

## Block 2: Zusammenfassung

Inferenz, Unsicherheit & Entscheidungslogik

January 11, 2026

**Kernidee:** Von Stichprobe auf Population schliessen

- **Standardfehler:**  $SE = s/\sqrt{n}$  – Unsicherheit der Schätzung
- **Konfidenzintervall:**  $\bar{x} \pm 1.96 \times SE$  (95%)
- **Signal vs. Rauschen:** Ist der Effekt echt oder Zufall?

**Mehr Daten** → **kleinerer SE** → **engere KI** → **präzisere Schätzung**

---

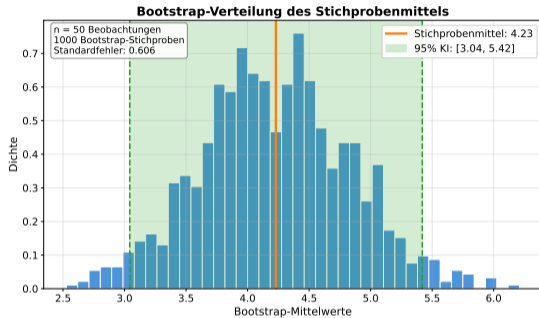
Jede Schätzung aus Daten ist mit Unsicherheit behaftet.

## 2. Bootstrapping

### Algorithmus:

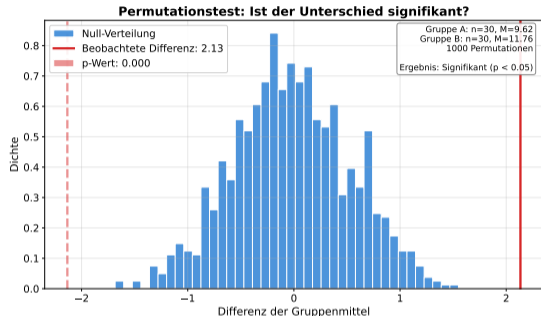
1. Ziehe  $n$  Werte mit Zurücklegen
2. Berechne Statistik
3. Wiederhole 1000+ Mal
4. KI = 2.5% und 97.5% Perzentil

**Vorteil:** Keine Verteilungsannahmen!



### 3. Permutationstest

**Frage:** Ist Gruppenunterschied echt?  
**Idee:** Unter  $H_0$  sind Labels irrelevant  
**p-Wert:** Anteil extremerer Werte



Both Bootstrap and Permutation: "Only one test" – konsistentes Framework.

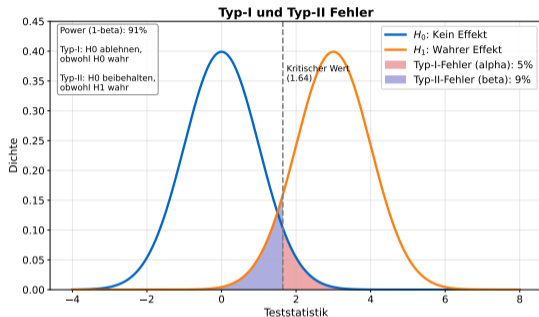
## 4. Fehler 1. und 2. Art

### Typ-I ( $\alpha = 0.05$ )

- $H_0$  ablehnen, obwohl wahr
- "Falsch Positiv"
- Bsp: Unwirksame Kampagne starten

### Typ-II ( $\beta$ , Power = $1 - \beta$ )

- $H_0$  behalten, obwohl  $H_1$  wahr
- "Falsch Negativ"
- Bsp: Wirksame Kampagne verpassen



### Expected Value: Entscheidung unter Unsicherheit

$$P(\text{Erfolg}) = 60\% \quad P(\text{Misserfolg}) = 40\%$$

	Kampagne erfolgreich	Kampagne nicht erfolgreich
Kampagne starten	+100k CHF	-30k CHF
Kampagne nicht starten	0k CHF	0k CHF

$$E[\text{Starten}] = 0.6 \cdot 100 + 0.4 \cdot (-30) = 48 \text{ k CHF}$$
$$E[\text{Nicht starten}] = 0 \text{ k CHF}$$

**Empfehlung: Kampagne starten (EV = +48k CHF)**

**Formel:**  $E[\text{Aktion}] = P(\text{Erfolg}) \times \text{Gewinn} + P(\text{Misserfolg}) \times \text{Verlust}$

Business-Entscheidungen basieren auf Kosten-Nutzen, nicht nur p-Werten.

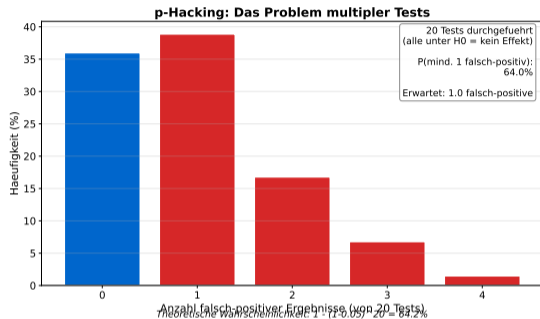
## 6. p-Hacking und FDR

**p-Hacking:** Analyse manipulieren bis  $p < 0.05$

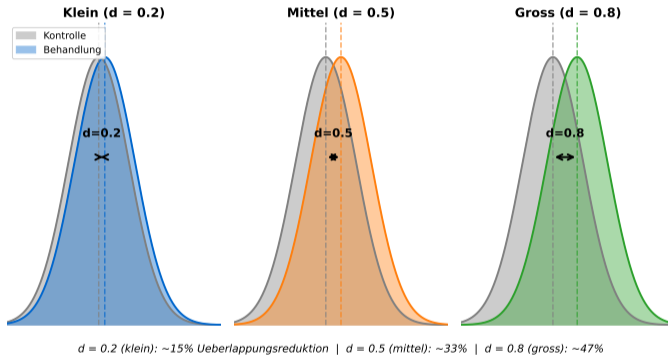
**Problem:** Bei 20 Tests ist 1 zufällig signifikant

**Lösungen:**

- Bonferroni:  $\alpha/m$
- FDR (Benjamini-Hochberg)



### Cohen's d: Effektstaerken-Interpretation



**Cohen's d:** 0.2 klein, 0.5 mittel, 0.8 gross – unabhängig von n!

## 8. Signifikanz vs. Relevanz

### Statistisch signifikant

- $p < 0.05$
- Hängt von  $n$  ab
- Bei Big Data: alles signifikant

**Regel:** Immer beide betrachten!

### Praktisch relevant

- Effektstärke ( $d$ ,  $r$ , etc.)
- Unabhängig von  $n$
- Business Impact

---

Ein winziger Effekt kann signifikant sein, aber irrelevant.

### Konzepte verstanden?

- Standardfehler und KI
- Bootstrap-Prinzip
- Permutationstest-Logik
- Typ-I/II Fehler
- Expected Value
- p-Hacking Problem
- Effektstärken

### R-Skills anwendbar?

- infer-Paket nutzen
- Bootstrap KI berechnen
- Permutationstest durchführen
- Regressionstabelle lesen
- `p.adjust()` anwenden
- Cohen's d berechnen