Stochastik

Statistischer Test

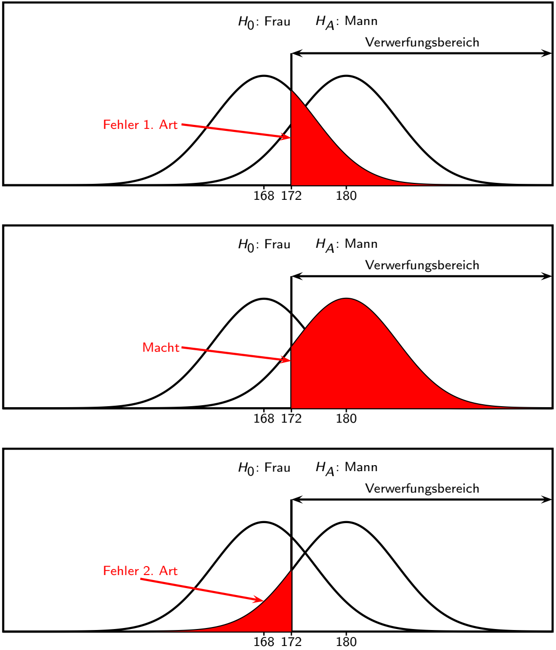
Mirko Birbaumer

Hochschule Luzern Technik & Architektur

Verschiedene Fehlerarten

Entscheidung / **Wahrheit**	H_0	H_A
H_0	✓	Fehler 1. Art
H_A	Fehler 2. Art	✓

Sie entscheiden sich für H_0, aber H_A wäre richtig $\longrightarrow$ Fehler 2. Art
Sie entscheiden sich für H_A, aber H_0 wäre richtig $\longrightarrow$ Fehler 1. Art



H_0: Frau
H_A: Mann
Verwerfungsbereich
Fehler 1. Art
168
172
180
H_0: Frau
H_A: Mann
Verwerfungsbereich
Macht
168
172
180
H_0: Frau
H_A: Mann
Verwerfungsbereich
Fehler 2. Art
168
172
180

Macht eines statistischen Tests

Macht

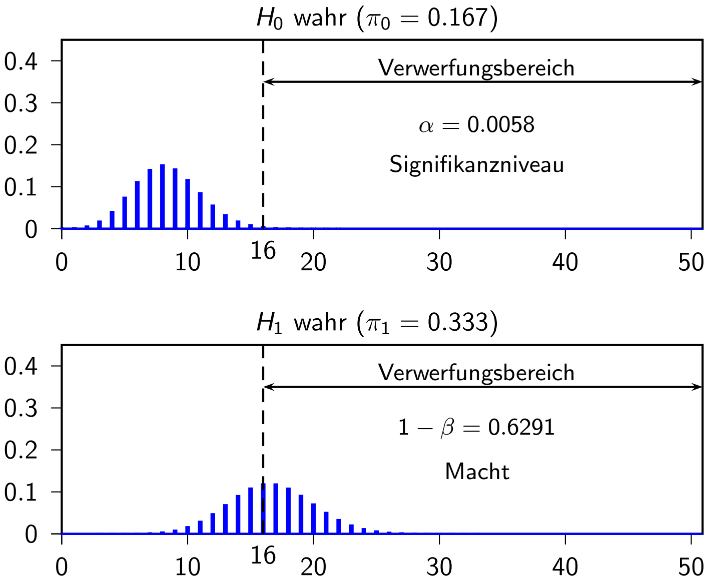
Die **Macht** gibt die Wahrscheinlichkeit an, H_A zu entdecken, falls H_A richtig ist:

