Statistik für Data Scientists

Vorlesung 6: Schätzen & Konfidenzintervalle

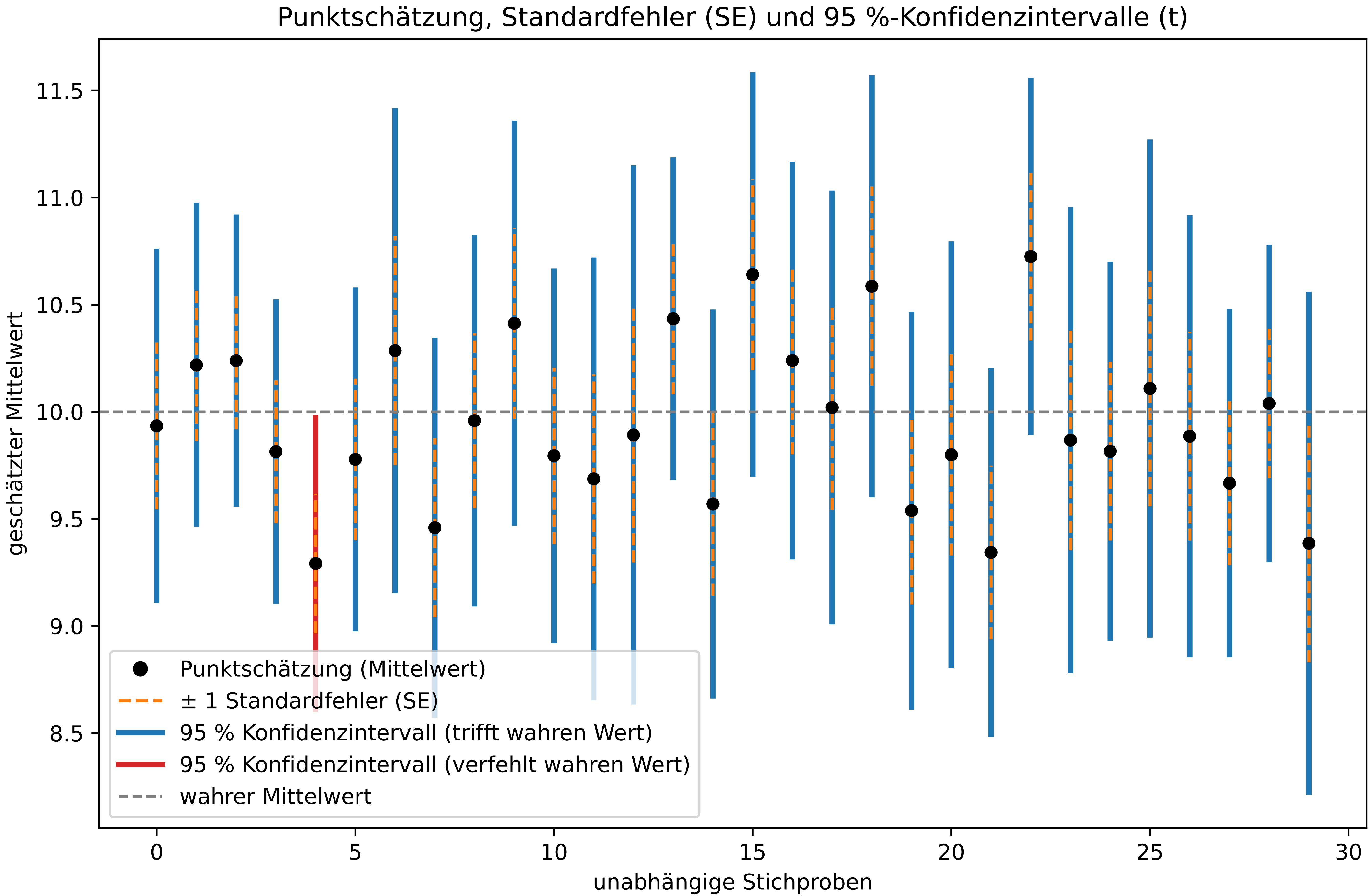
Prof. Dr. Siegfried Handschuh
DS-NLP
Universität St. Gallen

Recap & Ziele heute

- Recap: Wahrscheinlichkeit & Verteilungen.
- Heute: Punktschätzung, Standardfehler, Konfidenzintervalle.
- Anteile: Wald vs. Wilson.
- Bootstrap als praxisnahe Alternative.
- Planung: Stichprobengrösse via Margin of Error.

Punktschätzung, Standardfehler, Konfidenzintervalle.

- Wir können selten alle Menschen oder Dinge messen.
- Darum nehmen wir eine Stichprobe – eine kleine Gruppe – und schätzen daraus das Ganze.
- Beispiele:
  - Wie hoch ist der durchschnittliche Puls in eurer Klasse?
  - Wie gross ist der Anteil der Menschen, die TikTok täglich nutzen?
- Unsere Schätzung ist also ein bestmöglicher Tipp, basierend auf begrenzten Daten.



Punktschätzung, Standardfehler (SE) und 95 %-Konfidenzintervalle (t)
geschätzter Mittelwert
11.5
11.0
10.5
10.0
9.5
9.0
8.5
0
5
10
15
20
25
30
unabhängige Stichproben
Punktschätzung (Mittelwert)
± 1 Standardfehler (SE)
95 % Konfidenzintervall (trifft wahren Wert)
95 % Konfidenzintervall (verfehlt wahren Wert)
wahrer Mittelwert

Beispiel: TikTok in der Vorlesung (n = 5)

- In der Vorlesung sitzen 32 Studierende.
- Wir wollen wissen, wie viele davon täglich TikTok schauen, aber wir fragen nur 5 zufällig ausgewählte Personen. $\to$ n
- Von diesen 5 sagen 3 „Ja“ $\to$ beobachteter Anteil $\hat{p} = 3 / 5 = 0.6$ (60 %).
- Das ist unsere Punktschätzung für den wahren Anteil p in der ganzen Vorlesung.
- Mit dem Standardfehler berechnen wir, wie unsicher diese Schätzung ist:

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.6(0.4)}{5}} \approx 0.22$$

- Das 95%-Konfidenzintervall (nach Abraham Wald) lautet:

$$\hat{p} \pm 1.96 \times SE(\hat{p}) \approx 0.6 \pm 0.43 \Rightarrow [0.17,\ 1.00]$$

$\to$ Wir schätzen, dass zwischen 17 % und 100 % täglich TikTok schauen.

Beispiel: TikTok in der Vorlesung (n = 10)

- In der Vorlesung sitzen 32 Studierende.
- Wir fragen 10 zufällig ausgewählte Personen, wie viele davon täglich TikTok schauen.
- 6 sagen „Ja“ → beobachteter Anteil p̂ = 6 / 10 = 0.6 (60 %).
- Das ist unsere Punktschätzung für den wahren Anteil p in der ganzen Vorlesung.
- Mit dem Standardfehler berechnen wir die Unsicherheit der Schätzung:

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.6(0.4)}{10}} \approx 0.155$$

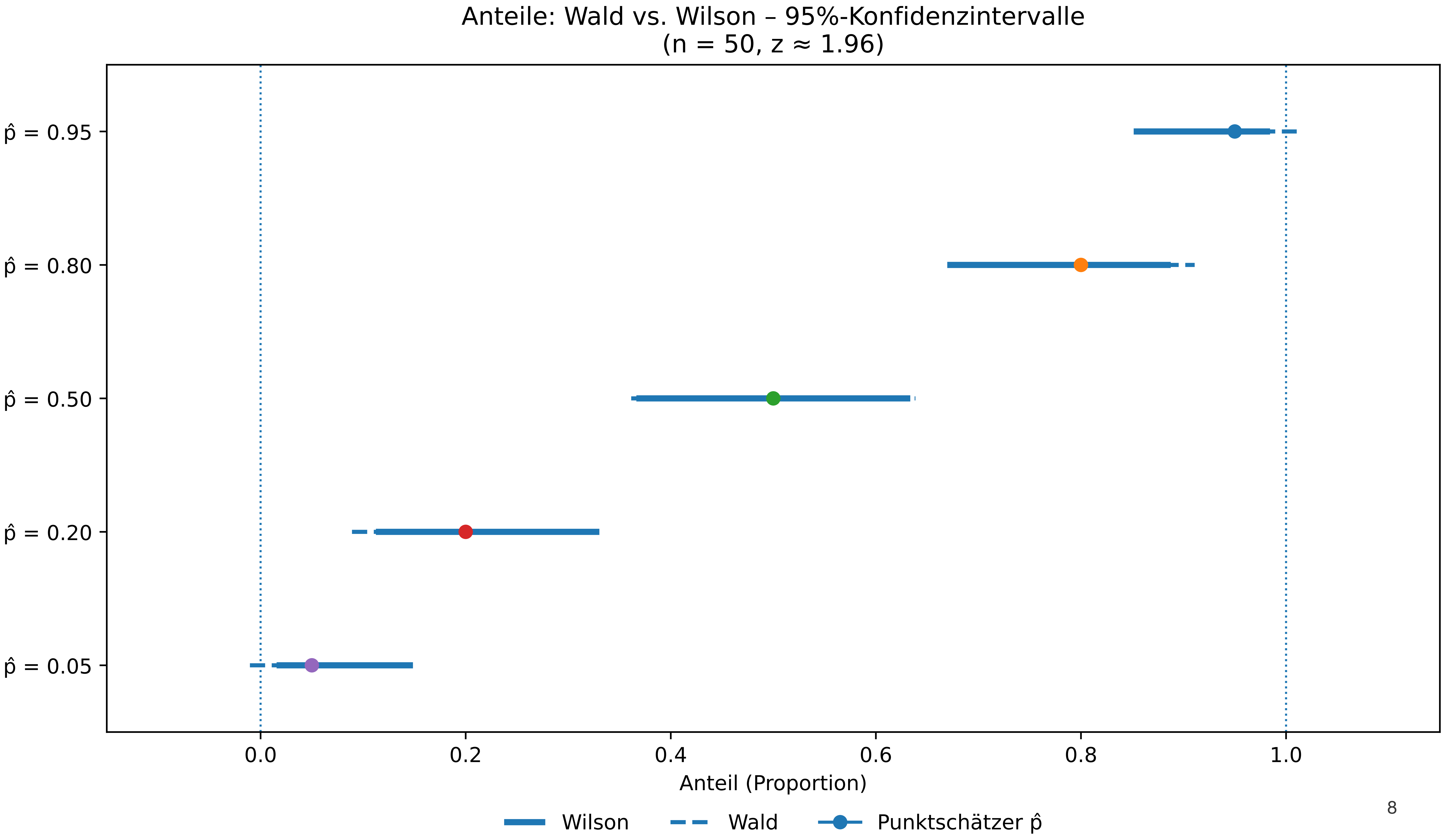
- Das 95%-Konfidenzintervall (nach Abraham Wald) lautet:

$$\hat{p} \pm 1.96 \times SE(\hat{p}) \approx 0.6 \pm 0.30 \Rightarrow [0.30,\ 0.90]$$

→ Wir schätzen, dass zwischen 30 % und 90 % der Studierenden täglich TikTok schauen.

