ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Binomialtest

Statistik 1 (Biol./Pharm./HST) - FS 2015

Wiederholung: Panini-Bilder

Beobachtung von Vorjahren

- Ganze Box: Wenige doppelte Bilder
- Einzelne Packungen an verschiedenen Kiosks:
  Viele doppelte Bilder

- “Null”hypothese: Bilder werden zufällig verpackt
  (“Null”, weil kein System hinter dem Verpacken steckt)
- Alternativhypothese: Die Bilder werden systematisch verpackt, sodass man weniger doppelte hat

- Wie könnte man zwischen diesen beiden Hypothesen unterscheiden?

Hypothesentest

- Ich habe eine Box mit 500 Bildern gekauft. In eine leeres Album (661 mögliche Bilder) konnte ich 477 Bilder einkleben.

- Angenommen, die Nullhypothese stimmt:
  Ist es plausibel, dass ich dann 477 Bilder einkleben kann?

- Passen die Nullhypothese “zufällig verpackt” und die Beobachtung “477 Bilder eingeklebt” zusammen?

Problem: Was ist "normal"?

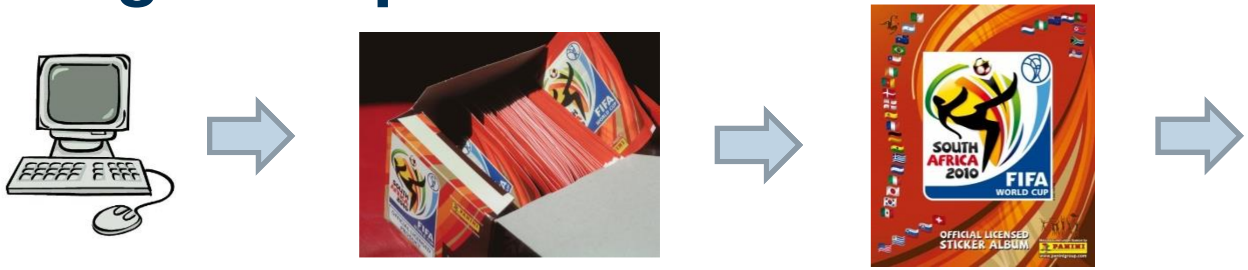
- Wenn wir viel mehr Bilder als "normal" einkleben konnten, wurden die Bilder wohl nicht zufällig verpackt.

- Angenommen, die Nullhypothese stimmt (Bilder zufällig verpackt):

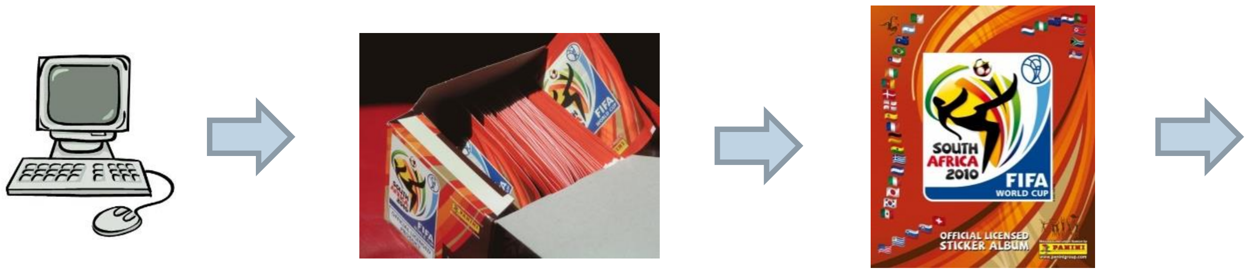
  Wie viele Bilder kann man normalerweise einkleben?

- Signifikanzniveau α: Wie "abnormal" muss die Beobachtung sein, damit wir der Nullhypothese nicht mehr glauben?
  Z.B.: $\alpha = 1/1.000.000$; wir lehnen die Nullhypothese ab, wenn wir etwas beobachten, das weniger wahrscheinlich als 1/1.000.000 ist.

