Hypothesentests I

Lektion B08 – BSc Wahrscheinlichkeit und Statistik

Digital Finance

Gliederung

Lernziele

Am Ende dieser Lektion werden Sie in der Lage sein:

1. Die Grundidee eines Hypothesentests und die Analogie zum Gerichtsverfahren zu erklären
2. Null- und Alternativhypothese (H_0, H_1) korrekt aus Fragestellungen zu formulieren
3. Einseitige und zweiseitige Tests zu unterscheiden und den passenden Test zu wählen
4. Fehler 1. und 2. Art (α, β) zu definieren und deren Konsequenzen zu beschreiben
5. Den p-Wert zu interpretieren und Testentscheidungen systematisch zu treffen

---

Diese Ziele leiten, was Sie aus dieser Lektion beherrschen sollten.

Warum Hypothesentests?

Problem: Wir beobachten Stichprobendaten und wollen eine **Behauptung über die Grundgesamtheit** prüfen.

- Hat sich die mittlere Tagesrendite einer Aktie nach einem Ereignis verändert?
- Ist der Marktanteil eines Fonds tatsächlich gestiegen?
- Liegt die Ausfallrate eines Portfolios über dem Branchenstandard?

Lösung: Ein **Hypothesentest** liefert eine systematische, reproduzierbare Entscheidungsregel auf Basis der Daten.

Kernfrage

Ist der beobachtete Unterschied **statistisch signifikant** – oder nur zufällige Schwankung?

Hypothesentests sind das Standardwerkzeug der schliessenden Statistik.

Analogie: Gerichtsverfahren

Gericht:

- Angeklagter gilt als **unschuldig** (Ausgangslage)
- Anklage muss Schuld **beweisen**
- Freispruch $\neq$ Beweis der Unschuld
- Beweislast liegt bei der Anklage

Hypothesentest:

- H_0 gilt als **wahr** (Ausgangslage)
- Daten müssen **genug Evidenz** gegen H_0 liefern
- H_0 nicht ablehnen $\neq$ H_0 beweisen
- Beweislast liegt bei den Daten

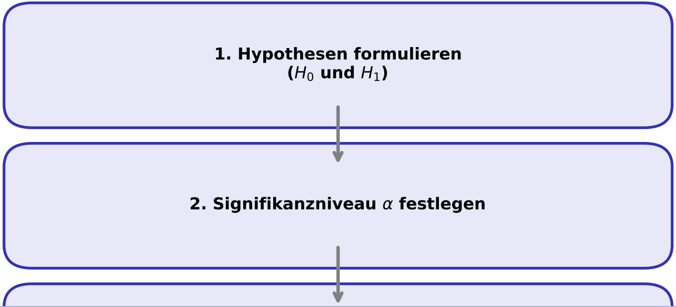
Wichtiges Prinzip

Wir lehnen H_0 nur ab, wenn die Daten **stark genug dagegen sprechen**. Ansonsten behalten wir H_0 bei.

„Im Zweifel fuer den Angeklagten“ = „Im Zweifel fuer H_0“

1. **Hypothesen formulieren:** H_0 (Nullhypothese) und H_1 (Alternativhypothese) aufstellen
2. **Signifikanzniveau festlegen:** α wählen (typisch: 0,05 oder 0,01)
3. **Teststatistik berechnen:** Aus den Stichprobendaten einen Prüfwert bestimmen
4. **p-Wert bestimmen** (oder mit kritischem Wert vergleichen)
5. **Entscheidung treffen:** H_0 ablehnen oder nicht ablehnen

Ablauf eines Hypothesentests (5 Schritte)



Nullhypothese und Alternativhypothese

Nullhypothese H_0:

- Status quo, „kein Effekt“
- Enthält immer ein **Gleichheitszeichen**
- Beispiele: $\mu = \mu_0$, $p = p_0$, $\mu \leq \mu_0$

Alternativhypothese H_1:

- Die Behauptung, die wir prüfen wollen
- Enthält **kein Gleichheitszeichen**
- Beispiele: $\mu \neq \mu_0$, $p > p_0$, $\mu > \mu_0$

Merkreger

H_0 und H_1 sind **komplementär** – zusammen decken sie alle Möglichkeiten ab. Genau eine der beiden ist wahr.