Anteile: Wald vs. Wilson

- Abraham Wald (1902–1950) vs Edwin Bidwell Wilson (1879–1964)
- Wir wollen wissen, wie sicher ein beobachteter Prozentwert (Anteil) ist.
- Beispiel: 30 % der Befragten wählen Partei A, aber wie genau ist das?
- Dafür gibt es zwei Wege:
    - Wald: schnell gerechnet, aber kann unrealistische Werte liefern.
    - Wilson: etwas aufwendiger, aber stabil und immer zwischen 0 % und 100 %.
- Moderne Statistik nutzt Wilson $\rightarrow$ weil er Unsicherheit ehrlicher darstellt.


Anteile: Wald vs. Wilson – 95%-Konfidenzintervalle
(n = 50, z ≈ 1.96)
p̂ = 0.95
p̂ = 0.80
p̂ = 0.50
p̂ = 0.20
p̂ = 0.05
0.0
0.2
0.4
0.6
0.8
1.0
Anteil (Proportion)
Wilson
Wald
Punktschätzer p̂

Bootstrap: praxisnahe Alternative

- Klassische Formeln brauchen oft strenge Annahmen (z. B. Normalverteilung).
- Der Bootstrap funktioniert auch dann, wenn diese Annahmen nicht gelten.
- Idee: Die vorhandene Stichprobe als Mini-Welt behandeln.
- Viele neue Stichproben mit Zurücklegen ziehen $\rightarrow$ jedes Mal neu schätzen.
- So entsteht eine Verteilung der Schätzwerte, aus der man SE und CI direkt ablesen kann.
- Vorteil: funktioniert bei kleinen n, schiefen Daten und komplizierten Schätzern.

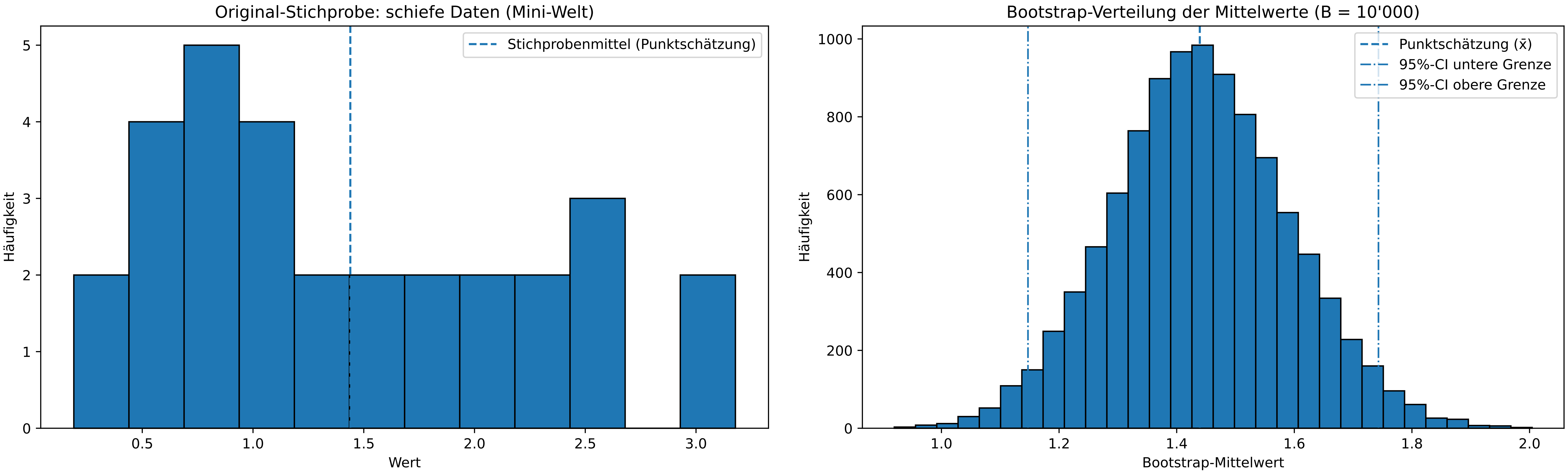
Bootstrap: Intuition

Du misst die Schlafdauer von 30 Studierenden.

Du kennst nicht die Schlafgewohnheiten aller 9879 Studierenden der HSG.

Aber du tust so, als wäre deine Stichprobe (die 30 Werte) eine kleine Version der gesamten Uni.

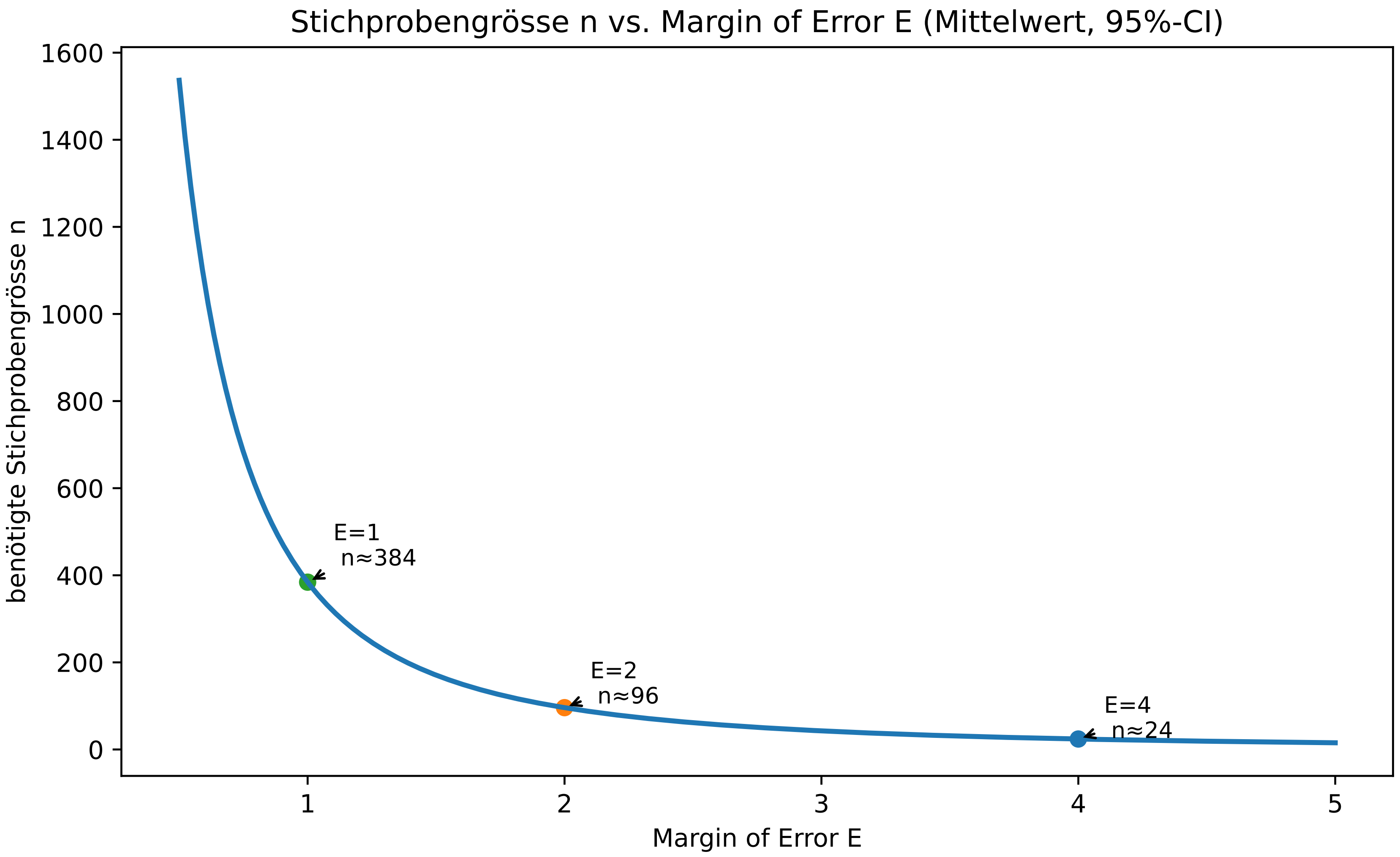
Dann ziehst du tausendmal neue Stichproben aus diesen 30 Werten → und siehst, wie stark der durchschnittliche Schlaf schwankt.

