

# **Grundprinzip des statistischen Testens**

---

**Albert Rosenberger, Cornelia Frömke, Antonia Zapf**

---

# Angeblich hat jeder Mensch einen Doppelgänger

---

Fotograf François Brunelle

- fotografierte nicht miteinander verwandte Menschen
- Projekt: „Je ne suis pas un sosie“

... manche Menschen sehen sich wirklich ähnlich.

Hier bietet sich ein Foto mit unechten Zwillingen an.

**Wie gut können Sie echte von unechten Zwillingen unterscheiden?**

# Gesamtablauf eines Experiments

---

1. Fragestellung und Hypothesen formulieren
  2. Signifikanzniveau  $\alpha$  festlegen
  3. Daten sammeln
  4. Teststatistik berechnen bzw.  $p$ -Wert bestimmen
  5. Testentscheidung
  6. Interpretation der Ergebnisse
-

# Mein Experiment mit Ihnen

---

- ich zeige Ihnen später elf Fotos mit jeweils zwei Personen
- auf manchen Fotos sind nicht verwandte Doppelgänger
- auf manchen Fotos sind echte Zwillinge  
(vielleicht auch zweieiige Zwillinge)
- Sie raten bei jedem Foto, ob es sich um Zwillinge oder Doppelgänger handelt
- Welche Ergebnisse erwarten Sie?

# Sind Sie besser als eine Münze?

---

- Sie können als Entscheidung eine Münze werfen und es dem **Zufall** überlassen
  - ✓ dann haben Sie eine Wahrscheinlichkeit von 50%, das Foto richtig zu bewerten
- Sie achten aber auf Form der Augen, Nase usw.
  - ✓ Sie könnten also mehr als 50% der Bilder richtig bewerten

Vermutung: Sie sind **systematisch** besser als der Münzwurf.

# Vorgehensweise

---

1. Hypothesen formulieren
2. Signifikanzniveau  $\alpha$  festlegen
3. Daten sammeln
4. Teststatistik bzw.  $p$ -Wert berechnen
5. Testentscheidung
6. Interpretation der Ergebnisse

# Statistische Hypothesen

---

## Fall 1

- Der wahre Anteil  $\pi$  der richtigen Zuordnungen **unterscheidet sich nicht** vom Münzwurf
- Man ist nicht in der Lage, Doppelgänger von Zwillingen zu unterscheiden.
- $H_0: \pi = 50\%$

Das nennt man Nullhypothese.

## Fall 2

- Der wahre Anteil (Erwartungswert)  $\pi$  der richtigen Zuordnungen **ist größer als** beim Münzwurf
- Man ist in der Lage, Doppelgänger von Zwillingen zu unterscheiden.
- $H_A: \pi > 50\%$

Das nennt man Alternativhypothese.

---

# Vorgehensweise

---

1. Hypothesen formulieren
2. Signifikanzniveau  $\alpha$  festlegen
3. Daten sammeln
4. Teststatistik bzw.  $p$ -Wert berechnen
5. Testentscheidung
6. Interpretation der Ergebnisse



# Entscheidung mit statistischer Unsicherheit

---

Zwei mögliche Fehler:

		statistischer Test	
		nicht besser als mit Münzwurf	besser als mit Münzwurf
unbekannte Wahrheit	nicht besser als mit Münzwurf	richtig	Fehler 1. Art, $\alpha$
	besser als mit Münzwurf	Fehler 2. Art, $\beta$	richtig

- der Fehler 1. Art: maximale Wahrscheinlichkeit, die Nullhypothese fälschlicherweise abzulehnen
- der Fehler 1. Art darf in den meisten Fragestellungen maximal 2,5% einseitig sein → das Signifikanzniveau wird somit auf 2,5% gesetzt

# Vorgehensweise

---

1. Hypothesen formulieren
2. Signifikanzniveau  $\alpha$  festlegen
3. **Daten sammeln**
4. Teststatistik bzw.  $p$ -Wert berechnen
5. Testentscheidung
6. Interpretation der Ergebnisse

# Zwillinge oder Doppelgänger?

---

Bitte entscheiden Sie bei jeder  
der folgenden Aufnahmen,  
ob die Personen Zwillinge sind  
oder eben nicht.

Für Ihre Einschätzung können Sie das  
Handout verwenden.  
Kreuzen Sie „Zwillinge“ oder „Doppelgänger“ an.

