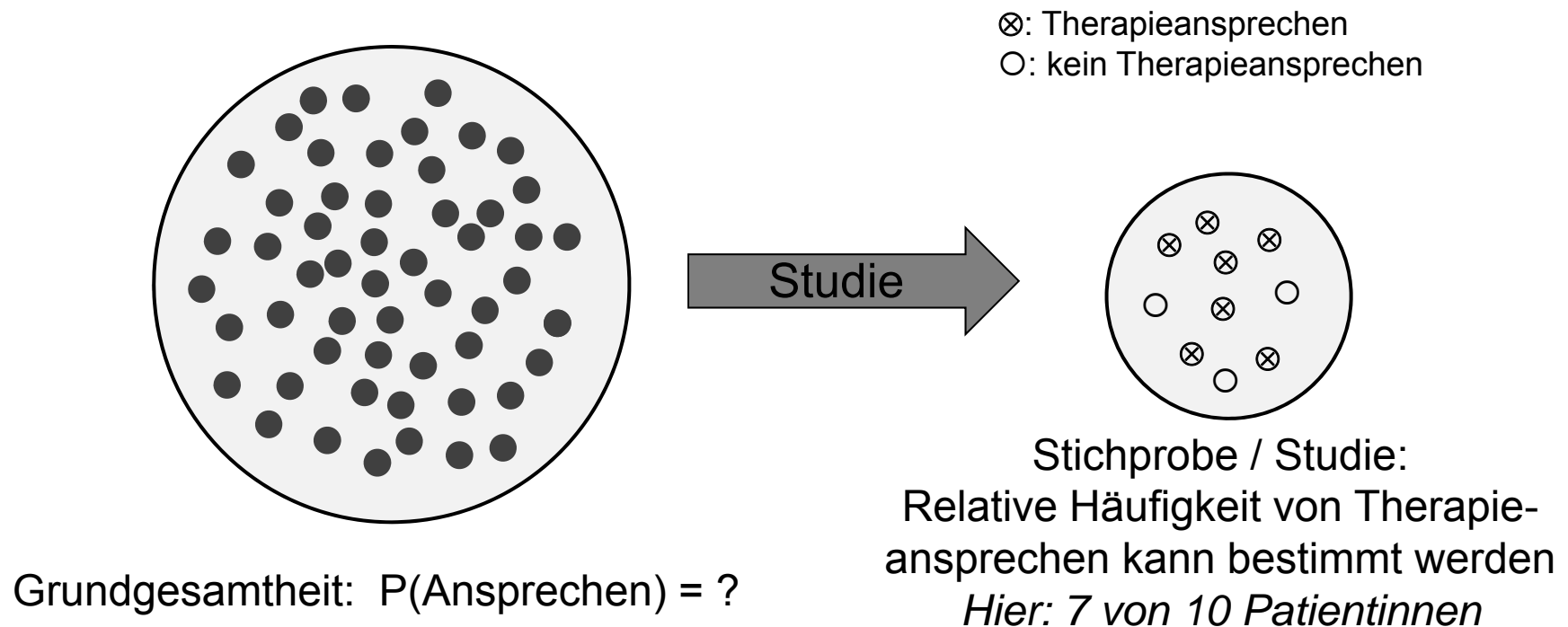

Übung

Ein Münzwurfexperiment

zur Herleitung und Erläuterung der Verteilung der
Teststatistik unter der Nullhypothese und des p-Werts

Statistischer Test – Ein fiktives Beispiel

Es soll überprüft werden, ob die Ansprechrate einer neuen Therapie bei Mammakarzinompatientinnen größer ist als die unter Standardtherapie erwarteten 50%.



