei Einzelmessungen $\mathrm{df} = n - 1$, also hier $\mathrm{df} = 10 - 1 = 9$.

Beispiel: Berechnung des Konfidenzintervalls

In Ihrer Stichprobe erhalten Sie für die Nasenlängen einen Mittelwert $\bar{x} = 4.5cm$ und eine Streuung $\hat{\sigma} = 0.5$. Das Konfidenzniveau $1 - \alpha$ solle 95% sein (also $\alpha = 0.05$).

In einem ersten Schritt muss der kritische t-Wert $t_{(1-\frac{\alpha}{2},\ \mathrm{df})}$ bestimmt werden. Wir erhalten ihn beispielsweise aus der t-Tabelle:

$$t_{(1-\frac{\alpha}{2},\ \mathrm{df})} = t_{(0.975,\ 9)} = 2.262$$

	Fläche				
df	0,8	0,85	0,9	0,95	0,975
1	1,377	1,964	3,078	6,314	12,706
2	1,001	1,386	1,886	2,92	4,303
3	0,978	1,25	1,638	2,353	3,182
4	0,941	1,19	1,533	2,132	2,776
5	0,92	1,156	1,476	2,015	2,571
6	0,906	1,134	1,44	1,943	2,447
7	0,896	1,119	1,415	1,895	2,305
8	0,889	1,108	1,397	1,86	2,306
9	0,883	1,1	1,383	1,833	2,262
10	0,879	1,093	1,372	1,813	2,228

Wir setzen alle Werte ein, um das Konfidenzintervall für diese Mittelwertschätzung zu erhalten:

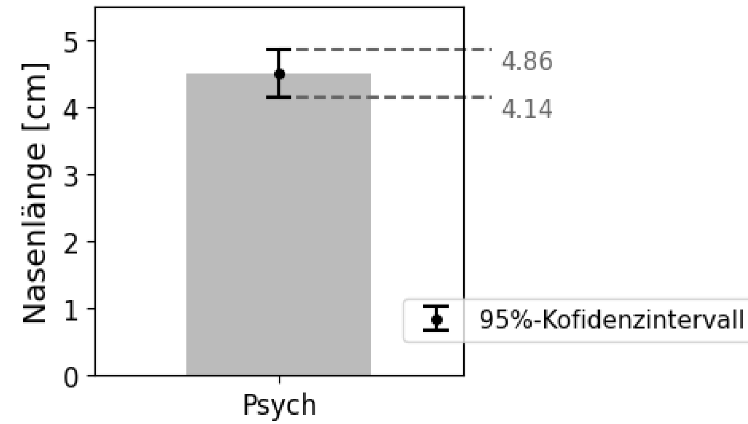
$$CI_{95} = 4.5 \pm 2.262 \cdot \frac{0.5}{\sqrt{10}} \approx 4.5 \pm 0.36$$

$$\xrightarrow[\text{darstellung}]{\text{Intervall-}} \quad CI_{95} = [4.14; 4.86]$$

$\Rightarrow$ **Ergebnis:** Der Mittelwert $\bar{x} = 4.5cm$ ist mit einem 95%-Konfidenzintervall $[4.14cm; 4.86cm]$ verbunden.

Beispiel: Bericht des Konfidenzintervalls

- 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 ±-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 $4.5cm$ ($CI_{95} = [4.14cm; 4.86cm]$).

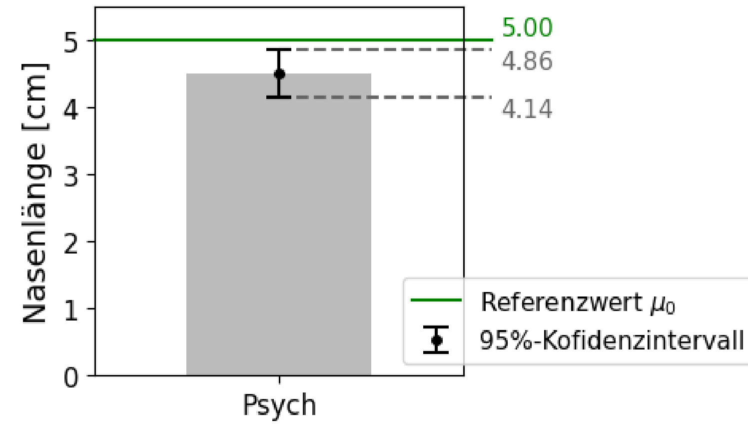
- Oder ausführlicher (wenn Sie auf der Suche nach Worten für Ihre Bachelorarbeit sind):

> Die Messung der Nasenlängen ergab einen Mittelwert von $4.5cm$. Das 95%-Konfidenzintervall der Mittelwertschätzung zeigte eine untere Grenze von $4.14cm$ und eine obere Grenze von $4.86cm$ an.

Beispiel: Signifikanztestung

Mithilfe des Konfidenzintervalls können wir implizit eine weitere Frage beantworten: **ist die mittlere Nasenlänge von Psychologiestudierenden signifikant verschieden vom Bundesnasenlängendurchschnitt $\mu_0 = 5cm$?**

Antwort: ja, da der Referenzwert $\mu_0 = 5cm$ außerhalb des Konfidenzintervalls $[4.14cm; 4.86cm]$ liegt.



Ein paar Aspekte gilt es dabei zu beachten:

- Eine Aussage zur Signifikanz ist nur möglich, wenn wir für den Signifikanztest das dem Konfidenzniveau korrespondierende Signifikanzniveau anlegen. Im Beispiel ist das Konfidenzniveau $1 - \alpha = 0.95$, daher ist das korrespondierende Signifikanzniveau $\alpha = 0.05$.
- Konfidenzintervalle basieren immer auf einem zweiseitigen Test, entsprechend gelten Aussagen zur Signifikanz auch nur für den zweiseitigen Test und damit für die ungerichtete Hypothese "Stichprobeneffekt $\hat{\theta}$ *verschieden* vom Referenzwert".
- Anhand des Konfidenzintervalls kann unmittelbar nur eine Aussage über *signifikant / nicht signifikant* getroffen werden, jedoch nicht über den genauen p-Wert.

Einflussfaktoren auf das Konfidenzintervall

Einflussfaktoren auf das Konfidenzintervall

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

Für das Konfidenzintervall eines einfachen Mittelwertes $\bar{x}$ gilt:

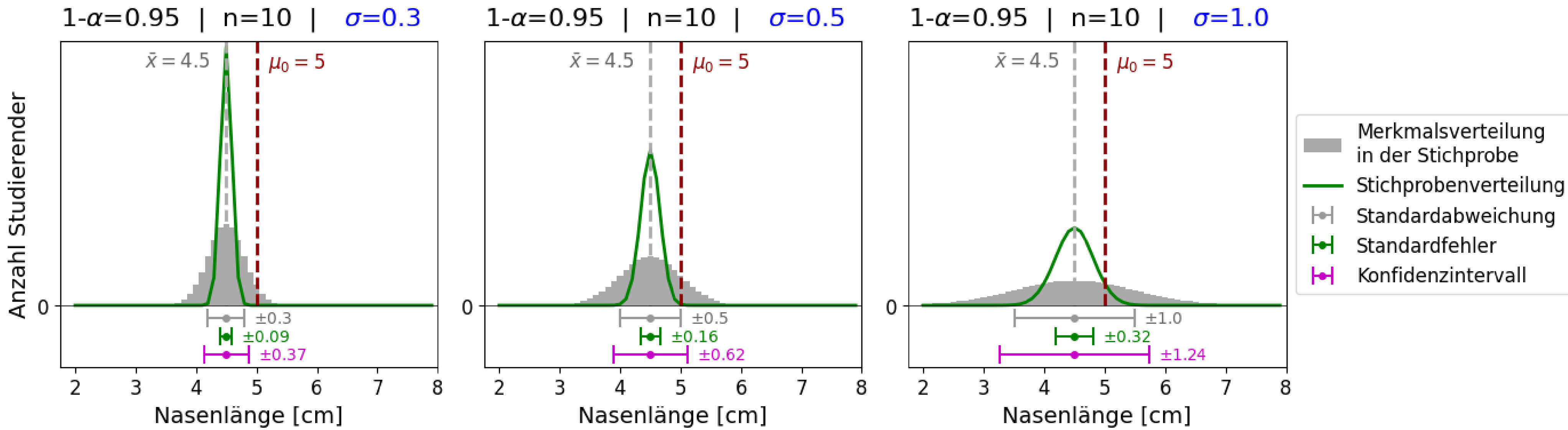
$$CI = \hat{\theta} \pm z_{(1-\frac{\alpha}{2})} \cdot se = \hat{\bar{x}} \pm z_{(1-\frac{\alpha}{2})} \cdot \frac{\sigma}{\sqrt{n}}$$