H_0 ist die „Standardannahme“, H_1 ist das, was wir zeigen wollen.

Beispiele: Hypothesen aus Fragestellungen

Verbale Beschreibung → Mathematische Hypothesen:

Fragestellung	H_0	H_1
Hat sich die mittlere Rendite verändert?	$\mu = \mu_0$	$\mu \neq \mu_0$
Ist die mittlere Rendite gestiegen?	$\mu \leq \mu_0$	$\mu > \mu_0$
Liegt die Ausfallrate unter 5%?	$p \geq 0{,}05$	$p < 0{,}05$
Ist der Marktanteil grösser als 20%?	$p \leq 0{,}20$	$p > 0{,}20$
Weicht die Volatil. vom Benchmark ab?	$\sigma = \sigma_0$	$\sigma \neq \sigma_0$

Praxistipp

Die Behauptung, die man **belegen** möchte, gehört in H_1. H_0 formuliert man als Gegenteil.

Die Formulierung der Hypothesen bestimmt den gesamten weiteren Testverlauf.

Finanzbeispiel: Hypothesen

Situation: Ein Fondsmanager behauptet, seine Strategie liefere eine mittlere Jahresrendite von **mehr als 8%**.

Schritt 1 – Hypothesen:

$$H_0 : \mu \leq 8\% \qquad \text{vs.} \qquad H_1 : \mu > 8\%$$

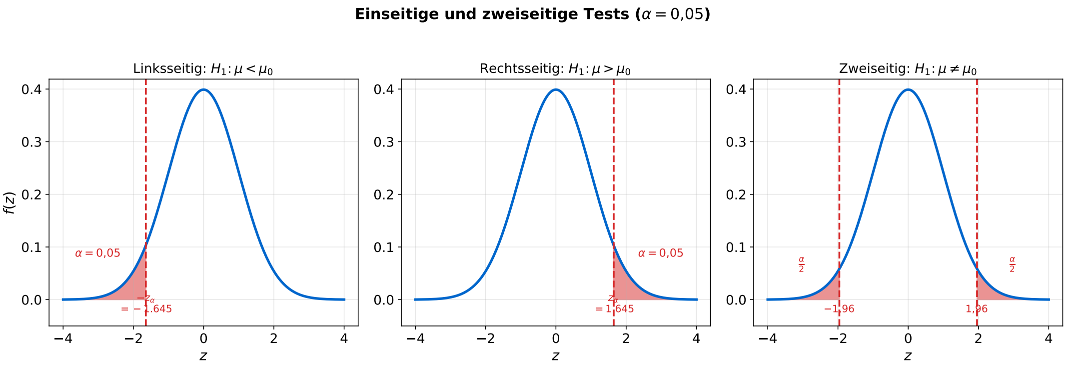
- H_0: Die Rendite ist höchstens 8% (kein überlegener Fonds)
- H_1: Die Rendite ist tatsächlich grösser als 8% (Behauptung des Managers)
- Dies ist ein **rechtsseitiger** Test, da H_1 eine Richtung vorgibt ($>$)

Warum so herum?

Der Manager trägt die **Beweislast**. Ohne genügend Evidenz nehmen wir an, dass seine Strategie nicht besser als 8% ist.

Die Beweislast liegt immer bei H_1 – genau wie im Gerichtsverfahren.

Einseitige und zweiseitige Tests



- **Linksseitig** (H_1: $\mu < \mu_0$): Ablehnungsbereich im linken Rand
- **Rechtsseitig** (H_1: $\mu > \mu_0$): Ablehnungsbereich im rechten Rand
- **Zweiseitig** (H_1: $\mu \neq \mu_0$): Ablehnungsbereich in **beiden** Rändern

Die Wahl hängt von H_1 ab – vor der Datenerhebung festlegen!

Wann welcher Test?

Fragestellung	**Testrichtung**	H_1
„Ist der Wert **anders** als μ_0?“	Zweiseitig	$\mu \neq \mu_0$
„Ist der Wert **grösser** als μ_0?“	Rechtsseitig	$\mu > \mu_0$
„Ist der Wert **kleiner** als μ_0?“	Linksseitig	$\mu < \mu_0$

Wichtig

- Einseitige Tests sind **mächtiger** in der gewählten Richtung (kleinerer kritischer Wert)
- Aber: Sie können Abweichungen in der **anderen** Richtung nicht erkennen
- Die Richtung muss **vor Sichtung der Daten** feststehen!