Lösung: Computersimulation

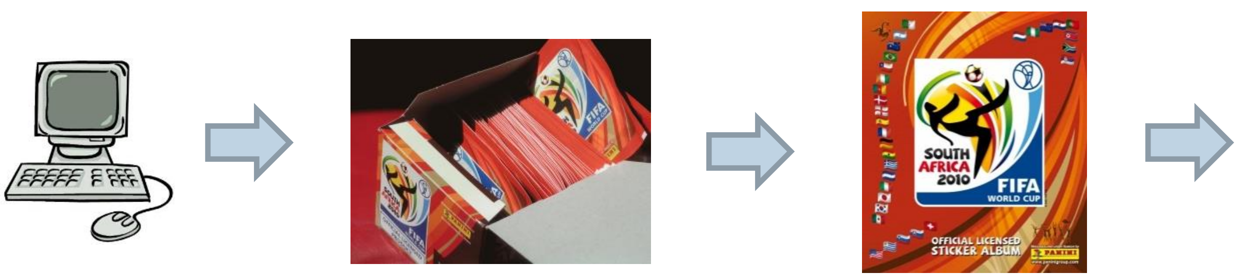
1:


350 Bilder eingeklebt

2:


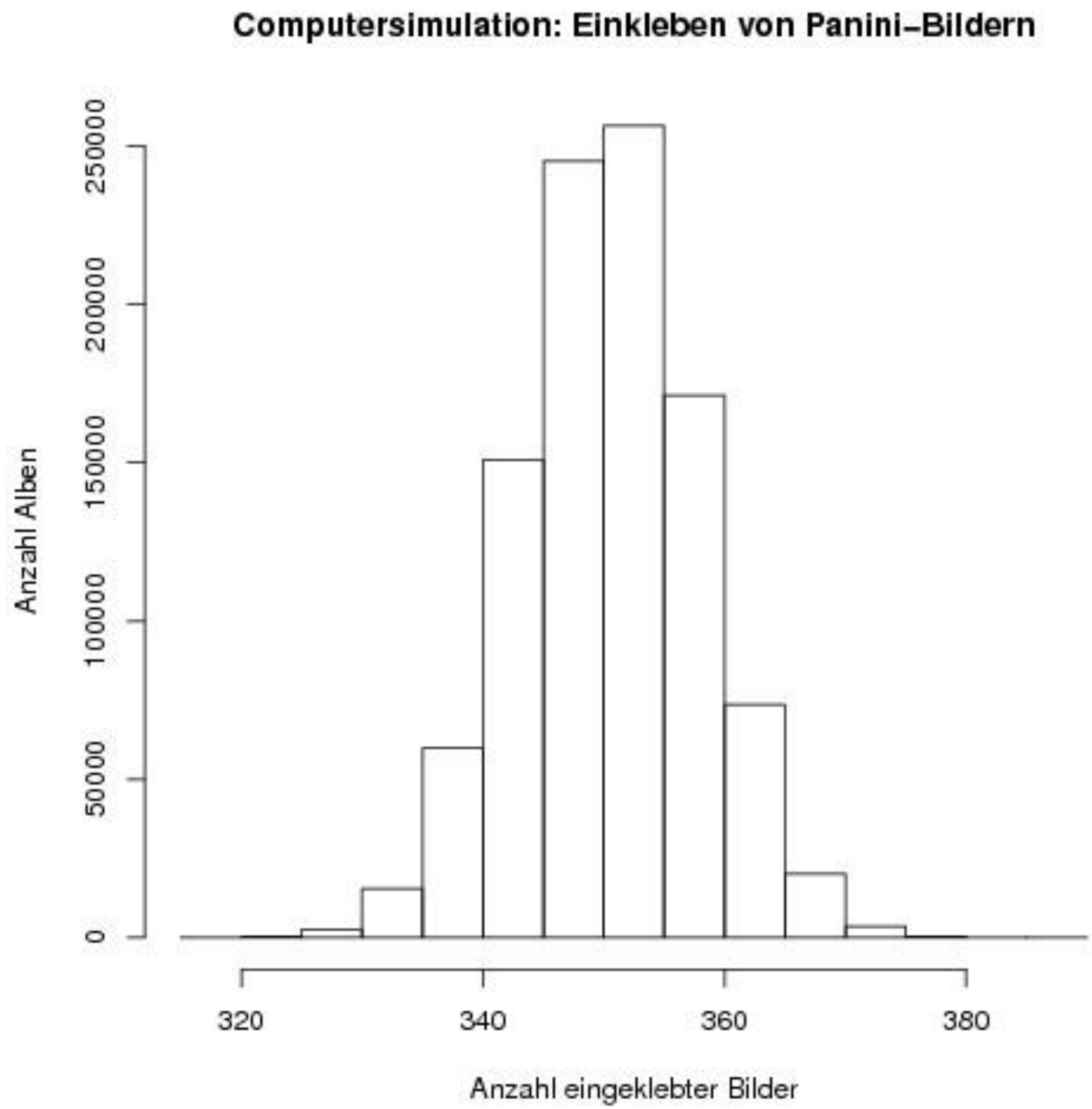
361 Bilder eingeklebt