Damit 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 Konfidenzintervall.



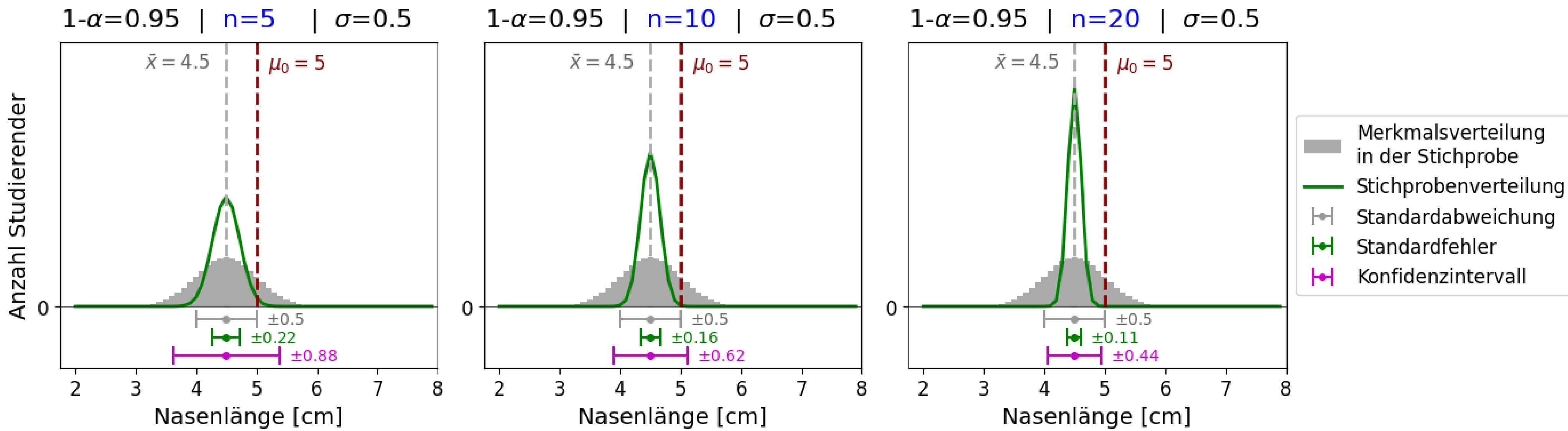
Interpretation: Je größer die Streuung σ des Merkmals, desto größer auch der Standardfehler, desto unsicherer ist man bezüglich der Schätzung des Mittelwertes, desto größer das Konfidenzintervall.

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

Größere Merkmalstreuung σ → Schätzung des Mittelwertes ist mit mehr Unsicherheit behaftet.

Einflussfaktoren auf das Konfidenzintervall

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



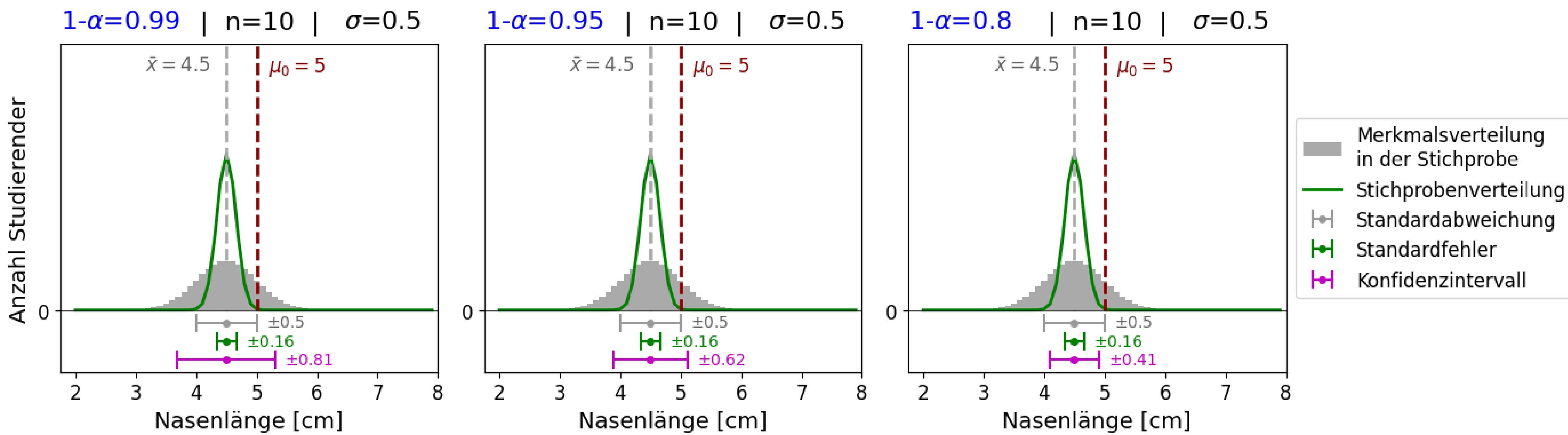
Interpretation: Je größer die Stichprobe, desto kleiner ist der Standardfehler, desto sicherer ist man bezüglich der Schätzung des Mittelwertes, desto kleiner das Konfidenzintervall.

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

Größere Stichprobe n → Schätzung des Mittelwertes ist mit weniger Unsicherheit behaftet.

Einflussfaktoren auf das Konfidenzintervall

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



Interpretation:

Je größer das Konfidenzniveau $1-\alpha$, desto größer der kritische z-Wert $z_{(1-\frac{\alpha}{2})}$, desto größer ist das Intervall potentieller Populationsparameter θ, die statitisch kompatibel mit dem Stichprobenkennwerwert $\hat{\theta}$ sind.

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

Höheres Konfidenzniveau $1-\alpha$ → Hat nichts mit der Schätzung des Mittelwertes zu tun 😉

Konfidenzintervalle für Mittelwerte und Mittelwertdifferenzen

Mittelwerte und Mittelwertdifferenzen

Konfidenzintervalle für Mittelwerte $\bar{x}$:

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

- $\hat{\sigma}$ = Standardabweichung des Merkmals X in der Stichprobe
- Es gilt $\mathrm{df} = n - 1$.