Testen: Formulieren der Hypothesen

Nullhypothese

$$H_0: P(\text{Ansprechen}) \leq 50\%$$

Alternativhypothese

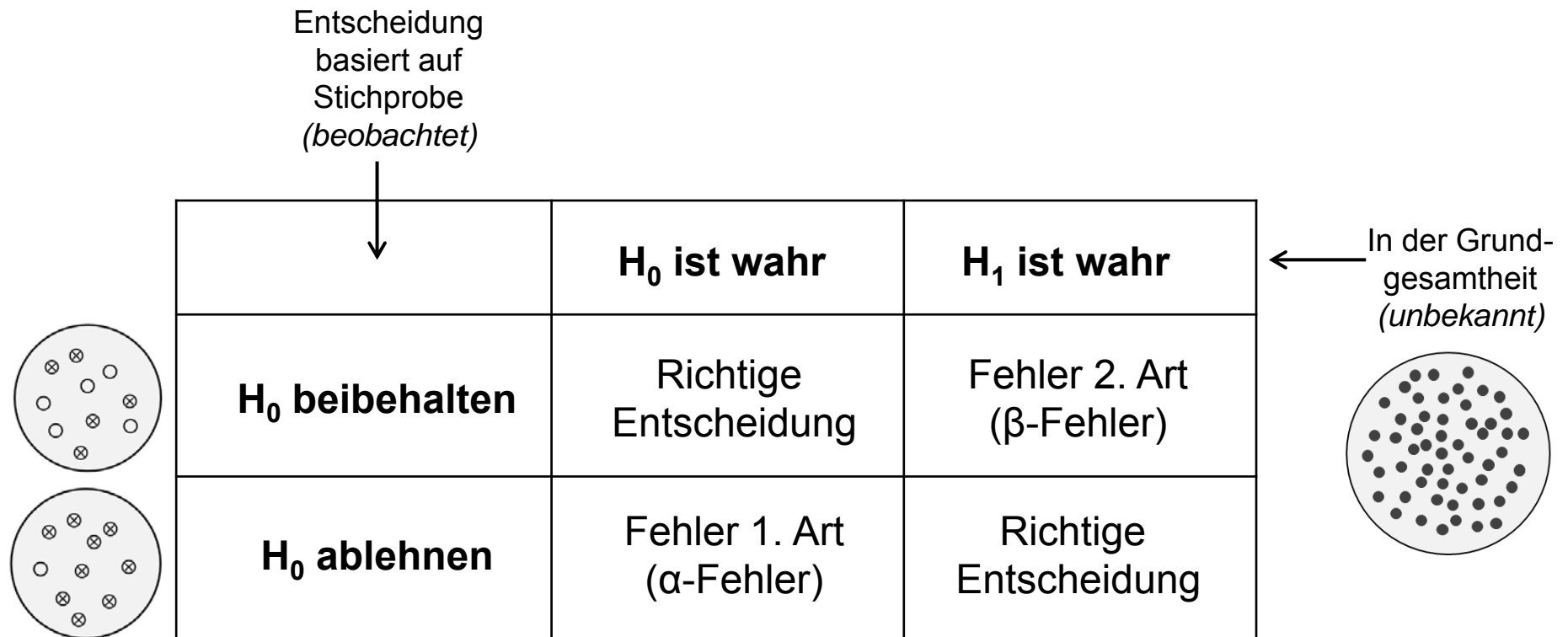
$$H_1: P(\text{Ansprechen}) > 50\%$$

Beobachtet: 7 Erfolge bei 10 Patientinnen (70%)

Für welche der beiden Hypothesen sollte man sich entscheiden?

Was sind mögliche Fehlentscheidungen?