Zweiseitig ist die „sichere“ Wahl; einseitig nur bei klarer theoretischer Richtung.

Kritische Werte: Einseitig vs. Zweiseitig

Bei gleichem $\alpha = 0{,}05$:

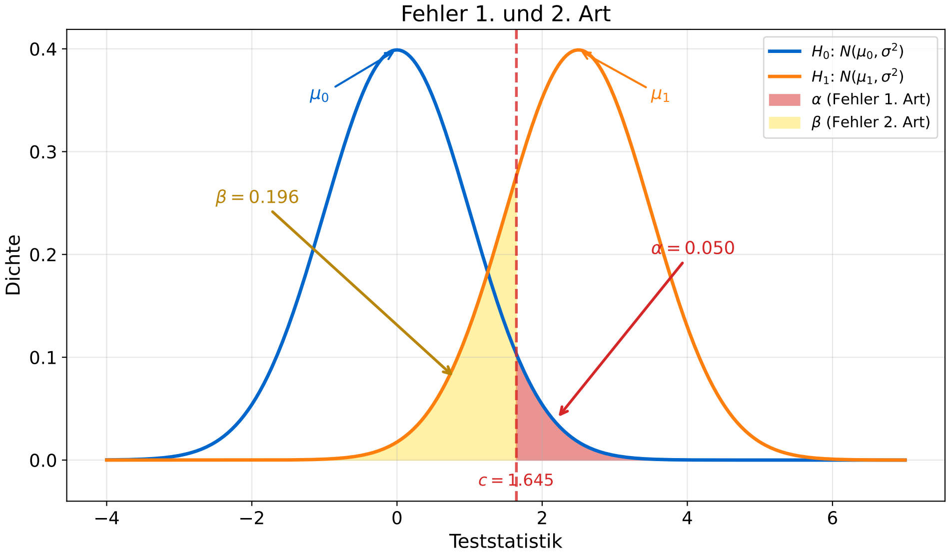
Testart	**Kritischer Wert**	**α-Aufteilung**
Rechtsseitig	$z_{0,05} = 1{,}645$	Gesamtes α rechts
Linksseitig	$-z_{0,05} = -1{,}645$	Gesamtes α links
Zweiseitig	$\pm z_{0,025} = \pm 1{,}960$	Je $\alpha/2$ pro Seite

Konsequenz

Beim zweiseitigen Test braucht man einen **extremeren** Wert der Teststatistik, um H_0 abzulehnen (1,960 statt 1,645), da das α auf beide Seiten aufgeteilt wird.

Einseitig: leichter signifikant in einer Richtung. Zweiseitig: konservativer, aber erkennt beide Richtungen.

Fehler bei Testentscheidungen



Entscheidungstabelle

	H_0 **ist wahr**	H_1 **ist wahr**
H_0 **nicht ablehnen**	Korrekt $(1-\alpha)$	Fehler 2. Art (β)
H_0 **ablehnen**	Fehler 1. Art (α)	Korrekt $(1-\beta)$

Finanzbeispiele Fehler 1. Art:

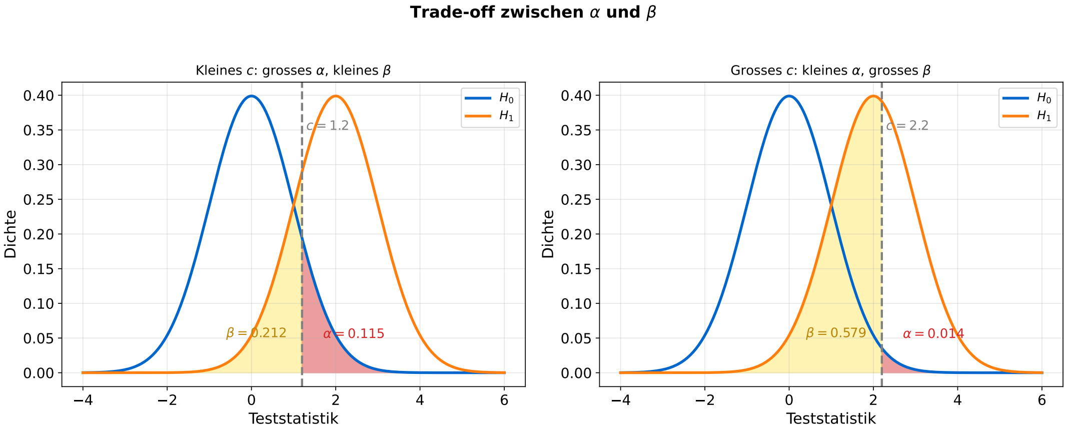
- Fondsmanager hat keinen Skill, wird aber als erfolgreich eingestuft
- Gesunde Aktie wird als riskant klassifiziert