Konfidenzintervalle für Mittelwertdifferenzen $\Delta\bar{x}$:

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

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

Mittelwerte und Mittelwertdifferenzen

Mittelwertdifferenzen: Beispiel aus der Praxis

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

Konfidenzintervall je Mittelwert

Konfidenzintervall für den Unterschied

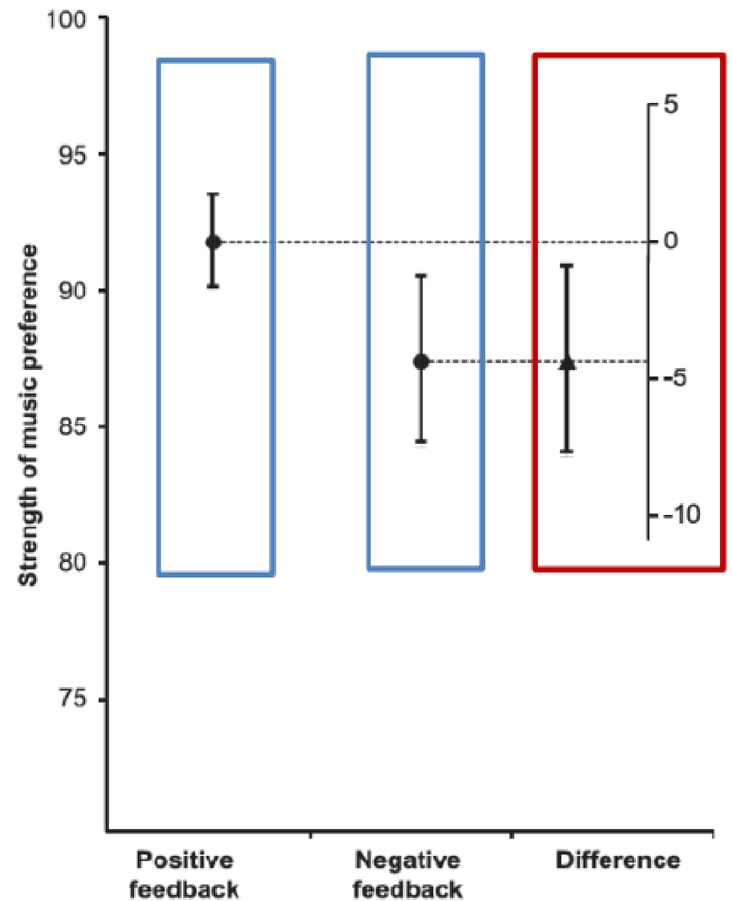


Figure 1. Means and 95% confidence intervals (CIs) for the positive and negative social feedback conditions in Session 3, together with the mean difference between these conditions and its 95% CI on a floating axis (see Cumming, 2013, for a detailed explanation of CIs of effects).

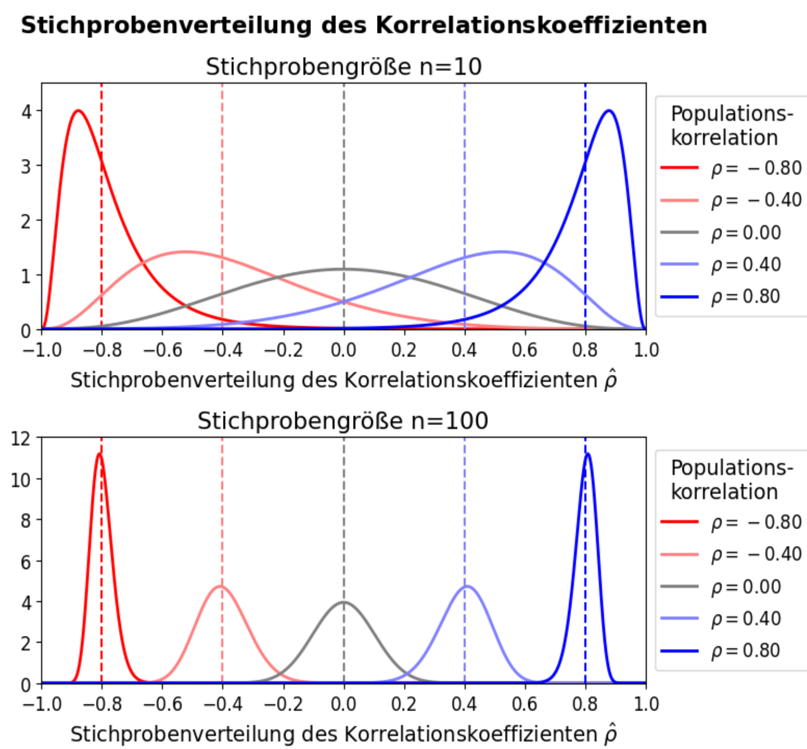
Konfidenzintervalle für die Korrelation

Konfidenzintervalle für die Korrelation

- Auch für den **Pearson-Korrelationskoeffizient** $\hat{\rho}$ lassen sich Konfidenzintervalle bestimmen.
- Den dafür notwendigen Standardfehler haben wir bereits kennengelernt:

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

- Wie aber sieht die Stichprobenverteilung des Korrelationskoeffizienten aus?
- Antwort: kompliziert[2]. Grund: der Korrelations-koeffizient ist auf die Grenzen -1 und $+1$ beschränkt, was insbesondere für Populations-korrelationen ρ nahe ± 1 zu einer stark asymmetrischen Stichprobenverteilung führt.
- Die Annahme einer Normalverteilung ist hier zumindest bei kleinen bis mittleren Stichproben nicht sinnvoll. Was also tun?



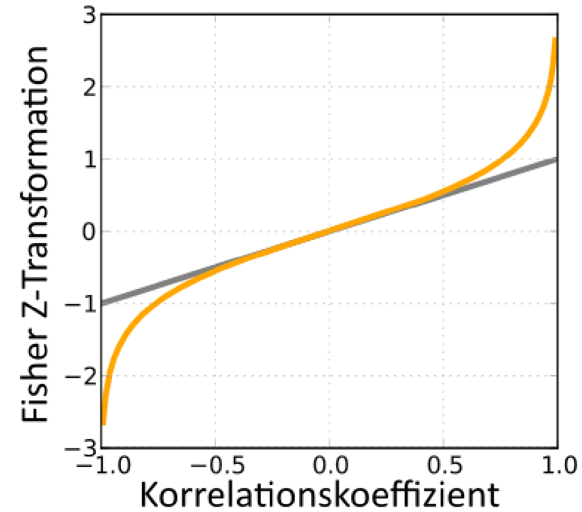
Die theoretische Stichprobenverteilung des Korrelations-koeffizienten $\hat{\theta}$ ist bei kleineren Stichprobengrößen n deutlich verschieden von einer Normalverteilung.

Konfidenzintervalle für die Korrelation

- Die Lösung für diese Problematik stammt von Ronald Fisher: die Korrelationskoeffizienten werden **Z-transformiert** (in Abgrenzung zum z-Wert verwenden wir hier ein großes Z!)

