M24 Statistik 1: Wintersemester 23/24

Vorlesung 08: Inferenzstatistik

Prof. Matthias Guggenmos

Health and Medical University Potsdam

Statistischer Kennwert $\hat{\theta}$

- Wir führen ein neues Symbol ein, das von nun an Platzhalter für einen statistischen Kennwert steht, der auf Basis einer Stichprobe berechnet wurde:

$$\text{Statistischer Kennwert:} \qquad \hat{\theta}$$

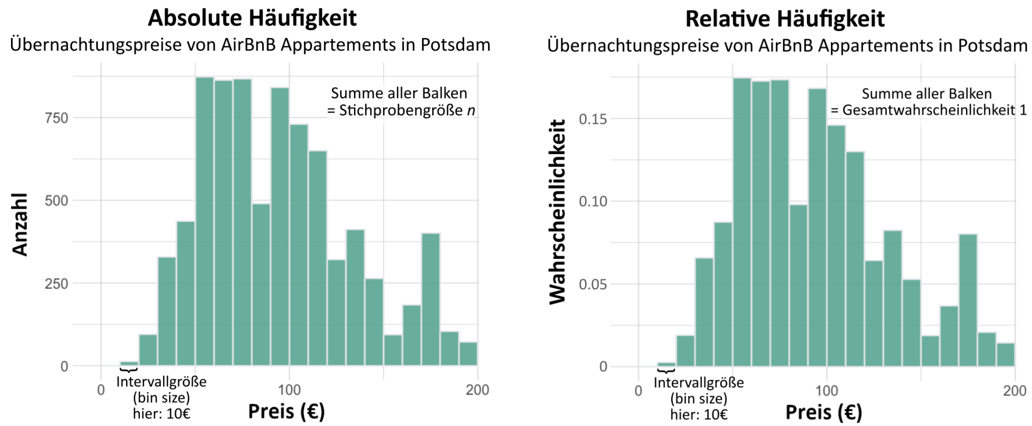
Definition	Als statistische Kennwerte werden quantitative Maße bezeichnet, die eine Eigenschaft von Stichprobendaten in einer Zahl zusammenfassen. Das Symbol $\hat{\theta}$ dient dabei als allgemeines Symbol statistische Kennwerte.

- Zu statischen Kennwerten zählen nicht nur Lage- und Streuungsmaße, die *eine* Variable beschreiben (z.B. Mittelwert $\bar{x}$, Varianz σ^2), sondern auch *zwei* Variablen (z.B. Korrelation $\hat{\rho}$, Regressionskoeffizient $\hat{b}_1$) oder noch mehr Variablen ($\Rightarrow$ Statistik 2).

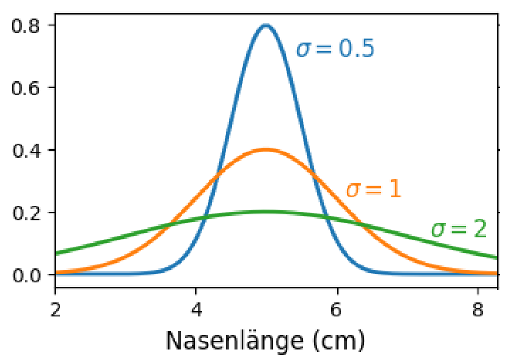
Stichprobenverteilung

Verteilung

- **Verteilungen** kennen wir bereits aus der Vorlesung 04 zu Lage- und Streuungsmaßen.
- Wir unterscheiden zwischen **empirischen Verteilungen**, die etwa in der Form von Histogrammen dargestellt werden…



- … und **theoretischen Verteilungen**, die durch eine mathematische Funktion $f(x)$ definiert sind und für jede Merkmalsausprägung x die Häufigkeit $f(x)$ angeben:



Normalverteilung:

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

Stichprobenverteilung

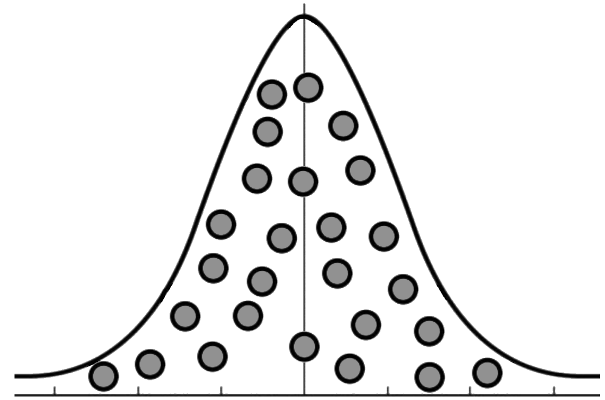
Empirische und theoretische Verteilungen gibt es **nicht nur für Merkmale X, sondern auch für statistische Kennwerte** $\hat{\theta}$, die für das Merkmal X bestimmt wurden — zum Beispiel Mittelwert $\bar{x}$.

