ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

DMATH
Department of Mathematics

Binomialtest

Mathematik IV: Statistik
für D-UWIS, D-ERDW, D-USYS und D-HEST – SS15

Helfen Sie uns jetzt, das Leben an der ETH zu verbessern!

→ **Auf diese E-Mail mit Link achten: *studierendenbefragung@ethz.ch***

→ **Online-Fragebogen ausfüllen.**

→ **Es geht es um *Ihre* Zufriedenheit an der ETH!**

Vielen Dank für Ihre Mithilfe,
Sarah Springmann

Die Umfrage läuft vom 16. März bis zum 6. April 2015
Ausführliche Informationen unter www.ethz.ch/studierendenbefragung

Lernziele heute

- Binomialtest
- Fehler 1. und 2. Art
- Sensitivität und Spezifizität

Hausaufgaben

- Skript: Kapitel 3.2.2 lessen
- Serie 5 lösen
- Quiz 5 bearbeiten
- bis etutoR 6 anschauen

FIFA WORLD CUP
Brasil

Bauchgefühl und Hypothesen

- Komplette Box → wenige doppelte Sticker
- Einzelne Blister an verschiedenen Kiosks → viele doppelte

«Null», weil kein System hinter dem Verpacken steckt

- «Nullhypothese»:
  - Sticker werden **zufällig** in Boxen gepackt
- Alternativhypothese:
  - Sticker werden **systematisch** in Boxen gepackt, sodass es weniger doppelte gibt

FIFA WORLD CUP
Brasil

Wie könnte man zwischen diesen beiden Hypothesen unterscheiden?

Hypothesentest

- Ich habe eine Box mit 350 Stickern gekauft und konnte in ein leeres Album (600 mögliche Bilder) 339 Sticker einkleben.

- Angenommen die Nullhypothese $\mathcal{H}_0$ stimmt:
  - Ist es plausibel, dass ich dann 339 Bilder einkleben kann?

- Passen die Nullhypothese $\mathcal{H}_0$: «*zufällig verpackt*» und die Beobachtung «*339 Bilder eingeklebt*» zusammen?

Problem: Was ist «*normal*»?

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

- Angenommen die Nullhypothese stimmt (d.h. $\mathcal{H}_0$: «*die Bilder sind zufällig verpackt worden*»):
  - Wie viele Bilder kann man normalerweise einkleben?

- Signifikanzniveau α : Wie «*abnormal*» muss eine Beobachtung sein, damit wir der Nullhypothese nicht mehr glauben?
  - z.B. α = 1/1’000’000 → wir lehnen $\mathcal{H}_0$ ab, wenn wir etwas beobachten, das weniger wahrscheinlich als 1/1’000’000 ist.