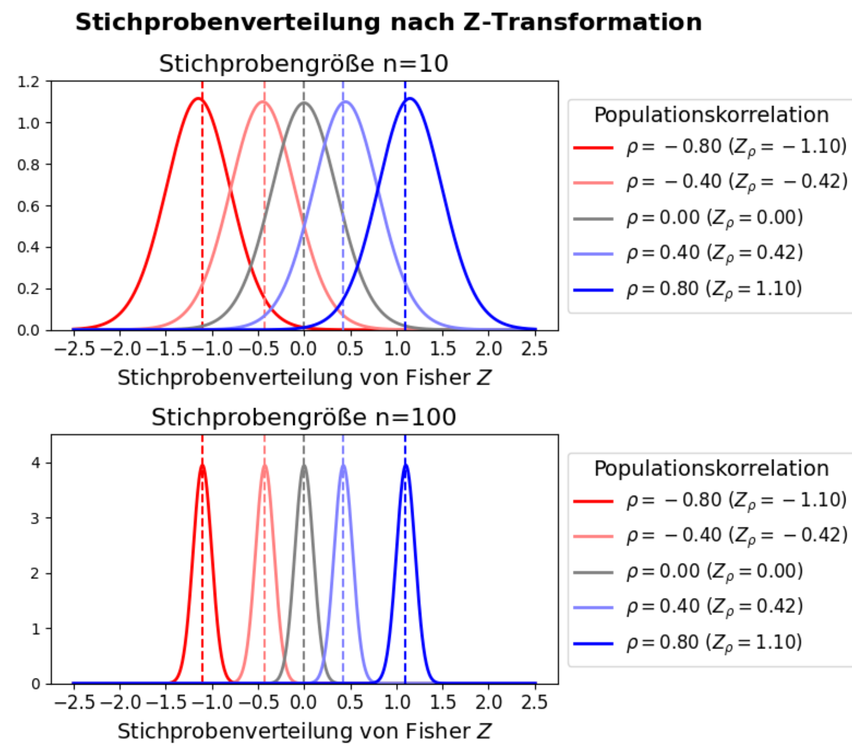
$$\text{Fisher Z-Transformation:} \quad Z = \text{artanh}(\hat{\rho})$$

(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:

$$Z = \operatorname{artanh}(\hat{\rho})$$

2. Berechnung des Konfidenzintervalls

Konfidenzintervalle für Fisher Z-Werte werden auf Basis von kritischen z-Werten $z_{(1-\frac{\alpha}{2})}$ berechnet (auch dann, wenn die Populationsstreuungen unbekannt sind!):

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

bzw. nach Einsetzen des Standardfehlers:

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

3. Rücktransformation in Korrelation

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

$$CI = [\tanh{(Z^u)}; \tanh{(Z^o)}]$$

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 $\hat{\rho} = 0.4$ gemessen hatten.
- Ein Signifikanztest hatte gezeigt, dass die Korrelation bei Anlegen eines Signifikanznievaus $\alpha = 0.05$ nicht signifikant war. Da der Test darüber hinaus sogar einseitig war, erwarten wir, dass das 95%-Konfidenzintervall den Wert Null in jedem Fall einschließen wird.

1. Fisher z-Transformation

$$Z = \text{artanh}(\hat{\rho}) = \text{artanh}(0.4) = 0.424$$

2. Berechnung des Konfidenzintervalls

Standardfehler: $\hat{se} = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{12-3}} = \frac{1}{\sqrt{9}} = \frac{1}{3}$

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:

$$CI_Z = \hat{\rho} \pm z_{(1-\frac{\alpha}{2})} \cdot \hat{se} = 0.424 \pm 1.96 \cdot \frac{1}{3} = [-0.230; 1.077]$$

Konfidenzintervalle für die Korrelation: Beispiel

3. Rücktransformation in Korrelation

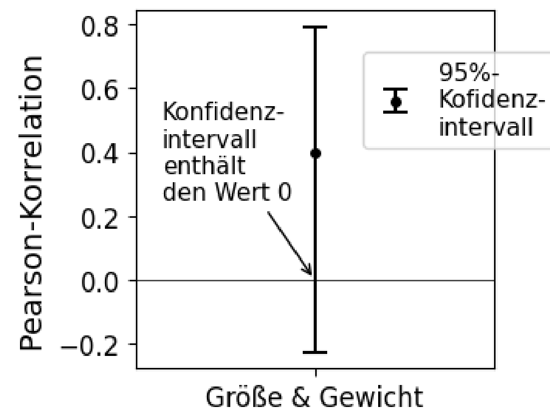
Mithilfe des *Tangens Hyperbolicus* transformieren wir das Konfidenzintervall CI_Z zurück in Korrelationswerte:

$$CI = [\tanh(Z^u); \tanh(Z^o)] = [\tanh(-0.230); \tanh(1.077)] = [-0.226; 0.792]$$

Das Konfidenzintervall für unsere Korrelation $\hat{\rho} = 0.4$ lautet also $[-0.226; 0.792]$. Das ist ein sehr großer Bereich, der sowohl negative als auch positive Werte umfasst. **Und wir sehen: bei der Korrelation können Konfidenzintervalle asymmetrisch sein!**

Damit sind auch Korrelationswerte ≤ 0 im Konfidenzintervall enthalten und wir können direkt ablesen, dass ein Signifikanztest mit $\alpha = 0.05$ für die Hypothese $\rho > 0$ nicht signifikant wäre.

Warum? Weil das Konfidenzintervall jene Korrelationsbereiche umfasst, in denen mögliche Populationskorrelationen ρ statistisch kompatibel mit dem Stichprobenwert $\hat{\rho}$ sind. Insbesondere schließt das Konfidenzintervall den Wert 0 ein und die Nullhypothese ist damit definitiv statistisch kompatibel mit unserer Stichprobenkorrelation. Die Nullhypothese ist also **nicht abgelehnt** und die Korrelation ist **nicht signifikant** .



Zu sehen ist in dieser Abbildung auch, dass das Konfidenzintervall um den empirischen Korrelationskoeffizienten $\hat{\rho} = 0.4$ asymmetrisch ist.

Weiteres WissensWertes

Konfidenzintervall: what it is and what it isn’t

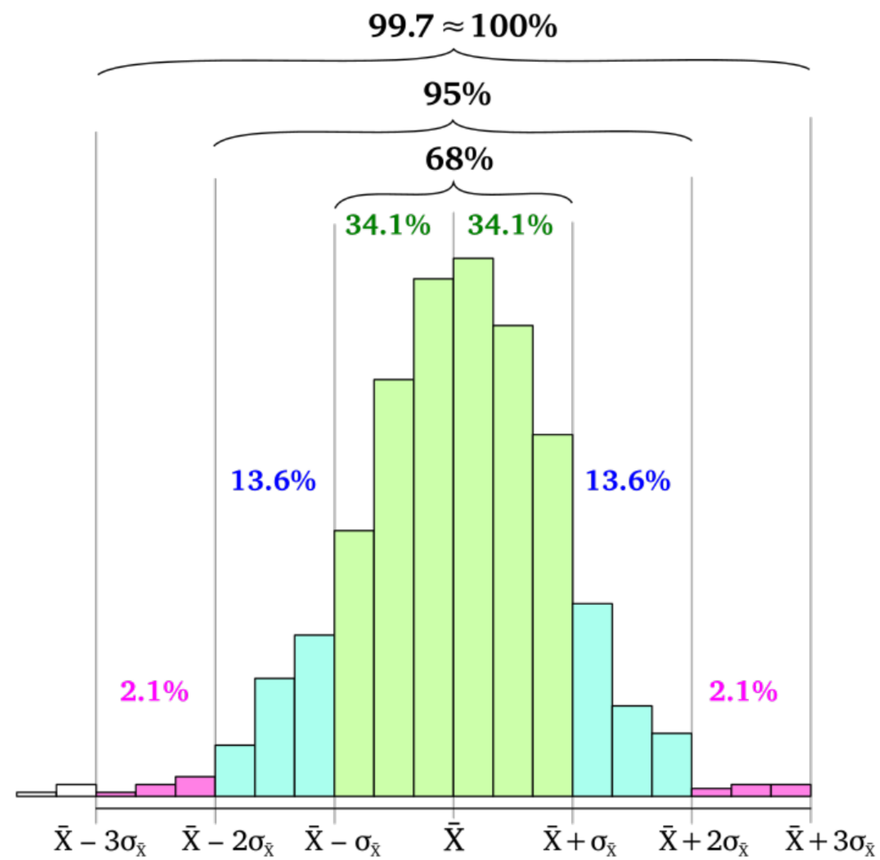
Korrekte Aussagen zu einem 95%-Konfidenzintervall	
Schließt 95% der Fläche unter der Stichprobenverteilung ein.	✓
Definiert einen Bereich möglicher Populationsparameter θ, die bei Anlegen eines 5%-Signifikanzniveaus nicht signifikant verschieden von der Punktschätzung $\hat{\theta}$ sind.	✓
Entspricht einem “Kompatibilitätsbereich” — Populationswerte innerhalb des Intervalls sind kompatibel mit dem Stichprobenkennwert.	✓
Auf lange Sicht werden 95% der erhobenen Stichproben ein Konfidenzintervall aufweisen, das den Populationsparameter einschließt.	✓