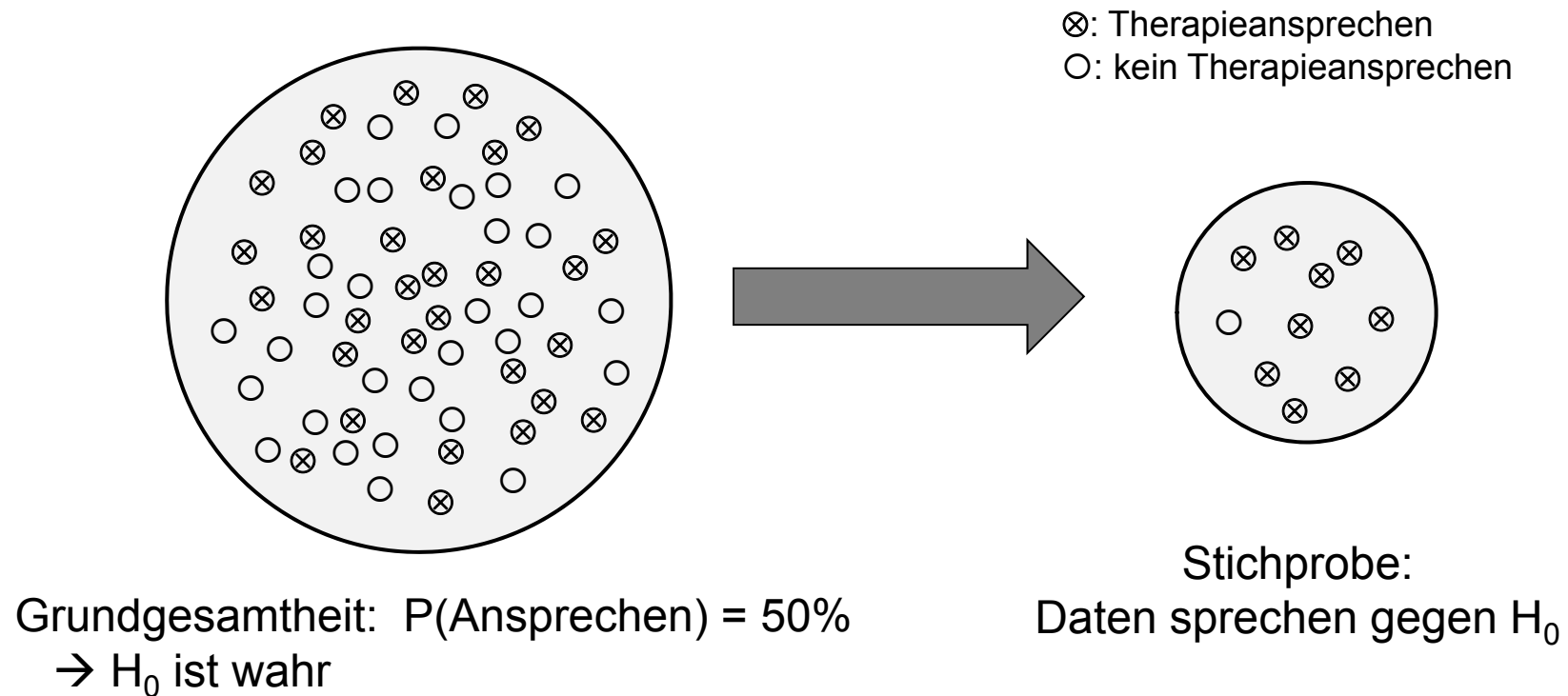
Testen – mögliche Fehlentscheidungen



Testen – mögliche Fehlentscheidungen

Fehler 1. Art:

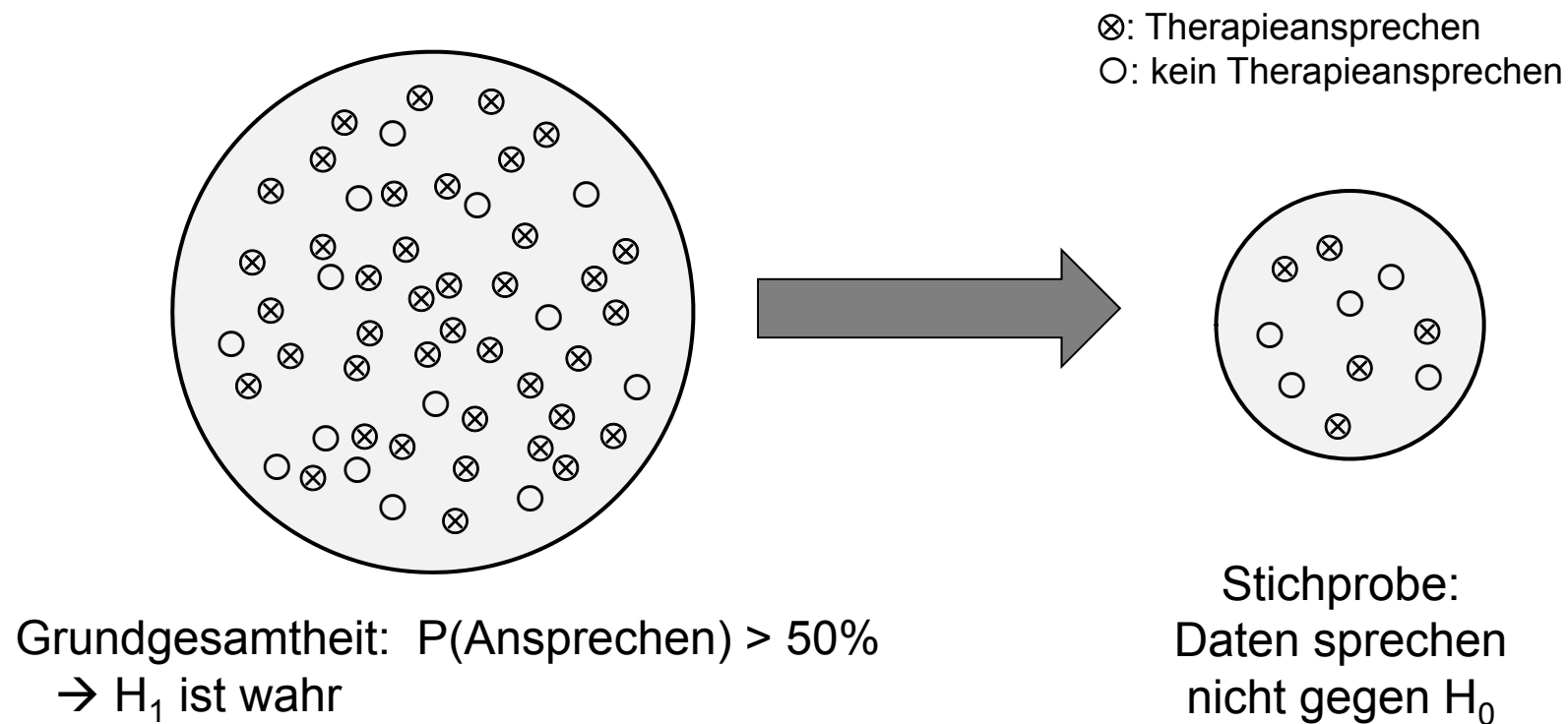
H_0 ist wahr, aber H_0 wird (zugunsten von H_1) abgelehnt



Testen – mögliche Fehlentscheidungen

Fehler 2. Art:

H_1 ist wahr, aber H_0 wird nicht abgelehnt



Testen – wie entscheiden?