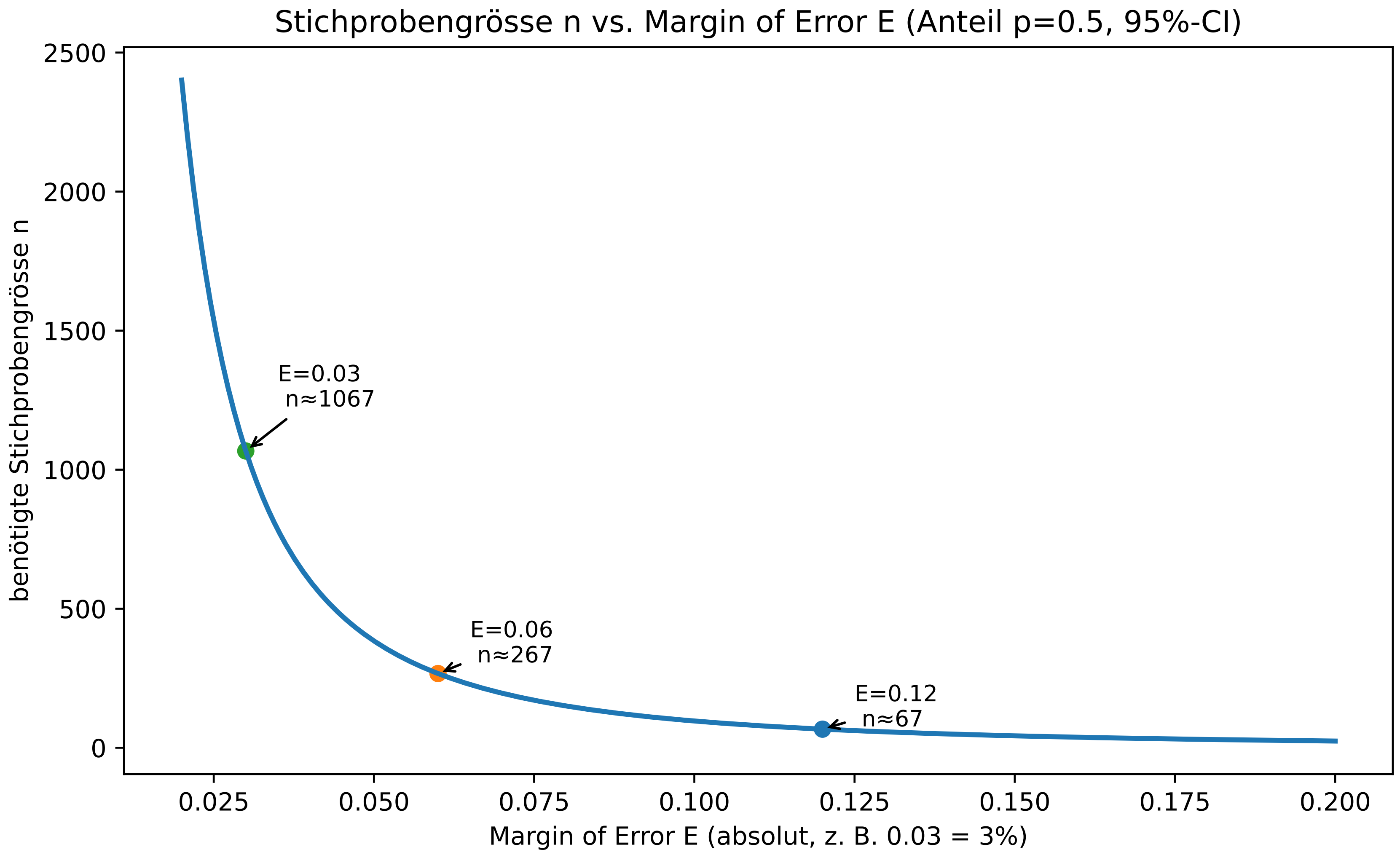


Original-Stichprobe: schiefe Daten (Mini-Welt)
Stichprobenmittel (Punktschätzung)
Häufigkeit
Wert
0
1
2
3
4
5
0.5
1.0
1.5
2.0
2.5
3.0
Bootstrap-Verteilung der Mittelwerte (B = 10'000)
Punktschätzung (x̄)
95%-CI untere Grenze
95%-CI obere Grenze
Häufigkeit
Bootstrap-Mittelwert
0
200
400
600
800
1000
1.0
1.2
1.4
1.6
1.8
2.0

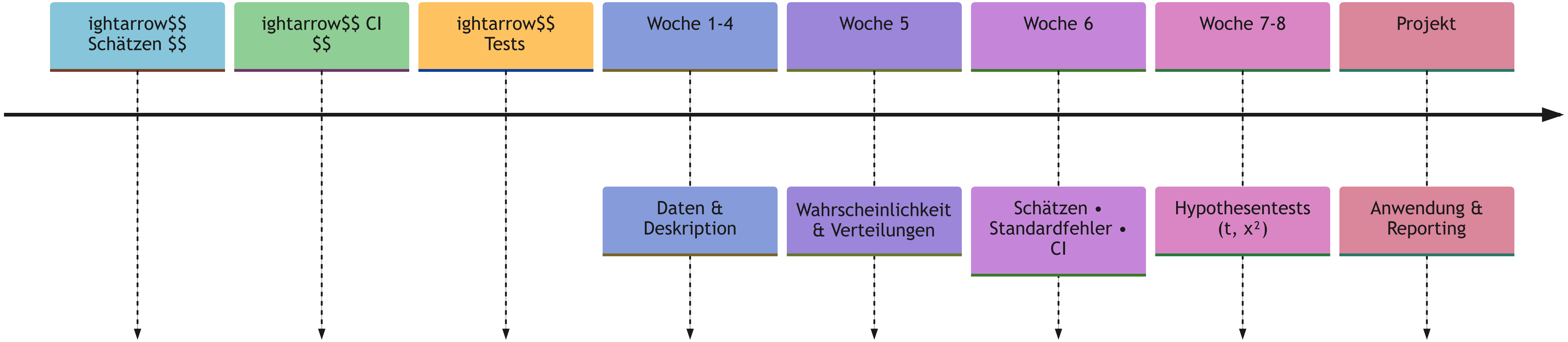
Planung: Stichprobengrösse via Margin of Error

- Bevor wir Daten erheben oder messen, müssen wir wissen: Wie viele Beobachtungen brauchen wir?
- Der Margin of Error (Fehlergrenze) zeigt, wie genau unsere Schätzung am Ende sein soll.
- Je kleiner der erlaubte Fehler, desto mehr Daten müssen wir sammeln – und zwar quadratisch mehr.
- Beispiel: Wenn der Fehler halb so gross sein soll, braucht man viermal so viele Messungen oder Befragungen.
- Ziel: nicht «so viel wie möglich», sondern so viel wie nötig, um verlässlich zu schätzen.

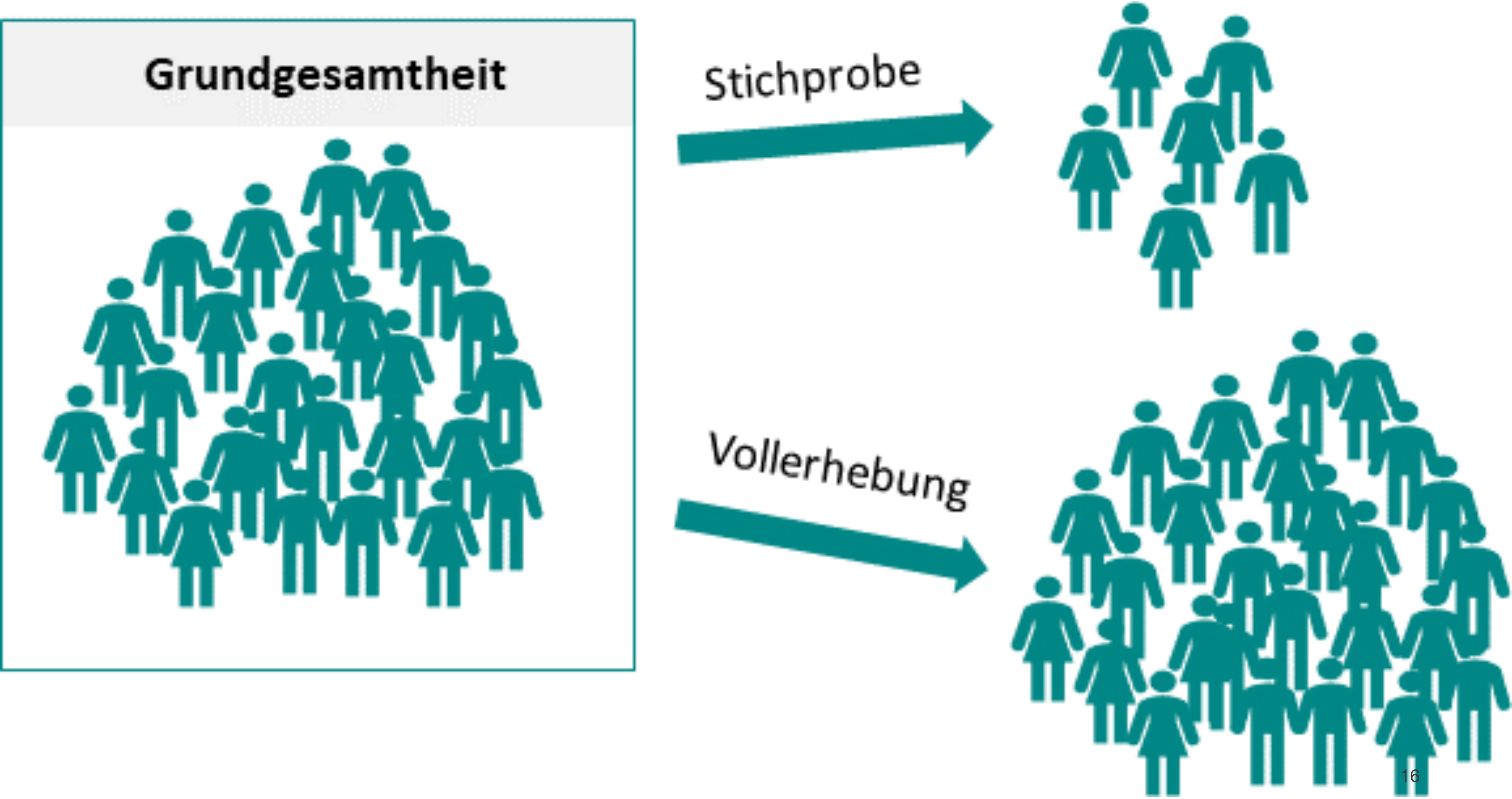

Stichprobengrösse n vs. Margin of Error E (Mittelwert, 95%-CI)
benötigte Stichprobengrösse n
Margin of Error E
E=1
n≈384
E=2
n≈96
E=4
n≈24
0
200
400
600
800
1000
1200
1400
1600
1
2
3
4
5


Stichprobengrösse n vs. Margin of Error E (Anteil p=0.5, 95%-CI)
benötigte Stichprobengrösse n
Margin of Error E (absolut, z. B. 0.03 = 3%)
E=0.03
n≈1067
E=0.06
n≈267
E=0.12
n≈67
0
500
1000
1500
2000
2500
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200

Daten $$


ightarrow$$ Schätzen $$
ightarrow$$ CI $$
ightarrow$$ Tests
Woche 1-4
Woche 5
Woche 6
Woche 7-8
Projekt
Daten & Deskription
Wahrscheinlichkeit & Verteilungen
Schätzen • Standardfehler • CI
Hypothesentests (t, x²)
Anwendung & Reporting

Primer


Grundgesamtheit
Stichprobe
Vollerhebung

Population, Stichprobe & Parameter

- Wir wollen etwas über eine grosse Gruppe (Population) herausfinden, z. B. den durchschnittlichen Puls aller Studierenden.
- Weil wir nicht alle messen können, nehmen wir eine kleinere Gruppe (Stichprobe).
- Die Population hat feste, aber unbekannte Werte (Parameter):
  - Mittelwert μ, Varianz σ^2, Anteil p.
- Aus der Stichprobe berechnen wir eine Schätzung dieser Werte, z. B. den Mittelwert $\bar{x}$.
- Der wahre Wert μ ist fix, die Schätzung $\bar{X}$ zufällig, sie hängt von der Stichprobe ab.