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.	✗

±Standardfehler ≙ Konfidenzintervall mit 68%-Konfidenzniveau

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.
  - Reminder: Standardfehler = Standardabweichung der Stichprobenverteilung
- Der **Bereich [Kennwert $\hat{\theta}$ ± 1 Standardfehler]** umfasst gemäß der *68-95-99.7-Regel* ungefähr 68% der Stichprobenverteilung und entspricht damit einem recht laxen **Konfidenzintervall mit einem Konfidenzniveau von 68%.**



Zwei Anwendungsfälle für Konfidenzintervalle

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

Anwendung als Signifikanzmaß

Bei Effekten wie z.B. Mittelwert*unterschieden* 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 $\Delta\bar{x} = 0.2cm; CI_{95} = [-0.1cm; 0.4cm]$" → impliziert einen nicht-signifikanten Effekt auf einem Signifikanzniveau $\alpha = 0.05$.

"Der Gruppenunterschied betrug $\Delta\bar{x} = 0.2cm; CI_{95} = [0.1cm; 0.3cm]$" → 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.

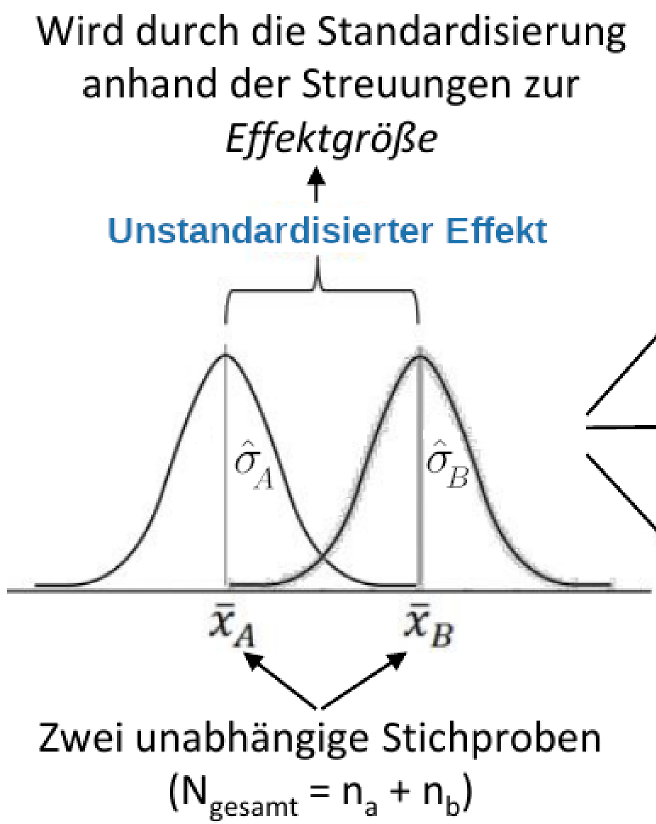
Konfidenzintervalle für weitere Effektarten

- Im Rahmen von Statistik 1 lernen wir Konfidenzintervalle für Mittelwerte $\bar{x}$, Mittelwertdifferenzen $\Delta\bar{x}$ und Korrelationskoeffizienten $\hat{\rho}$ kennen.
- Es gilt jedoch: **Konfidenzintervalle lassen sich für jegliche Arten von Stichprobeneffekten berechnen.**

Informationen zu Konfidenzintervallen von zwei weiteren Effektarten finden sich im Bonuscontent:

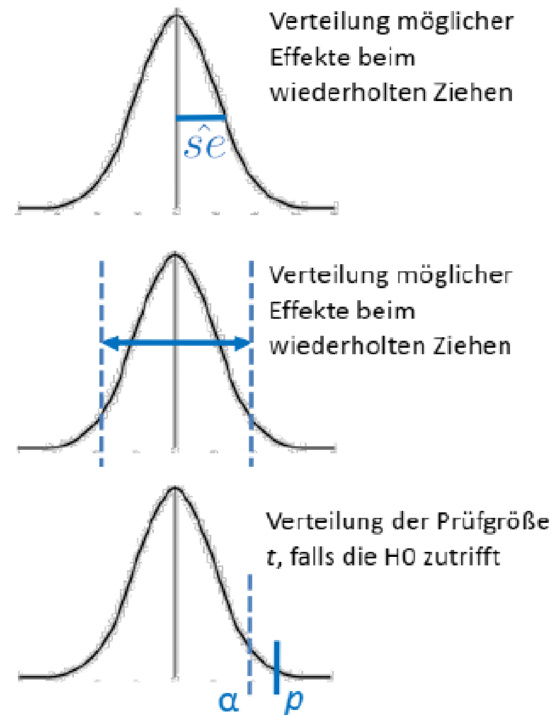
- Regressionskoeffizient $\hat{b}_1$
- Anteile / relative Häufigkeiten $\hat{p}$

Inferenzstatistik: Übersicht



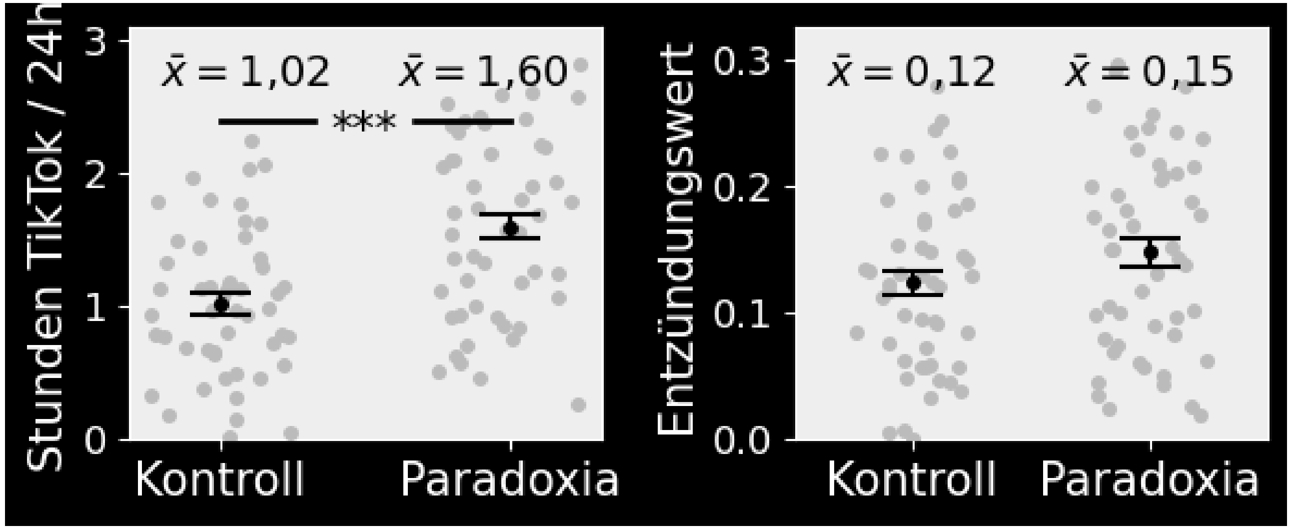
Drei wichtige inferenzstatistische Aussagen:

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 ($p < \alpha$?)