Beispiel:

$$H_0: P(\text{Ansp.}) \leq 50\%$$

$$H_1: P(\text{Ansp.}) > 50\%$$

Mögliches Studienergebnis:

Bei 7 von 10 behandelten Patientinnen zeigte sich Therapieansprechen. Für H_0 oder H_1 entscheiden?

→ Statistischer Test

Testen – wie entscheiden?

Beispiel:

$$H_0: P(\text{Ansp.}) \leq 50\%$$

$$H_1: P(\text{Ansp.}) > 50\%$$

Ergebnis:

Beobachtetes Ansprechen

bei 7 von 10 Patientinnen

Priorität bei einem statistischen Test ist, die **Wahrscheinlichkeit für den Fehler 1. Art** (=Wkt. für falsch positives Testergebnis) zu kontrollieren

Idee: Wie wahrscheinlich sind 7 oder mehr Therapieansprecher bei 10 Patientinnen unter H_0 (d.h. wenn $P(\text{Ansp.})$ tatsächlich = 50% ist)?

- Diese Wahrscheinlichkeit wird als **p-Wert** bezeichnet
 - Ist diese Wkt. klein, so wird H_0 abgelehnt.
 - Die Schranke (Wkt. für Fehler 1. Art) wird vorab festgelegt und als Signifikanzniveau α bezeichnet (häufig $\alpha = 5\%$)
-

Verteilung der Teststatistik

Beispiel:

$$H_0: p \leq 50\%$$

$$H_1: p > 50\%$$

Wie lässt sich die Verteilung der Anzahl an Therapieansprechern unter H_0 - bzw. hier am Rand von H_0 : $P(\text{Ansprechen}=50\%)$ - bestimmen?

- Theoretische Herleitung (Binomialverteilung)
 - Simulation / experimentell
- z.B. sehr häufiger Wurf einer fairen Münze

Experiment: Abschätzung der Verteilung von Therapieansprechern unter H_0 , hier $P(\text{Ansp.}) = 50\%$

Werfen Sie eine Münze 10 mal und notieren Sie, wie häufig dabei „Zahl“ eingetreten ist:



Wiederholen Sie das Experiment ... mal!

Beispiel

1)  $6 \times \text{Zahl}$

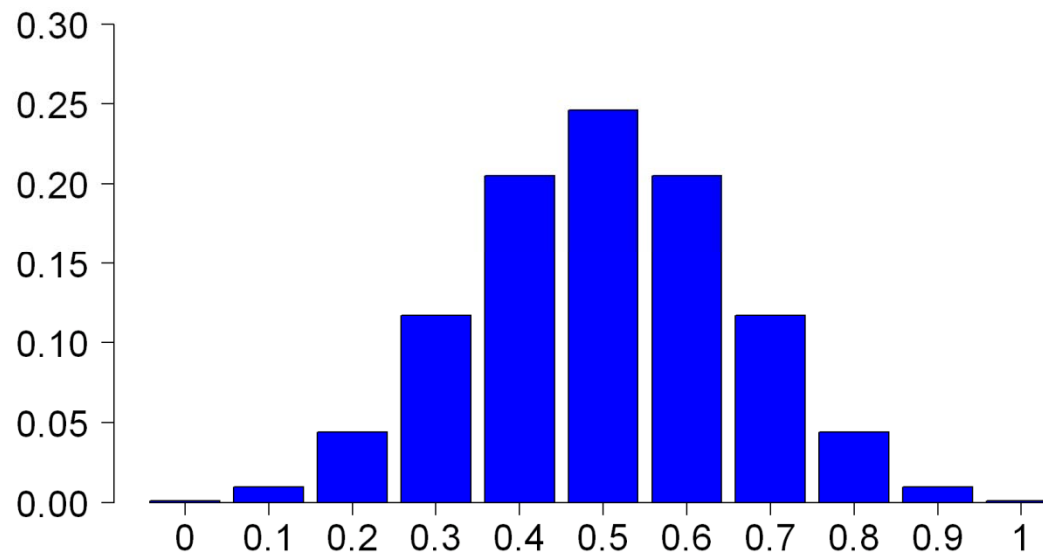
2)  $4 \times \text{Zahl}$

3)  $6 \times \text{Zahl}$

4)  $3 \times \text{Zahl}$

Verteilung unter H_0 ($p=50\%$)