Primer: Population, Stichprobe & Parameter

Schätzer basieren auf Stichproben und zielen auf Populationsparameter.

- Population: Gesamtheit; Parameter: μ, σ^2, p (fix, unbekannt).
- Stichprobe: n Beobachtungen; ideal: zufällig und unabhängig.
- Statistik/Schätzer: $T(X_1, \ldots, X_n)$; Schätzwert: $t = T(x_1, \ldots, x_n)$.
- Beispiele: $\bar{X} \to \mu,\ \hat{p} \to p$.

Mini-Check: Was ist zufällig, was fix: $\bar{X},\ \mu,\ \bar{x}$?

Symbole verstehen: μ, σ^2, p, n, T, t, $\bar{x}$, $\hat{p}$ ...

- $T(X_1,\ldots,X_n)$ = allgemeiner Schätzer, also die Rechenregel (z. B. „bilde den Durchschnitt")
- $t = T(x_1,\ldots,x_n)$ = Schätzwert, also das konkrete Ergebnis, wenn wir Zahlen einsetzen

- In der Population (der ganzen Gruppe):
  - μ (sprich: mü) = wahrer Mittelwert
  - σ^2 (sigma-Quadrat) = wahre Varianz (zeigt, wie stark die Werte streuen)
  - p = wahrer Anteil oder Wahrscheinlichkeit

- In der Stichprobe (den gemessenen Daten):
  - n = Anzahl der Beobachtungen
  - $\bar{x}$ (sprich: x quer) = Mittelwert der Stichprobe (unsere Schätzung für μ)
  - $\hat{p}$ (sprich: p-Dach) = beobachteter Anteil (unsere Schätzung für p)

Unterschied: Grosses $\bar{X}$ vs. kleines $\bar{x}$

- Grosses $\bar{X}$ steht für die Zufallsvariable des Mittelwerts.
  $\rightarrow$ Sie beschreibt, wie der Mittelwert über viele mögliche Stichproben schwankt.
- Kleines $\bar{x}$ steht für den beobachteten Mittelwert aus einer konkreten Stichprobe.
  $\rightarrow$ Das ist der Wert, den du tatsächlich ausgerechnet hast.
- Wenn man das Experiment viele Male wiederholt, bekommt man viele verschiedene $\bar{x}$-Werte $\rightarrow$
  die Gesamtheit dieser Werte ist die Verteilung von $\bar{X}$.
- Kurz gesagt:
  - $\bar{X}$ = theoretisch, zufällig
  - $\bar{x}$ = beobachtet, konkret

SD vs. SE – zwei Arten von Streuung

- Beide messen Streuung, aber auf unterschiedlichen Ebenen.

Kürzel	Bedeutung	Wofür sie steht
SD (Standard Deviation)	Standardabweichung	Wie stark die Daten selbst streuen
SE (Standard Error)	Standardfehler	Wie stark ein Schätzer (z. B. der Mittelwert) zwischen Stichproben schwankt

SD $\to$ beschreibt die Variabilität innerhalb einer Stichprobe.
SE $\to$ beschreibt die Unsicherheit des Mittelwerts über viele Stichproben hinweg.

Beispiel: $\bar{X}$ vs. $\bar{x}$ mit Zahlen

- In unserer Stichprobe messen wir $\bar{x} = 73.1$ $$\rightarrow$ das ist der beobachtete Schätzwert.
- Bei einer anderen Stichprobe könnte $\bar{x}$ z. B. 71.5 oder 74.2 sein.

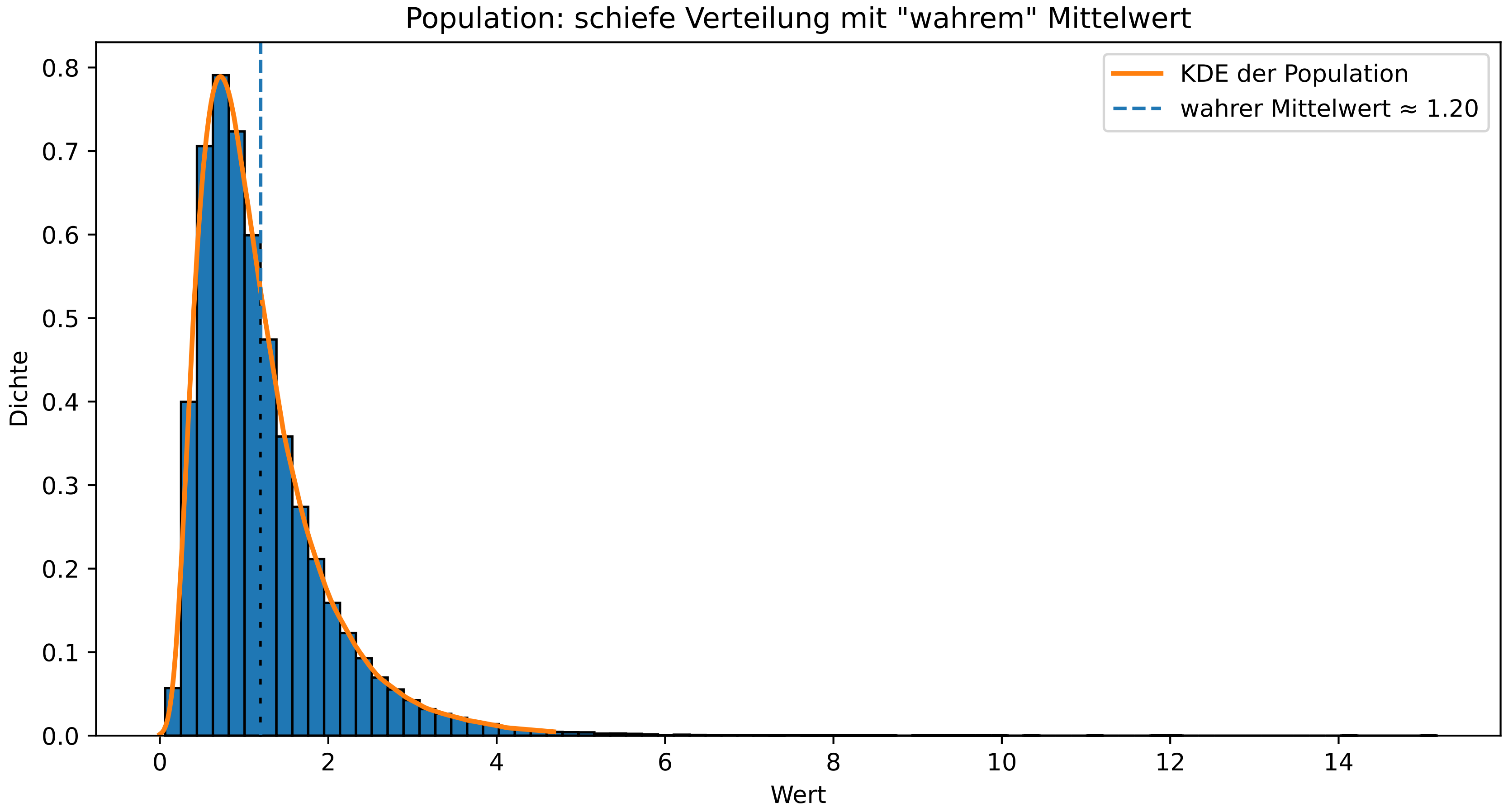
Wir schätzen den durchschnittlichen Puls aller Studierenden an der Uni.

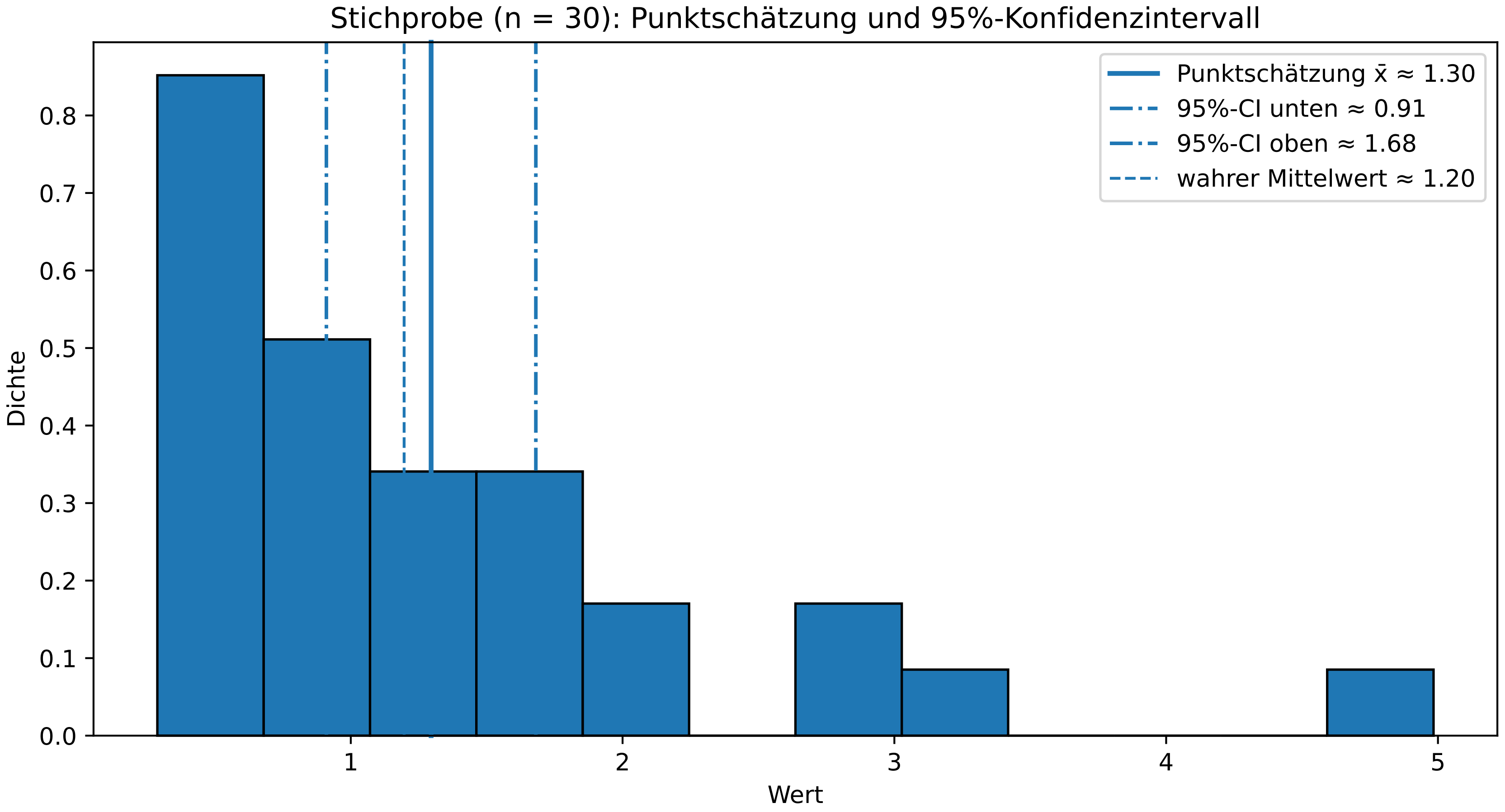
In der Population gilt:

- Wahrer Mittelwert $\mu = 72$
- Standardabweichung $\sigma = 8$
- Stichprobengrösse $n = 25$

Theoretisch gilt für den Mittelwert $\bar{X}$:

- Erwartungswert $E[\bar{X}] = 72$
- Standardfehler $SE = SD(\bar{X}) = \sigma / \sqrt{n} = 1.6$
- Mögliche Werte von $\bar{X}$ liegen meist zwischen 68.8 und 75.2


Population: schiefe Verteilung mit "wahrem" Mittelwert
KDE der Population
wahrer Mittelwert ≈ 1.20
Dichte
Wert
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
2
4
6
8
10
12
14


Stichprobe (n = 30): Punktschätzung und 95%-Konfidenzintervall
Punktschätzung x̄ ≈ 1.30
95%-CI unten ≈ 0.91
95%-CI oben ≈ 1.68
wahrer Mittelwert ≈ 1.20
Dichte
Wert
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
1
2
3
4
5

Sampling & Bias

- Gute Konfidenzintervalle brauchen gute Stichproben: zufällig, unabhängig, repräsentativ.
- Eine Zufallsstichprobe verhindert systematische Verzerrungen und hält die Formeln gültig.
- Typische Bias-Fallen:
  - Selektionsbias: Nur bestimmte Personen werden befragt.
  - Nonresponse: Wer nicht antwortet, unterscheidet sich vom Rest.
  - Abhängigkeiten: Daten stammen aus Clustern oder Zeitreihen.

Zufällig, unabhängig & repräsentativ

- Eine zufällige Stichprobe bedeutet: Jede Person oder jedes Objekt hat die gleiche Chance, ausgewählt zu werden.
  $\rightarrow$ Beispiel: Zufallsgenerator wählt 50 Studenten aus der ganzen Uni.
- Unabhängig heisst: Das Ergebnis einer Messung beeinflusst keine andere.
  $\rightarrow$ Beispiel: Wenn du den Puls misst, darf der Puls des ersten Studenten keinen Einfluss auf den zweiten haben.
- Repräsentativ bedeutet: Die Stichprobe spiegelt die Vielfalt der Population wider.
  $\rightarrow$ Beispiel: Wenn du nur deine Freunde befragst, ist das nicht repräsentativ für die ganze Schule.
- Nur wenn alle drei Bedingungen erfüllt sind, können wir verlässlich schätzen und Konfidenzintervalle korrekt berechnen.

Notation: Population & Stichprobe

In der Population (ganze Uni):

- μ = wahrer Durchschnitt (z. B. aller Studierenden)
- σ^2 = wahre Varianz → zeigt, wie stark die Werte streuen
- p = wahrer Anteil (z. B. Anteil der Studierenden, die täglich Kaffee trinken)

In der Stichprobe (unsere Messdaten):

- $\bar{x}$ = Mittelwert der Stichprobe → unsere Schätzung für μ
- s^2 = Stichprobenvarianz → Schätzung für σ^2
- ddof = 1 („Delta Degrees of Freedom") korrigiert die Berechnung, damit s^2 im Mittel unverzerrt ist
- $\hat{p}$ (sprich: p-Dach) = beobachteter Anteil aus der Stichprobe

Beispiel:

Wir befragen 50 Studierende über ihren täglichen Kaffeekonsum:
→ μ = wahrer Durchschnitt (unbekannt)
→ $\bar{x}$ = Durchschnitt aus den 50 Antworten
→ $\hat{p}$ = Anteil der Kaffeetrinker in dieser Stichprobe

Notation: SE & Quantile verstehen

SE (Standardfehler) beschreibt, wie stark ein Schätzer schwankt, wenn man viele Stichproben zieht:

Für den Mittelwert: $SE(\bar{x}) = \frac{\sigma}{\sqrt{n}}$

Für Anteile: $SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Je grösser n, desto kleiner der Standardfehler $\rightarrow$ unsere Schätzungen werden stabiler.

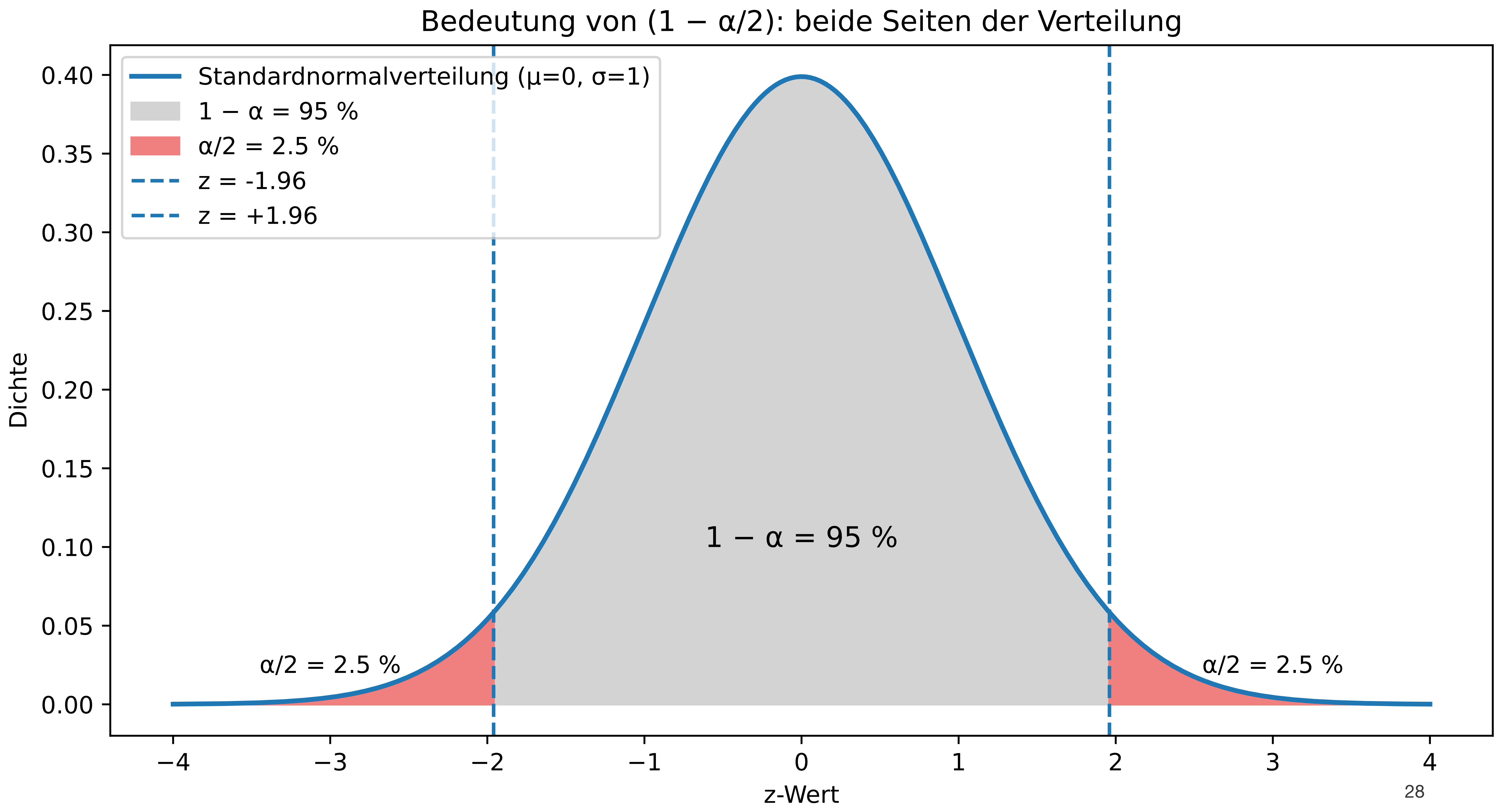
Quantile bestimmen die Grenzen von Konfidenzintervallen:
- $z_{1-\alpha/2} \approx 1.96 \rightarrow$ 95 % der Werte liegen innerhalb von ±1.96 Standardabweichungen um den Mittelwert.
- $t_{\mathrm{df},\,1-\alpha/2} \rightarrow$ analog für kleinere Stichproben mit der t-Verteilung, wobei $\mathrm{df} = n - 1$.

$1 - \frac{\alpha}{2}$ bedeutet, dass wir beide Seiten der Verteilung betrachten:

- α ist die Irrtumswahrscheinlichkeit (z. B. 0.05 bei 95 %)
- $1 - 0.05/2 = 0.975 \Rightarrow 97.5\% \rightarrow$ ergibt $z \approx 1.96$

Beispiel:
Bei 95 % Konfidenz: $\bar{x} \pm 1.96 \times SE(\bar{x})$

$\rightarrow$ Das Intervall enthält mit grosser Wahrscheinlichkeit den wahren μ.



Bedeutung von (1 − α/2): beide Seiten der Verteilung
Standardnormalverteilung (μ=0, σ=1)
1 − α = 95 %
α/2 = 2.5 %
z = -1.96
z = +1.96
1 − α = 95 %
α/2 = 2.5 %
α/2 = 2.5 %
Dichte
z-Wert
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
−4
−3
−2
−1
0
1
2
3
4

Die z-Verteilung

- Die z-Verteilung ist eine Standardnormalverteilung mit Mittelwert $\mu = 0$ und Standardabweichung $\sigma = 1$.
- Sie beschreibt, wie stark standardisierte Werte (z-Werte) vom Mittelwert abweichen:

$$z = \frac{x - \mu}{\sigma}$$

$\rightarrow$ z zeigt, wie viele Standardabweichungen ein Wert vom Mittelwert entfernt ist.