⋮

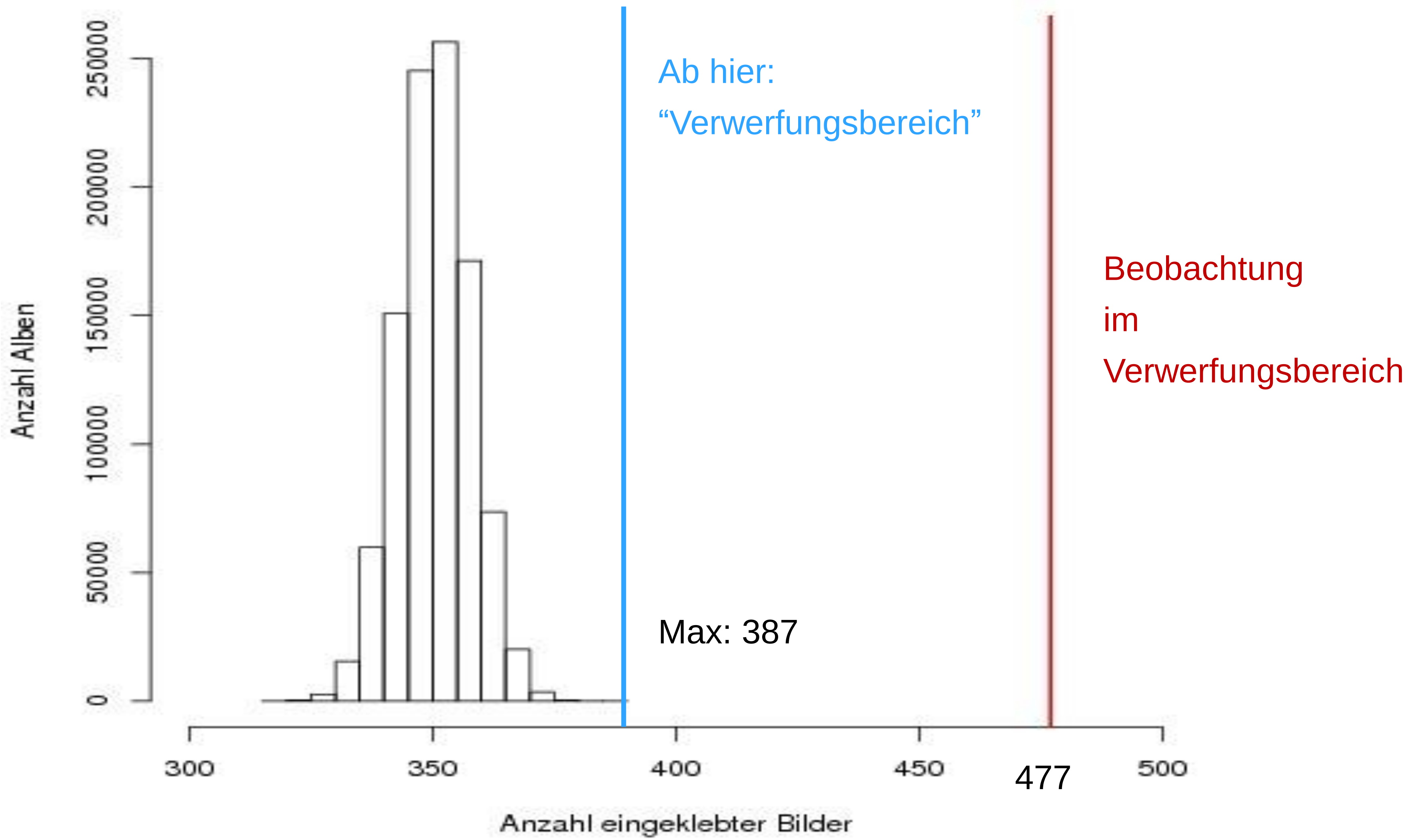
1 Mio:


358 Bilder eingeklebt

Ergebnis der Computersimulation



Passt unsere Beobachtung zur Computersimulation?



