M24 Statistik 1: Wintersemester 23/24

Vorlesung 08: Inferenzstatistik

Prof. Matthias Guggenmos

Health and Medical University Potsdam

Der Forschungsprozess

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

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

↓

Hypothese

↓

Operationalisierung

↓

Experiment / Studie

↓

Daten

↓

Auswertung / Ergebnis

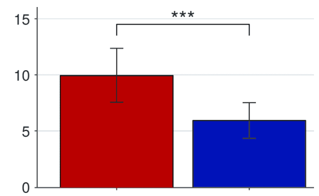
↓

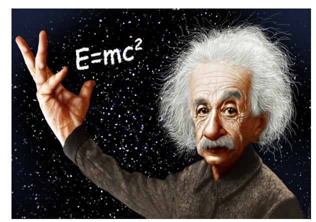
Schlussfolgerung / Interpretation

↓

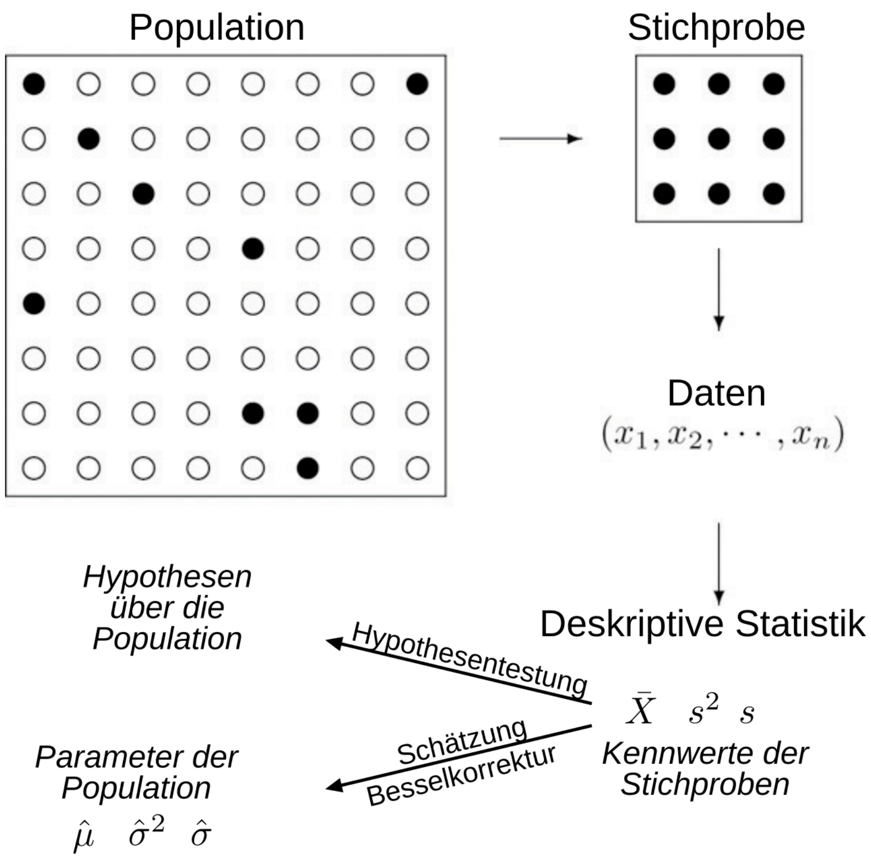
Aufstellen oder Update einer Theorie

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





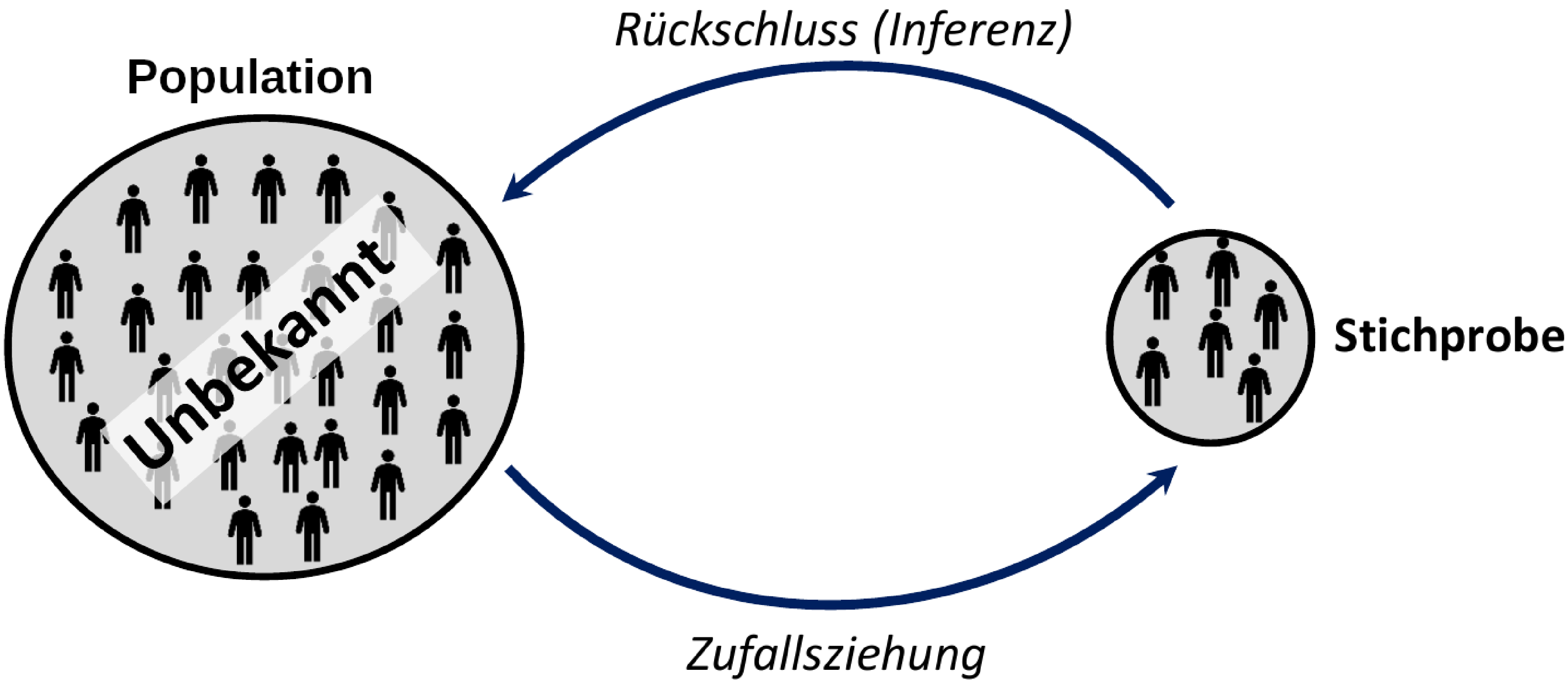
Was ist Inferenzstatistik?



Die statistische Prozess in der Psychologie auf einen Blick: eine Teilmenge von Versuchspersonen wird aus der Population gezogen — die Stichprobe. In der Stichprobe werden Daten $(x_1, x_2, \ldots, x_n)$ eines Merkmals gemessen. Mit Methoden der deskriptiven Statistik werden **Kennwerte** in der Stichprobe beschrieben $(\bar{X}, s^2, s)$. Letztendlich ist das Ziel ein Rückschluss auf die wahren **Parameter** der Population (μ, σ^2, σ) und das Testen von Hypothesen über die Population. Bildnachweis[1]

- In den vergangenen Vorlesungen haben wir kennengelernt, wie Merkmale in Stichproben quantitativ beschrieben werden können — wir haben **deskriptive Statistik** betrieben.
- In den meisten Fällen ist das Ziel in der Psychologie allerdings nicht die Beschreibung der spezifischen Stichprobe, sondern ein Rückschluss – eine Inferenz, eine Verallgemeinerung – auf die Population.
- Die Mathematik hinter diesem Inferenzprozess ist Gegenstand der **Inferenzstatistik**.

Was ist Inferenzstatistik?



Zentrale Frage: Ist das Ergebnis meiner Studie eine gute Schätzung für die wahren Verhältnisse in der Population?

Das Gesetz der großen Zahlen

Was passiert wenn meine Stichprobe immer größer wird ($n \longrightarrow \infty$)?