Erwartete Verteilung



Ansprecher	0	1	2	3	4	5	6	7	8	9	10
P(#Ansp.)	0,1%	1,0%	4,4%	11,7%	20,5%	24,6%	20,5%	11,7%	4,4%	1,0%	0,1%

Testentscheidung

Bei einem statistischen Test gilt:

Die Wahrscheinlichkeit für den Fehler erster Art – die Nullhypothese wird abgelehnt, obwohl sie richtig ist – soll klein sein

→ in der Regel (per Konvention): $\leq 5\%$ = **Signifikanzniveau (α)**

Wahrscheinlichkeit für den Fehler 2. Art wird im Test nicht kontrolliert – sie hängt z.B. maßgeblich von der Fallzahl ab.

Testentscheidung - Beispiel

Hier: Therapieansprechen bei 7 von 10 Patientinnen.

→ Wie groß ist die Wahrscheinlichkeit 7 Therapieansprecher oder mehr unter H_0 zu beobachten?

Ansprecher	0	1	2	3	4	5	6	7	8	9	10
P(#Ansp.)	0,1%	1,0%	4,4%	11,7%	20,5%	24,6%	20,5%	11,7%	4,4%	1,0%	0,1%

$$P(\text{\#Ansp. unter } H_0 \geq 7) = 11,7\% + 4,4\% + 1,0\% + 0,1\% = 17,2\%$$

Diese Wahrscheinlichkeit wird als p-Wert bezeichnet!

Wenn $p \leq \alpha$ (häufig: $\alpha=5\%$) → H_0 ablehnen

$p > 0,05$ → H_0 kann nicht abgelehnt werden

In der Stichprobe/Studie
beobachtetes Ergebnis
und in Richtung der
Alternativhypothese (H_1)
extremere Ergebnisse

Entscheidung treffen - Beispiel

Hier: $p = 0,172 > 0,05$

→ H_0 kann nicht abgelehnt werden

Wichtig:

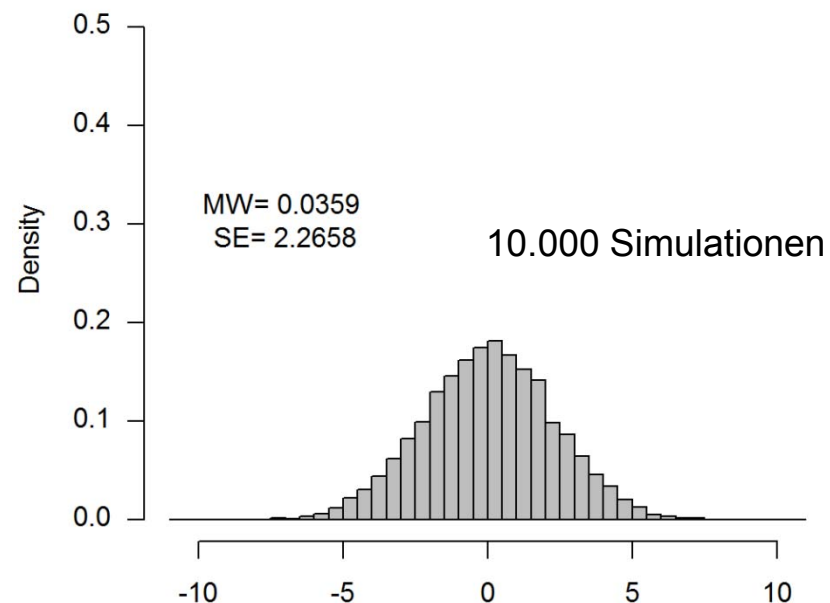
- Dies ist kein „Nachweis“ für H_0 !
 - Es ist nicht genug Evidenz vorhanden, um die Nullhypothese (hier: $p \leq 50\%$) abzulehnen.
 - Wahrscheinlichkeit für den Fehler 2. Art wird im Test nicht kontrolliert.
 - „Absence of evidence is not evidence of absence“ (Altman DG & Bland JM, *BMJ*, 1995)
-