Schlussfolgerung

- Angenommen, die Bilder werden zufällig verpackt. Die Wa. 477 oder mehr Bilder einkleben zu können ist kleiner als ein Millionstel !

- Beobachtung und Simulation passen nicht zusammen:

  Die Bilder werden wohl NICHT zufällig eingepackt.

Zusammenfassung: Hypothesentest

1. Modell: Ziehen 500 Bilder mit Zurücklegen aus 661 Bildern
2. Nullhypothese: “Panini-Bilder in Kiste zufällig eingepackt”
   Alternative: “Systematisch eingepackt, sodass weniger Doppelte”
3. Teststatistik: Anz. Bilder, die man in eine leeres Album einkleben kann, wenn man eine Kiste mit 500 Bildern hat
   Verteilung der Teststatistik, wenn Nullhypothese stimmt: Computersimulation
4. Signifikanzniveau $\alpha = 1/1.000.000$
5. Verwerfungsbereich der Teststatistik:
   Computer beobachtet bei 1 Mio Simulationen nie mehr als 387 eingeklebte Bilder
   Verwerfungsbereich: K = {388, 389, …, 500}
6. Testentscheid: Der beobachtete Wert (477) liegt im Verwerfungsbereich der Teststatistik. Daher wird die Nullhypothese auf dem Signifikanzniveau 1/1.000.000 verworfen.

Binomialtest: Zauberwürfel

Binomialtest: Bsp Zauberwürfel

1. Modell: X: Anzahl 6er bei 50 Würfen; $X \sim Bin(n = 50, \pi = 1/6)$
2. Nullhypothese: $H_0: \pi = 1/6$
   Alternative: $H_A: \pi > 1/6$ (einseitig)
3. Teststatistik T: Anz. 6er bei 50 Würfen
   Verteilung der Teststatistik, wenn Nullhypothese stimmt:
   $$T \sim Bin(50, 1/6)$$
4. Signifikanzniveau: $\alpha = 0.05$ (Konvention)
5. Verwerfungsbereich der Teststatistik:
   $P[T = t] = \binom{n}{t}\pi^t(1-\pi)^{n-t}$; berechne $P[T \geq t]$

