

Hypothesentest und Konfidenzintervall

Eine kompakte Zusammenfassung

Reimund Vehling

15. Januar 2026

1. Realität und Modell

Realität: Wir beobachten endlich viele Patient:innen. Es gibt biologische Variabilität, Messfehler, Bias und Confounding. Der wahre Parameter (z. B. ein Mittelwert oder Anteil) ist *fix*, aber unbekannt.

Modell: Für statistische Aussagen verwenden wir ein vereinfachtes Zufallsmodell (z. B. Normal-/t-Modell, Binomialmodell). *Wahrscheinlichkeiten existieren nur im Modell, nicht in der Realität.*

Merksatz: Statistik spricht über Daten *unter Annahmen*, nicht über Wahrheit.

2. Was ist ein Hypothesentest?

Ein Hypothesentest prüft, ob die beobachteten Daten *unter Annahme der Nullhypothese H_0* (im Modell) ungewöhnlich sind.

Der Test entscheidet nicht, ob H_0 wahr ist. Er bewertet ausschließlich die Daten.

Typische H_0 :

- kein Unterschied zwischen Gruppen
- kein Effekt einer Therapie
- Mittelwert = Referenzwert

3. Der p-Wert

Definition: Der p-Wert ist die Wahrscheinlichkeit, *mindestens so extreme Daten* zu beobachten, *wenn H_0 wahr wäre* (im Modell).

Der p-Wert ist nicht:

- die Wahrscheinlichkeit, dass H_0 wahr ist
- ein Maß für die Effektgröße
- eine Aussage über klinische Relevanz

4. Konfidenzintervall (KI)

Ein 95%-Konfidenzintervall beschreibt die Menge aller Parameterwerte, die mit den beobachteten Daten *im Modell* auf diesem Sicherheitsniveau vereinbar sind.

Zweiseitiger Zusammenhang:

H_0 wird auf $\alpha = 0,05$ verworfen \iff Nullwert liegt nicht im 95%-KI.

Achtung: Das Intervall enthält Parameterwerte – keine p-Werte.

5. Warum zweiseitig testen?

Zweiseitige Tests

- machen Unsicherheit sichtbar,
- passen direkt zu Konfidenzintervallen,
- zeigen Richtung und Größenordnung des Effekts.

Einseitige Tests erfordern starke Vorannahmen und sind didaktisch weniger transparent.

6. Drei typische Denkfehler

- Nicht signifikant \neq kein Effekt
- Signifikant \neq klinisch relevant
- Statistik entscheidet nicht über Wahrheit

7. Ein Satz, der immer richtig ist

Statistische Tests bewerten Daten unter Annahmen.