Ausblick: Der Zweistichproben-t-Test

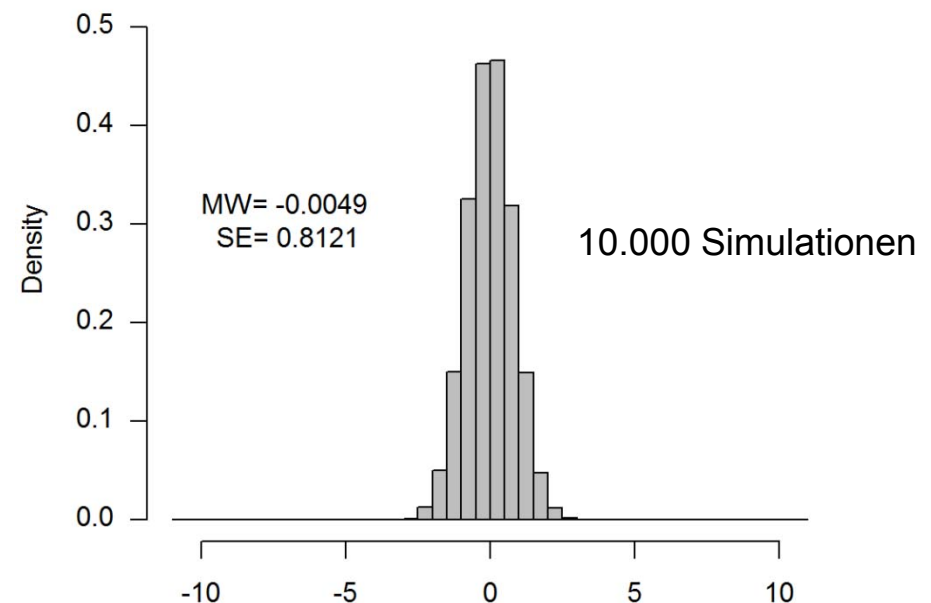
- Mittelwertvergleich zweier Gruppen (A vs. B)
 - $H_0: \mu_A = \mu_B$ $H_0: \mu_A - \mu_B = 0$
 - $H_1: \mu_A \neq \mu_B$ $H_1: \mu_A - \mu_B \neq 0$
 - Teststatistik:
$$T = \frac{\text{Mittelwert}_A - \text{Mittelwert}_B}{\sqrt{\frac{SD_A^2}{n_A} + \frac{SD_B^2}{n_B}}}$$
 - $T \sim t(k)$ mit $k = \left(\frac{SD_A^2}{n_A} + \frac{SD_B^2}{n_B} \right)^2 / \left(\frac{1}{n_A - 1} \left(\frac{SD_A^2}{n_A} \right)^2 + \frac{1}{n_B - 1} \left(\frac{SD_B^2}{n_B} \right)^2 \right)$
 - Für große n_A, n_B : $T^{\text{approx.}} \sim N(0;1)$
-

Simulation unter H_0 ($\mu_A = \mu_B$)

- Zwei Gruppen, $n=40$ pro Gruppe,
 $\mu_A = \mu_B = 50$, $\sigma_A = \sigma_B = 10$
- Verteilung von $MW_A - MW_B$



- Zwei Gruppen, $n=1500$ pro Gruppe,
 $\mu_A = \mu_B = 200$, $\sigma_A = \sigma_B = 20$
- Verteilung von $MW_A - MW_B$

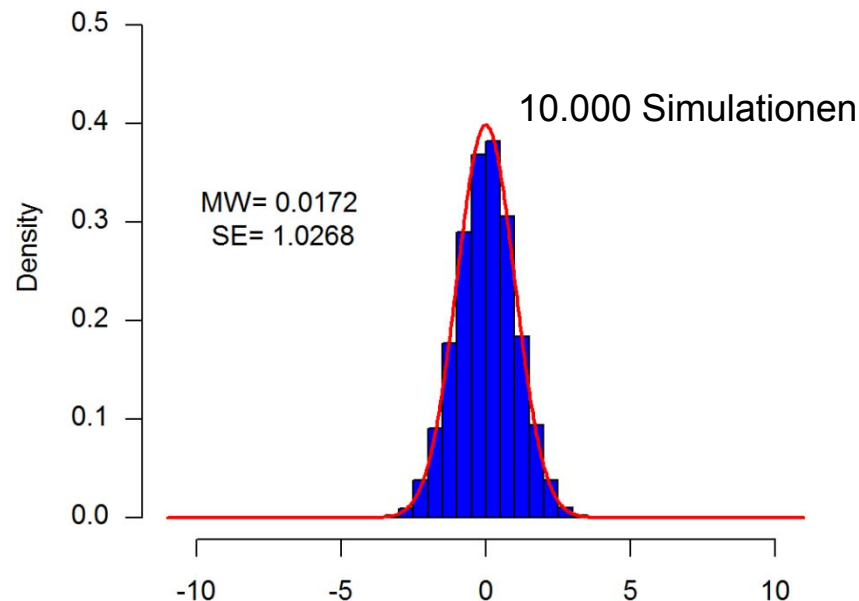


Verteilung der Differenzen hängt von der Fallzahl und der Standardabweichung in den Gruppen ab → Standardisierung