Das Gesetz der großen Zahlen

Beispiel: gezinkter Würfel

- Sie wollen überprüfen, ob ein Würfel zu Ihrem Nachteil gezinkt ist, so dass die "Eins" überproportional häufig kommt. Dazu führen Sie eine Reihe von Probewürfen durch.
- Sollte der Würfel *nicht gezinkt* sein, sollte die Wahrscheinlichkeit einer "Eins" genau $\frac{1}{6}$ ist.
- Sie zählen die Häufigkeit einer "Eins" nach 10, 100, 1000 und 10000 Würfen.

Analogie:

Anzahl Würfe =
Stichprobengröße

Ergebnis:

	10 Würfe	100 Würfe	1000 Würfe	10000 Würfe
Anzahl Einsen	3	20	172	1669
Relative Häufigkeit	$\frac{3}{10} = 0.3$	$\frac{20}{100} = 0.2$	$\frac{172}{1000} = 0.172$	$\frac{1669}{10000} = 0.1669$

- Mit zunehmender Zahl der Würfe nähert sich also Ihre gemessene Wahrscheinlichkeit der erwarteten Wahrscheinlichkeit $\frac{1}{6} = 0{,}1666..$ an — dies ist das **Gesetz der großen Zahlen**!

($\Rightarrow$ … und offenkundig ist der Würfel nicht gezinkt!)

Das Gesetz der großen Zahlen

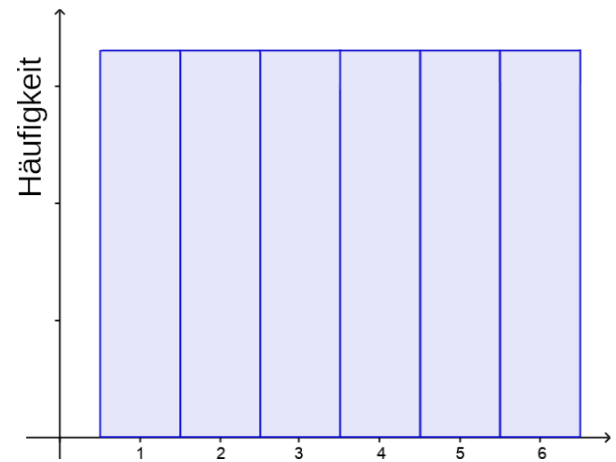
Definition	**Gesetz der großen Zahlen:** Die gemessene relative Häufigkeit eines Zufallsereignisses nähert sich immer weiter an die tatsächliche Wahrscheinlichkeit dieses Ereignisses an, je häufiger das Zufallsexperiment durchgeführt wird.

Das Gesetz der großen Zahlen ist ein mathematischer **Grenzwertsatz**, weil er das Verhalten einer Größe in einem Grenzfall ($\text{Anzahl Zufallsexperimente } \rightarrow \infty$) beschreibt.

Das Gesetz der großen Zahlen

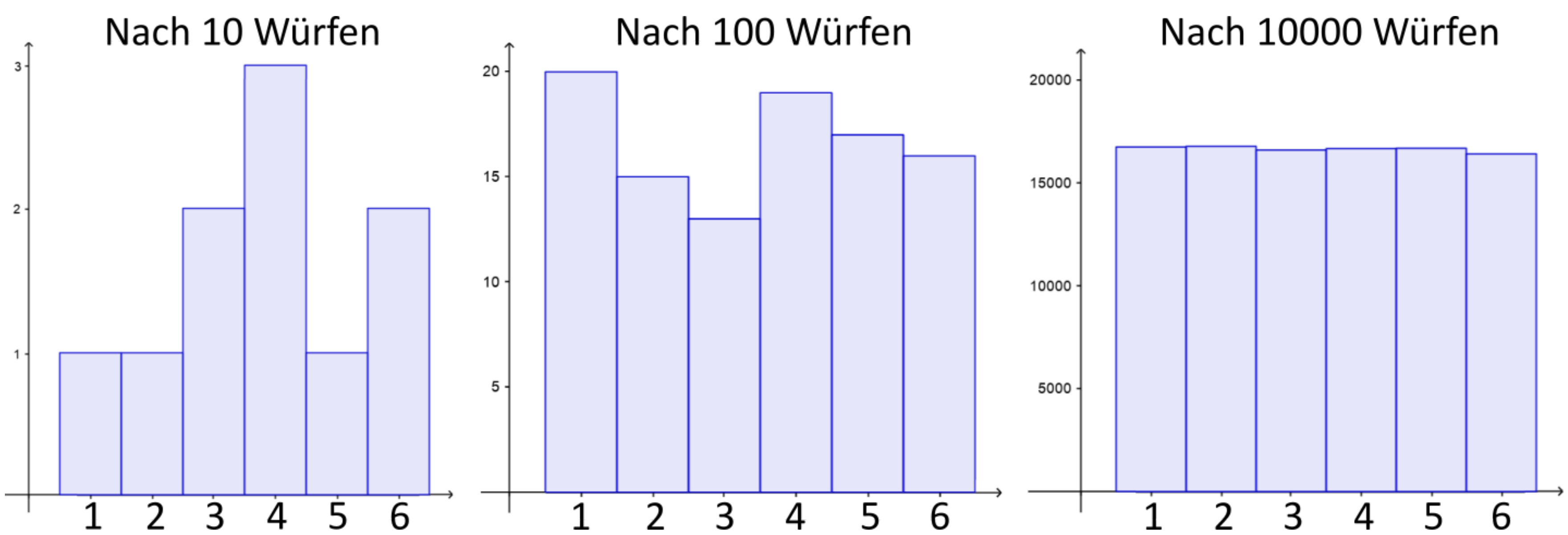
Im nächsten Schritt übertragen wir das ganze auf **Verteilungen**:

- Im Würfelbeispiel haben wir bislang die relative Häufigkeit einer *einzelnen* Merkmalsausprägung (Würfelzahl 1) betrachtet
- Natürlich gilt das Gesetz der großen Zahlen aber für alle Merkmalsausprägungen – im Würfelbeispiel also auch für die Zahlen 2 bis 5.
- Die gemessene relative Häufigkeit *aller* Merkmalsausprägungen einer Zufallsvariablen X wird auch als die **empirische Häufigkeitsverteilung** oder **empirische Stichprobenverteilung** von X bezeichnet.
- Im Würfelbeispiel ist die Zufallsvariable die gewürfelte Zahl. Wir erwarten, dass alle Zahlen 1-6 gleich wahrscheinlich sind, oder mit anderen Worten: wir erwarten dass sich die empirische Häufigkeitsverteilung einer **uniformen Verteilung** annähert:



Das Gesetz der großen Zahlen

- Analog nähert sich hier mit steigender Zahl von Würfen die gemessene Häufigkeitsverteilung der erwarteten Häufigkeitsverteilung an:



Das Gesetz der großen Zahlen kann deshalb auch auf Verteilungen bezogen werden:

Definition

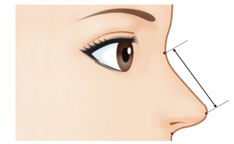
Gesetz der großen Zahlen (2): Die gemessene relative Häufigkeits*verteilung* (=Wahrscheinlichkeitsverteilung) einer Zufallsvariablen nähert sich immer weiter an die tatsächliche Wahrscheinlichkeitverteilung der Variable an, je häufiger das Zufallsexperiment durchgeführt wird.

Das Gesetz der großen Zahlen in der Psychologie

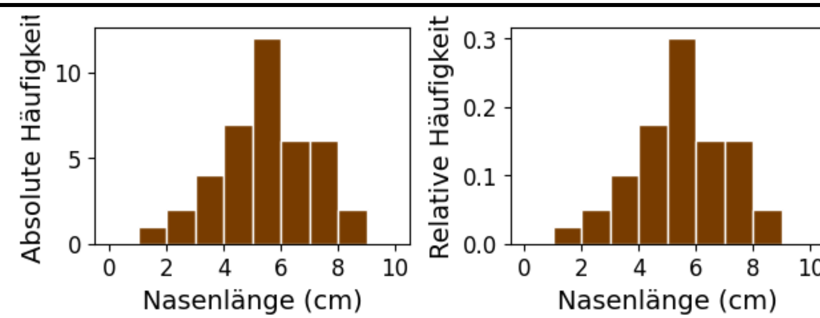
- Im Kontext der Psychologie begegnet uns das Gesetz der großen Zahlen, wenn wir die **relative Häufigkeitsverteilung eines Merkmals in der Stichprobe** betrachten.