- Die z-Verteilung ist symmetrisch und glockenförmig.
  - Etwa 68 % aller Werte liegen zwischen z = –1 und +1
  - Etwa 95 % zwischen z = –1.96 und +1.96
- Der Buchstabe z steht für Zufallsvariable der standardisierten Normalverteilung.
  $\rightarrow$ Man kann sagen: «z ist die Sprache der Standardabweichungen».

Die t-Verteilung

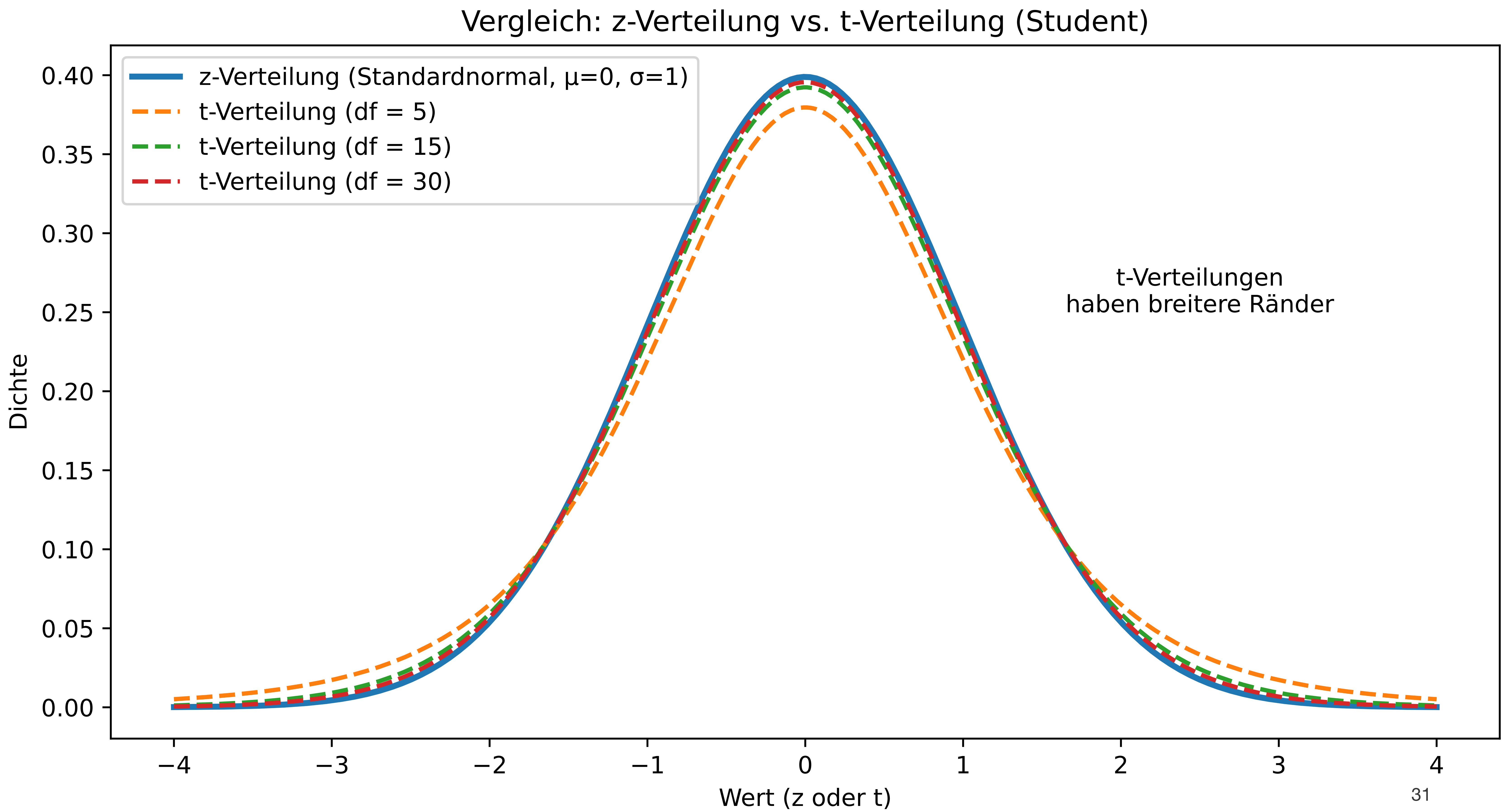
- Die t-Verteilung sieht ähnlich aus wie die z-Verteilung, aber sie hat breitere Flanken (dickere «Tails»).
  $\rightarrow$ Sie berücksichtigt, dass wir σ nicht kennen, sondern nur aus der Stichprobe schätzen.

  $$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

  $\rightarrow$ Statt σ (Populationsstreuung) verwenden wir s (Stichprobenstreuung).

- Der Buchstabe t steht für die t-Statistik, eingeführt von William Sealy Gosset,
  der unter dem Pseudonym „Student“ veröffentlichte $\rightarrow$ daher auch „Student’s t-Verteilung“.

- Die Form der Verteilung hängt von den Freiheitsgraden (df = n – 1) ab:
  - Bei kleinen n $\rightarrow$ breiter, unsicherer.
  - Bei grossem n $\rightarrow$ nähert sich der z-Verteilung an.



Vergleich: z-Verteilung vs. t-Verteilung (Student)
z-Verteilung (Standardnormal, μ=0, σ=1)
t-Verteilung (df = 5)
t-Verteilung (df = 15)
t-Verteilung (df = 30)
t-Verteilungen
haben breitere Ränder
Dichte
Wert (z oder t)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
−4
−3
−2
−1
0
1
2
3
4

Notation: Quick-Ref

Konsistente Symbole senken kognitive Last.

- Population: μ (Mittelwert), σ^2 (Varianz), p (Anteil).
- Stichprobe: $\bar{x},\ s^2\ (\mathrm{ddof}=1),\ \hat{p}$.
- SE: $\mathrm{SE}(\bar{X}) = \sigma/\sqrt{n},\ \ \mathrm{SE}(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n}.$
- Quantile: $z_{1-\alpha/2} \approx 1.96$ (95 %), $t_{\mathrm{df},\,1-\alpha/2}$.

Mini-Check: Welche Grössen ändern sich, wenn n wächst: $\mu,\ \bar{x},\ s,\ \mathrm{SE}$?

Punktschätzung

Warum schätzen wir überhaupt?

Schätzen liefert handlungsrelevante Zahlen unter Unsicherheit.

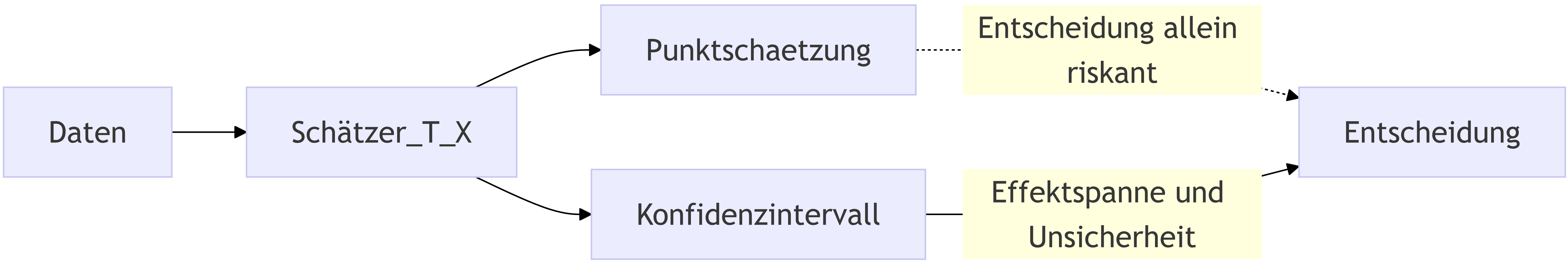
- Beispiele: Impf-Effekt, Conversion-Rate, Produktionsausbeute.
- Punktschätzung = Kompass; CI = Sicherheitsgurt.
- Lernziel: Punkt- und Intervallschätzer korrekt nutzen.

Mini-Check: Nenne ein Entscheidungsproblem, das ohne CI riskant wäre.

Beispiele

- Impf-Effekt: → Wie stark schützt eine Impfung wirklich?
  Beispiel: Von 100 geimpften Personen werden 5 krank, von 100 ungeimpften 20.
  Impf-Effekt = Unterschied = 15 % weniger Erkrankte.
- Conversion-Rate: → Wie viele Besucher einer Webseite tun das, was man will, z. B. etwas kaufen oder sich anmelden?
  Beispiel: 1000 Besucher, 50 Käufe → Conversion-Rate = 50 / 1000 = 5 %.
- Produktionsausbeute: → Wie viele hergestellte Produkte sind wirklich in Ordnung?
  Beispiel: 100 produzierte Teile, 92 fehlerfrei → Ausbeute = 92 %.
- In allen drei Fällen schätzen wir einen Anteil oder eine Wahrscheinlichkeit → also wie oft etwas «erfolgreich» ist.

Weg von Daten zur Entscheidungsfindung



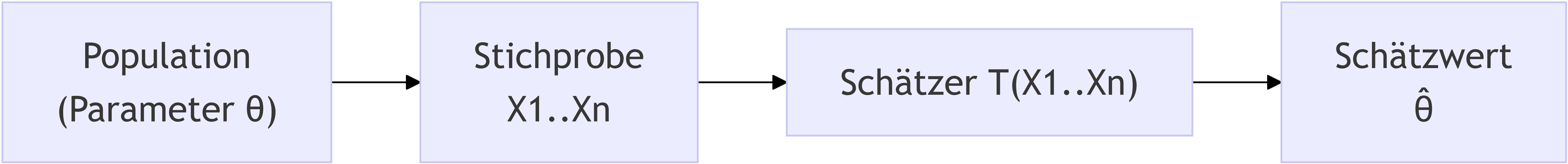
Schätzer: Definition & Notation

Ein Schätzer ist eine Funktion der Daten für einen Parameter.