[[Zusammenfassung]]

- Konfidenzintervalle geben an, **wie sicher wir uns bei der Schätzung eines Stichprobenkennwertes $\hat{\theta}$ sind**.
- Sie stellen damit eine **Verallgemeinerung des Standardfehlers $\hat{se}$** dar, der selbst einem 68%-Konfidenzintervall entspricht.
- Konfidenzintervalle **basieren auf dem Prinzip der Signifikanztestung**:
  - Verwendung der kritischen Prüfgrößen $z_{(1-\frac{\alpha}{2})}$ und $t_{(1-\frac{\alpha}{2},\ \mathrm{df})}$.
  - Liegt die “Null” oder ein anderer Referenzwert θ_0 außerhalb des Konfidenzintervalls, so ist der Stichprobeneffekt $\hat{\theta}$ signifikant verschieden von θ_0.
- Konfidenzintervalle sind **Kompatibilitäsintervalle**: das Intervall gibt einen Bereich von wahren Populationskennwerten θ an, die statistisch kompatibel mit dem Stichprobenkennwert $\hat{\theta}$ sind.
- Je höher das **Konfidenzniveau** $1-\alpha$, desto strenger das Signifikanzniveau α, desto mehr Populationskennwerte θ sind statistisch nicht verschieden von (also kompatibel mit) dem Stichprobenkennwert $\hat{\theta}$, desto größer das Konfidenzintervall.
- Konfidenzintervalle der Korrelation erfordern eine vorherige Transformation der Korrelationskoeffizienten $\hat{\rho}$: die **Fisher-Z-Transformation** $Z = \operatorname{artanh}(\hat{\rho})$.



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

$$CI = \Delta\bar{x} \pm t_{(1-\frac{\alpha}{2},\,\mathrm{df})} \cdot \hat{se} \qquad \text{mit} \quad \alpha = 0.05$$

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

	Fallzahl n	$\Delta\bar{x}$	**Standardfehler** $\hat{se}$	**Freiheitsgrade** df	$t_{(0.975,\,\mathrm{df})}$
TikTok	2×50	0.577	0.123	93.7	1.986
Entzündung	2×50	0.0243	0.0147	96.2	1.985

Neu hinzugekommen sind hier lediglich die kritischen t-Werte $t_{(1-\frac{\alpha}{2},\,\mathrm{df})}$ für $\alpha = 0.05$, die jedoch für dieses Beispiel aufgrund der ungeraden Freiheitsgrade ($\mathrm{df} = 93.7$ und 96.2) mit dem Computer und nicht der t-Tabelle bestimmt wurden. Prüfen Sie jedoch für sich, ob die angegebenen kritischen t-Werte auf Basis der t-Tabelle plausibel erscheinen!

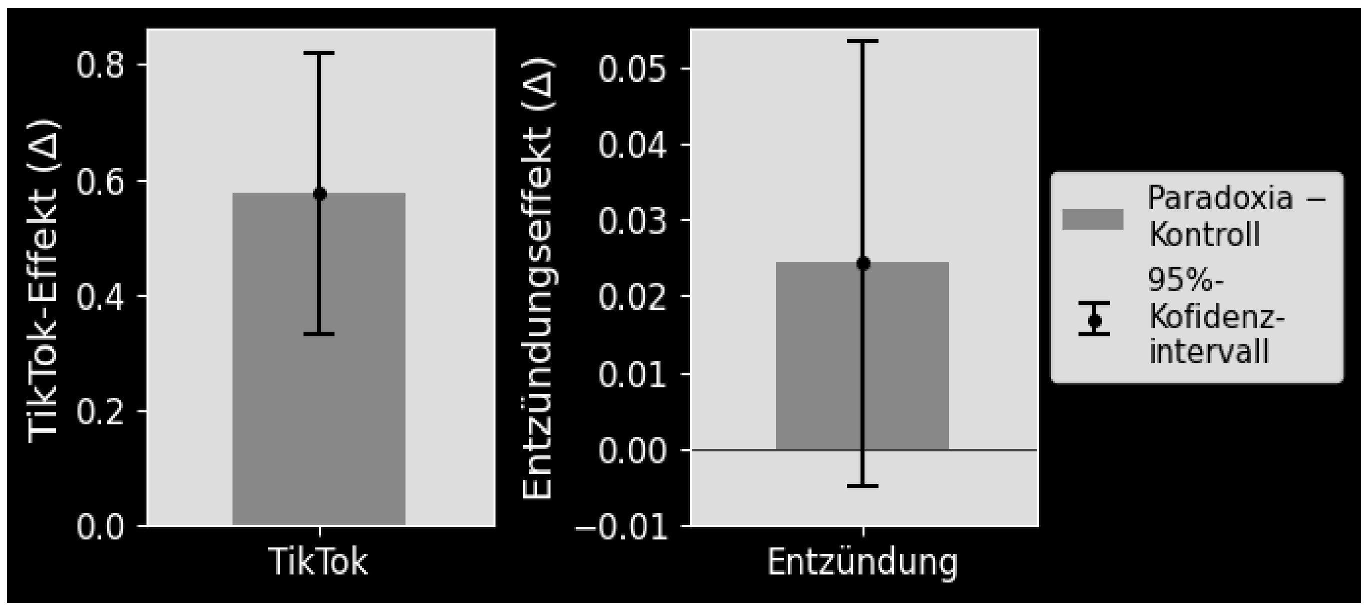
Damit lauten die Konfidenzintervalle:

$$\text{TikTok:} \quad CI = \Delta\bar{x} \pm t_{(1-\frac{\alpha}{2},\,\text{df})} \cdot \hat{se} = $$
$$= 0.577 \pm 1.986 \cdot 0.123 = [0.333; 0.821]$$

$$\text{Entzündung:} \quad CI = \Delta\bar{x} \pm t_{(1-\frac{\alpha}{2},\,\text{df})} \cdot \hat{se} = $$
$$= 0.0243 \pm 1.985 \cdot 0.0147 = [-0.0049; 0.0535]$$

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 $\hat{\rho} = 0.517$:

$$Z = \text{artanh}(\hat{\rho}) = \text{artanh}(0.517) = 0.572$$

Als kritische Prüfgröße fungiert bei Konfidenzintervallen der Korrelation der kritische z-Wert $z_{(1-\frac{\alpha}{2})}$:

$$CI_Z = Z \pm z_{(1-\frac{\alpha}{2})} \cdot \frac{1}{\sqrt{n-3}} = 0.572 \pm z_{(1-\frac{0.05}{2})} \cdot \frac{1}{\sqrt{50-3}} =$$

$$= 0.572 \pm 1.96 \cdot \frac{1}{\sqrt{47}} = [0.286; 0.858]$$

Rücktransformation zu Korrelation: $CI = [\tanh(0.286); \tanh(0.858)] = [0.278; 0.695]$

Das 95%-Konfidenzintervall enthält die Null <u>deutlich nicht</u> — die Korrelation ist also signifikant.