# Fotos

---

- Fügen Sie jetzt die Fotos ein (ein Foto je Folie)

## Auflösung: Zählen Sie, wie oft Sie richtig bewertet haben.

---

Bild	Zwillinge	Doppelgänger
1		X
2	X	
3		X
4		X
5	X	
6		X
7	X	
8		X
9	X	
10		X
11	X	

# Wie bewertet man das Ergebnis?

---

- Wer hat 11 richtige Bewertungen?
- Mit dem Münzwurf (zufälliges Ergebnis) hätten Sie ungefähr 5,5 (=50%) richtige Bewertungen.
- Aber bei welchem Ergebnis sind Sie systematisch besser als der Münzwurf? Also ein **statistisch signifikantes** Ergebnis?
- Diese Frage beantwortet man mit einem **statistischen Test**.

# Vorgehensweise

---

1. Hypothesen formulieren
2. Signifikanzniveau  $\alpha$  festlegen
3. Daten sammeln
4. Teststatistik bzw.  $p$ -Wert berechnen
5. Testentscheidung
6. Interpretation der Ergebnisse

# Die Teststatistik

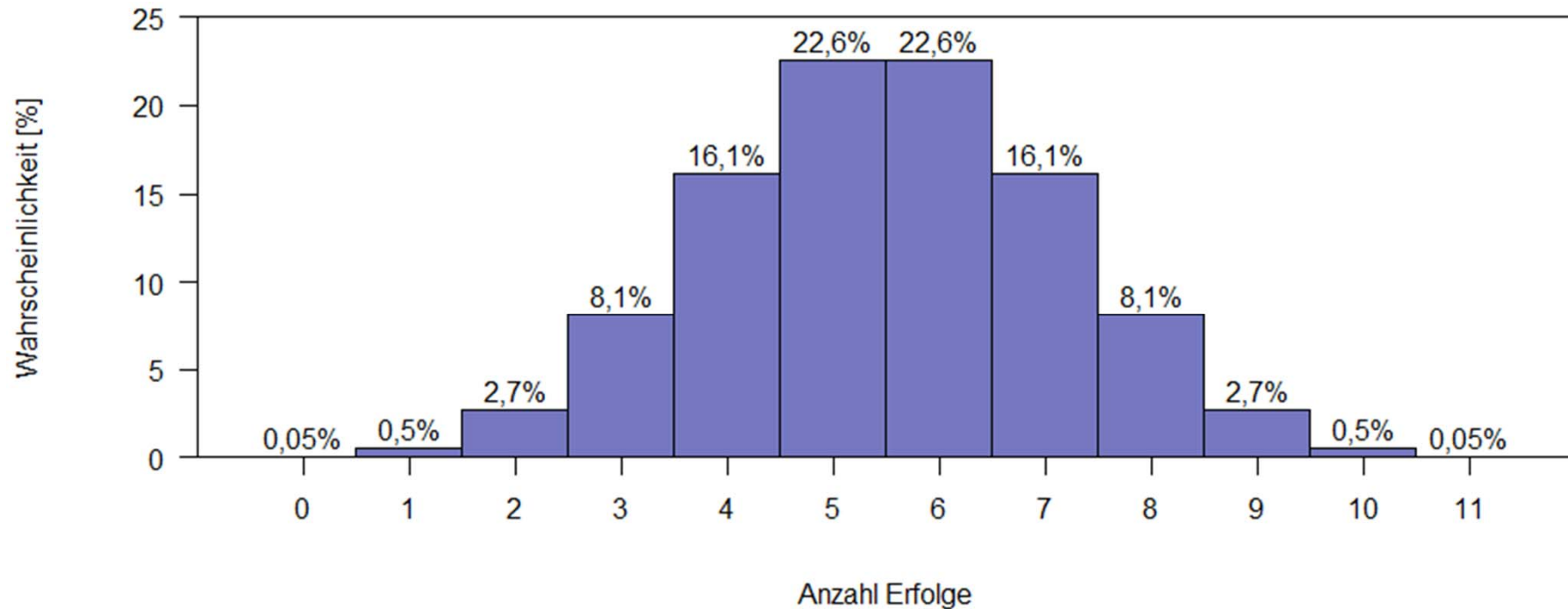
---

- Eine Maßzahl zur Zusammenfassung der Ergebnisse des Experiments
- Die Wahl des statistischen Tests und somit der Teststatistik  $T$  ist abhängig von den Daten und der Fragestellung
- Hier:  **$T$  = Anzahl der richtigen Bewertungen**
- Als Beispiel wählen wir  $T = 8$ .
- Wie wird die Teststatistik bewertet?
- Was hätte man bei Zufallsergebnissen (Münzwürfe) erwartet?