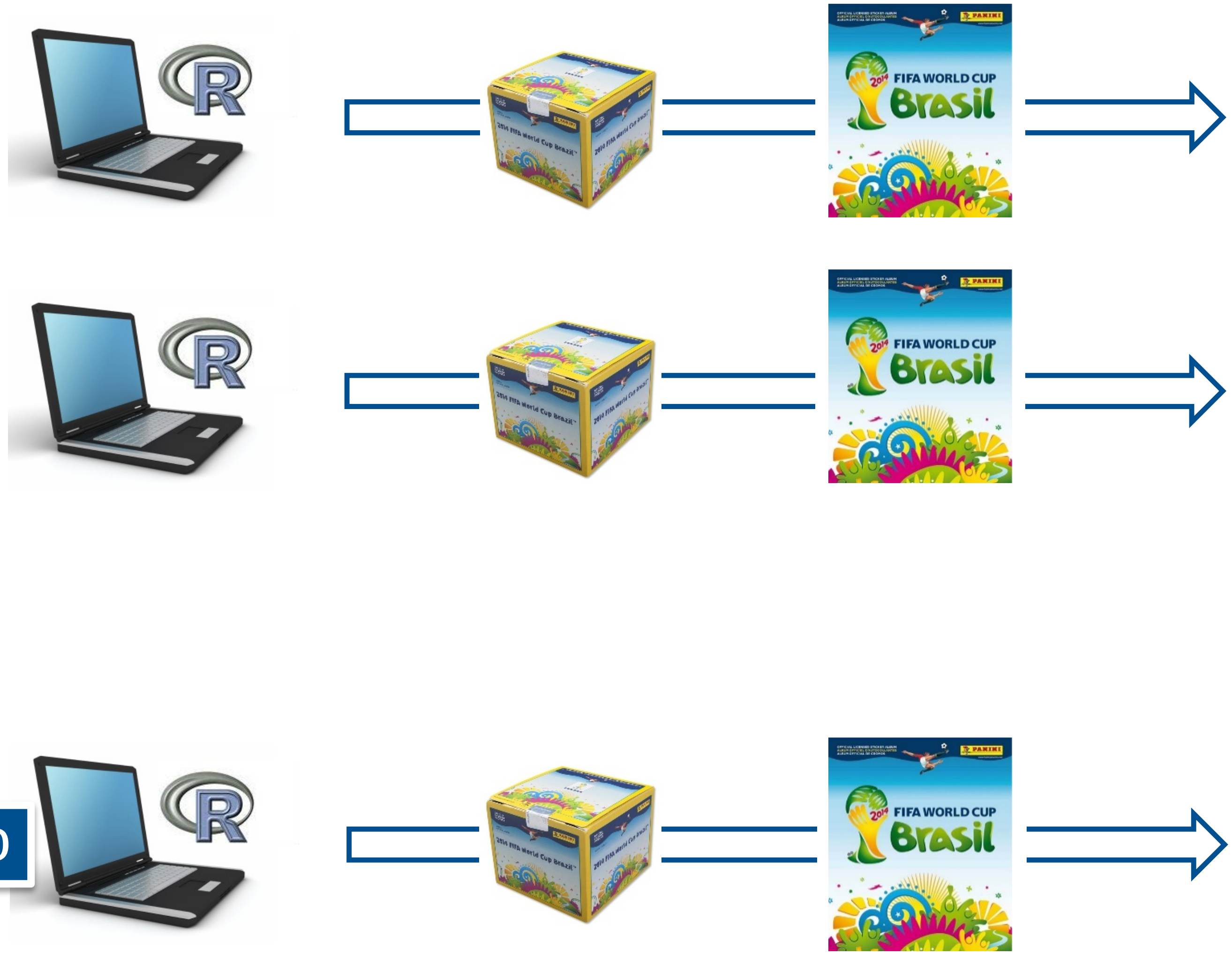
Lösung: Computersimulation

1 → 264

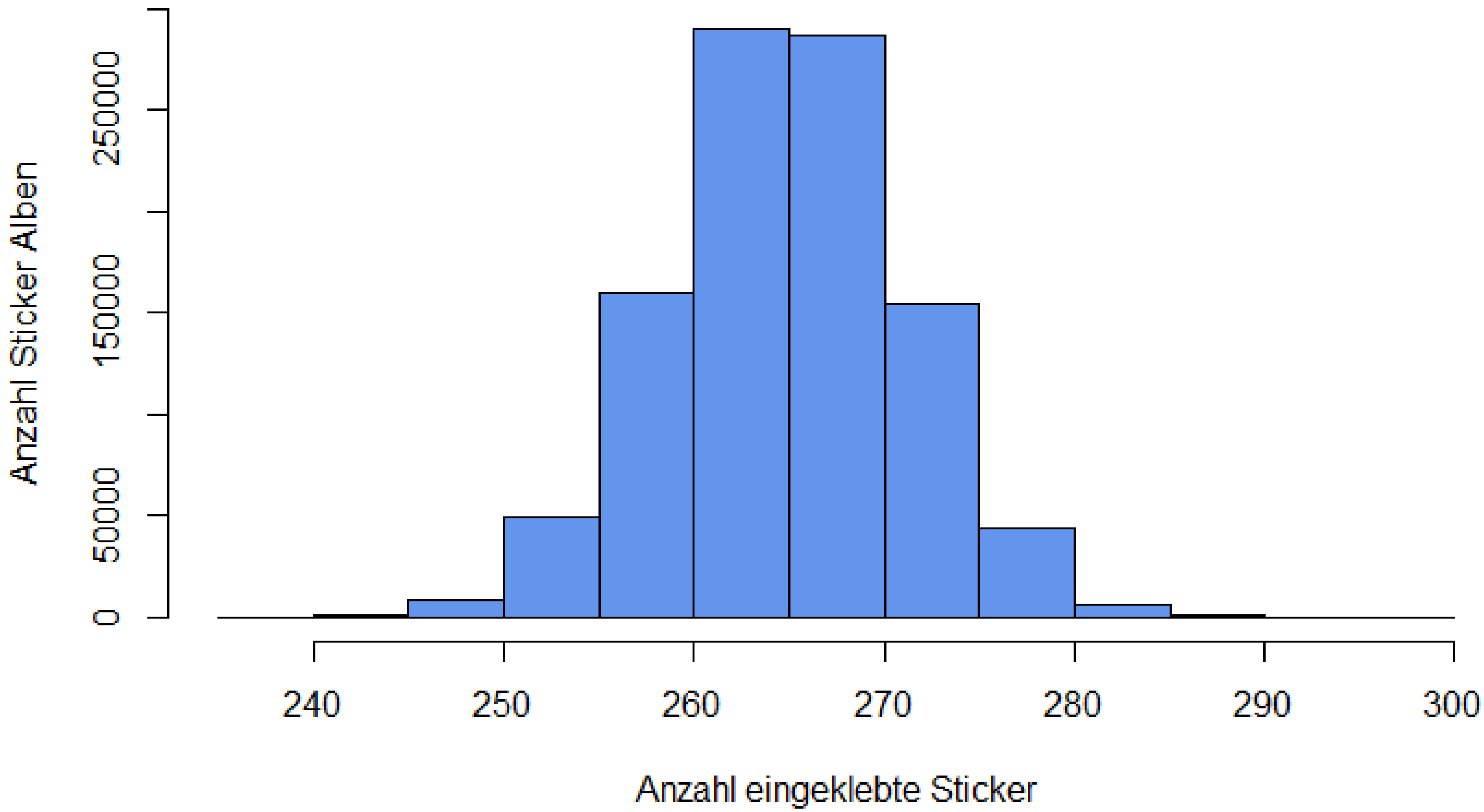
2 → 259

⋮

1 Mio → 271

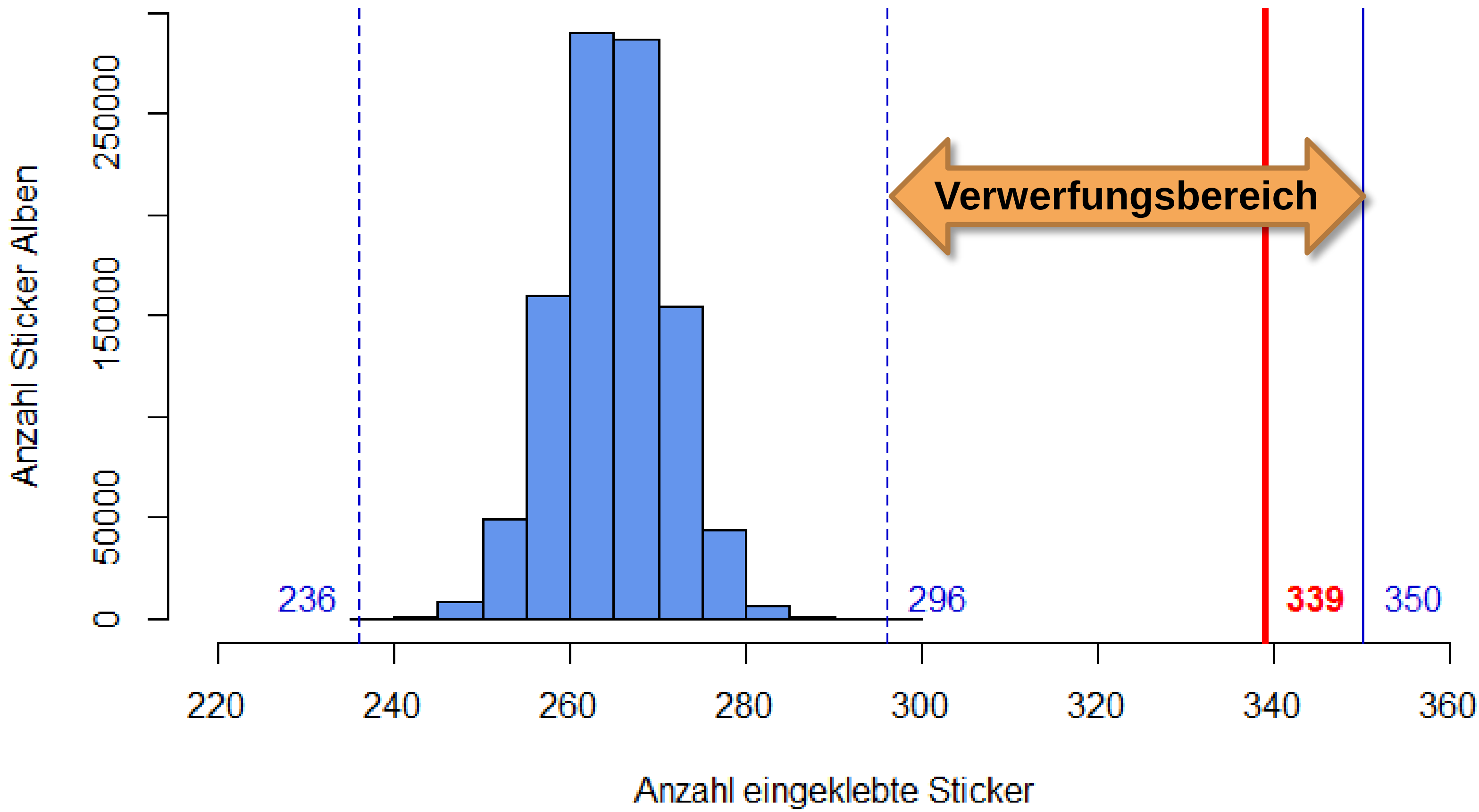


Resultat der Computersimulation



→ *Den R Code für diese Grafiken können Sie von der Homepage herunterladen*

Wie «*abnormal*» ist unsere Beobachtung nun?



Schlussfolgerung

- Angenommen, die Sticker werden zufällig verpackt:
  - Die W'keit 339 oder mehr Sticker einkleben zu können wäre kleiner als ein Millionstel!
  - ➢ Unsere Beobachtung und die Simulation passen nicht zu einander!

Sticker werden **nicht** **zufällig verpackt**

Zusammenfassung: Hypothesentest

1. Modell: Ziehe 350 Sticker mit Zurücklegen aus 600 möglichen Stickern
2. Nullhypothese $\mathcal{H}_0$: «*die Sticker sind **zufällig** verpackt worden*»
   Alternative $\mathcal{H}_A$: «***systematisch** verpackt, sodass weniger Doppelte*»
3. Teststatistik: Anzahl der Sticker, die man in ein leeres Album einkleben kann, wenn man eine Box mit 350 Stickern kauft.
   Verteilung der *Teststatistik*, wenn die *Nullhypothese* stimmt: Computersimulation
4. Signifikanzniveau α = 1/1'000'000
5. Verwerfungsbereich der *Teststatistik*:
   Computer beobachtet bei 1 Mio. Simulationen nie mehr als 296 eingeklebte Sticker → *Verwerfungsbereich*: K={297, 298, …, 350}
6. Testentscheid: Der beobachtete Wert (339) liegt im *Verwerfungsbereich* der *Teststatistik*. Daher wird die *Nullhypothese* auf dem *Signifikanzniveau* α verworfen.

Binomialtest

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

DMATH