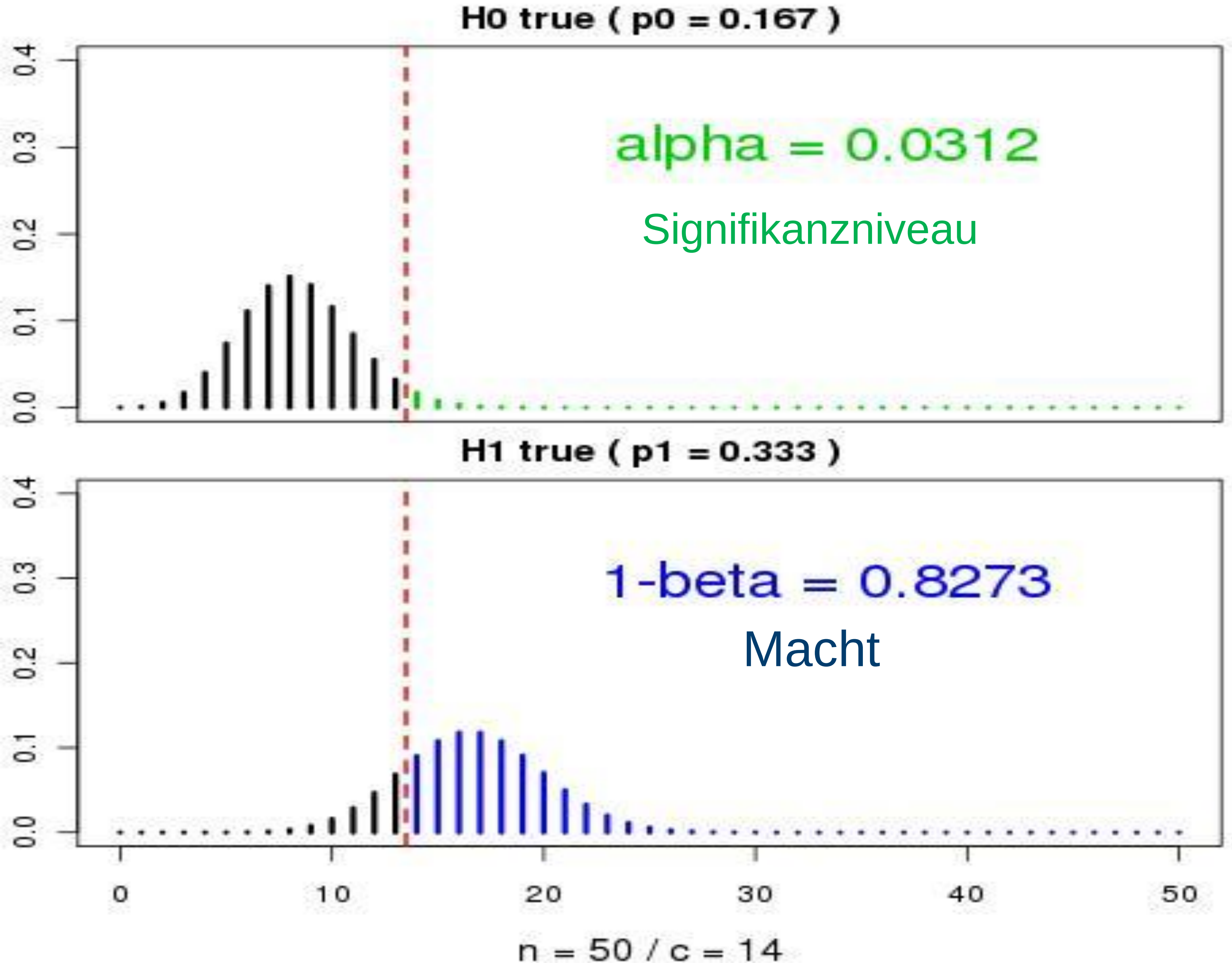
Verwerfungsbereich

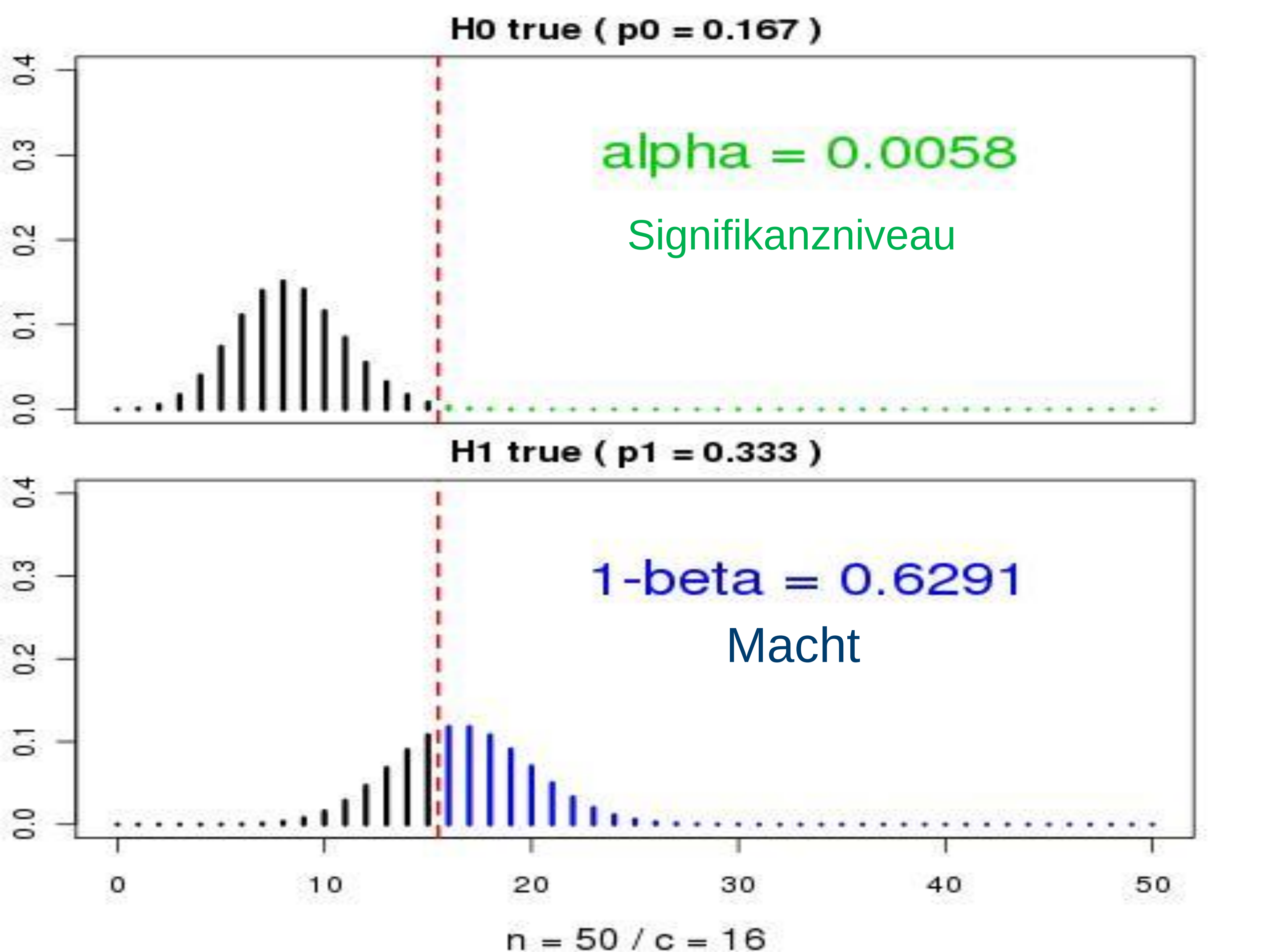
Grenze: Kleinste Zahl t,

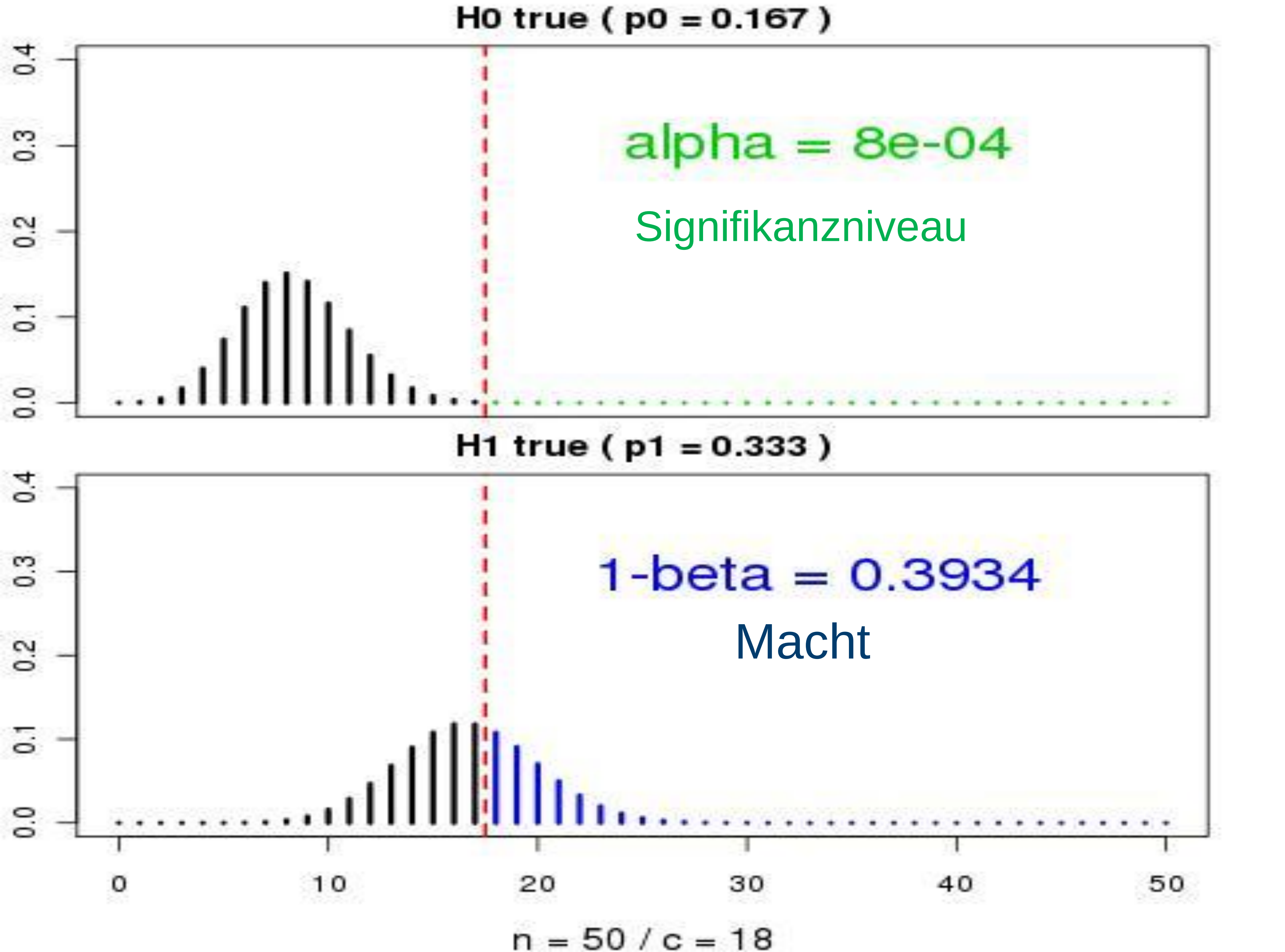
sodass $P[T \geq t] \leq \alpha$

t	…	13	14	15	…
$P[T \geq t]$	…	0.06	0.03	0.01	…

6. Testentscheid: Liegt die beobachtete Anzahl 6er bei 50 Würfen im Verwerfungsbereich der Nullhypothese?
   Falls ja: H_0 wird auf dem 5% Niveau verworfen
   Falls nein: H_0 kann auf dem 5% Niveau nicht verworfen werden


H0 true (p0 = 0.167)
alpha = 0.0312
Signifikanzniveau
H1 true (p1 = 0.333)
1-beta = 0.8273
Macht
0.0
0.1
0.2
0.3
0.4
0
10
20
30
40
50