- $T(X_1, \ldots, X_n) \to \theta$, z. B. $\bar{X} \to \mu$; $\hat{p} \to p$.
- T ist Zufallsvariable mit Erwartungswert und Varianz.
- $\mathrm{MSE}(T) = \mathrm{Bias}(T)^2 + \mathrm{Var}(T)$.

Mini-Check: Was ist hier zufällig: $\bar{X}$ oder μ?

Population zur Schätzung



Bias & Varianz – zwei Arten von Fehlern

- Wenn wir schätzen, wollen wir den wahren Wert treffen, wie beim Dartspiel die Mitte.
- Dabei passieren zwei Arten von Fehlern:

🎯 Bias (systematischer Fehler)

- Bedeutet: im Durchschnitt daneben zielen.
- Beispiel: Eine Waage zeigt immer 2 kg zu viel $\rightarrow$ Schätzer ist systematisch zu hoch.
- Bias misst, wie weit der Durchschnitt der Schätzungen vom wahren Wert entfernt ist.
- Ein unverzerrter Schätzer hat Bias = 0.

🎲 Varianz (Zufallsschwankung)

- Bedeutet: Die Schätzungen springen hin und her.
- Beispiel: Eine Messung schwankt stark – mal 68, mal 74, mal 70.
- Varianz zeigt, wie stark die Ergebnisse zwischen Stichproben schwanken.

Modell-Bias: Vereinfachung mit Absicht

- Modell-Bias entsteht, wenn wir die Welt bewusst vereinfachen, um stabile und verständliche Schätzungen zu bekommen.

🧠 Was heisst Modell bei Schätzern?

Ein Modell ist eine Annahme über die Struktur der Daten:

- Wie sehen die Werte aus? (z. B. normalverteilt oder schief?)
- Wie hängen sie zusammen? (z. B. linear oder gekrümmt?)
- Welche Grössen sind fix, welche zufällig?

🔍 Beispiele für Modell-Annahmen:

- Normalverteilung: Viele Schätzer (z. B. t- und z-Intervalle) nehmen an, dass Daten symmetrisch verteilt sind.
- Linearität: Regressionsschätzer gehen davon aus, dass Zusammenhänge gerade Linien sind.
- Unabhängigkeit: Jeder Messwert ist eigenständig und beeinflusst keinen anderen.
- Homogene Varianz: Die Streuung der Daten ist überall etwa gleich.

Varianz: Warum Schätzungen zufällig sind

Jede Stichprobe ist anders $\rightarrow$ Schätzungen schwanken zufällig. Diese Streuung nennt man Varianz des Schätzers.

🎲 Zufallsquellen beim Schätzen

1. Zufällige Auswahl:
   Wer in der Stichprobe landet, ist Zufall $\rightarrow$ andere Personen $\rightarrow$ anderer Mittelwert.
2. Messrauschen:
   Kleine Messfehler verändern die Werte leicht.
3. Modellunsicherheit:
   Wenn das Modell nicht perfekt passt, schwanken Schätzungen stärker.

📊 Folge:

Schätzer haben keine feste Zahl, sondern eine Verteilung von möglichen Werten.

Der Standardfehler (SE) misst diese Streuung:

$$SE = \frac{SD}{\sqrt{n}}$$

$\rightarrow$ Mehr Daten = kleinere Varianz = stabilere Schätzungen.

Gütekriterien: Bias, Varianz & Effizienz

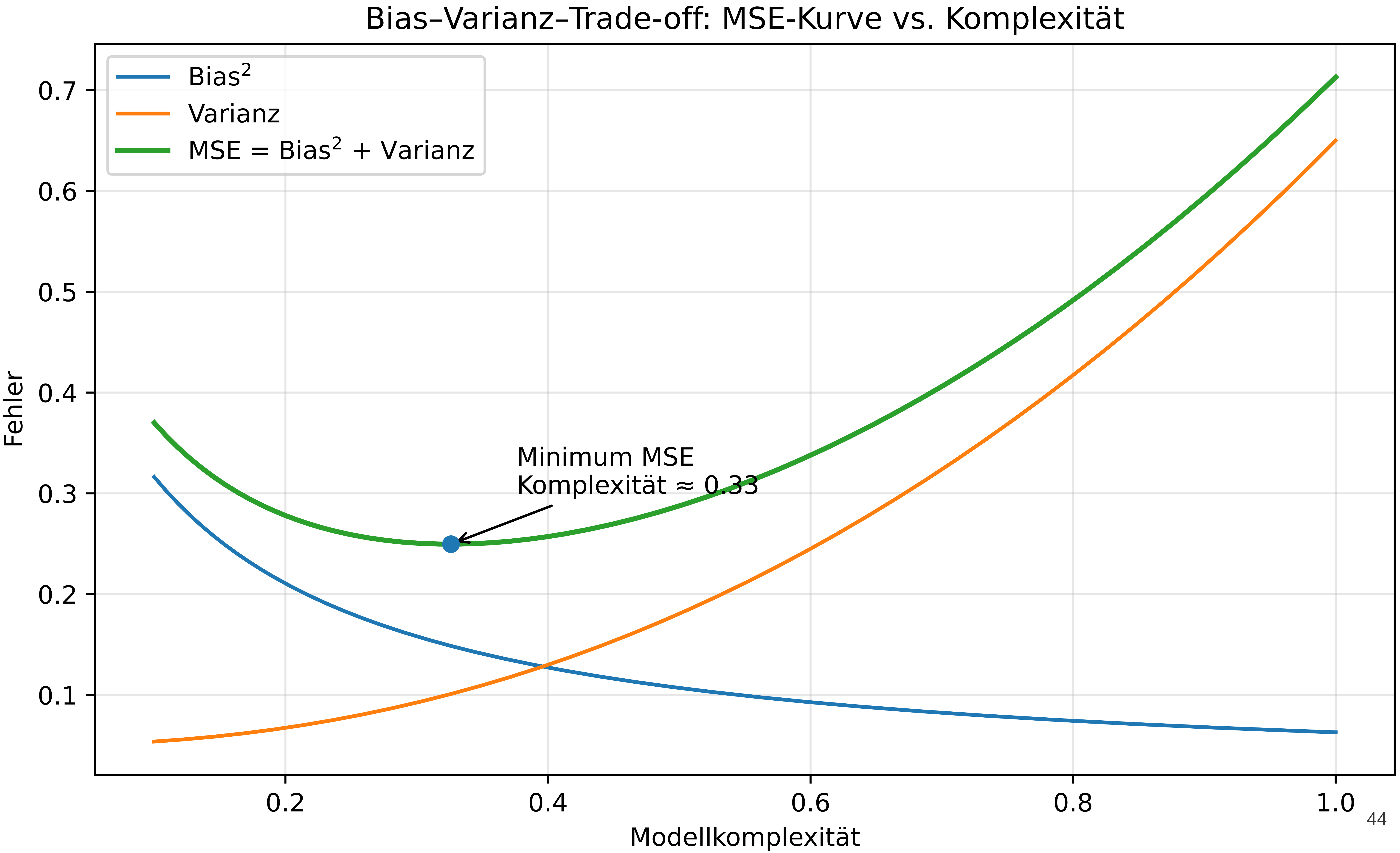
- Bias = Schätzer liegt im Durchschnitt systematisch daneben. $\rightarrow$ Beispiel: Waage zeigt immer +2 kg.
- Varianz = Schätzer schwankt stark zwischen Stichproben. $\rightarrow$ Beispiel: Messwerte springen zwischen 68 und 75.
- Erwartungstreue: $\mathbb{E}[T] = \theta \rightarrow$ Schätzer trifft den wahren Wert im Mittel.
- Effizienz: Unter allen erwartungstreuen Schätzern hat dieser die kleinste Varianz $\rightarrow$ stabilster Schätzer.
- Bias–Varianz-Trade-off: $\mathrm{MSE} = \mathrm{Bias}^2 + \mathrm{Varianz} \rightarrow$ Ein kleiner Bias kann gut sein, wenn die Varianz stark sinkt $\rightarrow$ stabiler Gesamtfehler.

Gütekriterien (Bias, Varianz, Effizienz)

Gute Schätzer balancieren Bias und Varianz.

- Erwartungstreue: $\mathbb{E}[T] = \theta$.
- Effizienz: kleinste Varianz unter erwartungstreuen Schätzern.
- Bias–Varianz-Trade-off: minimale MSE ist Ziel.

Mini-Check: Nenne ein Setting, in dem kleiner Bias akzeptabel ist.


Bias–Varianz–Trade-off: MSE-Kurve vs. Komplexität
Bias²
Varianz
MSE = Bias² + Varianz
Minimum MSE
Komplexität ≈ 0.33
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Fehler
0.2
0.4
0.6
0.8
1.0
Modellkomplexität

Mean Squared Error (MSE)

Der MSE misst, wie gut ein Schätzer im Durchschnitt trifft.

$$\text{MSE} = \text{Bias}^2 + \text{Varianz}$$

- Bias^2 $\rightarrow$ systematische Abweichung (immer leicht daneben).
- Varianz $\rightarrow$ Zufallsschwankung (mal zu tief, mal zu hoch).
- Beide zusammen ergeben den Gesamtfehler.

🎯 Ziel: Den MSE minimieren $\rightarrow$ bestmöglicher Kompromiss zwischen

«nicht schief» (kleiner Bias) und «nicht zitterig» (kleine Varianz).