Department of Mathematics

Binomialtest – Beispiel Würfel

- Modell: X = Anzahl 6er bei 50 Würfen;
- Nullhypothese $\mathcal{H}_0$: $\pi = \frac{1}{6}$
- Alternative $\mathcal{H}_A$: $\pi > \frac{1}{6}$ (einseitig)
- Teststatistik: T = Anzahl 6er bei 50 Würfen
- Verteilung der Teststatistik, wenn $\mathcal{H}_0$ stimmt:

$$T \sim \mathrm{Bin}(50, \frac{1}{6})$$

- Signifikanzniveau α = 0.05 («*Konvention*»)

Binomialtest – Beispiel Würfel

- Teststatistik: T = Anzahl 6er bei 50 Würfen
- Verwerfungsbereich der Teststatistik:

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

Grenze des Verwerfungsbereiches: Kleinste Zahl t, sodass $P[T \geq t] \leq \alpha$

- Testentscheid: Liegt die beobachtete Anzahl 6er bei 50 Würfen im Verwerfungsbereich der Nullhypothese $\mathcal{H}_0$?
  - Falls **ja**: $\mathcal{H}_0$ wird auf dem 5% Niveau verworfen
  - Falls **nein**: $\mathcal{H}_0$ **kann** auf dem 5% Niveau **nicht** verworfen werden

Fehlertypen

		Wahrheit / Realität	
	Total (n)	Richtig (T)	Falsch (F)
Testergebnis	$\mathcal{H}_0$ verwerfen (P, positiv)		
	$\mathcal{H}_0$ nicht verwerfen (N, negativ)		
		Sensitivität TP/P	Spezifizität TN/P

Was ist schlimmer? 1. Typ (FP) oder 2. Typ (FN)?

Feueralarm		Brennt es?	
		Ja	Nein
Alarm?	Alarm! (positiv)		
	Kein Alarm (negativ)		

Signifikanzlevel α vergrössern

- Fehler 2. Art <u>viel</u> schlimmer!
- Optimiere, indem man den Test auflockert (Test verwirft früher → mehr Fehler 1. Art)

Was ist schlimmer? 1. Typ (FP) oder 2. Typ (FN)?

Spamfilter

		Spam?	
		Ja	Nein
Filter?	Entfernen (positiv)		
	Lassen (negativ)		