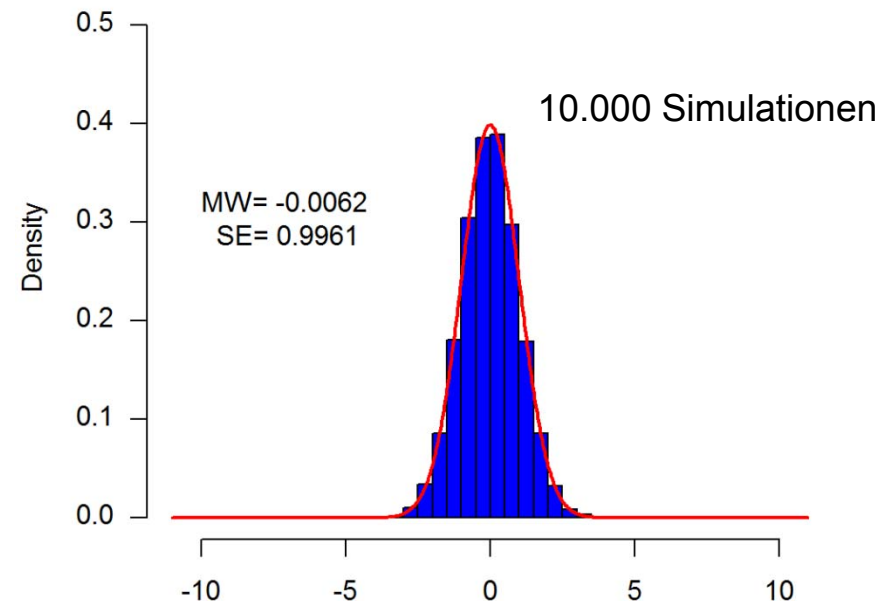
t-Test: Teststatistik

$$T = \frac{\text{Mittelwert}_A - \text{Mittelwert}_B}{\sqrt{\frac{SD_A^2}{n_A} + \frac{SD_B^2}{n_B}}}$$

- Zwei Gruppen, n=40 pro Gruppe, $\mu_A = \mu_B = 50$, $\sigma_A = \sigma_B = 10$
- Verteilung von T



- Zwei Gruppen, n=1500 pro Gruppe, $\mu_A = \mu_B = 200$, $\sigma_A = \sigma_B = 20$
- Verteilung von T



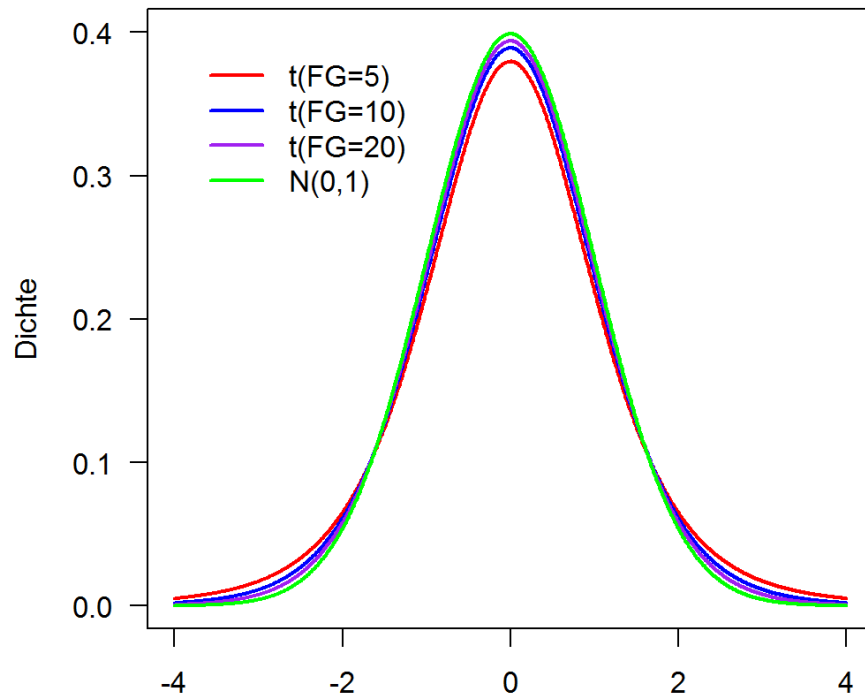
Die Teststatistik T folgt unter H_0 (unabhängig von Fallzahl und SD) approximativ (d.h. für großes n) einer Standardnormalverteilung [$T \sim N(0;1)$]

t-Test – Verteilung unter H_0

- Für die Teststatistik T ist die Verteilung unter H_0 bekannt
- T kann zur Berechnung des p-Wertes und entsprechend zur Testentscheidung herangezogen werden

Hinweise:

- T folgt eigentlich einer t-Verteilung mit k Freiheitsgraden
- Für große n nähert sich diese Verteilung der Standardnormalverteilung ($\mu=0$, $\sigma=1$) an.