# Die Binomialverteilung für 11 Versuche und Erfolgswahrscheinlichkeit 50%

---



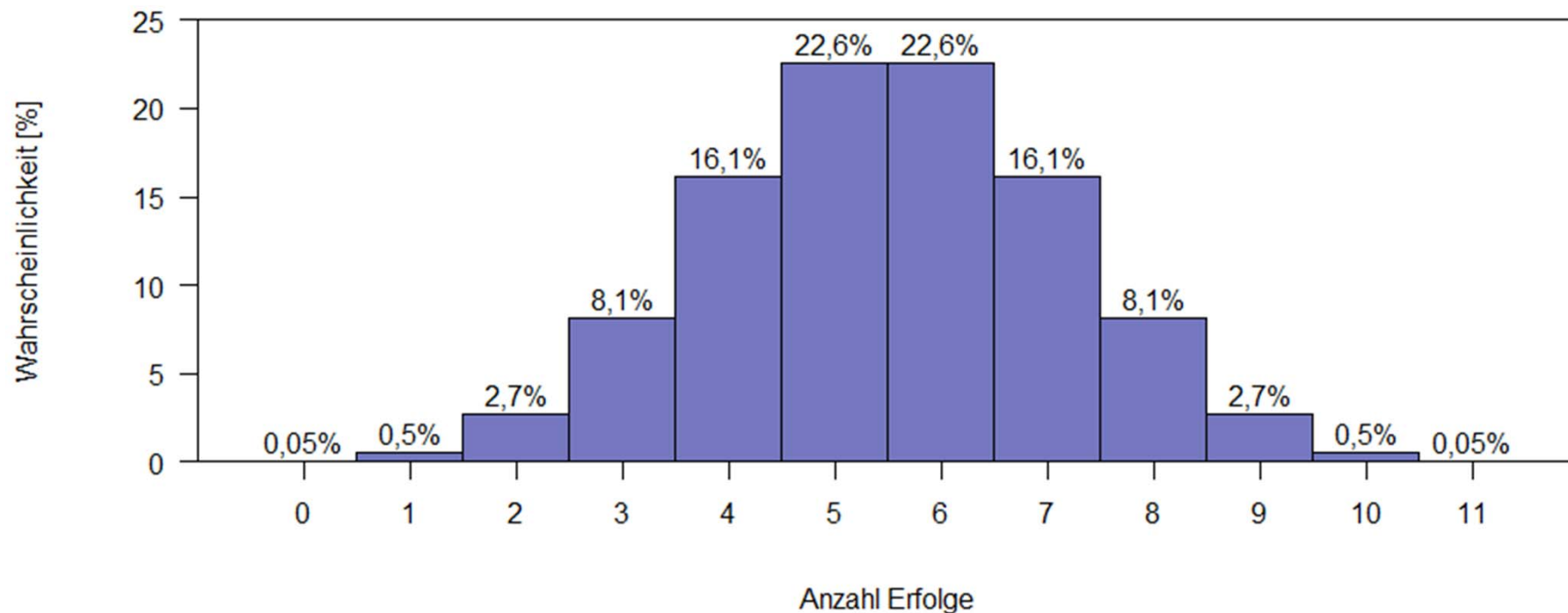
$$p_k = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

$$p_2 = \binom{11}{2} \cdot 0,5^2 \cdot (1 - 0,5)^{11-2} = \frac{11!}{2! \cdot 9!} \cdot 0,5^2 \cdot 0,5^9 = 0,027 = 2,7\%$$