Finanzbeispiele Fehler 2. Art:

- Fondsmanager hat Skill, wird aber nicht erkannt
- Riskante Aktie wird als sicher eingestuft

$1-\beta$ **heisst Macht (Power) des Tests – die Wahrscheinlichkeit, einen echten Effekt zu erkennen.**

Trade-off: α und β

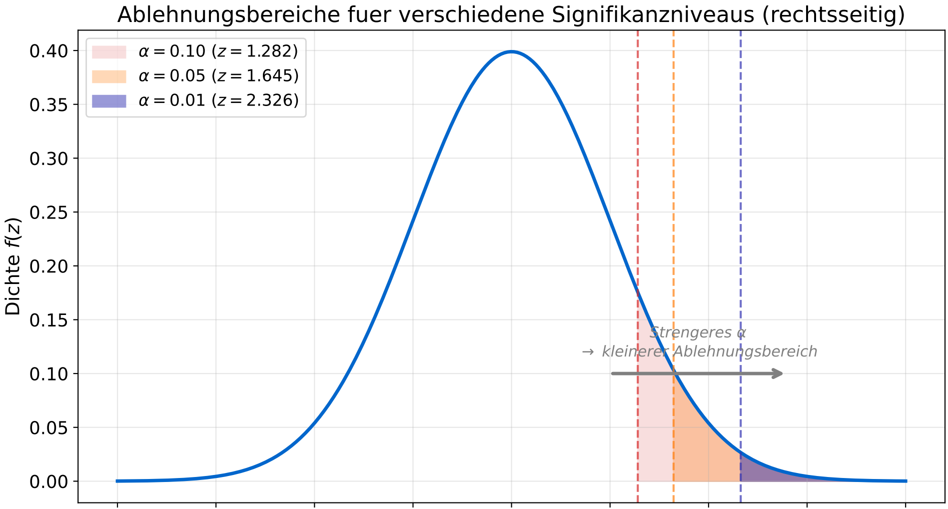


Fundamentaler Trade-off

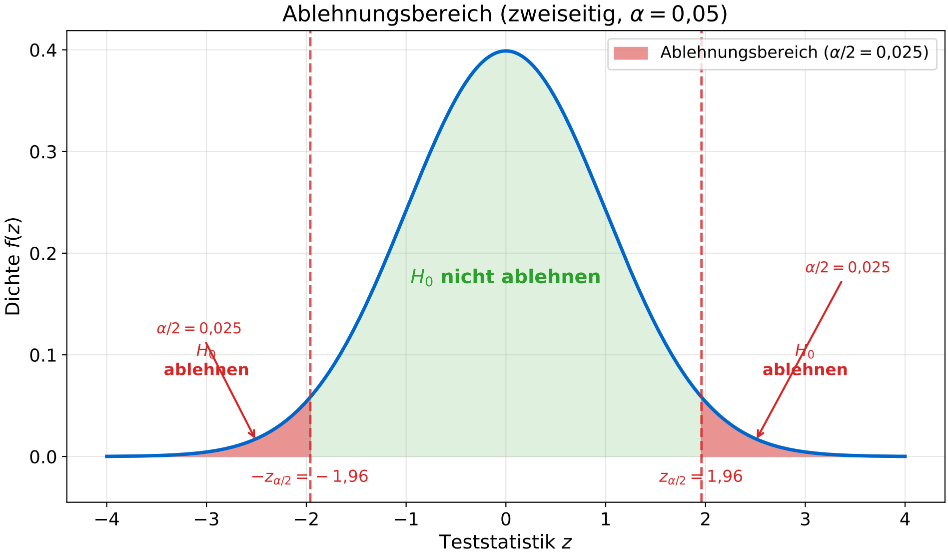
- Strengeres α (weniger Fehler 1. Art) $\Rightarrow$ grösseres β (mehr Fehler 2. Art)
- Man kann **nicht** beide gleichzeitig beliebig klein machen (bei fester Stichprobengrösse)