Signifikanzlevel α verkleinern

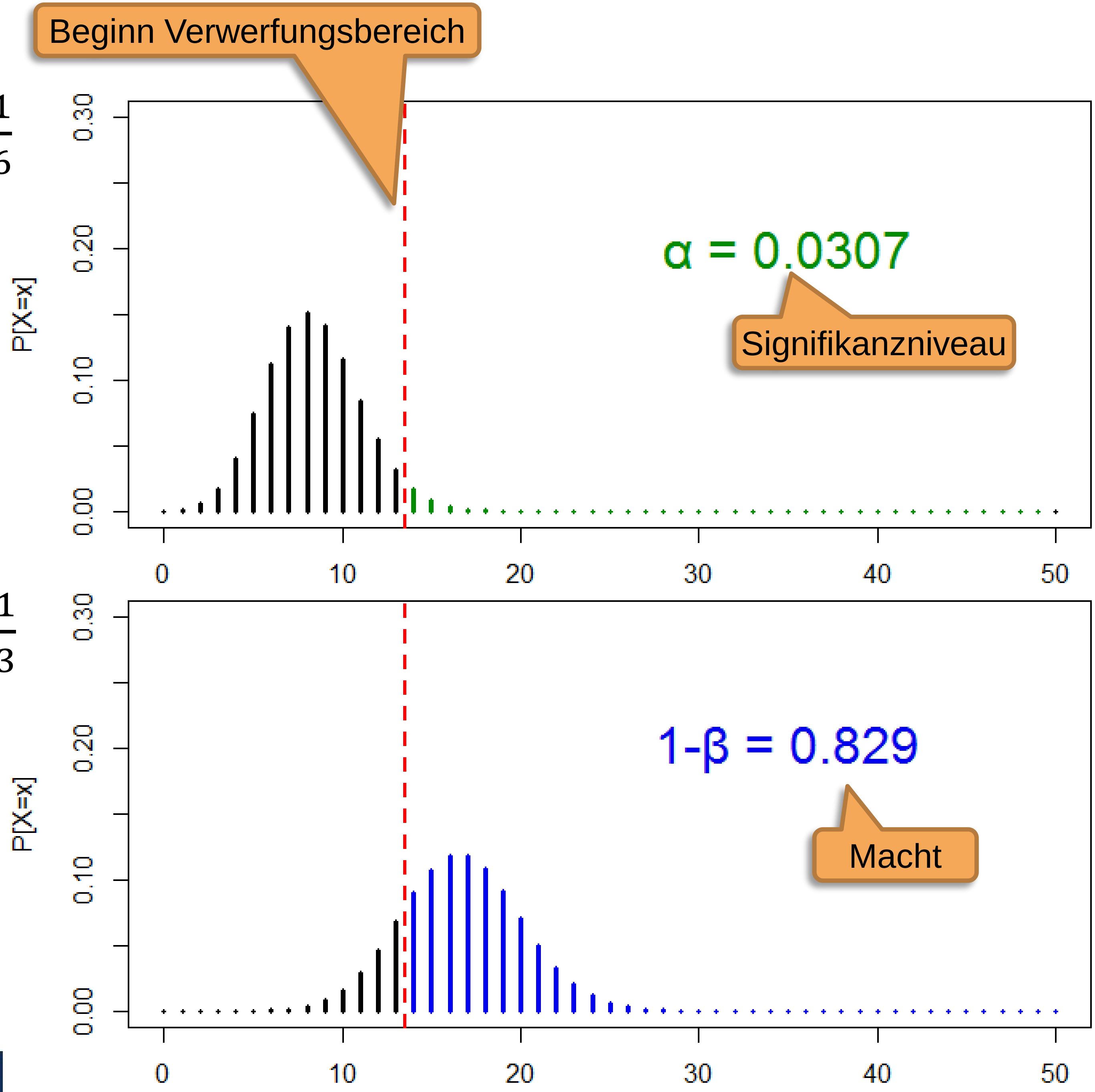
- Fehler 1. Art viel mühseliger!
- Optimiere, indem der Test strenger wird
  (Test verwirft viel später → mehr Fehler 2. Art)