Prinzip: wir nehmen nicht nur eine einzelne Studie an, sondern viele Studien $i = 1...k$, die jeweils Mittelwerte $\bar{x}_i$ bestimmt haben. Die Mittelwerte $\bar{x}_i$ folgen ebenfalls einer Verteilung — der **Stichprobenverteilung**. Dieses Prinzip gilt nicht nur für den Mittelwert, sondern für alle statistischen Kennwerte $\hat{\theta}$ (Median, Varianz, Kovarianz, Korrelation, usw.).

Stichprobenverteilung

Empirische Stichprobenverteilung

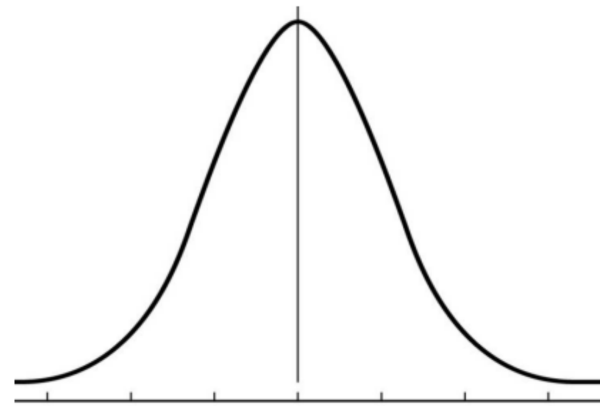
- Führe ich dieselbe Studie mehrmals durch und notiere jeweils den statistischen Kennwert (z.B. Mittelwert), erhalte ich eine empirische Stichprobenverteilung.
- Dies ist die Idee der **Metaanalyse**, die eine Vielzahl empirischer Studien zusammenfasst und analysiert (⇒ Vorlesung 13).



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.

Theoretische Stichprobenverteilung

- Ich habe nur *eine* Studie, aber überlege, was theoretisch passieren würde, wenn ich diese Studie immer wieder wiederholen würde.
- Die resultierende Verteilung ist die theoretische Stichprobenverteilung.
- Die theoretische Stichprobenverteilung erlaubt uns eine Einschätzung darüber, wie stabil unser Ergebnis bei einer (hypothetischen) Wiederholung der Studie sein würde.
- Dies ist der Ansatz der **Inferenzstatistik**.



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

Gedankenexperiment: Wir nehmen an, die Population besteht nur aus 9 Männern und wir kennen von allen Männern die Nasenlänge:

In diesem Gedankenexperiment kennen wir also den wahren Mittelwert der Population. Er beträgt $\mu = 6cm$.

Nun betrachten wir eine Studie, in der 3 Männer untersucht werden. Wir ziehen also eine zufällige **Stichprobe** $n = 3$ 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:

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

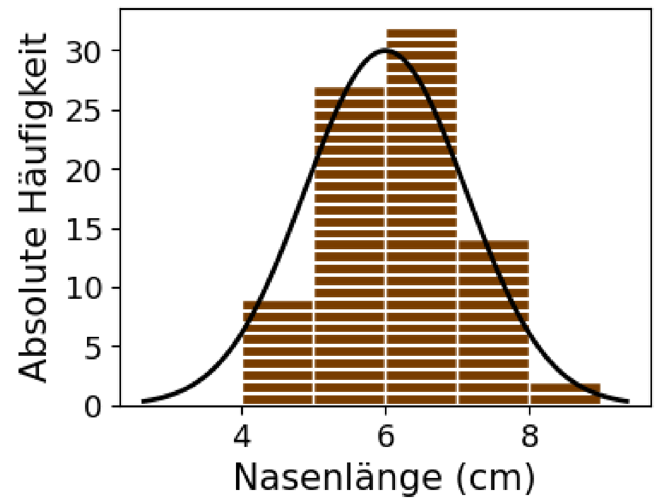
Theoretische Stichprobenverteilung: Beispiel

... oder diese drei Männer:

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

… und so weiter

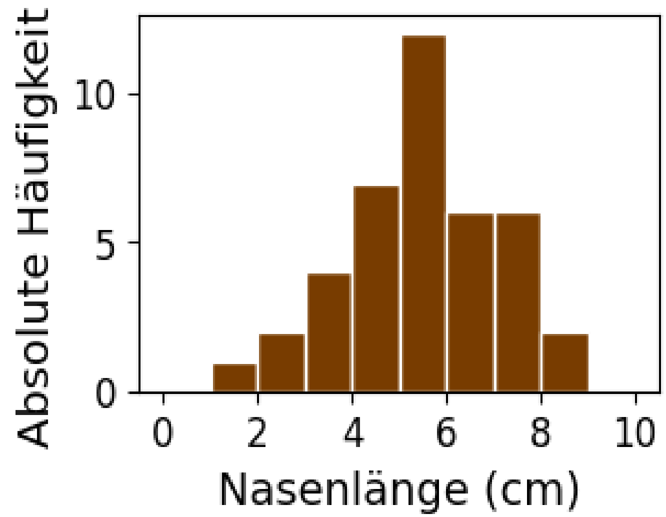
- Jeder Mittelwert wäre eine Schätzung für den wahren Populationswert.
- Die gesammelten Mittelwerte all dieser hypothetischen Studien können nun ebenfalls in ein Histogramm eingetragen werden (in braun).
- Wie wir noch sehen werden, lässt sich diese **Stichprobenverteilung** auch mathematisch beschreiben — im Bild rechts schon einmal durch die schwarze Kurve angedeutet.



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 eine *Bandbreite*.
- Die Bandbreite gibt einen Anhaltspunkt für die Genauigkeit der Schätzung des Populationsmittelwertes, die mit einer einzelnen Stichprobe erzielt werden kann.
- Ein wichtiges Ziel der Inferenzstatistik ist, diese Bandbreite mit mathematischen Methoden abzuschätzen, so dass nicht wie im Gedankenexperiment tatsächlich viele Wiederholungen einer Studie notwendig sind.

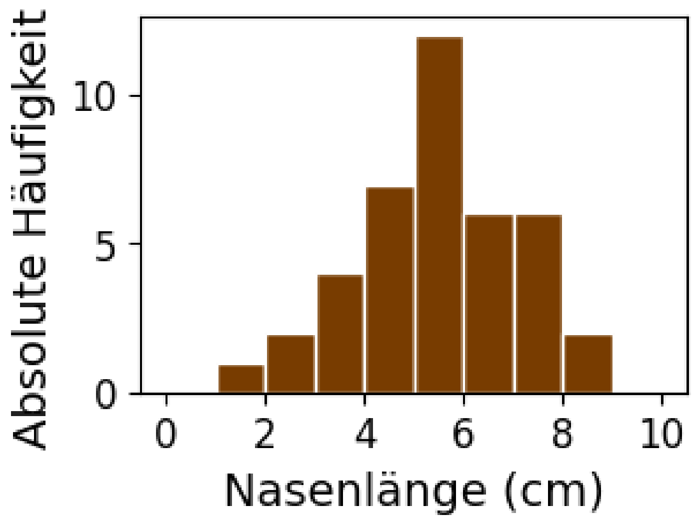


Stichprobenverteilung der Mittelwerte

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).



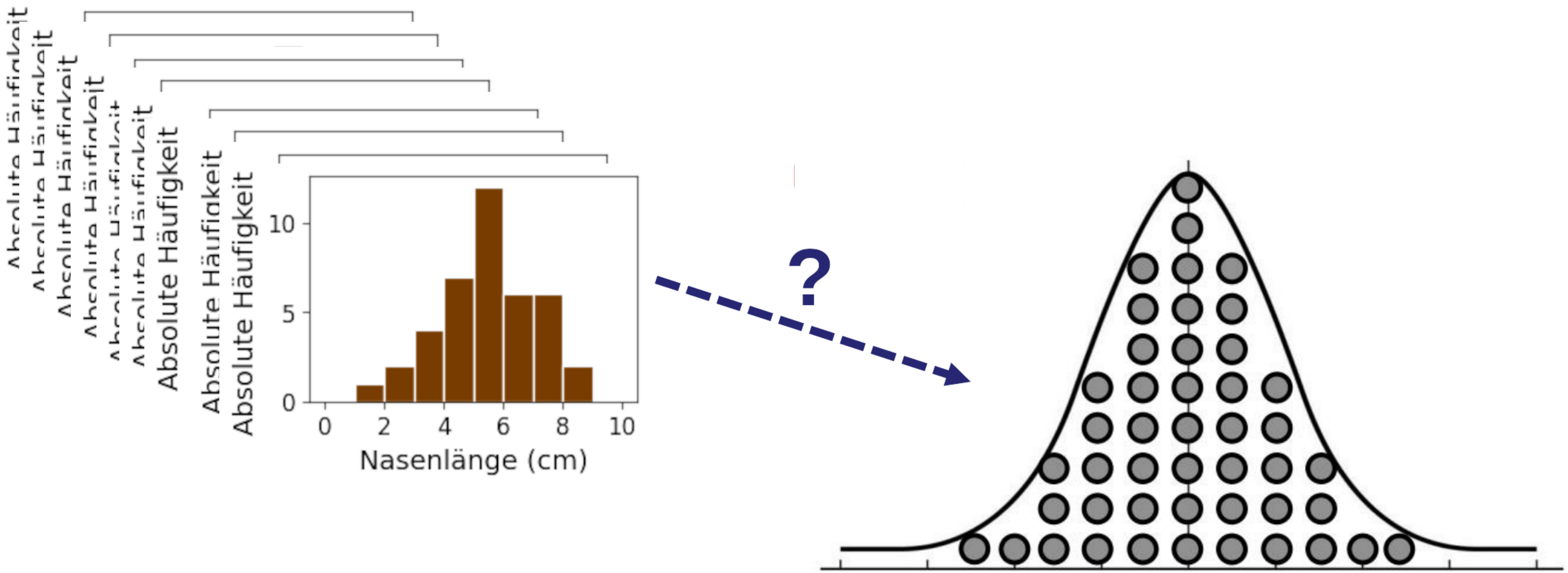
Achtung: das Histogramm zeigt nun im ersten Schritt wieder die Verteilung der Daten in einer einzelnen Studie!