Das Signifikanzniveau α

Definition: $\alpha = P(H_0 \text{ ablehnen} \mid H_0 \text{ wahr})$ – die maximal tolerierte Wahrscheinlichkeit eines Fehlers 1. Art.



Kritische Werte und Ablehnungsbereich



Beispiel: Kritischer Wert berechnen

Situation: Zweiseitiger Test, $\alpha = 0{,}05$, Teststatistik ist standardnormalverteilt.

Gesucht: Kritische Werte $z_{\alpha/2}$

$$\begin{aligned} \alpha/2 &= 0{,}025 \\ P(Z > z_{\alpha/2}) &= 0{,}025 \\ z_{\alpha/2} &= \Phi^{-1}(1 - 0{,}025) = \Phi^{-1}(0{,}975) = 1{,}96 \end{aligned}$$

Ablehnungsbereich: $z_{\text{obs}} < -1{,}96$ **oder** $z_{\text{obs}} > 1{,}96$

Wichtige kritische Werte der Standardnormalverteilung

α	0,10	0,05	0,02	0,01
z_α (einseitig)	1,282	1,645	2,054	2,326
$z_{\alpha/2}$ (zweiseitig)	1,645	1,960	2,326	2,576

Diese Werte sollten Sie auswendig kennen oder schnell nachschlagen können.

Was ist der p-Wert?

Definition:

p-Wert

Der p-Wert ist die Wahrscheinlichkeit, unter der Annahme dass H_0 wahr ist, eine Teststatistik zu beobachten, die **mindestens so extrem** ist wie der beobachtete Wert.

Formell:

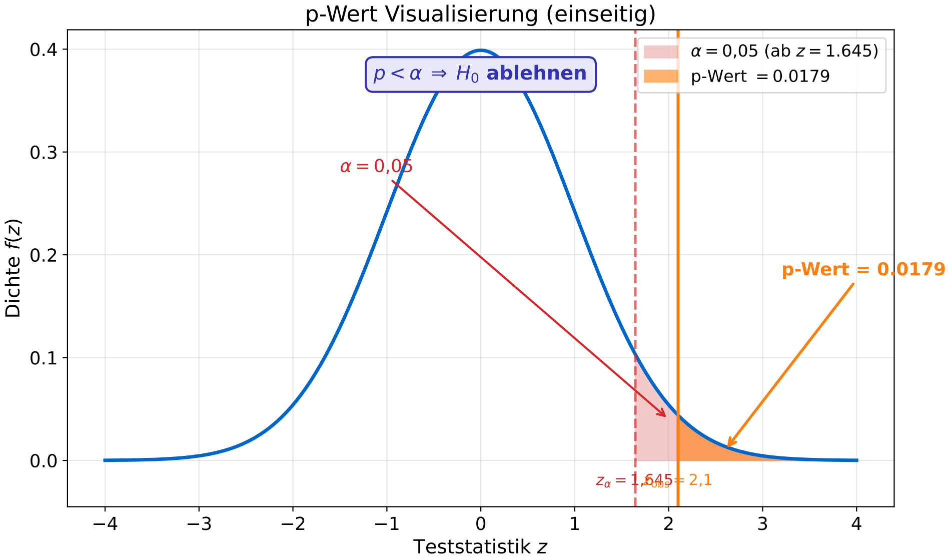
- Rechtsseitig: $p = P(Z \geq z_{\text{obs}} \mid H_0)$
- Linksseitig: $p = P(Z \leq z_{\text{obs}} \mid H_0)$
- Zweiseitig: $p = 2 \cdot P(Z \geq |z_{\text{obs}}| \mid H_0)$

Entscheidungsregel:

$$\boxed{p < \alpha \implies H_0 \text{ ablehnen}}$$

Je kleiner der p-Wert, desto stärker die Evidenz **gegen** H_0.