$\mathcal{H}_0$richtig: $p_0 = \frac{1}{6}$

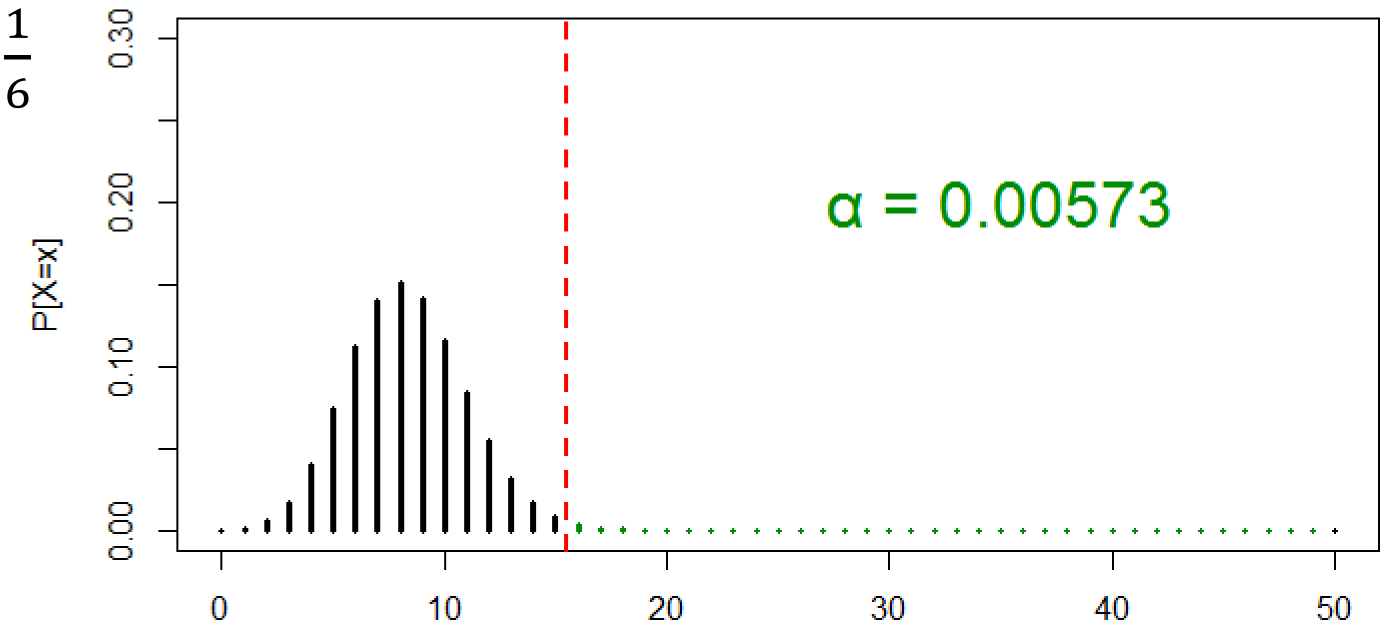
Fairer Würfel

Falscher Würfel

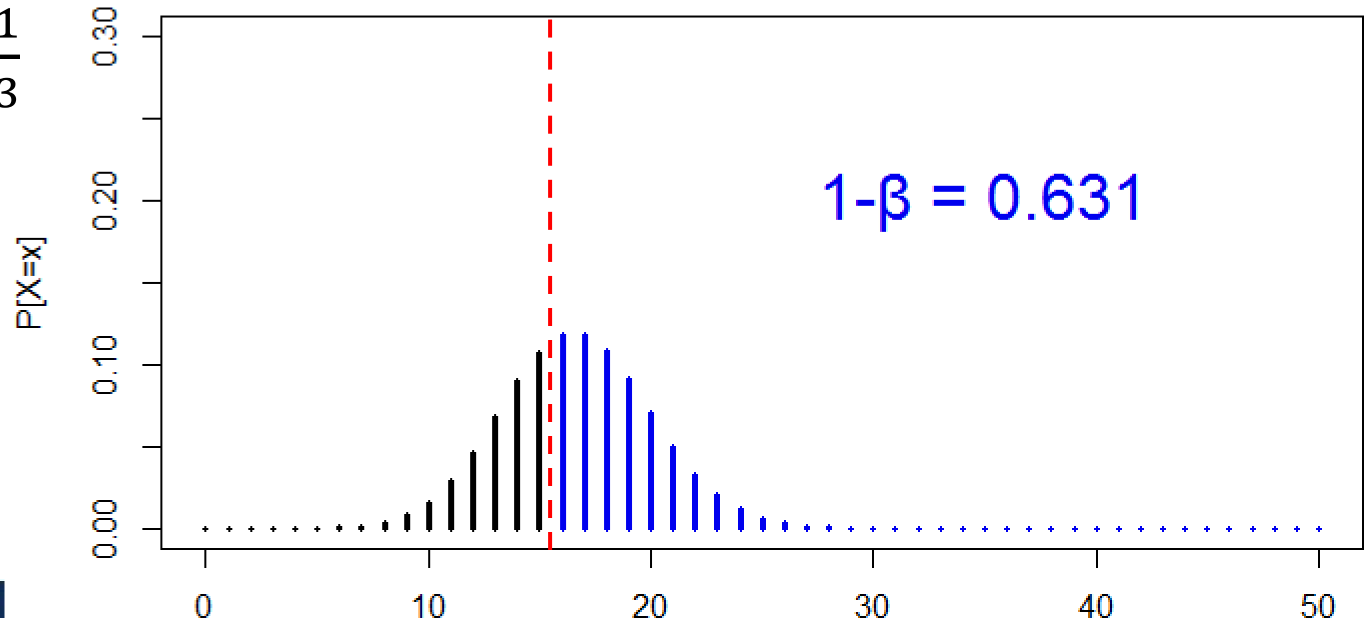
$\mathcal{H}_A$richtig: $p_A = \frac{1}{3}$