# Zentrale Fragestellung

---

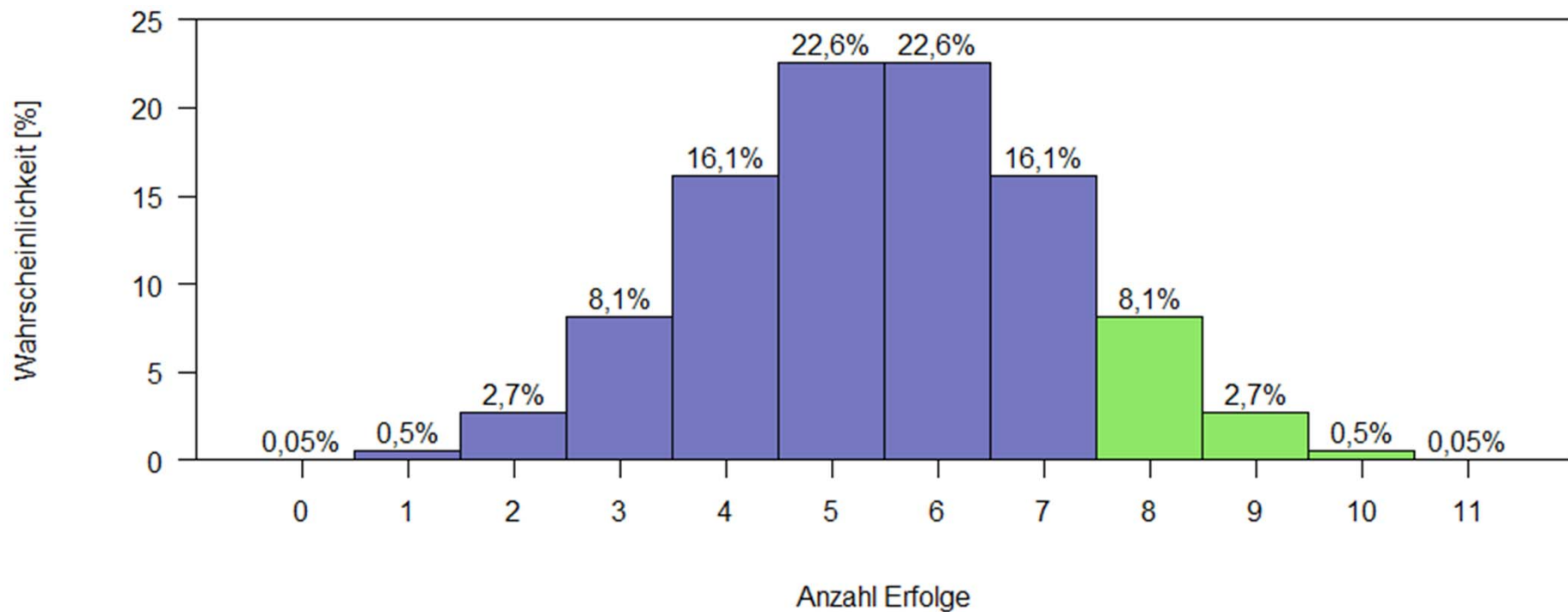
Wenn die Nullhypothese stimmt, dann ist der Anteil richtiger Zuordnungen 50% – somit 5,5 Erfolge. Wie groß ist vor diesem Hintergrund die Wahrscheinlichkeit, die beobachtete Teststatistik  $T = 8$  oder eine noch größere zu beobachten?