$$P_{H_A}(T \in K)$$

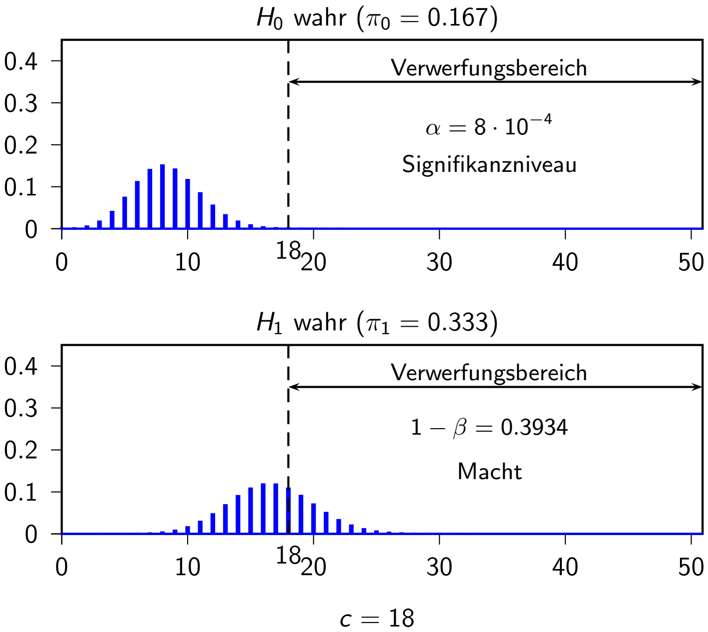
- **Beispiel:** Sie spielen Würfel mit einem Trickbetrüger. Der weiss, dass die Gewinnwahrscheinlichkeit für einen 6er $\pi = 1/3$ ist.

- Wenn Sie nun zu vorsichtig sind (aus Angst vor einer Schlğerei...), den Trickbetrüger des Betrugs zu beschuldigen und α (Fehler 1. Art) sehr klein machen, freut es den Trickbetrüger!

- Wird α kleiner gemacht, sinkt Ihre Chance, den Betrug aufzudecken: die Macht des statistischen Tests nimmt ab.


H_0 wahr ($\pi_0 = 0.167$)
Verwerfungsbereich
$\alpha = 0.0058$
Signifikanzniveau
0.4
0.3
0.2
0.1
0
0
10
16
20
30
40
50
H_1 wahr ($\pi_1 = 0.333$)
Verwerfungsbereich
$1 - \beta = 0.6291$
Macht
0.4
0.3
0.2
0.1
0
0
10
16
20
30
40
50

$c = 16$


H_0 wahr ($\pi_0 = 0.167$)
Verwerfungsbereich
$\alpha = 8 \cdot 10^{-4}$
Signifikanzniveau
0.4
0.3
0.2
0.1
0
0
10
18
20
30
40
50
H_1 wahr ($\pi_1 = 0.333$)
Verwerfungsbereich
$1 - \beta = 0.3934$
Macht
0.4
0.3
0.2
0.1
0
0
10
18
20
30
40
50
$c = 18$

P-Wert

P-Wert

Der **P-Wert** ist die Wahrscheinlichkeit, unter der Nullhypothese ein mindestens so extremes Ereignis (in Richtung der Alternative) zu beobachten wie das aktuell beobachtete. Man kann anhand des P-Werts direkt den Testentscheid ablesen: Wenn der P-Wert kleiner als das Niveau ist, so verwirft man H_0, ansonsten nicht.

- **Beispiel** Bei einer Binomialverteilung mit $n = 10$ wollen wir die Nullhypothese

$$H_0 \ : \ \pi = \pi_0 = 0.5$$

gegen die Alternative

$$H_A \ : \ \pi > 0.5$$

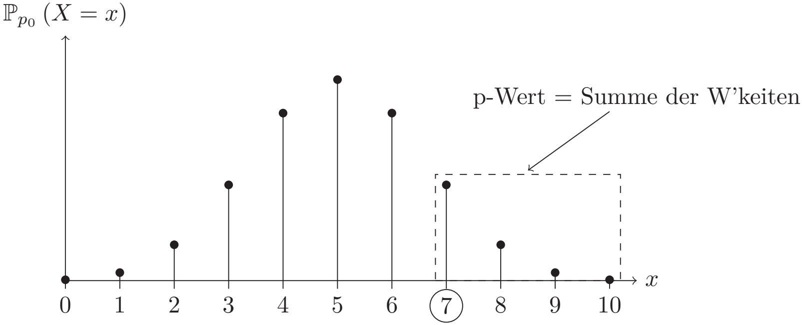
testen (π ist z.B. die Wahrscheinlichkeit für Kopf bei einer Münze).