Der p-Wert misst, wie „überraschend“ die Daten unter H_0 sind.



p-Wert Visualisierung (einseitig)
$p < \alpha \Rightarrow H_0$ **ablehnen**
$\alpha = 0{,}05$ (ab $z = 1.645$)
p-Wert $= 0.0179$
$\alpha = 0{,}05$
p-Wert = 0.0179
$z_\alpha = 1{,}645$
$z_{obs} = 2{,}1$
0.4
0.3
0.2
0.1
0.0
Dichte $f(z)$
−4
−3
−2
−1
0
1
2
3
4
Teststatistik z

p-Wert: Bewertungsskala

p-Wert	Interpretation
$p > 0{,}10$	Kein Hinweis gegen H_0
$0{,}05 < p \leq 0{,}10$	Schwache Evidenz gegen H_0 („marginal signifikant“)
$0{,}01 < p \leq 0{,}05$	Moderate Evidenz gegen H_0 („signifikant“)
$0{,}001 < p \leq 0{,}01$	Starke Evidenz gegen H_0 („hoch signifikant“)
$p \leq 0{,}001$	Sehr starke Evidenz gegen H_0

Achtung

Diese Abstufung ist nur eine Orientierungshilfe. Formal gilt: bei gewähltem α wird H_0 genau dann abgelehnt, wenn $p < \alpha$.

Berichten Sie immer den exakten p-Wert, nicht nur „signifikant“ oder „nicht signifikant“.

Beispiel: Mittlere Rendite testen (1/3)

Situation: Ein Analyst behauptet, die mittlere Tagesrendite einer Aktie sei $\mu_0 = 0{,}05\%$. In einer Stichprobe von $n = 64$ Handelstagen wird beobachtet:

$$\bar{x} = 0{,}09\%, \quad \sigma = 0{,}16\% \text{ (bekannt)}$$

Frage: Gibt es bei $\alpha = 0{,}05$ Evidenz, dass die mittlere Rendite von 0,05% abweicht?

Schritt 1 – Hypothesen:

$$H_0 : \mu = 0{,}05\% \qquad \text{vs.} \qquad H_1 : \mu \neq 0{,}05\%$$

(Zweiseitiger Test, da „abweicht" keine Richtung vorgibt.)

Schritt 2 – Signifikanzniveau:

$$\alpha = 0{,}05$$

Erst Hypothesen und α festlegen, dann rechnen!

Beispiel: Mittlere Rendite testen (2/3)

Schritt 3 – Teststatistik berechnen:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{0{,}09 - 0{,}05}{0{,}16/\sqrt{64}} = \frac{0{,}04}{0{,}02} = 2{,}0$$

Schritt 4 – p-Wert bestimmen:
Da zweiseitig:

$$p = 2 \cdot P(Z \geq |z_{\text{obs}}|) = 2 \cdot P(Z \geq 2{,}0) = 2 \cdot 0{,}0228 = 0{,}0456$$

Alternativ mit kritischem Wert:

$$z_{\alpha/2} = z_{0{,}025} = 1{,}96 \qquad \text{und} \qquad |z_{\text{obs}}| = 2{,}0 > 1{,}96$$

Beide Wege – p-Wert und kritischer Wert – fuehren zur selben Entscheidung.

Beispiel: Mittlere Rendite testen (3/3)

Schritt 5 – Entscheidung:

Über den p-Wert:

$$p = 0{,}0456 < 0{,}05 = \alpha \quad \Longrightarrow \quad H_0 \text{ ablehnen}$$

Über den kritischen Wert:

$$|z_{\text{obs}}| = 2{,}0 > 1{,}96 = z_{\alpha/2} \quad \Longrightarrow \quad H_0 \text{ ablehnen}$$

Interpretation

Bei einem Signifikanzniveau von 5% gibt es genügend statistische Evidenz, dass die mittlere Tagesrendite von 0,05% abweicht. Der beobachtete Mittelwert von 0,09% ist statistisch signifikant verschieden von $\mu_0 = 0{,}05\%$.

Beachte: Bei $\alpha = 0{,}01$ wäre $p = 0{,}0456 > 0{,}01$ – wir würden H_0 **nicht** ablehnen! Die Entscheidung hängt vom gewählten α ab.