- 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, nicht *alle möglichen* Stichproben betrachten kann?

Mit anderen Worten: kann man abschätzen, wie die Verteilung von Stichprobenkennwerten erwartbar aussehen würde, würden wir die Studie – rein hypothetisch – **unendlich oft wiederholen**? Die Antwort lautet JA und führt über die mathematischer Herleitung der **theoretischen Stichprobenverteilung**.



Theoretische Stichprobenverteilung

- Als erste Frage stellt sich: durch welche grundlegende **mathematische Funktion** lässt sich die theoretische Stichprobenverteilung beschreiben?
- Die Antwort auf diese Frage lässt sich aus dem **zentralen Grenzwertsatz** ableiten, demzufolge viele natürliche Merkmale **normalverteilt** sind, weil sie sich aus einer **Summe von Zufallseffekten** (Genetik, Umwelt, Erziehung, usw.) zusammensetzen.
- Die zentrale Erkenntnis ist nun, dass sich die gleiche Logik — **Summe von Zufallseffekten** — auf statistische Kennwerte $\hat{\theta}$ wie den Mittelwert übertragen lässt!

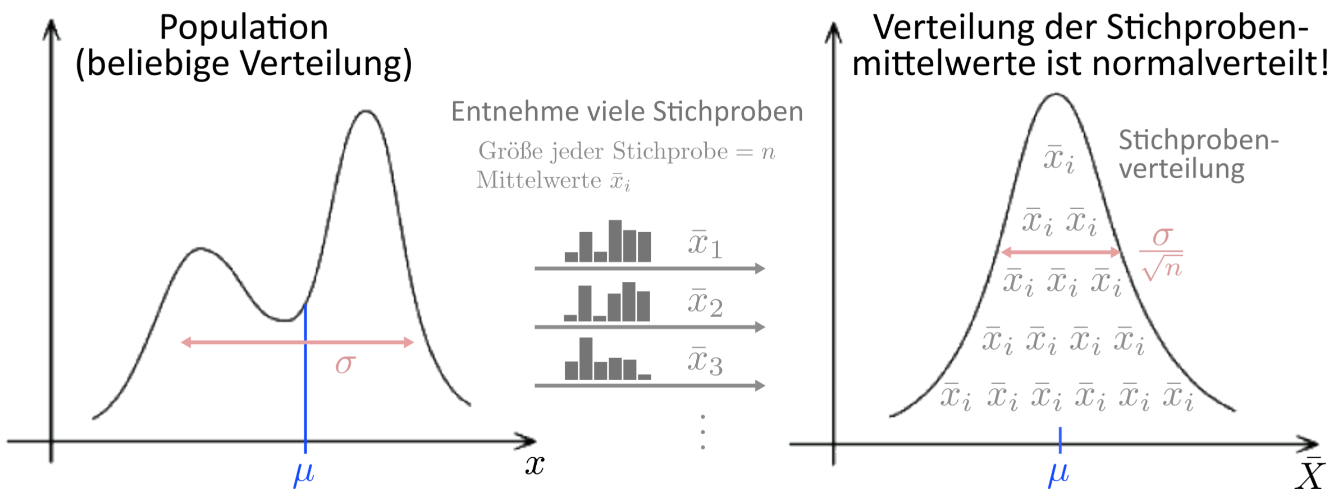
Beispiel: statistischer Kennwert $\hat{\theta}$ = Mittelwert $\bar{x}$