# Der $p$ -Wert

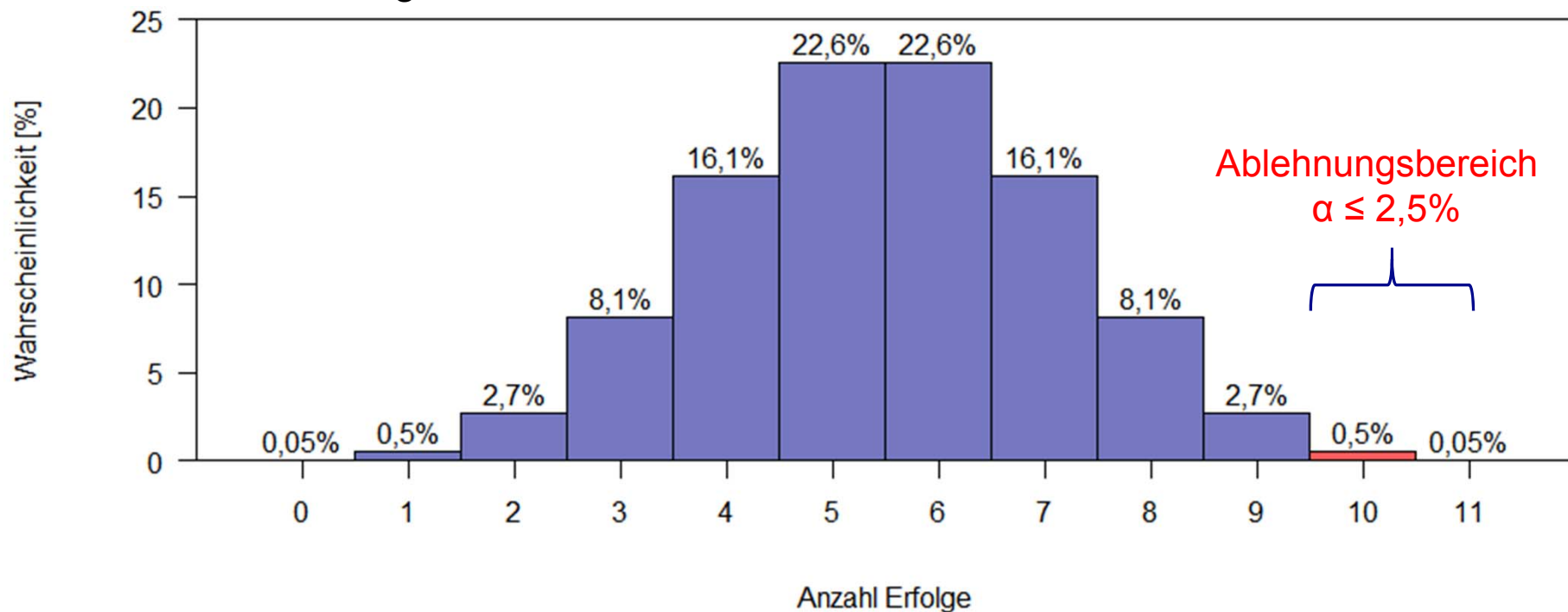
---

- Der  $p$ -Wert ist die Wahrscheinlichkeit für die beobachtete oder eine noch extremere Teststatistik unter der Annahme der Nullhypothese.
- Für  $T = 8$  ist der  $p$ -Wert 11,39%.



# Bestimmung des Ablehnungsbereiches

- Das Signifikanzniveau von 2,5% wird auf dem oberen (rechten) Ende der Verteilung abgetragen.
- Der Ablehnungsbereich darf maximal die Wahrscheinlichkeit  $\alpha$  haben.



# Vorgehensweise

---

1. Hypothesen formulieren
2. Signifikanzniveau  $\alpha$  festlegen
3. Daten sammeln
4. Teststatistik bzw.  $p$ -Wert berechnen
5. Testentscheidung
6. Interpretation der Ergebnisse

# Testentscheidung

---

- **Über die Teststatistik:** Wenn  $T$  im Ablehnungsbereich liegt, kann die Nullhypothese abgelehnt werden.
- **Über den  $p$ -Wert:** Wenn der  $p$ -Wert kleiner als  $\alpha$  ist, kann die Nullhypothese abgelehnt werden.
- Mit 8 richtigen Zuordnungen liegt die Teststatistik nicht im Ablehnungsbereich. Außerdem ist der zugehörige  $p$ -Wert = 11,3% größer als  $\alpha = 2,5\%$ .

# Wie ist Ihr Ergebnis?

---

- Berechnen Sie den  $p$ -Wert für das Ergebnis Ihres Experiments.
- Füllen Sie, basierend auf diesen Ergebnissen, eine Testentscheidung.

$T$	0	1	2	3	4	5	6	7	8	9	10	11
$p$ -Wert in %	100	99,9	96,4	96,7	88,7	72,6	50,0	27,4	11,3	3,3	0,6	0,1
$H_0$ ablehnen	nein										ja	

# Vorgehensweise

---

1. Hypothesen formulieren
2. Signifikanzniveau  $\alpha$  festlegen
3. Daten sammeln
4. Teststatistik bzw.  $p$ -Wert berechnen
5. Testentscheidung
6. Interpretation der Ergebnisse



# Interpretation der Ergebnisse

---

## **für Teststatistiken von $T = 0$ bis $T = 9$**

- Die Nullhypothese kann nicht abgelehnt werden.
- Der Anteil der richtigen Zuordnungen ist zu einem einseitigen Fehler 1. Art von 2,5% nicht statistisch signifikant größer als 50%.

## **für Teststatistiken $T = 10$ und $T = 11$**

- Die Nullhypothese wird abgelehnt.
  - Zu einem einseitigen Fehler 1. Art von 2,5% ist der Anteil der richtigen Zuordnungen statistisch signifikant größer als 50%.
-

# Was Sie aus dieser Veranstaltung mitnehmen sollten

---

- Mit einem Test kann eine Annahme überprüft werden.
- Die Testauswahl ist abhängig von der Fragestellung und den Daten.
- Die „Vorgehensweise“ ist unabhängig von dem Testverfahren und muss eingehalten werden.
- Prinzip der Testentscheidung:  $p$ -Wert,  $\alpha$ , Teststatistik, Ablehnungsbereich
- Grundidee des Testens: Wenn die Nullhypothese gilt, wie wahrscheinlich ist das beobachtete Ergebnis?