Immer den p-Wert berichten, damit der Leser selbst urteilen kann.

Häufige Fehler bei Hypothesentests

Fehler 1: „Der p-Wert ist die Wahrscheinlichkeit, dass H_0 wahr ist“

FALSCH! Der p-Wert ist $P(\text{Daten} \mid H_0)$, **nicht** $P(H_0 \mid \text{Daten})$.
Er sagt: „Wie wahrscheinlich sind so extreme Daten, **wenn** H_0 stimmt?“

Fehler 2: „H_0 nicht ablehnen $=$ H_0 ist bewiesen“

FALSCH! Nicht-Ablehnung bedeutet nur: Es gibt **nicht genügend Evidenz** gegen H_0.
Analog: Ein Freispruch beweist nicht die Unschuld.

Fehler 3: „Statistisch signifikant $=$ praktisch relevant“

FALSCH! Bei grossen n können auch winzige Effekte signifikant werden.
Immer auch die **Effektgrösse** betrachten!

Diese Missverstaendnisse sind extrem verbreitet – auch in Fachpublikationen.

Weitere Fallstricke

- α **nachträglich anpassen:** Das Signifikanzniveau muss **vor** der Datenanalyse feststehen. Wer α so wählt, dass das Ergebnis gerade signifikant wird, betreibt „*p*-Hacking“.

- **Multiples Testen:** Bei 20 Tests mit $\alpha = 0{,}05$ erwartet man ca. 1 falsch-signifikantes Ergebnis, auch wenn kein echter Effekt vorliegt (Bonferroni-Korrektur nötig).

- **Verwechslung von α und β:**
  - α: Fehler 1. Art (falsch-positiv) – **kontrolliert** durch den Forscher
  - β: Fehler 2. Art (falsch-negativ) – hängt von Effektgrösse und n ab

- **Konfidenzintervall ignoriert:** Ein Hypothesentest liefert nur „Ja/Nein“. Das Konfidenzintervall zeigt zusätzlich die **Grösse und Richtung** des Effekts.

Gute Praxis: p-Wert + Konfidenzintervall + Effektgroesse gemeinsam berichten.

Zusammenfassung: Wichtige Formeln

z-**Teststatistik** (bei bekanntem σ):

$$z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

p-Wert Berechnung:

Testart	**p-Wert**
Rechtsseitig (H_1: $\mu > \mu_0$)	$p = P(Z \geq z_{\text{obs}}) = 1 - \Phi(z_{\text{obs}})$
Linksseitig (H_1: $\mu < \mu_0$)	$p = P(Z \leq z_{\text{obs}}) = \Phi(z_{\text{obs}})$
Zweiseitig (H_1: $\mu \neq \mu_0$)	$p = 2 \cdot P(Z \geq \lvert z_{\text{obs}} \rvert) = 2\big(1 - \Phi(\lvert z_{\text{obs}} \rvert)\big)$

Entscheidungsregel:

$$p < \alpha \implies H_0 \text{ ablehnen} \qquad\qquad p \geq \alpha \implies H_0 \text{ nicht ablehnen}$$

Alle Formeln gelten fuer den z-Test bei bekannter Standardabweichung σ.

Zusammenfassung: Begriffe und Konzepte

Begriff	**Beschreibung**
Nullhypothese H_0	Status quo, enthält Gleichheitszeichen
Alternativhypothese H_1	Behauptung, die belegt werden soll
Signifikanzniveau α	Max. Fehlerwahrscheinlichkeit 1. Art
Fehler 1. Art	H_0 ablehnen, obwohl H_0 wahr (α)
Fehler 2. Art	H_0 nicht ablehnen, obwohl H_1 wahr (β)
Macht (Power)	$1 - \beta$, Wahrscheinlichkeit, echten Effekt zu erkennen
p-Wert	$P(\text{so extrem oder extremer} \mid H_0)$
Kritischer Wert	Grenze des Ablehnungsbereichs
Teststatistik	Aus Daten berechneter Prüfwert

Nächste Lektion: Hypothesentests II – t-Test, Varianztest, Tests fuer Anteile.

Zusammenfassung: Ablauf

1. H_0 und H_1 formulieren

↓

2. α festlegen

↓

3. Teststatistik berechnen

↓

4. p-Wert bestimmen

↓