$\mathcal{H}_0$richtig: $p_0 = \frac{1}{6}$

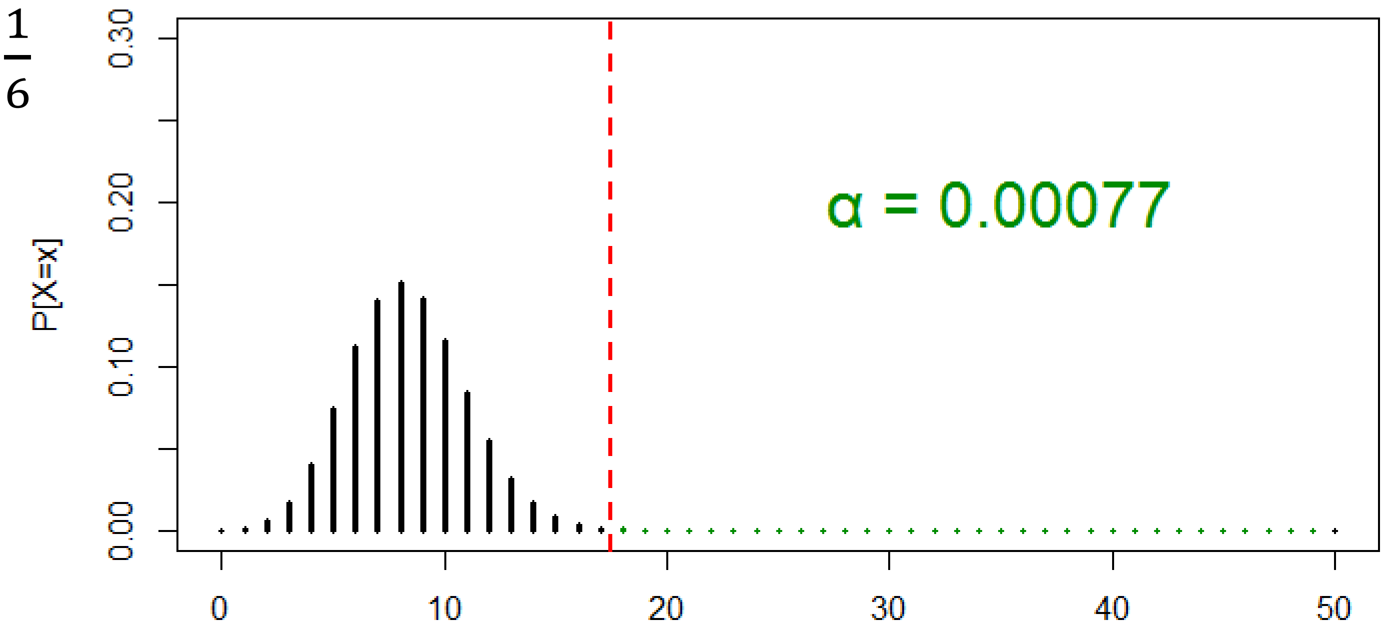


$\mathcal{H}_A$richtig: $p_A = \frac{1}{3}$

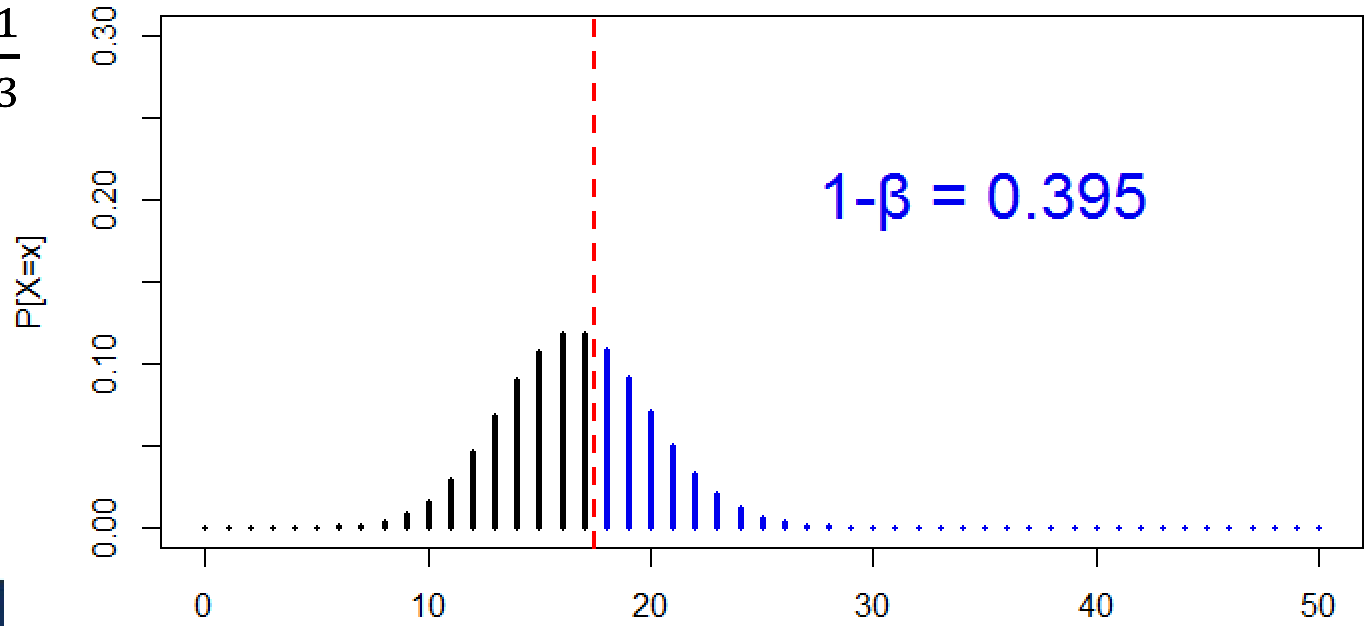


Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

$\mathcal{H}_0$richtig: $p_0 = \frac{1}{6}$



$\mathcal{H}_A$richtig: $p_A = \frac{1}{3}$



Animation zur Macht

- http://stat.ethz.ch/~kalisch/teaching/animations/binTestPower/

- Mehr Beobachtungen und Fehler 1. Art fix
  - Macht nimmt zu!

Poweranalyse

- Gegeben:
  - $\mathcal{H}_0, \mathcal{H}_A, \alpha$, z.B. $\mathcal{H}_0: \pi = \frac{1}{6}, \mathcal{H}_A: \pi > \frac{1}{6}, \alpha = 0.05$
  - Wert unter $\mathcal{H}_A$, den man mit Macht $1 - \beta$ entdecken will, z.B. $\pi = \frac{1}{3}$ soll mit einer Macht von 90% entdeckt werden.
- Gesucht:
  - Stichprobengrösse, d.h. wie gross muss meine Stichprobe sein, dass ich die obigen Voraussetzungen/Bedingungen erfüllen kann?
- Lösung:
  - Paper von A'Hern (auf Website)
  - Computer (z.B. mit R package clinfun)

Bsp. Lösung mit Paper von A'hern

Table I. Sample sizes and cut-offs for exact single-stage phase II trials.

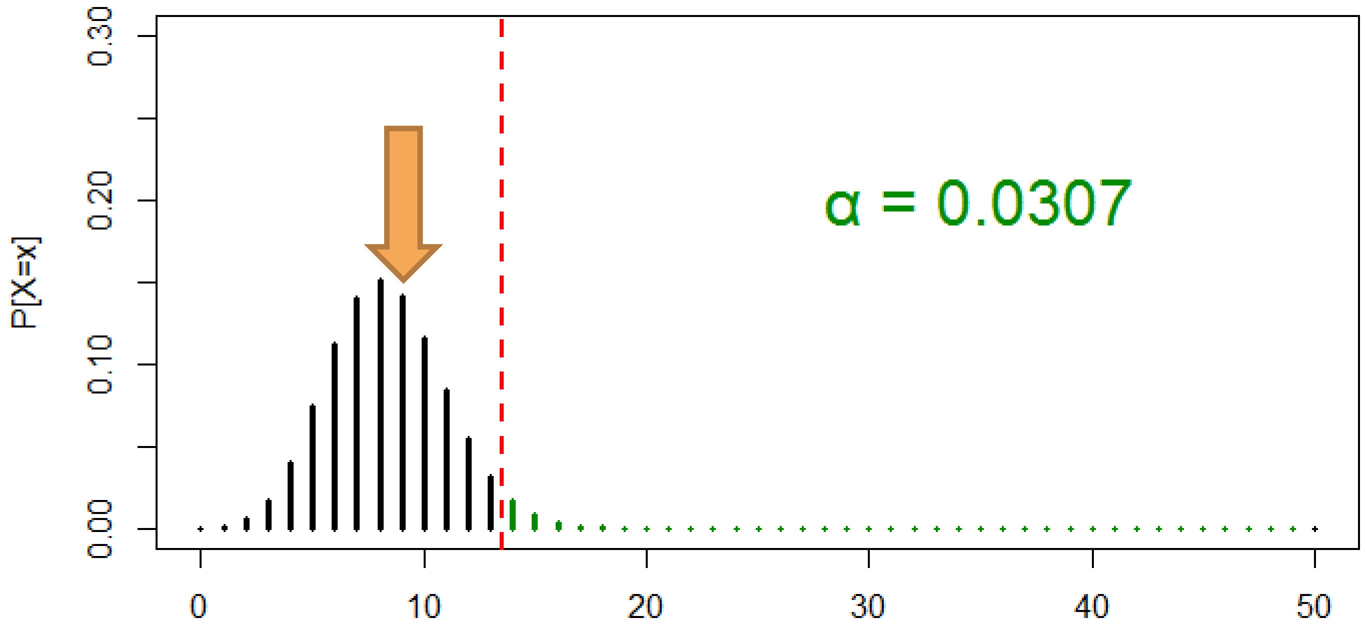
p_0	p_1	$\alpha=0.05;\ 1-\beta=0.8$	$\alpha=0.05;\ 1-\beta=0.9$	$\alpha=0.01;\ 1-\beta=0.8$	$\alpha=0.01;\ 1-\beta=0.9$
0.15	0.20	65/355	89/500	106/568	135/742
	0.25	22/101	28/136	35/157	44/206
	0.30	12/48	15/64	19/73	24/98
	0.35	8/28	10/38	14/47	16/58
	0.40	7/21	8/27	10/30	12/39
	0.45	5/14	7/21	8/21	10/29
	0.50	4/10	5/14	7/17	8/21
	0.55	4/9	5/13	6/13	7/17
	0.60	3/6	4/9	5/10	6/13
	0.65	3/6	4/9	5/9	5/10
	0.70	3/5	3/6	5/8	5/9
	0.75	3/5	3/6	4/6	5/9
	0.80	3/4	3/5	4/6	4/6