n = 50 / c = 14


H0 true (p0 = 0.167)
alpha = 0.0058
Signifikanzniveau
H1 true (p1 = 0.333)
1-beta = 0.6291
Macht
0.0
0.1
0.2
0.3
0.4
0
10
20
30
40
50
n = 50 / c = 16


H0 true (p0 = 0.167)
alpha = 8e-04
Signifikanzniveau
H1 true (p1 = 0.333)
1-beta = 0.3934
Macht
0.0
0.1
0.2
0.3
0.4
0
10
20
30
40
50
n = 50 / c = 18

Phase 1:
Maximal Tolerierbare Dosis MTD

- Keine Statistik – fixe Regeln
- **„3 + 3 Design“**
  3 Patienten mit LD10/10
  - Alle drei OK: Drei neue Patienten mit höherer Dosis
  - Sonst: Abbruch (oder ähnliche Variante)
- “Berühmter” Fall: TGN1412
  http://de.wikipedia.org/wiki/TGN1412

Phase 2:
Effektivität

- Ist das Medikament bei Menschen erfolgsversprechend?
- Lohnt sich eine extrem teure Phase 3 Studie?
- Uninteressante / interessante Wirkwahrscheinlichkeit?

- **Einseitiger Binomialtest**


Pancreatic
Cancer

Phase 2 Studie

- „Phase II trial of S-1 and concurrent radiotherapy in patients with locally advanced pancreatic cancer“
- (Kim et.al., Cancer Chemotherapy and Pharmacology (2009); 63: 535 - 541)

Verwerfungsbereich - Zwischenrechnungen

- $P(T \geq 0) = 1$
- $P(T \geq 1) = 1 - P(T = 0) = 1 - \binom{25}{0} 0.1^0 0.9^{25} \approx 1 - 0.07 = 0.93$
- $P(T \geq 2) = 1 - P(T \leq 1) = 1 - [P(T =$