Nehmen wir $j = 1..k$ hypothetische Studien an, die jeweils einen Mittelwert $\bar{x}^{(j)}$ berechnen. Jeder Mittelwert basiert auf der **Summe** (Σ) von **zufällig gezogenen Daten** ($x_i^{(j)}$) aus einer Stichprobe. Gemäß dem zentralen Grenzwertsatz erwarten wir daher im Grenzfall (d.h. Stichprobengröße gegen ∞), dass die Mittelwerte einer **Normalverteilung** folgen.

$$\bar{x}^{(j)} = \frac{1}{n} \sum x_i^{(j)}$$

Theoretische Stichprobenverteilung

- Die theoretische Stichprobenverteilung eines statistischen Kennwertes $\hat{\theta}$ ist aus diesem Grund häufig die **Normalverteilung**.
- Nochmals in 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 häufig normalverteilt — und zwar **unabhängig von der Verteilung der Merkmalsvariable X in der Population**!



Theoretische Stichprobenverteilung

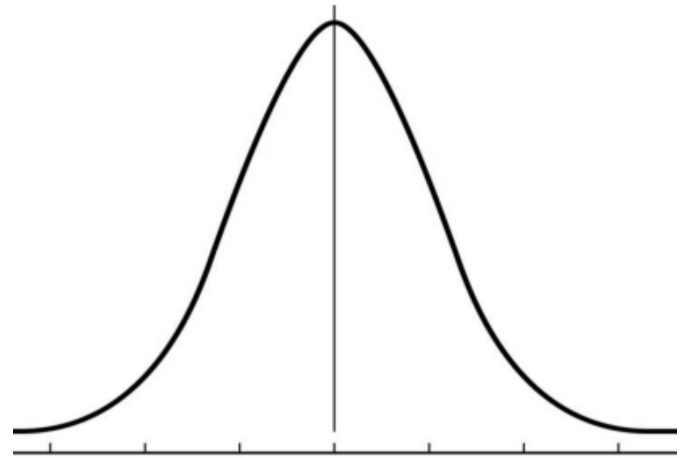
Zu beachten ist, dass der zentrale Grenzwertsatz streng genommen nur für den *Grenzwert* gilt, d.h. wenn die Stichprobengröße n sehr groß wird. Als Faustregel gilt für den Mittelwert etwa gilt, dass ab $n = 30$ die Stichprobenverteilung hinreichend genau durch die Normalverteilung beschrieben werden kann.

Bei anderen statistischen Kennwerten, 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 typischen Stichprobengrößen wie $n = 30$ nicht ohne Weiteres.
In diesem Fall werden andere — asymmetrische — Funktionen als die Normalverteilung für die Stichprobenverteilung angenommen ($\Rightarrow$ Vorlesung 12).

Theoretische Stichprobenverteilung

Die Form der theoretischen Stichprobenverteilung (SV) ist also geklärt (zumindest im Grenzfall $n \to \infty$): **Normalverteilung.**

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



Dass eine Zufallsvariable X normalverteilt ist, wird häufig auch mit folgender Notation zum Ausdruck gebracht

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

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

Zwei Informationen fehlen nun noch:

1. Was ist der **Mittelwert** (μ_{SV}) der theoretischen Stichprobenverteilung?
2. Was ist die **Streuung** (σ_{SV}) der theoretischen Stichprobenverteilung?

Mittelwert der theoretischen Stichprobenverteilung

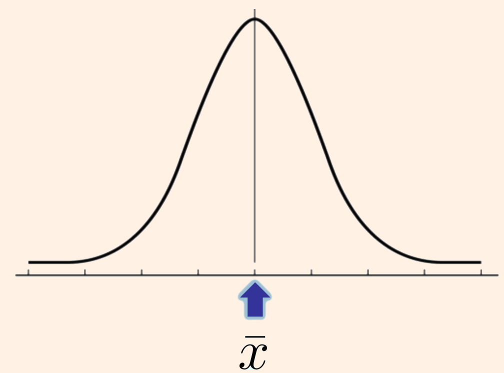
- Kann von einer Normalverteilung für die Form der Stichprobenverteilung ausgegangen werden, ist die beste Schätzung $\hat{\mu}_{\mathrm{SV}}$ für den Mittelwertsparameter μ_{SV} der Stichprobenverteilung der statistische Kennwert selbst (z.B. $\bar{x}$, s).