$\approx \frac{1}{6}$

$\approx \frac{1}{3}$

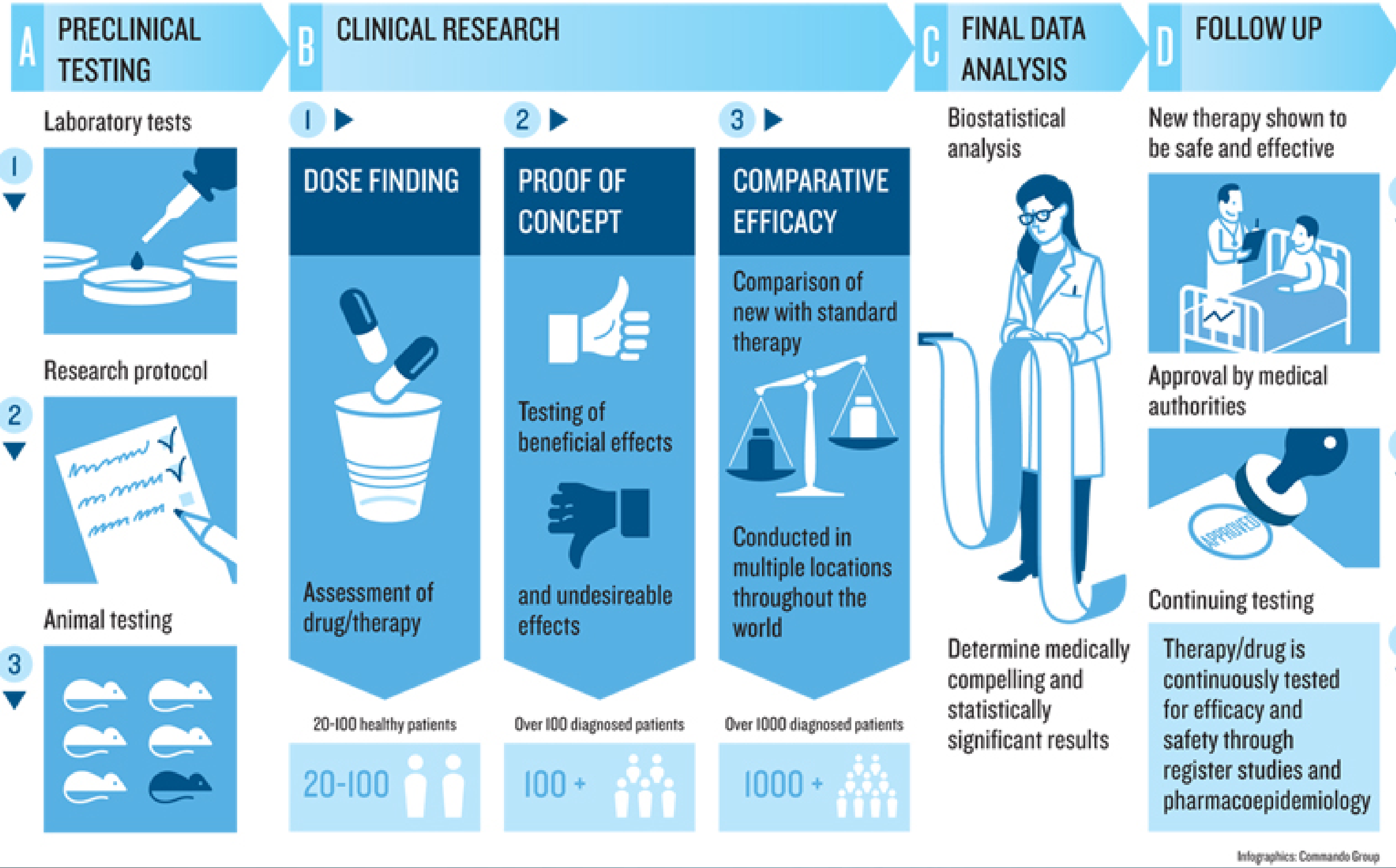
- Stichprobengrösse $n = 38$
- Verwerfungsbereich $K = \{10, 11, 12, \ldots, 38\}$

Nullhypothese nicht verwerfen...

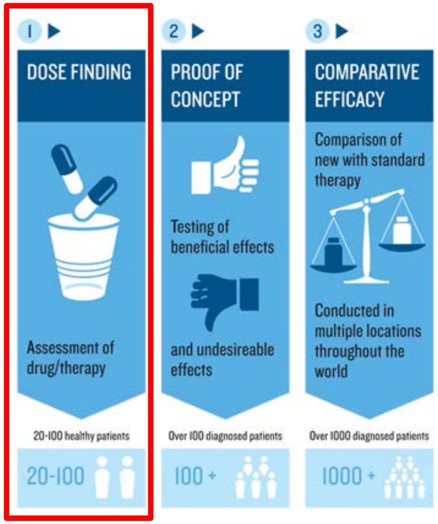


- Nullhypothese ist richtig.
- Sowohl Nullhypothese, als auch die Alternative sind möglich.
- Alternative ist falsch.

Klinische Studie revisited



Phase 1 – Maximal Tolerierbare Dosis



- Keine Statistik / fixe Regeln
- «3+3 Design» (single ascending dose)
  3 gesunde Probanden erhalten Startdosis
  - alle drei OK: 3 neue Patienten erhalten höhere Dosis
  - sonst: Abbruch (oder Konsolidierung)
- «Berüchtigte» Phase I Studie: TGN1412
  http://de.wikipedia.org/wiki/TGN1412


Bleibende Schäden nach Medikamententests

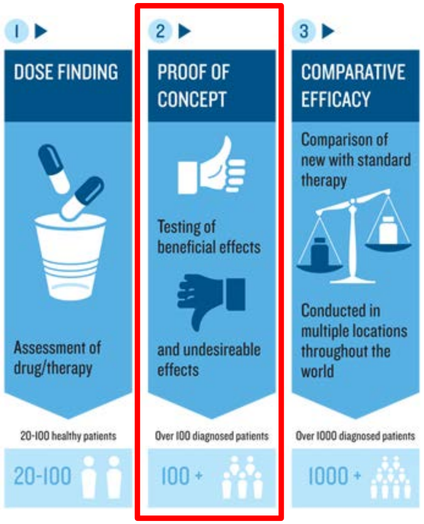
Rätselraten in London über den exakten Wirkungsmechanismus