Beispiel

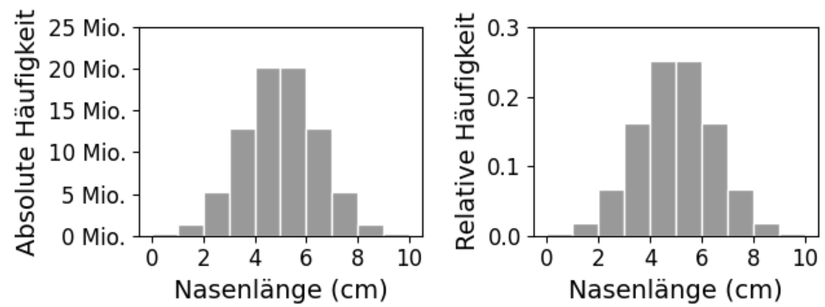
Sie messen die Nasenlänge in ihrer Kohorte. Von jeder Person notieren Sie die Nasenlänge und tragen diese in einem Histogramm auf.



Ihr Ergebnis in Ihrer Stichprobe sieht folgendermaßen aus:



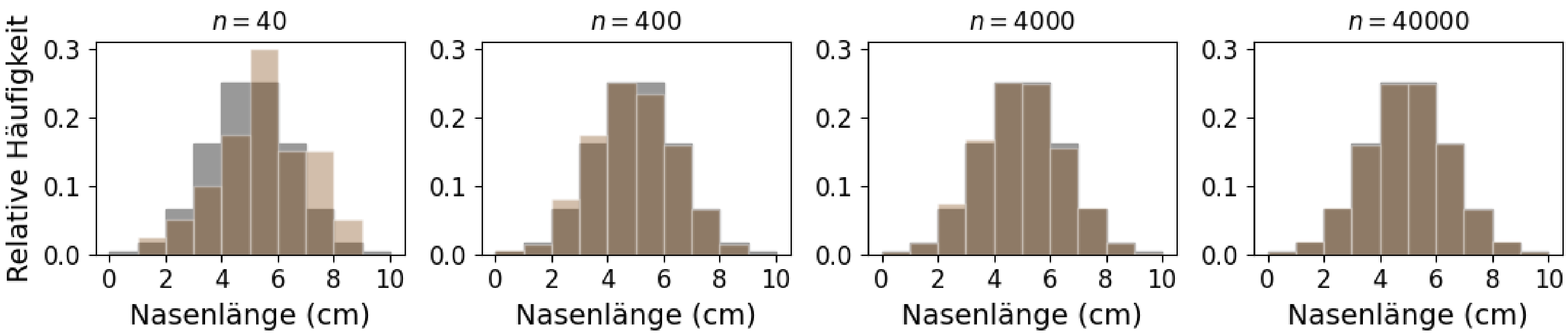
Der Gott der Nasen hat uns zum Vergleich die **tatsächliche** Häufigkeitssverteilung von Nasenlängen in Deutschland(=Population) anvertraut:



- Auch hier würden wir erwarten, dass sich die Verteilung des Merkmals in der Stichprobe an die tatsächliche Verteilung in der Population annähert, je größer die Stichprobe ist.

Gesetz der großen Zahlen in der Psychologie

- Jede zusätzliche Versuchsperson entspricht einem neuen **Zufallsexperiment** ("Würfelwurf").
  - Annahme: Versuchspersonen werden *zufällig* aus der Population gezogen.
- Wir können die Annäherung der Stichprobenverteilung an die Populationverteilung mit $n \to N$ simulieren (N = Umfang der Population, 80 Mio. Deutsche in diesem Beispiel):



Die Stichprobenverteilungen sind in braun schattiert, die (gleichbleibende) Häufigkeitsverteilung der Population in grau, die in diesem Fall alle 80 Mio. Deutsche umfasst. Beachte, dass in diesem Fall die *relative* Häufigkeit dargestellt ist.

- Ähnlich wie beim Würfelwurf nähert sich die Verteilung des Merkmals in der Stichprobe an die tatsächliche Verteilung in der Population an.

Theoretische Stichprobenverteilung

Theoretische Stichprobenverteilung

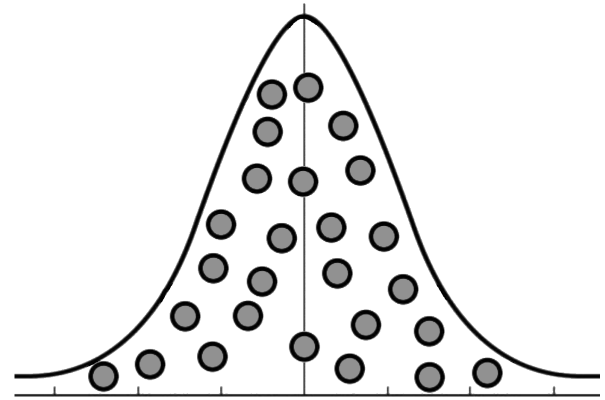
Empirische Stichprobenverteilung

- Führe ich dieselbe Studie mehrmals durch und notiere jeweils den statistischen Kennwert, erhalte ich eine **empirische Stichprobenverteilung**.
- Dies ist die Idee der **Metaanalyse**, die eine Vielzahl empirischer Studien zusammenfasst und analysiert ($\Rightarrow$ Vorlesung 13).

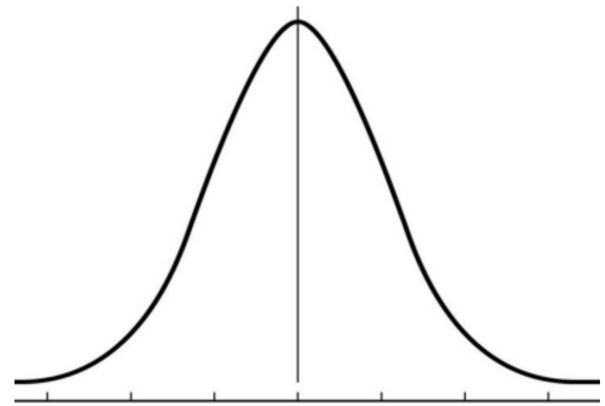
Die Idee der Inferenzstatistik ist ähnlich:

- Ich habe nur *eine* Studie, aber überlege, was theoretisch passieren würde, wenn ich diese Studie immer wieder wiederholen würde.
- Dieser Ansatz führt zur **theoretischen Stichprobenverteilung**.

Zentrale Idee: Die theoretische Stichprobenverteilung erlaubt uns eine Einschätzung darüber, wie stabil unser Ergebnis bei einer (hypothetischen) Wiederholung der Studie sein würde.



Die empirische Stichprobenverteilung besteht aus tatsächlich erhobenen Studien. Die Verteilung empirischer Stichprobenkennwerte folgt im Idealfall (u.a. kein Publikationsbias, großes n pro Stichprobe) einer Normalverteilung.



Die theoretische Stichprobenverteilung ist durch eine Funktion gegeben. Ist die Stichprobengröße n, für die die theoretische Stichprobenverteilung angenommen wird, groß, folgt die Stichprobenverteilung einer Normalverteilung.

Theoretische Stichprobenverteilung: Beispiel

Was würde passieren, wenn ich meine Studie ganz oft wiederholen würde?

- Gehen wir das Problem zunächst mit einem Gedankenexperiment (Simulation) an und tun so, als ob wir die wahren Verhältnisse in der Population kennen.

Gedankenexperiment: wir interessieren uns für die durchschnittliche Nasenlängen von Männern. Wir nehmen an, die Population besteht nur aus 9 Männern:

$\Rightarrow$ Der wahre Mittelwert (d.h. der Mittelwert der Population) ist in diesem Beispiel $\mu = 6cm$.

Wir nehmen weiterhin an, dass wir in unserer Studie 3 Männer untersuchen, d.h. wir ziehen eine **Stichprobe** $n = 3$ mit Zufallsauswahl aus der Population.