Beispiel: statistischer Kennwert $\hat{\theta}$ = Mittelwert $\bar{x}$

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 $\hat{\mu}_{\mathrm{SV}}$ der Stichprobenverteilung ist identisch mit der besten Schätzung des statistischen Kennwertes $\hat{\theta}$ für die Population.

Experiment zur theoretischen Stichprobenverteilung: Würfeln

Streuung der theoretischen Stichprobenverteilung

Bleibt die Frage nach dem Streuungsparameter σ_{SV} der theoretischen Stichprobenverteilung: 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 hypothetischen Einzelstichproben verkleinern?

1. Wenn die Stichprobengröße höher ist als lediglich $n = 3$ Personen (z.B. $n = 6$)
   → damit lägen die Mittelwerte der Einzelstichproben idR näher am wahren Mittelwert!
2. Wenn die Population grundsätzlich eine geringere Streuung σ aufweist
   → damit würden auch die Mittelwerte der Einzelstichproben weniger streuen.

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

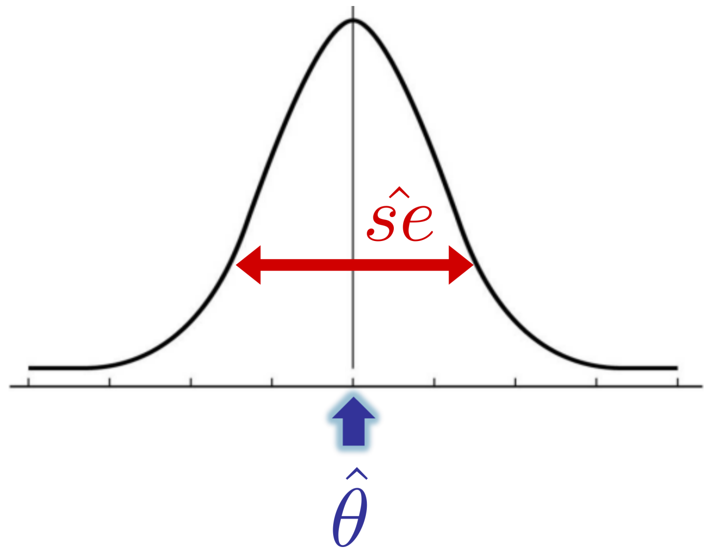
$$\hat{\sigma}_{\mathrm{SV}} = f(n, \sigma)$$

Streuung der theoretischen Stichprobenverteilung

- Kann für die Stichprobenverteilung eine Normalverteilung angenommen werden, so wird die Streuung $\hat{\sigma}_{\mathrm{SV}}$ als **Standardfehler** (engl. *standard error*) oder $\hat{se}$ bezeichnet.

$$\hat{\sigma}_{\mathrm{SV}} = \hat{se}$$

- Der Standard*fehler* ist die Standard*abweichung* der normalverteilten Stichprobenverteilung um den Mittelwert $\hat{\theta}$.



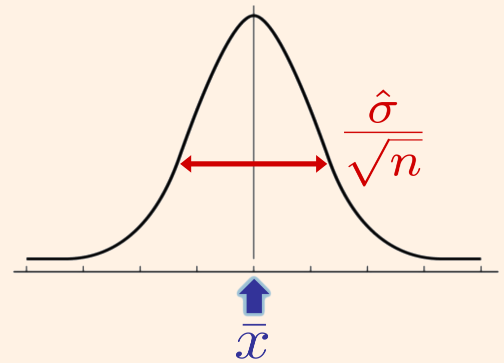
Beispiel: Standardfehler des Mittelwertes

Beispiel: statistischer Kennwert $\hat{\theta}$ = Mittelwert $\bar{x}$

Beim statistischen Kennwert “Mittelwert” berechnet sich der Standardfehler als Standardabweichung der Population σ geteilt durch die Wurzel aus der Stichprobengröße n (“Wurzel-N-Gesetz”):

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

In der Regel kennen wir die wahre Standardabweichung σ der Population nicht und schätzen sie deshalb (wie gehabt) als $\hat{\sigma}$ auf Basis der Stichprobe.