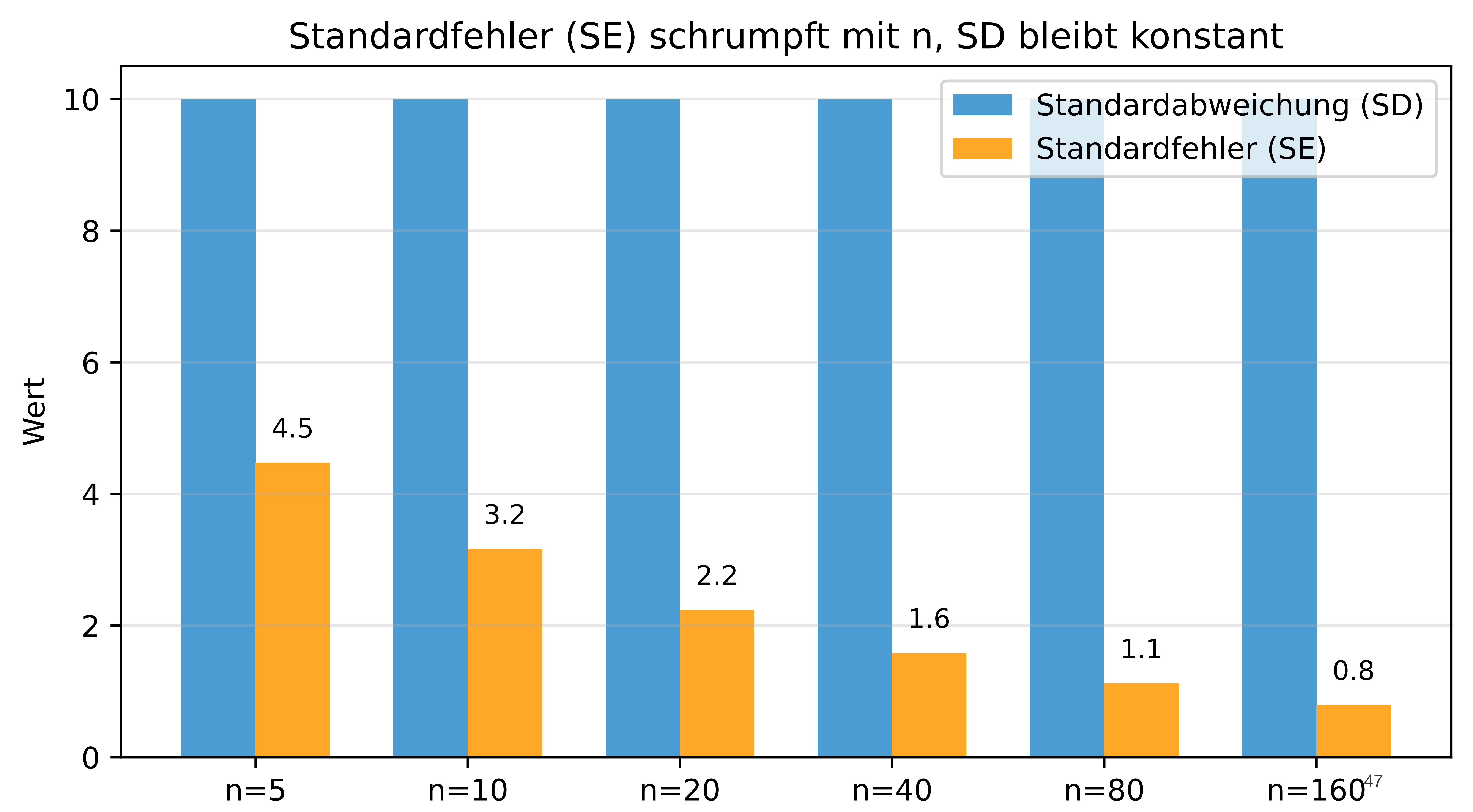
Standardfehler (SE) vs. Standardabweichung (SD)

SE misst die Streuung des Schätzers, SD die Streuung der Daten.

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}} \qquad\qquad SD(X) = \sigma$$

- Verdoppeln von n senkt den Standardfehler um den Faktor $\sqrt{2}$, die Standardabweichung der Daten bleibt gleich.
- In der Praxis bestimmt der Standardfehler die Breite des Konfidenzintervalls.

Mini-Check:
Warum sinkt SE mit n -- SD aber nicht?


Standardfehler (SE) schrumpft mit n, SD bleibt konstant
Standardabweichung (SD)
Standardfehler (SE)
Wert
10
8
6
4
2
0
4.5
3.2
2.2
1.6
1.1
0.8
n=5
n=10
n=20
n=40
n=80
n=160

Beispiel Punktschätzung: Münzwurf

- Wir wollen den Anteil „Kopf“ beim Münzwurf schätzen.
- 10 Würfe, 7 Köpfe $\rightarrow$ beobachteter Anteil $\hat{p} = 0.7$
  $\rightarrow$ das ist unsere Punktschätzung für den wahren Anteil p.
- Wie stark kann diese Schätzung schwanken? $\rightarrow$ das misst der Standardfehler:

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad SE(\hat{p}) = \sqrt{\frac{0.7 \cdot 0.3}{10}} \approx 0.145$$

- Interpretation: Wenn wir das Experiment oft wiederholen, schwankt $\hat{p}$ etwa um ±0.15.

Mini-Check: Warum $\hat{p}$ im SE statt p?

Beispiel: Münzwurf in Python

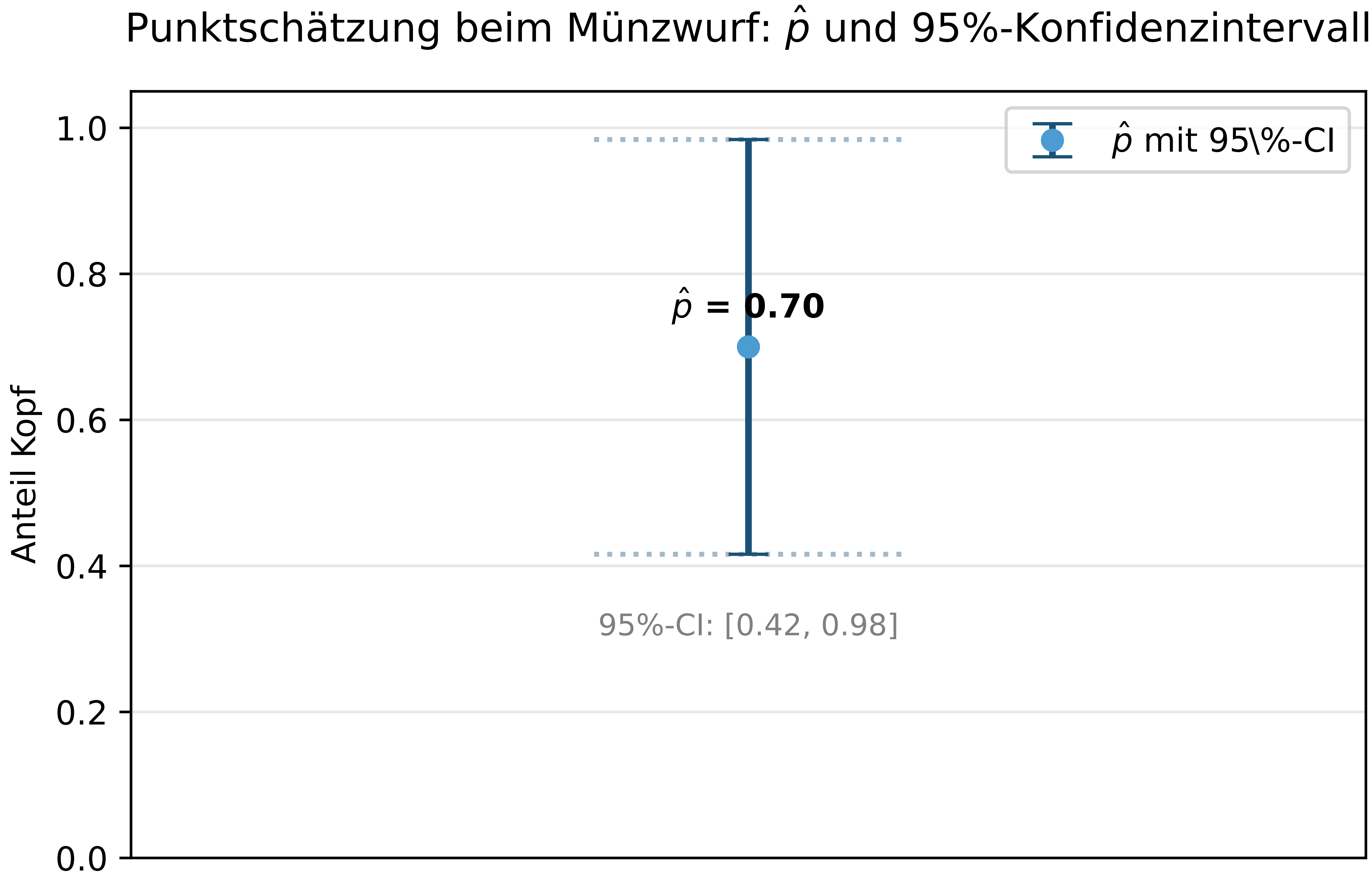
```
import numpy as np
rng = np.random.default_rng(123)

# 20 Münzwürfe, Wahrscheinlichkeit für "Kopf" = 0.6
x = rng.binomial(1, 0.6, 20)

# Punktschätzung und Standardfehler
p_hat = x.mean()
se = np.sqrt(p_hat * (1 - p_hat) / len(x))

p_hat, se
```

- Der Code simuliert 20 Münzwürfe, berechnet $\hat{p}$ und $SE(\hat{p})$.
- Beispielausgabe: (0.55, 0.111) $\rightarrow$ Anteil Kopf = 55 %, Standardfehler ≈ 0.11.
- Wenn du den Code mehrfach ausführst, ändern sich die Ergebnisse leicht $\rightarrow$ Zufall!
- Je grösser n, desto kleiner wird der Standardfehler.


Punktschätzung beim Münzwurf: $\hat{p}$ und 95%-Konfidenzintervall
1.0
0.8
0.6
0.4
0.2
0.0
Anteil Kopf
$\hat{p}$ mit 95\%-CI
$\hat{p}$ = **0.70**
95%-CI: [0.42, 0.98]

Zusammenfassung & Ausblick

Gesamt-Take-Away: Punkt, Spanne, Planung

Statistik ist die Kunst, Unsicherheit nützlich zu machen: mit Schätzung, Intervall und guter Planung.

- Punktschätzung ist der Kompass: eine beste Zahl aus begrenzten Daten.
- Konfidenzintervall ist der Sicherheitsgurt: zeigt die Spanne und Unsicherheit.
- Standardfehler ist der Taktgeber: bestimmt, wie breit die Spanne wird.
- Wilson statt Wald bei Anteilen: ehrlicher und immer zwischen 0 und 1.
- Bootstrap macht Unsicherheit sichtbar, wenn Formeln wackeln.
- Stichprobengrösse via Margin of Error: Präzision ist planbar, aber teuer im Quadrat.
- Bias–Varianz-Trade-off: lieber kleiner Bias und stabile Schätzung als perfekter, aber wackeliger Wert.

Quiz - Aktive Wiederholung

Kahoot Quiz VL5: Wahrscheinlichkeit & Verteilungen

Wrap-up & Nächstes Mal

Primer, Population, Stichprobe, Punktschätzung.

Online:
- Konfidenzintervalle
- Bootstrap

Nächste Sitzung:
- Hypothesentests (H_0 / H_1), Fehlerarten, t, (χ^2).