Theoretische Stichprobenverteilung: Beispiel

Unsere zufällige Stichprobe könnte folgende drei Männer aus der Population umfassen:

$$\bar{x} = 7cm$$

... oder diese drei Männer:

$$\bar{x} = 5cm$$

... oder diese drei Männer:

$$\bar{x} = 6cm$$

Theoretische Stichprobenverteilung: Beispiel

... oder diese drei Männer:

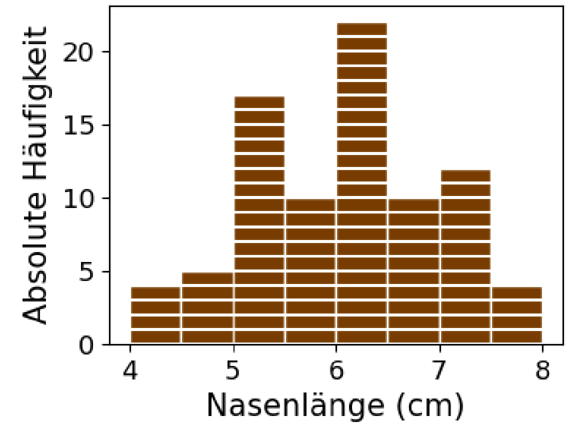
$$\bar{x} = 5{,}67cm$$

… und so weiter

Es gibt insgesamt 84 Möglichkeiten aus dieser Population von 9 Männern zufällig drei Männer auszuwählen.

→ Jeder Mittelwert wäre eine Schätzung für den wahren Populationswert.

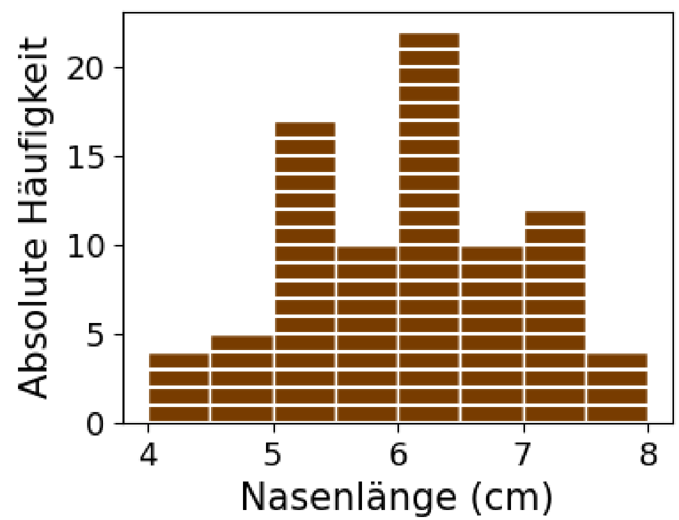
→ Die Abbilung rechts zeigt die Verteilung aller möglichen 84 Mittelwerte.



Theoretische Stichprobenverteilung: Beispiel

Was lernen wir aus dieser Verteilung?

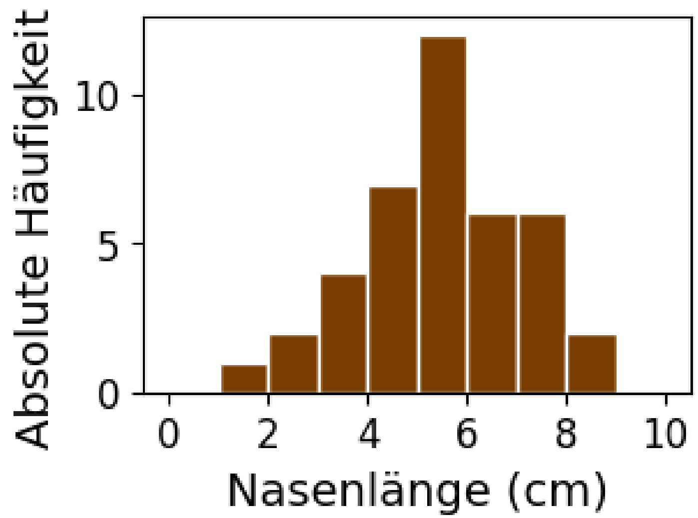
- Obwohl es nur einen *wahren Mittelwer*t gibt, weichen die *einzelnen Studienergebnisse* mehr oder weniger davon ab.
- Die Studienergebnisse haben also eine *Bandbreite*, in der offensichtlich eine Aussage über die Genauigkeit der Schätzung (von einer einzelnen Stichprobe auf die Population) steckt.
- Die Stichprobenergebnisse *schwanken zufällig* um den wahren Wert (siehe auch Metaanalyse).



Theoretische Stichprobenverteilung

Übertragen wir nun das Gedankenexperiment auf die Realität:

- Population seien nun **alle Männer in Deutschland**.
- Sie haben eine einzelne Studie durchgeführt (also eine Stichprobe aus der Population gezogen).

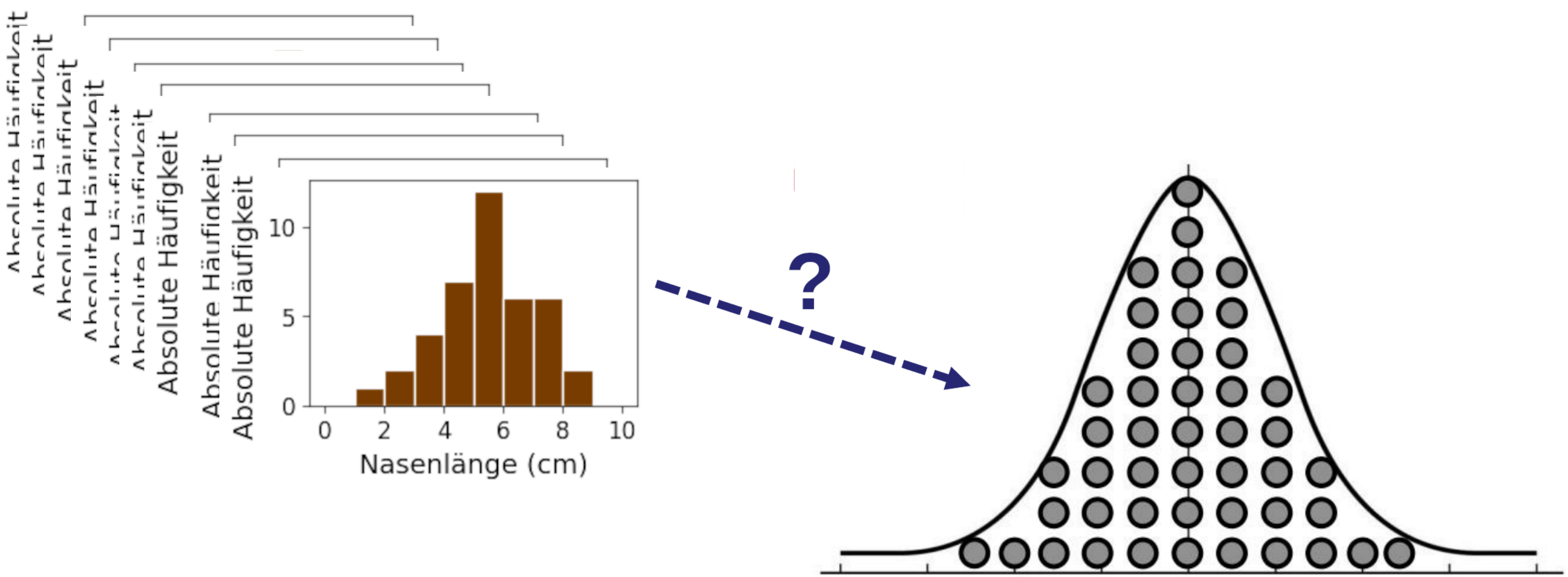


- Ihnen ist nun klar, dass das Ergebnis der Studie nur *eines von vielen möglichen Ergebnissen* ist.
- Beim Wiederholen derselben Studie würde also – rein zufallsbedingt – ein etwas anderes Ergebnis herauskommen.

Theoretische Stichprobenverteilung

Wie sieht die zu erwartende Stichprobenverteilung aus, wenn ich, anders als im Gedankenexperiment, die Population *nicht kenne*?

Dies führt uns zur **theoretischen Stichprobenverteilung**. Wie sich herausstellt, können wir anhand einer einzelnen Studie Aussagen darüber machen, wie die Verteilung von Stichprobenkennwerten erwartbar aussehen würde, würden wir die Studie – rein hypothetisch – **unendlich oft wiederholen**.

