

## Block 2: Statistische Inferenz

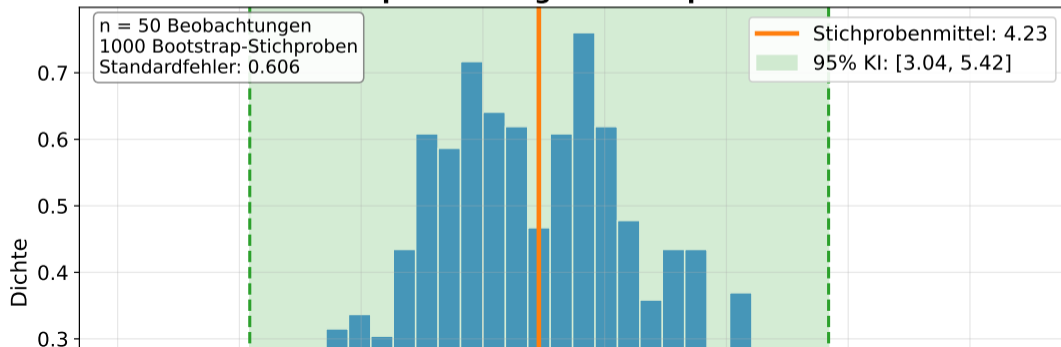
### Topic 2: Resampling-Methoden

**Grundprinzip:** Aus einer Stichprobe viele "neue" Stichproben ziehen

**Algorithmus:**

1. Ziehe zufällig  $n$  Beobachtungen **mit Zurücklegen** aus der ursprünglichen Stichprobe
2. Berechne die Statistik von Interesse (z.B. Mittelwert, Median)
3. Wiederhole Schritte 1–2 viele Male (z.B. 10.000 Mal)
4. Nutze die Verteilung der Bootstrap-Statistiken für Inferenz

## Bootstrap-Verteilung des Stichprobenmittels



## Perzentil-Methode:

- Berechne die Bootstrap-Verteilung der Statistik (z.B. 10.000 Bootstrap-Mittelwerte)
- 95%-Konfidenzintervall: Verwende das 2.5%- und 97.5%-Perzentil der Bootstrap-Verteilung
- Interpretation: Mit 95% Konfidenz liegt der wahre Parameter in diesem Intervall

## Vorteil:

- Keine Verteilungsannahmen nötig (z.B. keine Normalverteilung erforderlich)
- Funktioniert auch für komplexe Statistiken (Median, Korrelation, etc.)
- Besonders nützlich bei kleinen Stichproben oder schiefen Verteilungen

**Nachteil:** Rechenintensiv (aber mit moderner Hardware kein Problem)

---

Bootstrap-Konfidenzintervalle sind robust und flexibel

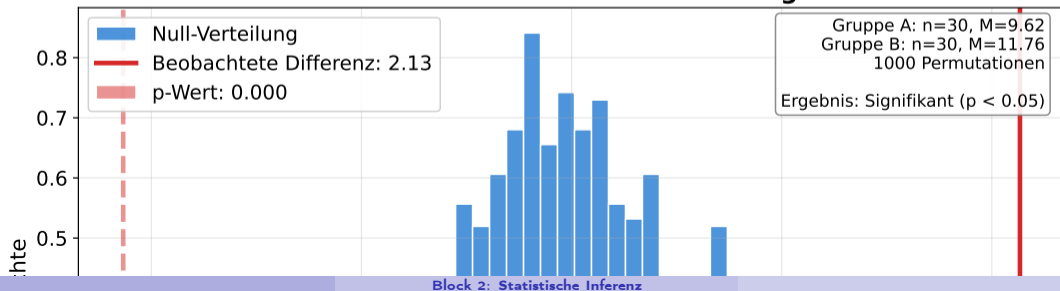
**Frage:** Ist der beobachtete Unterschied zwischen Gruppen echt oder nur Zufall?

**Idee:** Wenn  $H_0$  wahr ist (kein Unterschied), ist die Gruppenzugehörigkeit irrelevant

**Vorgehen:**

1. Berechne die beobachtete Teststatistik (z.B. Differenz der Mittelwerte)
2. Mische die Gruppenzugehörigkeit zufällig neu (Permutation)
3. Berechne die Teststatistik für die permutierte Stichprobe
4. Wiederhole Schritte 2–3 viele Male (z.B. 10.000 Mal)
5. p-Wert: Anteil der permutierten Statistiken, die extremer sind als die beobachtete

## Permutationstest: Ist der Unterschied signifikant?



**"Only one test" Philosophie:** Alle statistischen Tests folgen demselben Muster

**Konsistente Syntax für Bootstrap und Permutation:**

```
data %>%  
  specify(response ~ explanatory) %>%  
  generate(reps = 10000, type = "bootstrap") %>%  
  calculate(stat = "mean")
```

**Workflow:**

1. `specify()`: Definiere die Variablen und Hypothese
2. `generate()`: Erzeuge Bootstrap- oder Permutationsstichproben
3. `calculate()`: Berechne die Statistik für jede Stichprobe
4. `visualize()`: Visualisiere die Verteilung
5. `get_ci()`: Berechne Konfidenzintervall oder p-Wert

---

infer vereinheitlicht statistisches Testen in R