Beispiel t-Test

Es sollen zwei Behandlungsgruppen hinsichtlich der Lebensqualität (gemessen durch SF-36) verglichen werden.

Folgende Werte wurden beobachtet:

	Behandlung A	Behandlung B
Fallzahl	40	40
Mittelwert	53	47
Stdabw.	10	10

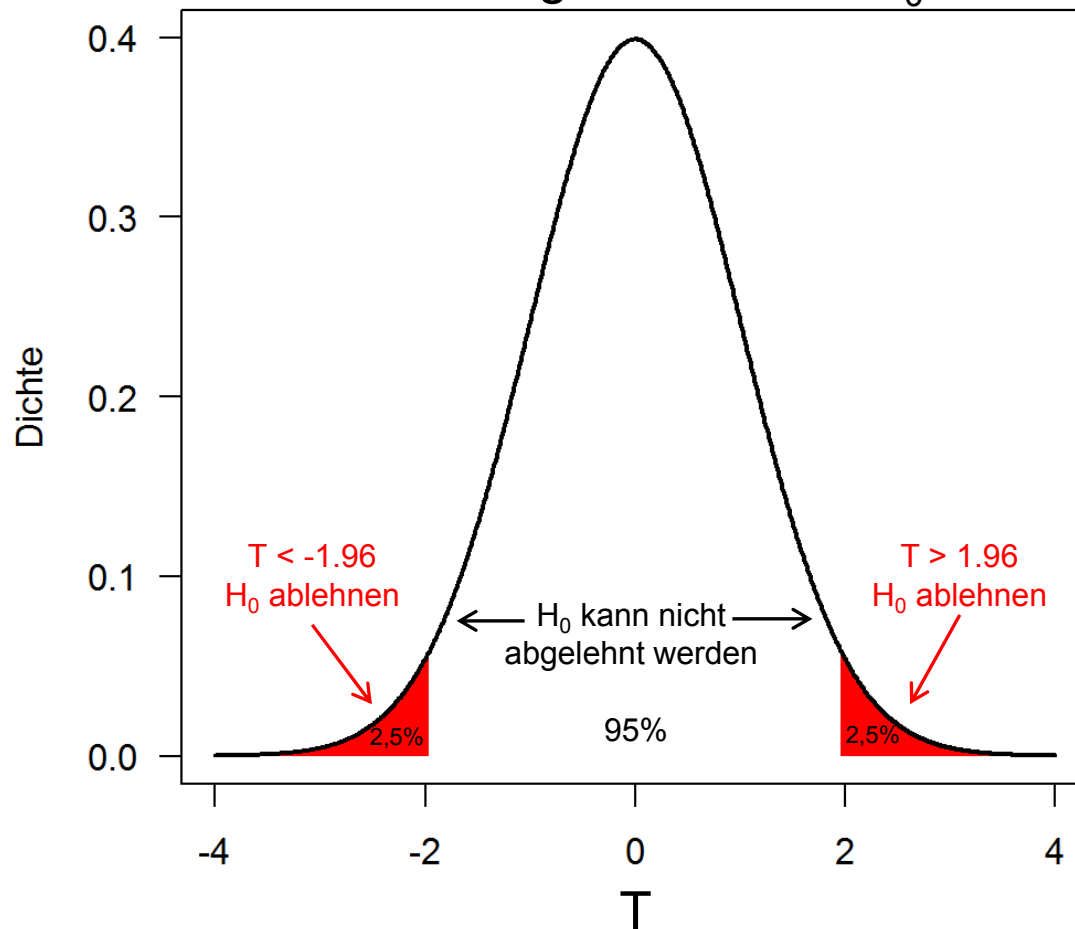
Wie lauten Null- und Alternativhypothese für diese Fragestellung?

$$H_0: \mu_A = \mu_B$$

$$H_1: \mu_A \neq \mu_B$$

t-Test: Testentscheidung

Verteilung von T unter H_0



	Beh. A	Beh. B
N	40	40
MW	53	47
SD	10	10

$$T = \frac{MW_A - MW_B}{\sqrt{\frac{SD_A^2}{n_A} + \frac{SD_B^2}{n_B}}}$$

$$T = \frac{53 - 47}{\sqrt{\frac{10^2}{40} + \frac{10^2}{40}}}$$

$$= \frac{6}{\sqrt{2,5 + 2,5}} = \frac{6}{2,24} = 2,68$$

$$T > 1,96 \Leftrightarrow p < 0,05 \quad (p = 0,009)$$

→ H_0 ablehnen