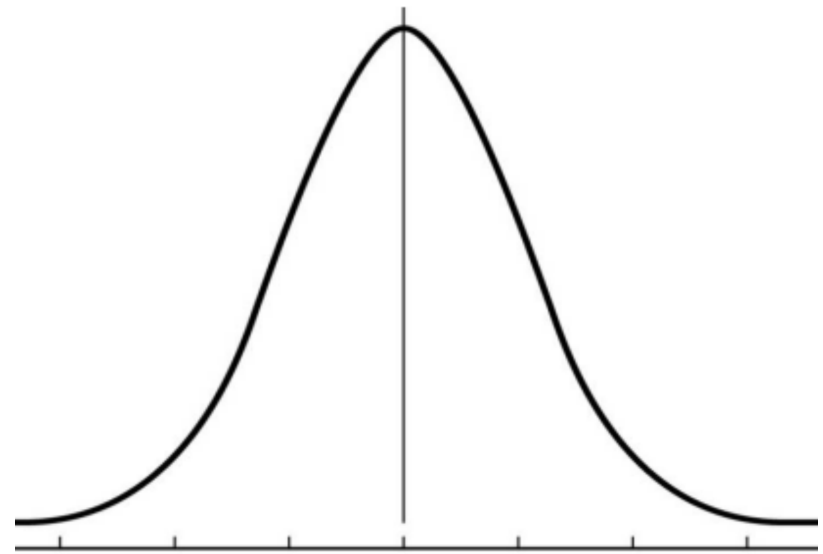


Theoretische Stichprobenverteilung

Wie gelangt man zu dieser theoretischen Stichprobenverteilung (obwohl man ja meist nur eine einzige Studie durchgeführt hat)?

Dazu benötigen wir drei Informationen:

1. Wie sieht die **Form** der Verteilung aus?
2. Was ist der **Mittelwert** der Verteilung?
3. Was ist die Breite (**Streuung**) der Verteilung?



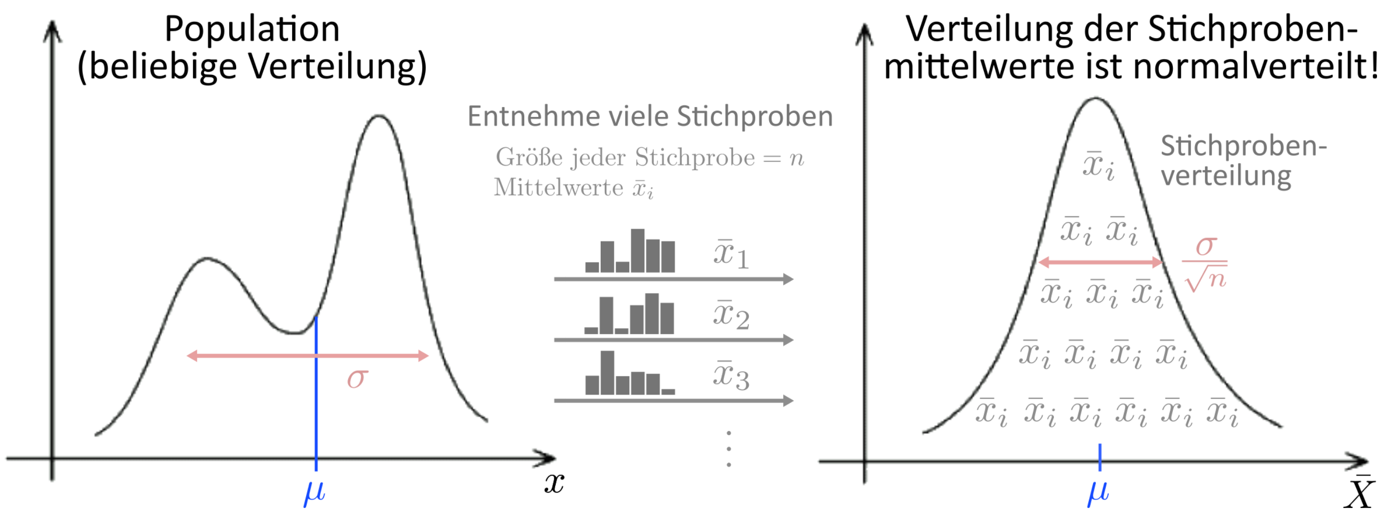
Form der theoretischen Stichprobenverteilung

- Wir haben bereits den **zentralen Grenzwertsatz** kennengelernt, demzufolge viele Merkmale in der Psychologie normalverteilt sind, weil sie sich aus einer Summe von Zufallseffekten (Genetik, Umwelt, Erziehung, usw.) zusammensetzen.
- Eine wesentliche Erkenntnis ist, dass statistische Kennwerte wie Mittelwert oder Varianz ebenfalls auf Summen von Zufallseffekten basieren.

Beispiel Mittelwert:

$$\bar{x} = \frac{1}{n} \sum x_i$$

Mit anderen Worten: ziehen wir sehr viele Stichproben aus der Population, berechnen für jede Stichprobe einen statistischen Kennwert (in Bezug auf die betrachtete Merkmalsvariable), so sind diese Kennwerte **normalverteilt** — und zwar unabhängig von der Verteilung der ursprünglichen Merkmalsvariable in der Population!



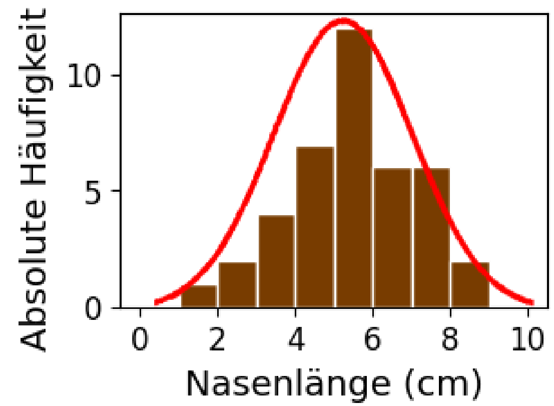
Form der theoretischen Stichprobenverteilung: Würfelexperiment

Form der theoretischen Stichprobenverteilung

Die Form der theoretischen Stichprobenverteilung (SV) ist also geklärt:

$$X \sim \mathcal{N}(\mu_{\mathrm{SV}}, \sigma_{\mathrm{SV}})$$

(in Worten: *wir nehmen an, dass unser Merkmal* X *aus einer Normalverteilung* $\mathcal{N}$ *mit Mittelwert* μ_{SV} *und Standardabweichung* σ_{SV} *gezogen ist*)



Auch wenn wir in den Beispielen vornehmlich den Mittelwert als Stichprobenkennwert betrachten, sind alle statistischen Kennwerte, die wir bis jetzt kennengelernt haben ($\bar{x}, s, r$, usw), zumindest *im Grenzfall* $n \to \infty$ normalverteilt.
Für unterschiedliche Kennwerte gibt es dabei unterschiedliche Faustregeln, ab welcher Stichprobengröße von einer Normalverteilung ausgegangen werden kann. Für den Mittelwert ist eine häufig genannte Regel $n = 30$.

Mittelwert der theoretischen Stichprobenverteilung

Die beste Schätzung $\hat{\mu}_{\mathrm{SV}}$ für den wahren Mittelwertsparameter μ_{SV} der Stichprobenverteilung (und damit den Mittelwert der Population) ist häufig der statistische Kennwert selbst (z.B. $\bar{x}$, s)!

Wichtig: diese Annahme gilt nur, wenn die Stichprobenverteilung tatsächlich einer Normalverteilung folgt!

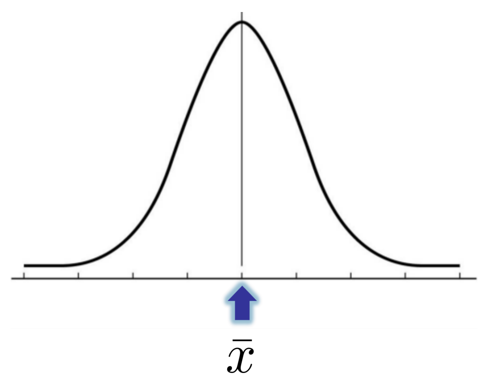
Für den Mittelwert kann man diese Normalverteilungs-Näherung bereits bei kleinem n annehmen (~30). Bei anderen Größen, v.a. solchen, die auf einen endlichen Bereich beschränkt sind (z.B. Korrelation −1 bis +1, relative Häufigkeiten 0 bis 1), gilt die Normalverteilungs-Näherung bei kleineren Stichprobengrößen nicht ohne Weiteres.
In diesem Fall werden andere Funktionen für die Stichprobenverteilung angenommen, bei denen der wahrscheinlichste Wert (Modus) dann u.U. *nicht* der statistische Kennwert selbst ist. (⇒ Vorlesung 12)

Beispiel

Beispiel Mittelwert. Ist der Mittelwert der betrachtete statistische Stichprobenkennwert so gilt:

$$\hat{\mu}_{\mathrm{SV}} = \bar{x}$$

→ Die theoretische Stichprobenverteilung wird also in diesem Fall um den Stichprobenmittelwert $\bar{x}$ herum konstruiert. Der Mittelwert ist die beste Schätzung für den Mittelwert der Stichprobenverteilung *auch dann* wenn die Stichprobengröße sehr klein ist und nicht mehr von einer Normalverteilung ausgegangen werden kann.