Bonuscontent

Konfidenzintervalle für Regressionskoeffizienten

Im Gegensatz zu Korrelationskoeffizienten sind Regressionskoeffizienten $\hat{b}_1$ nicht auf das Intervall $[0; 1]$ beschränkt. Das Konfidenzintervall kann daher wie beim Mittelwert berechnet werden:

$$CI = \hat{b}_1 \pm t_{(1-\frac{\alpha}{2},\, \mathrm{df})} \cdot \hat{se}$$

mit:

$$\hat{se}(\hat{b}_1) = \frac{\hat{\sigma}_Y}{\hat{\sigma}_X}\sqrt{\frac{1-\hat{\rho}^2}{n-2}}$$

Da für $\hat{se}(\hat{b}_1)$ genau 2 Mittelwerte berechnet werden müssen ($\bar{x}$ in $\hat{\sigma}_X$ und $\bar{y}$ in $\hat{\sigma}_Y$), ist die t-Prüfgröße der Regressionssteigung mit $\mathrm{df} = n - 2$ Freiheitsgraden verbunden, und damit:

$$\text{Konfidenzintervall Regression:} \qquad CI = \hat{b}_1 \pm t_{(1-\frac{\alpha}{2},\, n-2)} \cdot \hat{se}$$

Konfidenzintervalle für Regressionskoeffizienten

- 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 den Beispielzusammenhang Größe-Gewicht ($\hat{\rho} = 0.4$).
- Wir nehmen als Standardabweichungen an $\hat{\sigma}_{\text{Größe}} = 16.5$ und $\hat{\sigma}_{\text{Gewicht}} = 6.8$. Damit ergibt sich die Steigung $\hat{b}_1$ und der zugehörige Standardfehler:

$$\hat{b}_1 = \frac{\hat{\sigma}_{\text{Gewicht}}}{\hat{\sigma}_{\text{Größe}}} \cdot \hat{\rho} = \frac{6.8}{16.5} \cdot 0.4 = 0.16 \qquad \left(\text{Reminder: } \hat{b}_1 = \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} \hat{\rho}\right)$$

$$\hat{se}(\hat{b}_1) = \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} \sqrt{\frac{1 - \hat{\rho}^2}{n - 2}} = \frac{6.8}{16.5} \sqrt{\frac{1 - 0.4^2}{12 - 2}} = 0.12$$

- Der kritische t-Wert lautet:

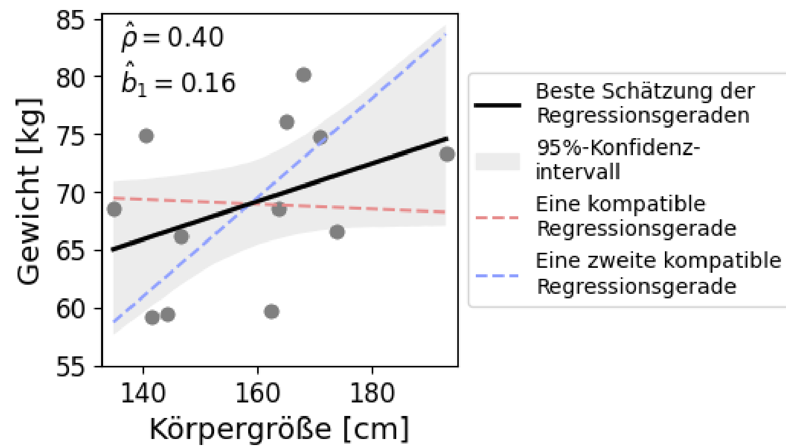
$$t_{(1-\frac{\alpha}{2},\, n-2)} = t_{(0.975,\, 10)} = 2.23$$

- Damit ergibt sich das Konfidenzintervall zu

$$CI = \hat{b}_1 \pm t_{(1-\frac{\alpha}{2},\, n-2)} \cdot \hat{se} = 0.16 \pm 2.23 \cdot 0.12 = 0.16 \pm 0.27 = [-0.11; 0.43]$$

Konfidenzintervalle für Regressionskoeffizienten

- Die maximale Steigung innerhalb des 95%-Konfidenzintervalls beträgt also -0.11 und die minimale Steigung 0.43.
- Auch der y-Achsenabschnitt $\hat{b}_0$ ist mit einer gewissen Unsicherheit verbunden.
- In Regressionsplots wird der Bereich *aller mit dem Konfidenzintervall kompatiblen Steigungen und y-Achsenabschnitte* häufig als farblich hinterlegte Fläche angedeutet (hellgrau in der Abbildung rechts).



Beispiel für einen Regressionsplot mit der geschätzten Regressionsgeraden (schwarz) und dem 95%-Konfidenzintervall (hellgrau). Die rote und blaue Line zeigen beispielhaft Regressionsgeraden, die mit dem 95%-Konfidenzintervall kompatibel wären.

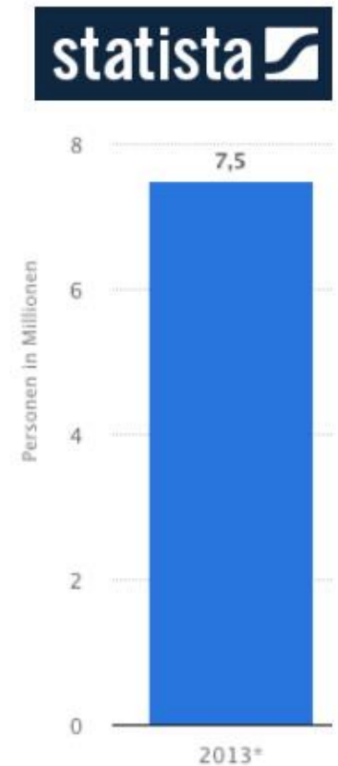
Konfidenzintervalle für Anteile / relative Häufigkeiten

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).

ROBERT KOCH INSTITUT

Nur vier Prozent der Menschen in Deutschland leben vegetarisch

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 ^ 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:

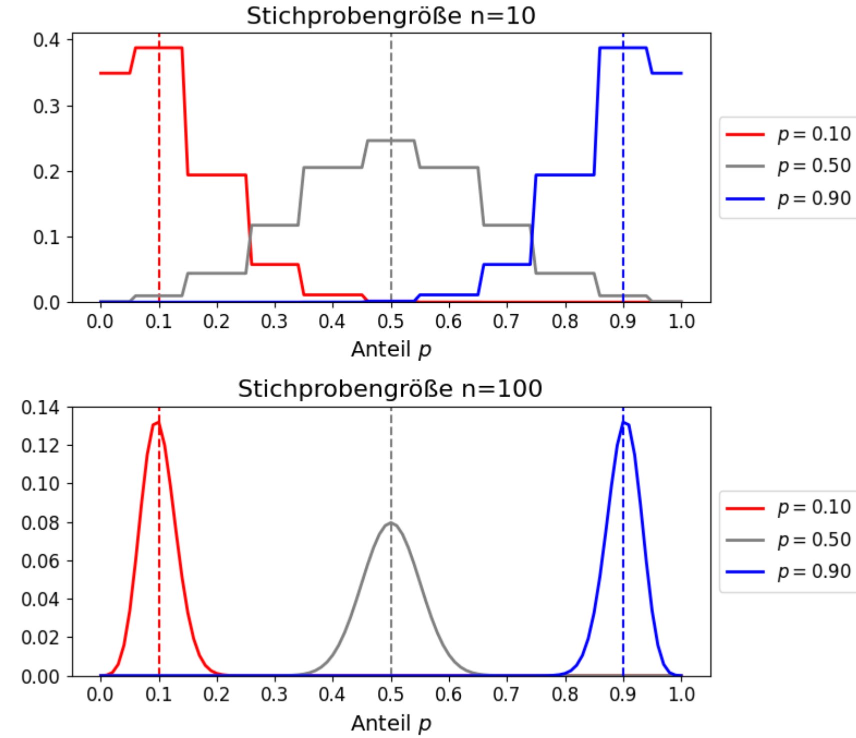
$$\hat{se}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{(Achtung: p steht hier für den Anteil ("proportion") und nicht den p-Wert!)}$$

- 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[3] [4]:

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



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

- Im Computerzeitalter sollten allerdings für die Konfidenzintervalle von Anteilen präzisere Schätzungen wie das **Wilson-Intervall**[5] verwendet werden.