- 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

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

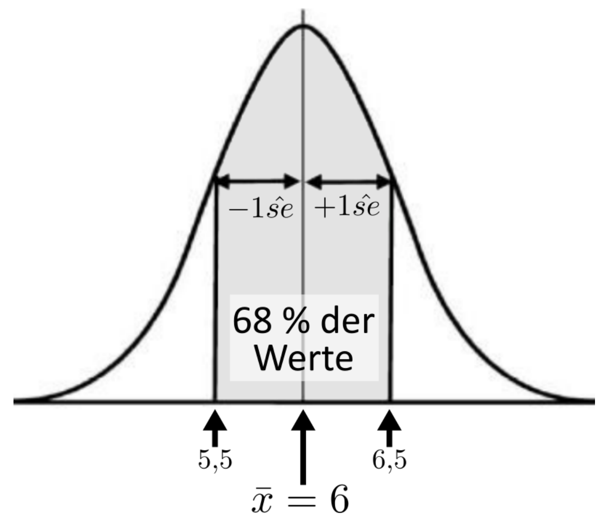
Nehmen wir an, die Nasenlängen der Männer in unserer Studie weisen eine durchschnittliche Länge von $6cm$ auf und einen Standardfehler (des Mittelwertes) von $0{,}5cm$.

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?

- Prinzipiell gilt: **je kleiner, desto präziser ist die Kennwertschätzung** auf Basis der Stichprobe.
- Allerdings ist der Standardfehler keine standardisierte Größe wie z.B. Cohen's d und hängt von den gewählten Einheiten der Variable X ab.
  - Interpretation ohne Kenntnis der Einheit/Messskala nicht möglich.
- Anhaltspunkt: Vergleich/Verhältnis des Standardfehlers zum Wertebereich Skala (z.B. Ratingskala 1-10) oder zur Standardabweichung in der Stichprobe:

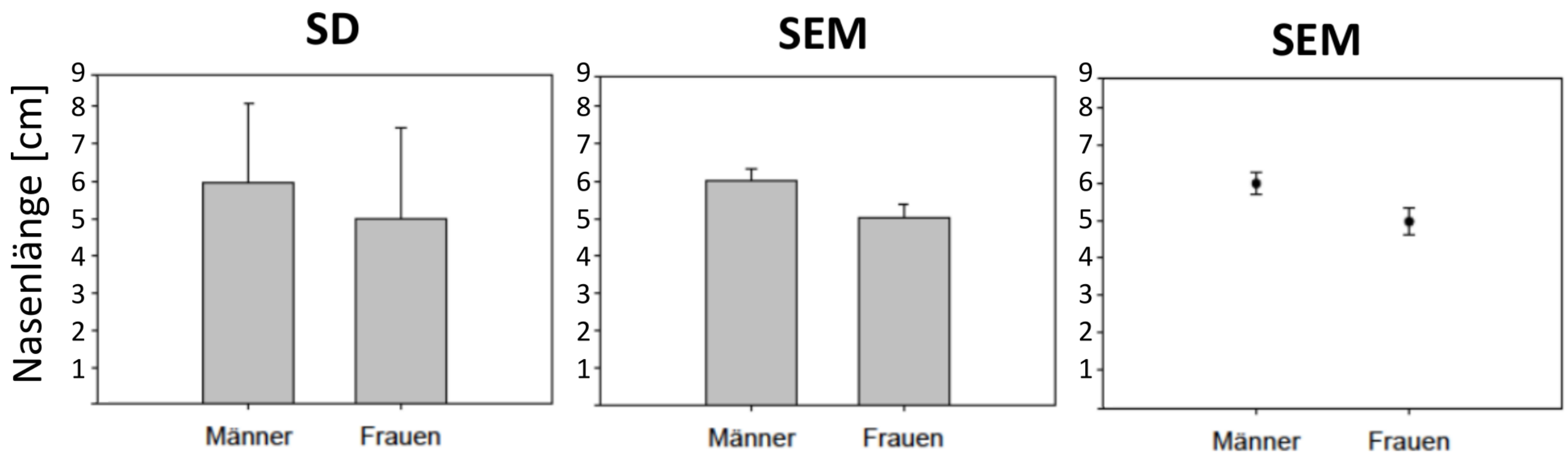
Beispiel

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: $M = 3.2 \pm 0.6 \ (\mathrm{SEM})$.
  - 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:

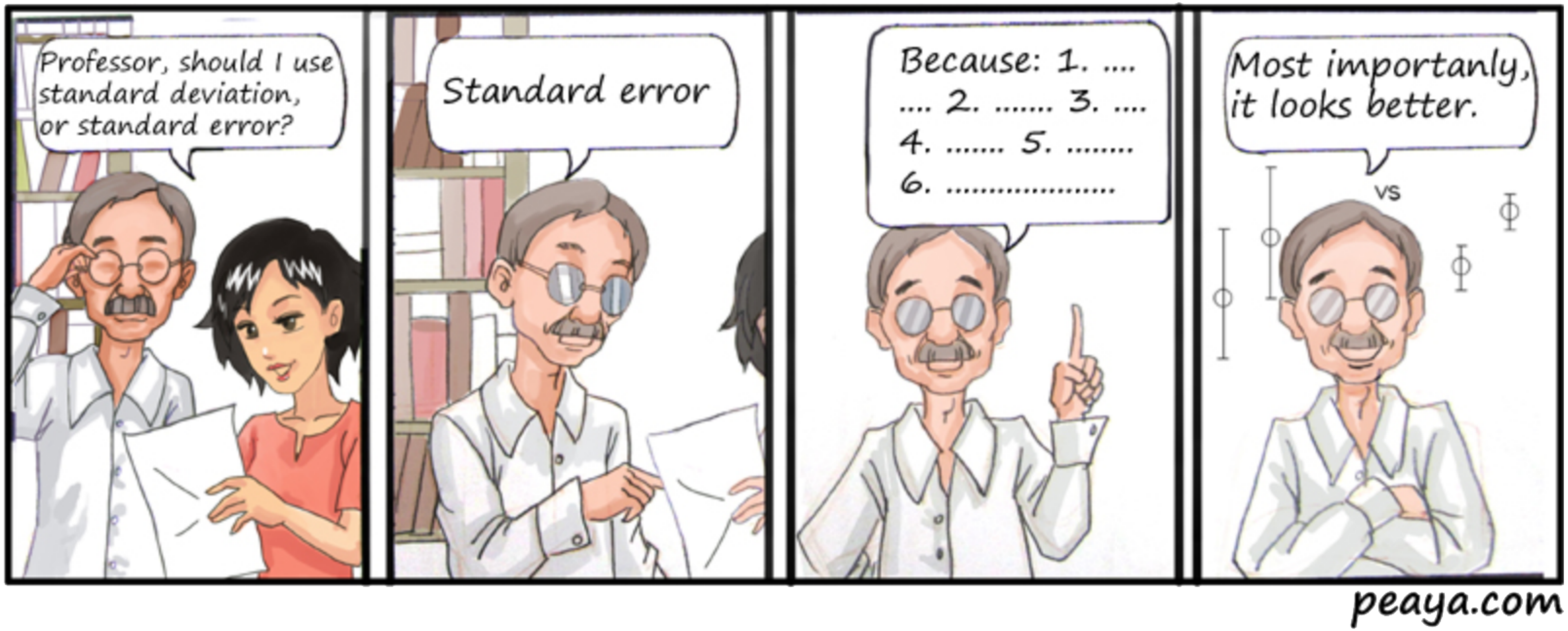


- Ist das Hauptinteresse ob sich Experimentalbedingungen **in ihrem Mittelwert unterscheiden**, ist der **Standardfehler aussagekräftiger** als die Varianz oder Standardabweichung
  - Aus diesem Grund ist der Standardfehler des Mittelwertes das vielleicht häufigste Streuungsmaß in der Psychologie

[[Zusammenfassung]]

- Das grundsätzliche Ziel der Inferenzstatistik ist es die **Verallgemeinerbarkeit** von Stichprobenkennwerten auf die Population zu untersuchen.
- Eine wichtige Frage ist dabei, wie präzise **Schätzungen von Populationskennwerten auf Basis von Stichprobenkennwerten** sind.
- Generelle Idee: Was würde passieren, wenn die Studie hypothetisch unendlich oft durchgeführt und jeweils der Kennwert bestimmt würde?
- Diese Idee wird durch die **theoretische Stichprobenverteilung** repräsentiert.
- Die Stichprobenverteilung von Kennwerten $\hat{\theta}$ kann aufgrund des Zentralen Grenzwertsatzes häufig als **normalverteilt** angenommen werden.
- Der Mittelwert $\hat{\mu}_{\mathrm{SV}}$ der normalverteilten Stichprobenverteilung ist der Kennwert $\hat{\theta}$ selbst und ihre Standardabweichung $\hat{\sigma}_{\mathrm{SV}}$ wird als **Standardfehler** $\hat{se}$ bezeichnet.
- Beispiel Kennwert $\hat{\theta}$ = Mittelwert $\bar{x}$: $\hat{\mu}_{\mathrm{SV}} = \bar{x}, \quad \hat{\sigma}_{\mathrm{SV}} = \frac{\hat{\sigma}}{\sqrt{n}}.$

Standard deviation or error?



Bildnachweis[1]

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:

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

- 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

Fußnoten

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