Streuung der theoretischen Stichprobenverteilung

Bleibt die Frage nach dem Streuungsparameter σ_{SV}: woher wissen wir, wie die Ergebnisse von hypothetischen Stichproben streuen würden?

Gehen wir dazu zu unserem Gedankenexperiment zurück:

Population (Nasenlängen in cm)

Was würde die Streuung der möglichen Ergebnisse verkleinern?

1. Wir ziehen nicht 3 Personen in unserer Studie, sondern 6
   → damit lägen fast alle Stichproben näher am wahren Mittelwert!
2. Wir hätten in der Population grundsätzlich eine geringere Streuung der Werte
   → damit würden auch die Mittelwerte einzelner Studien weniger streuen.

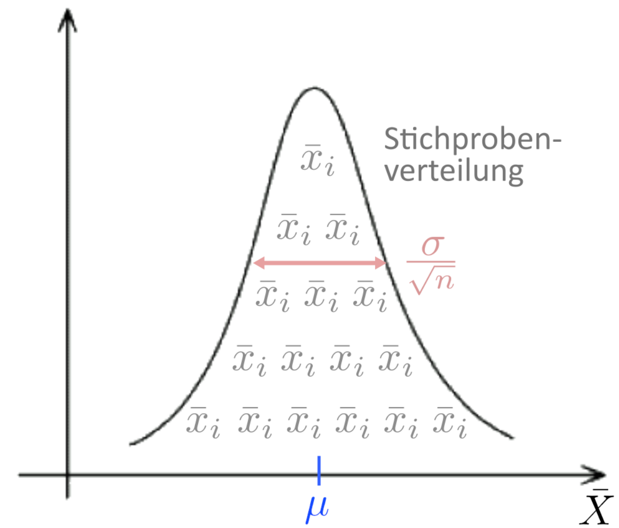
Der Streuungsparameter σ_{SV} der theoretischen Stichprobenverteilung muss also eine Funktion der Stichprobengröße n und der Streuung σ in der Population sein.

$$\hat{\sigma}_{\mathrm{SV}} = \hat{se} = f(n, \sigma)$$

Als Begriff für die Streuung der Stichprobenverteilung hat sich **Standardfehler** (engl. *standard error*) eingebürgert.

Beispiel: Standardfehler des Mittelwertes

- Ein für Statistik 1 besonders wichtiger Standardfehler ist der Standardfehler des Kennwertes *Mittelwert*.
- Er wird als **Standardfehler des Mittelwertes** oder engl. *standard error of the mean* (*SEM*) bezeichnet.
- Er berechnet sich als die Standardabweichung der Population σ geteilt durch die Wurzel aus der Stichprobengröße n ("Wurzel-N-Gesetz"):



$$\text{Standardfehler des Mittelwertes:} \qquad \hat{\sigma}_{\text{SV}} = \hat{se} = \frac{\hat{\sigma}}{\sqrt{n}}$$

- Da der wahre Streuungsparameter σ der Population unbekannt ist, verwenden wir die Schätzung $\hat{\sigma}$ auf Basis der Streuung der Stichprobendaten (→ dazu kommen wir gleich noch einmal).
- Intuitiv sagt der Standardfehler des Mittelwertes aus, wie sicher wir uns bei der Bestimmung des Mittelwertes sein können
  - Großer Standardfehler: Gemessener Mittelwert ist eher unsicher
  - Kleiner Standardfehler: Gemessener Mittelwert ist eher sicher

Beispiel: Standardfehler des Mittelwertes

Beachte: der Standardfehler des Mittelwertes allein sagt nichts über die tatsächliche Streubreite des Merkmals. Der Standardfehler kann klein sein, obwohl die Standardabweichung σ des gemessenen Merkmals groß ist, wenn im Gegenzug die Stichprobengröße n groß ist. Umgekehrt kann ein Standardfehler groß sein, obwohl die Standardabweichung σ klein ist, wenn n klein ist.

Herleitung des Standardfehlers

- Der Standardfehler ist ein Maß für die **Variabilität der Stichprobenmittelwerte** $\bar{x}$ — dies können wir zunächst über die Varianz zum Ausdruck bringen:

$$se^2 = Var(\bar{x})$$

- Wir wissen, dass $\bar{x} = \frac{1}{n} \sum X_i$, also:

$$se^2 = Var(\bar{x}) = Var\left(\frac{1}{n} \sum X_i\right)$$

- Um das $\frac{1}{n}$ aus der Varianz herausziehen zu können, versichern wir uns einer kleinen Rechenregel:

$$Var(aX) = \frac{1}{n}(aX_i - a\bar{x})^2 = \frac{a^2}{n}(X_i - \bar{x})^2 = a^2 Var(X)$$

- Daraus folgt:

$$se^2 = \frac{1}{n^2} Var\left(\sum X_i\right)$$

KEIN KLAUSUR STOFF

Herleitung des Standardfehlers

$$\text{Zwischenergebnis} \qquad se^2 = \frac{1}{n^2} Var\left(\sum X_i\right)$$

- Die Summe in der Varianz stört noch. Glücklicherweise gilt, dass die Varianz der Summe von unabhängigen Zufallsvariablen X_i gleich der Summe der Varianzen ist, d.h.

$$Var\left(\sum X_i\right) = \sum Var(X_i)$$

- Daraus folgt:

$$se^2 = \frac{1}{n^2} \sum Var(X_i) = \frac{1}{n^2} \left(n \cdot Var(X_i)\right) = \frac{1}{n} Var(X_i)$$

- Nun sind wir fast am Ziel. Da die Varianz der X_i nichts anderes als die quadrierte Standardabweichung σ^2 ist, gilt:

$$se^2 = \frac{\sigma^2}{n} \qquad \text{bzw.} \qquad se = \frac{\sigma}{\sqrt{n}}$$

KEIN KLAUSUR STOFF

Übersicht Standardfehler

Maß	Standardfehler	Einschränkung
Mittelwert	$\hat{se}(\bar{x}) = \frac{\hat{\sigma}}{\sqrt{n}}$	
Median	$\hat{se}(\tilde{x}) = \sqrt{\frac{\pi}{2}}\frac{\hat{\sigma}}{\sqrt{n}}$	Annahme: Normalverteilung von X
Varianz	$\hat{se}(s^2) = \sqrt{\frac{2}{n-1}}\hat{\sigma}^2$	Annahme: Normalverteilung von X
Standardabweichung	$\hat{se}(s) = \frac{\hat{\sigma}}{\sqrt{2(n-1)}}$	Näherung; Annahme: Normalverteilung von X
Korrelation	$\hat{se}(r) = \sqrt{\frac{1-r^2}{n-2}}$	Näherung; Hinweis: laut neuerer Forschung ist $\hat{se}(r) = \sqrt{\frac{1-r^2}{n-3}}$ sogar ein noch besserer Schätzer[2]
Cohen's d (abhängige Messungen)	$\hat{se}(d) = \sqrt{\frac{1}{n} + \frac{d^2}{2n}}$	Näherung
Cohen's d (unabhängige Messungen)	$\hat{se}(d) = \sqrt{\frac{n_1+n_2}{n_1 n_2} + \frac{d^2}{2(n_1+n_2)}}$	Näherung; Quelle[3]

Nützliches Paper[4]

KEIN KLAUSUR STOFF

Zwischenfazit

Die theoretische Stichprobenverteilung folgt einer **Normalverteilung** (falls n groß genug) mit einem **Mittelwert, der dem statistischen Kennwert entspricht**, und einer Standardabweichung, die sich aus der Populationsstreuung σ und der Stichprobengröße n berechnet (der sog. **Standardfehler**).

- Der Standardfehler gibt darüber Auskunft, wie verlässlich unsere Schätzung des statistischen Kennwertes ist.
- Wie wir noch sehen werden umfasst $1\hat{se}$ die mittleren 68% der möglichen Ergebnisse in der theoretischen Stichprobenverteilung.