Nota Bene

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.

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:

$$CI_{\text{RKI}} = \hat{p} \pm z_{(1-\frac{\alpha}{2})} \cdot \hat{se}(\hat{p}) = 0.04 \pm 1.96 \cdot 0.0024 = [0.0353; 0.0447]$$

$$CI_{\text{Statista}} = 0.075 \pm 1.96 \cdot 0.0016 = [0.0719; 0.0781]$$

- Da das Konfidenzintervall in beiden Fällen weniger als 1% umfasst (prüfe nach!), sind die Schätzung tatsächlich recht präzise. Zugleich widersprechen sich beide Schätzungen (warum?)

Konfidenzniveau und Wahrscheinlichkeit

Es ist verführerisch anzunehmen, dass z.B. ein 95%-Konfidenzniveau bedeutet, dass der wahre Populationskennwert mit einer Wahrscheinlichkeit von 95% innerhalb des Konfidenzintervalls liegt. **Dem ist nicht so.**

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 Nullhypthesentestung 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 Populationskennwert einschließt.

KEIN KLAUSUR STOFF

Konfidenzniveau und Wahrscheinlichkeit

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.

Alternativ können wir daher das gleiche Prinzip einfach auf zukünftige hypothetische Stichprobenerhebungen nach der aktuellen Stichprobe anwenden. Wir gehen dabei von der **Wahrscheinlichkeit** einer einzelen (noch vor uns liegenden) Stichprobenerhebung zu einer **Frequenz** bei zukünftigen Stichprobenerhebungen über. Dies führt zur (jedoch kaum intuitiveren 😉) **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.

Auch hier gilt die *Erste Regel von Konfidenzintervallen*:

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

KEIN KLAUSUR STOFF

https://bcdudek.net/confidence/

KEIN KLAUSUR STOFF

Confidence Interval for a Sample Mean: A simulation

This app randomly samples N data points from a Normal Distribution. The user specifies N. Sample means are computed for each simulated sample.

Choose wich Graphs

- (●) Confidence Interval Graph Only
- ( ) Confidence Interval Graph Plus Sampling Distribution of the Mean

Display Choice for CI Graph

- (●) Means Plus CI
- ( ) Means Only

Number of Simulated Samples: Five

Sample Size: 4

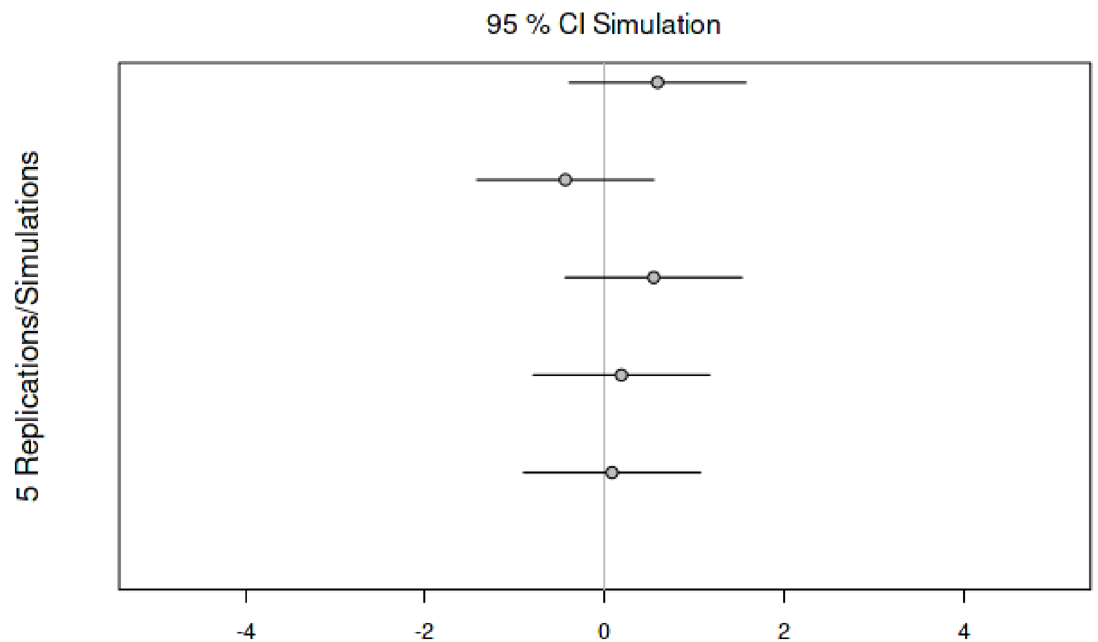
Confidence Level: 95

Population Mean: 0

Population Std Dev: 1

Plots | About

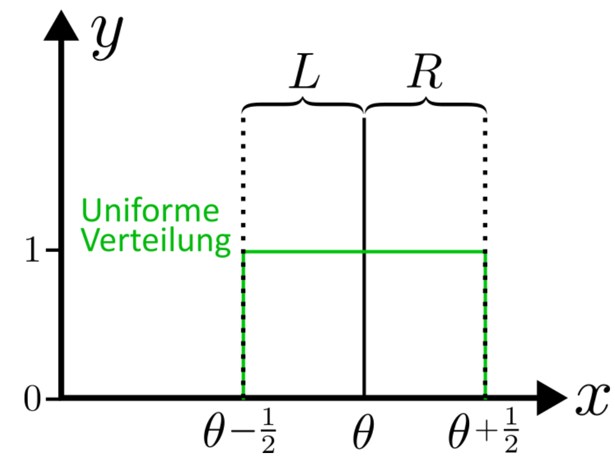
Click your mouse on the plot to see an explanation of the graph.



Konfidenzintervall-Beispiel von Welch

KEIN KLAUSUR STOFF

- 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.



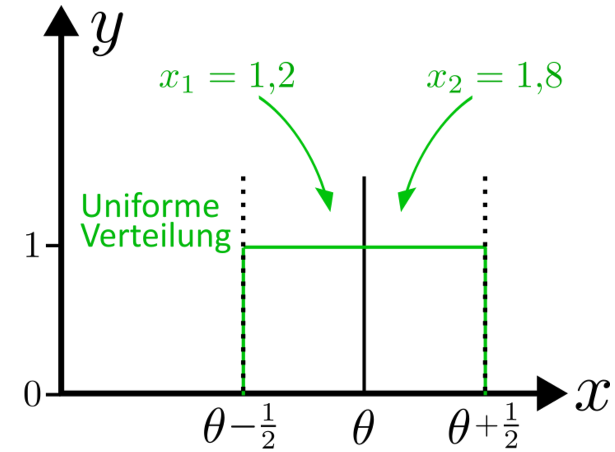
Vor der Messung:

- 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

KEIN KLAUSUR STOFF

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.

$\Rightarrow$ **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.

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. Die exakte Wahrscheinlichkeitsdichte der Stichprobenverteilung findet sich z.B. auf Wikipedia: https://en.wikipedia.org/wiki/Pearson_correlation_coefficient#Using_the_exact_distribution
3. Zu beachten ist allerdings, dass für die Verwendung des z-Wertes nicht die Verteilung von p relevant ist, sondern die Verteilung der Prüfgröße $\frac{\hat{p}}{\hat{se}}$. 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.
4. https://www.econometrics.blog/post/the-wilson-confidence-interval-for-a-proportion/
5. https://de.wikipedia.org/wiki/Konfidenzintervall_f%C3%BCr_die_Erfolgswahrscheinlichkeit_der_Binomialverteilung#Wilson-Intervall