22.4.2006, 02:05 Uhr

Mindestens ein Proband des Lond…
…rd nicht völlig gesund we…
…roffenen nun ih…
…n davo…

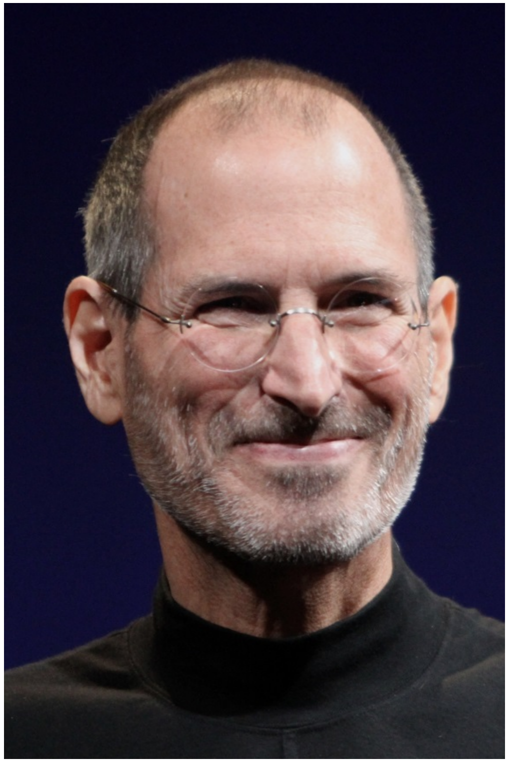
NZZ, 22.04.2006

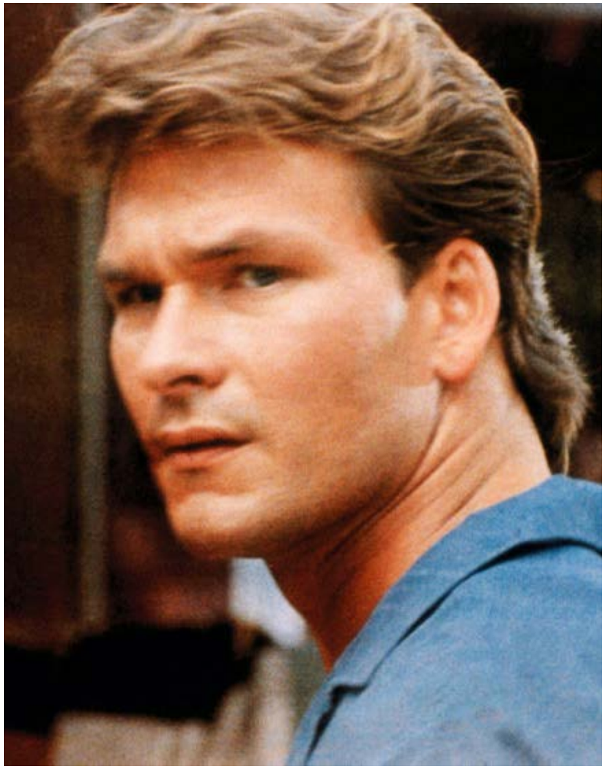
Phase 2 – Effektivität



- Ist das Medikament bei Menschen wirksam?
- Lohnt sich eine extrem teuere Phase 3 Studie?
- Üblicherweise scheitert die Entwicklung eines neuen Medikaments in dieser Phase

Einseitiger Binomialtest





Bauchspeicheldrüsenkrebs

(engl. 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

Phase 2 Studie – Bauchspeicheldrüsenkrebs

Statistical analysis

Recently, many studies of chemoradiotherapy for locally advanced pancreatic cancer have reported good response rates of up to 50% [24], and a phase I study of S-1 chemoradiotherapy by Sudo et al. [29] reported a response rate of 43.8%. Accordingly, we assumed that if the response rate was 30% or higher, the treatment would be beneficial. To test the alternative hypothesis that the minimum response rate was 30% with a null hypothesis that the response rate was 10% or lower, the required number of patients for a one-sided test was 25 with a type I error of 5%, and a power of 80%, according to single-stage phase II design [1]. If six or more patients were responsive, the treatment would be considered acceptable [1]. Tumor response and toxicity were evaluated with an intention-to-treat analysis, and patients who received at least a single dose of S-1 or a single fraction of radiotherapy were evaluated for tumor response and toxicity. The Kaplan–Meier method was used to estimate overall survival, time-to-progression and the 1-year survival rate. Statistical analysis was performed using SPSS version 11.0 for Windows.

Wunschliste

- $\alpha = 0.05$
- $1 - \beta = 0.8$ bei einer Effektgrösse von $\pi_A = 0.3$
- $\mathcal{H}_0 : \pi = 0.1$
- $\mathcal{H}_A : \pi > 0.1$

aus der Poweranalyse folgt:

- $n = 25$
- $c = 6$

Einseitiger Binomialtest

1. Modell:
   X: # Patienten mit *partial remission*, $X \sim Bin(25, \pi)$
2. $\mathcal{H}_0 : \pi = 0.1$
   $\mathcal{H}_A : \pi > 0.1$
3. Teststatistik T: gezählte # Patienten mit *partial remission*
   Falls $\mathcal{H}_0$ stimmt: $T \sim Bin(25, 0.1)$
4. Signifikanzniveau: $\alpha = 0.05$
5. Verwerfungsbereich:

Verwerfungsbereich - Zwischenrechnung

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

Einseitiger Binomialtest

1. Modell:
   X: # Patienten mit *partial remission*, $X \sim Bin(25, \pi)$
2. $\mathcal{H}_0: \pi = 0.1$
   $\mathcal{H}_A: \pi > 0.1$
3. Teststatistik T: gezählte # Patienten mit *partial remission*
   Falls $\mathcal{H}_0$ stimmt: $T \sim Bin(25, 0.1)$
4. Signifikanzniveau: $\alpha = 0.05$
5. Verwerfungsbereich:

t	0	1	2	3	4	5	6	7
$P[T \geq t]$	1	0.93	0.73	0.46	0.24	0.10	0.03	0.002

6. Testentscheid: $t = 6, t \in K \Rightarrow \mathcal{H}_0$ wird verworfen

Zusammenfassung

- Binomialtest: Wann ist ein Medikament wirksam?
- Fehler 1. und 2. Art ➔ Es brennt, kein Alarm?
- Spezifizität (Macht) nimmt zu, wenn n grösser und α fix

Hausaufgaben

- Skript: Kapitel 3.2.2 lessen
- Serie 5 lösen
- Quiz 5 bearbeiten
- bis etutoR 6 anschauen