Beispiel

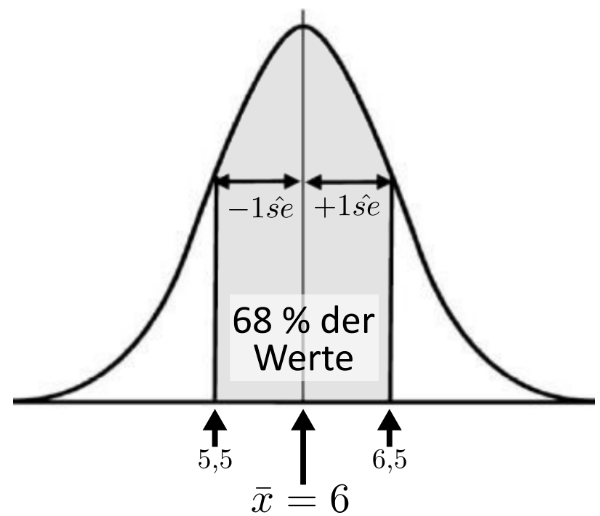
Nehmen wir an, die Nasenlängen der Männer in unserer Studie weisen eine durchschnittliche Länge von $6cm$ auf und einen Standardfehler (des Mittelwertes) von $0{,}5cm$.

Wir können damit sagen, dass der Bereich

$$\bar{x} \pm \hat{se} = 6 \pm 0{,}5 = [5{,}5; 6{,}5]$$

68% der Stichprobenverteilung umfasst.

In Vorlesung 12 werden wir noch feststellen, dass wir (leider) <u>nicht</u> schlussfolgern können, dass der wahre Populationsmittelwert μ mit 68% Wahrscheinlichkeit in diesem Intervall liegt.



Interpretation des Standardfehlers

Wie kann man den Wert eines Standardfehlers interpretieren?

- Lässt sich pauschal nicht beantworten, da der Standardfehler von der Messskala abhängt
  - Er ist keine standardisierte Größe wie z.B. der Korrelationskoeffizient.
- Prinzipiell gilt: je kleiner, desto besser (d.h. desto präziser ist unsere Kennwertschätzung)
- Anhaltspunkt: Vergleich/Verhältnis zur Skala (besonders einfach bei Ratingskalen wie 1-10) oder zur Standardabweichung in der Stichprobe:

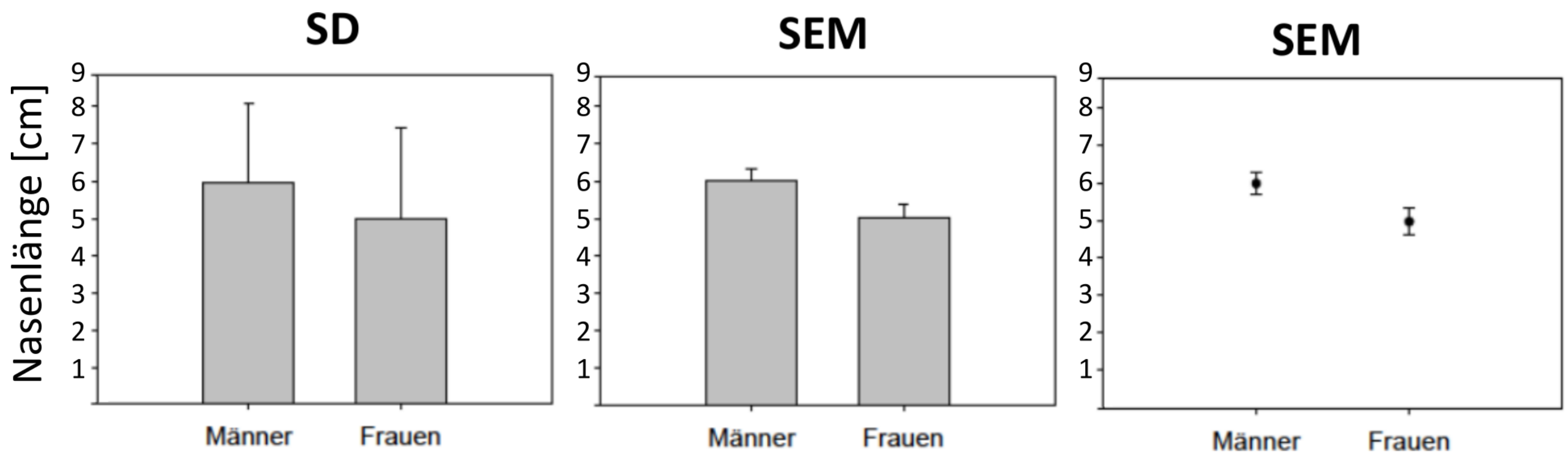
Beispiel

Nehmen wir wieder unser Nasenlängen-Beispiel mit $\bar{x} = 6cm$ und $\hat{se} = 0{,}5cm$, und nehmen wir an, die Standard*abweichung* von Nasenlängen in der Stichprobe betrug $5cm$. In diesem Fall hätten wir den Mittelwert mit einer Präzision von 10% der Streubreite in der Stichprobe geschätzt, was einer recht guten/präzisen Schätzung entspricht.

(Als kleine Übung: wie hoch müsste in diesem Beispiel die Stichprobenzahl gewesen sein? (Antwort: $n = 100$)

Verwendung des Standardfehlers in der Praxis

- Im Text wird der Standardfehler des Mittelwertes oft in folgender Form angegeben: $M = 3 \pm 0,6\ (\mathrm{SEM})$.
  - Wichtig: es sollte prinzipiell immer angegeben werden, um was für ein Streuungsmaß es sich handelt (SEM ist hier die geläufige englische Abkürzung für *standard error of the mean*).
- In Abbildungen wird der Standardfehler ähnlich wie die Standardabweichung häufig in Form von Fehlerbalken dargestellt:



- Ist das Hauptinteresse ob sich Experimentalbedingungen **in ihrem Mittelwert unterscheiden**, ist der **Standardfehler aussagekräftiger** als die Varianz oder Standardabweichung
  - Aus diesem Grund ist der Standardfehler des Mittelwertes das vielleicht häufigste Streuungsmaß in der Psychologie

Schätzung der Standardabweichung in der Population

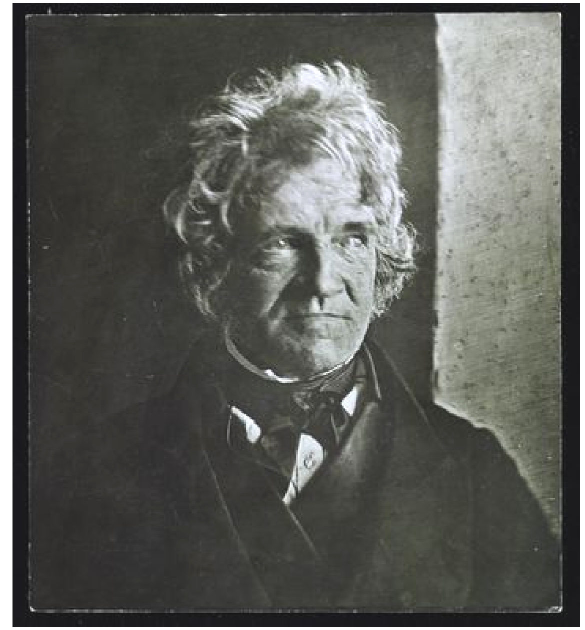
- Beim Mittelwert gilt, dass der Mittelwert $\bar{x}$ der Stichprobe die beste Schätzung $\hat{\mu}$ für den wahren Mittelwertsparameter μ der Population ist:

$$\text{Schätzung des Populationsmittelwertes:} \qquad \hat{\mu} = \bar{x}$$

- Bei der Varianz/Standardabweichung gilt dies nur (näherungsweise) für große Stichproben. Für kleinere Stichproben gilt leider nicht $\hat{\sigma}^2 = s^2$.
- Grund: um die Streuung in der Population zu schätzen, müsste eigentlich die Streuung der x_i **um den wahren Mittelwert** μ berechnet werden.
- Das wahre μ kennen wir aber nicht, sondern lediglich unsere Schätzung $\bar{x}$.
- Es stellt sich heraus, dass die Streuung der x_i um den Mittelwert $\bar{x}$ der Stichprobe immer etwas kleiner ist, als es die Streuung der x_i um den wahren Wert μ wäre.
  - Das ist durchaus intuitiv, da der Mittelwert $\bar{x}$ *aus den* x_i *selbst* bestimmt wird.
- Dies führt zur **Besselkorrektur**

Besselkorrektur

- Der deutsche Naturwissenschaftler Friedrich Wilhelm Bessel zeigte, dass für eine Schätzung $\hat{\sigma}^2$ der Populationsstreuung auf Basis der Stichprobenstreuung s^2, der Faktor $\frac{1}{n}$ in der Varianzformel durch $\frac{1}{n-1}$ ersetzt werden muss:



Friedrich Wilhelm Bessel (1748-1846)[5]

$$\text{Schätzung der Populationsvarianz:}$$
$$\hat{\sigma}^2 = \frac{n}{n-1}s = \frac{1}{n-1}\sum(x_i - \bar{x})^2$$

- Wir bezeichnen $\hat{\sigma}^2$ als **Schätzung der Populationsvarianz auf Basis der Stichprobe** oder kurz **biaskorrigierte Stichprobenvarianz**.
- Intution: die Varianz wird im Fall von $\frac{1}{n-1}$ *durch weniger geteilt* als beim ursprünglichen Faktor $\frac{1}{n}$, d.h. die resultierende Varianz wird *zu einem größeren Wert hin* korrigiert.
- Die "$n-1$"-Korrektur für Stichproben wird als **Besselkorrektur** bezeichnet.
- Ab ca. $n = 30$ spielt die Besselkorrektur kaum eine Rolle mehr ($\frac{1}{30} = 0.033$ vs. $\frac{1}{29} = 0.034$)

Herleitung des Faktors $\frac{1}{n-1}$

- Der Ausgangspunkt der Besselkorrektur die Beobachtung, dass die (unkorrigierte) Varianz s^2 der Stichprobe die Varianz σ^2 der Population systematisch unterschätzt.
- Der Grund ist liegt im Stichprobenmittelwert $\bar{x}$: er ist keine perfekte Schätzung für den wahren Mittelwert μ ist — er streut um das “wahre” μ.
- Diese Streuung haben wir gerade kennengelernt — es ist der Standardfehler $\hat{se} = \frac{\hat{\sigma}}{\sqrt{n}}$. Als Varianz formuliert:

$$\hat{se}^2 = \frac{\hat{\sigma}^2}{n}$$

- Für die Schätzung der Populationsvarianz $\hat{\sigma}^2$ muss nun die *zusätzliche* Variabilität $\hat{se}^2$ aufgrund der Unsicherheit des Mittelwertes auf die Stichprobenvarianz s^2 addiert werden:

$$\hat{\sigma}^2 = s^2 + \hat{se}^2 = s^2 + \frac{\hat{\sigma}^2}{n} \quad \longrightarrow \quad \hat{\sigma}^2 - \frac{\hat{\sigma}^2}{n} = s^2$$

- Linke Seite umformen:

$$\hat{\sigma}^2 - \frac{\hat{\sigma}^2}{n} = \hat{\sigma}^2 \left(1 - \frac{1}{n}\right) = \hat{\sigma}^2 \frac{n-1}{n} \overset{!}{=} s^2 \quad \longrightarrow \quad \hat{\sigma}^2 = \frac{n}{n-1} s^2$$

KEIN KLAUSUR STOFF

Herleitung des Faktors $\frac{1}{n-1}$

$$\text{Zwischenergebnis:} \qquad \hat{\sigma}^2 = \frac{n}{n-1} s^2$$

- Die unkorrigierte Varianz s^2 lautet:

$$s^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

- Einsetzen:

$$\hat{\sigma}^2 = \frac{n}{n-1} s^2 = \frac{n}{n-1} \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

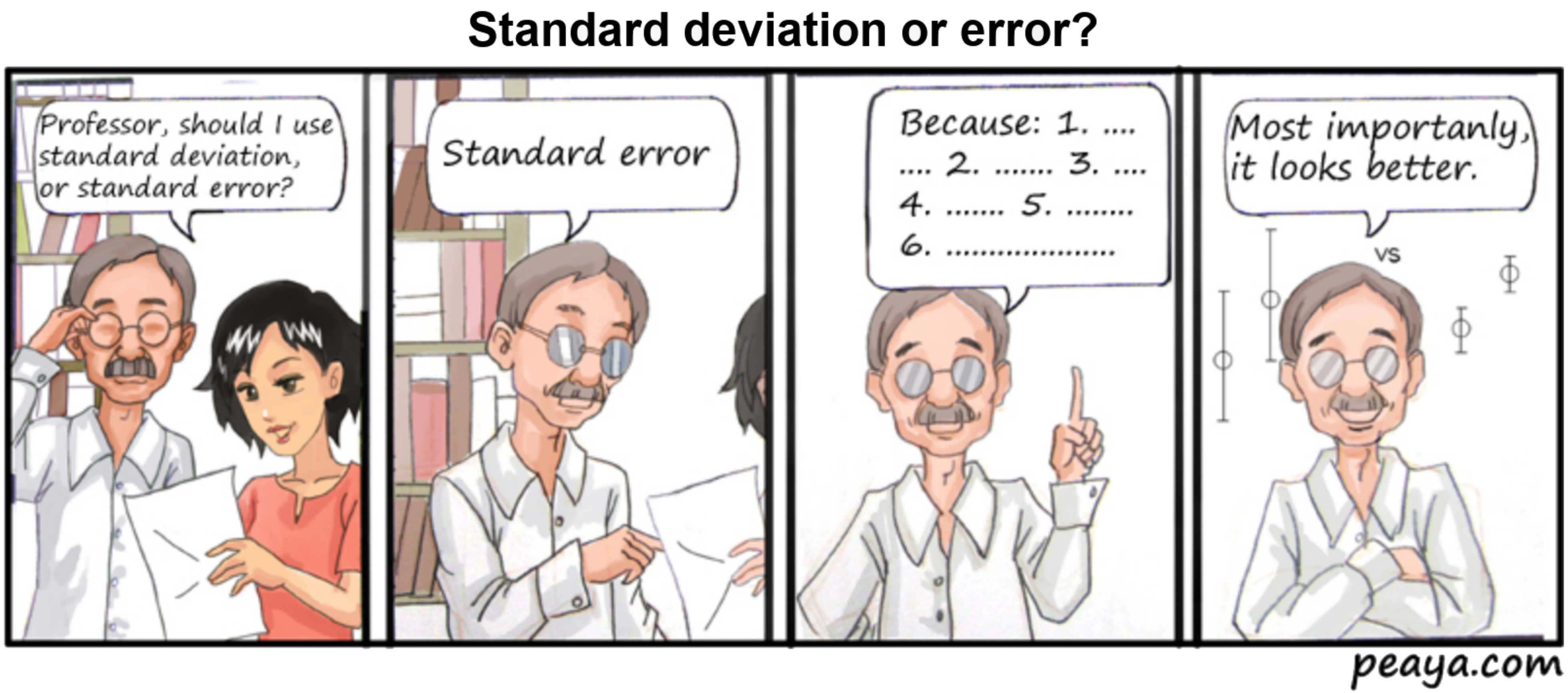
- Nun haben wir unseren Faktor $\frac{1}{n-1}$.
- Für die korrigierte Standardabweichung der Stichprobe wird häufig analog angenommen:

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

- Diese analoge Korrektur für die Standardabweichung ist nicht 100% gültig[6] — für alle praktischen Zwecke reicht sie aber aus.

[[Zusammenfassung]]

- Im Zuge der Inferenzstatistik wird untersucht, wie gut **Schätzungen von Parametern in der Population** auf Basis von **Stichprobenkennwerten** sind.
- Es geht also um die **Verallgemeinerbarkeit** von Stichprobendaten auf die Population
- Generelle Idee: Was würde passieren, wenn die Studie immer wieder durchgeführt und der Kennwert bestimmt würde?
- Diese Idee wird durch die **theoretische Stichprobenverteilung** repräsentiert.
- Die Stichprobenverteilung von Kennwerten ist laut dem Zentralen Grenzwertsatz **normalverteilt**.
- Die Standardabweichung dieser Stichprobenverteilung wird **Standardfehler** genannt.



Bildnachweis[7]

Fußnoten

1.

https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Introductory_Statistics_(Shafer_and_Zhang)/01%3A_Introduction_to_Statistics/1.01%3A_Basic_Definitions_and_Conc

2. Gnambs T. A Brief Note on the Standard Error of the Pearson Correlation. https://psyarxiv.com/uts98/

3. n[3]:

4. Harding B, Tremblay C, Cousineau D (2014) Standard errors: A review and evaluation of standard error estimators using Monte Carlo simulations. TQMP 10:107–123.

5. http://www.digiporta.net/index.php?id=658493715

6. https://en.wikipedia.org/wiki/Unbiased_estimation_of_standard_deviation

7. http://www.peaya.com/peaya.php?comicsid=1005