- X : Anzahl Würfe mit Kopf bei insgesamt 10 Würfen

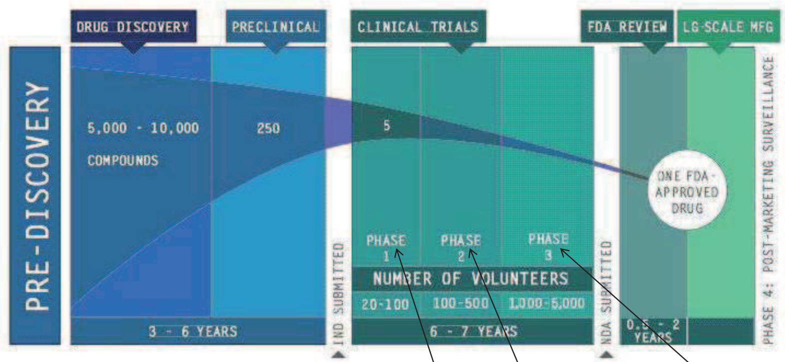
- Unter H_0 folgt die Zufallsvariable X der Verteilung: $X \sim \text{Bin}(10, 0.5)$

P-Wert

- Beobachtet wurde $x = 7$
- Der P-Wert ist hier die Summe aller Wahrscheinlichkeiten für X grösser gleich 7, d.h. P-Wert $= P_{\pi_0}(X \geq 7)$



Beispiel: Klinische Studien



- **Lose**: Alle denkbaren Patienten
- **n gezogene Lose**: n Patienten in Studie
- **Gewinn**: Patient wird gesund
- π: Anteil aller denkbaren Patienten, bei denen Medikament wirkt

Beispiel: Klinische Studie Phase 2

- Hersteller behauptet: Neues Medikament wirkt in 80% der Fällen (Nullhypothese)
- In einer Phase 2 Studie mit 100 Patienten werden aber nur 67 gesund.
- Ist das plausibel, wenn die Heilungswahrscheinlichkeit 80% ist? Wir vermuten, die Heilungswahrscheinlichkeit ist kleiner (Alternativhypothese)
- X: Anzahl geheilter Patienten
  Falls Hersteller (also unter Annahme der Nullhypothese) recht hat:

$$X \sim \text{Bin}(n = 100, \pi = 0.8)$$

Beispiel: Klinische Studie Phase 2

- "P-Wert" : $P(X \leq 67) = \ldots = 0.0016$

R-Befehl: pbinom()

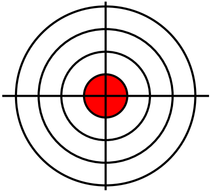
```
> pbinom(67,100,0.8)
[1] 0.001550441
```

- Falls der Hersteller recht hat, ist unsere Beobachtung sehr unwahrscheinlich (Verwerfen der Nullhypothese auf Signifikanzniveau $\alpha = 5\%$)

➔ Vermutlich ist die Heilungswahrscheinlichkeit kleiner als 80%

Grundidee Vertrauensintervall

- Sie sind ein guter Schütze und wissen, dass Sie in 95% der Fälle nicht allzu weit neben ihr persönlich gewähltes Ziel treffen.



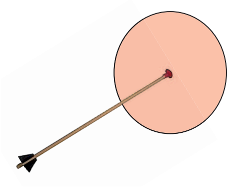
- D.h. in 95% der Fälle landen Sie in einem Bereich, der max. 5 cm von Ihrem Ziel entfernt ist (roter Kreis).

Grundidee

- Jemand anderes will nun herausfinden, auf was Sie wohl gezielt haben
- Die Person kennt Ihre Genauigkeit, d.h. sie weiss, dass Sie in 95% der Fälle max. 5 cm von ihrem Ziel entfernt liegen
- Die andere Person sieht aber leider nur noch den eingeschlagenen Pfeil und weiss nicht, auf was exakt Sie gezielt haben

Vertrauensintervalle: Grundidee

- Was macht die Person am besten? Sie zieht einen Kreis mit Radius 5 cm um die Einschlagstelle



- Dieser Kreis „fängt“ das wahre Ziel in 95% der Fälle ein
- Denn: In 95% der Fälle liegen Sie max. 5 cm vom Ziel entfernt. In diesen Fällen „fängt“ der Kreis das wahre Ziel ein

Vertrauensintervalle

Übertrag auf Statistik:

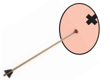
- Das Ziel, das wir treffen wollen, ist ein **unbekannter fixer Parameter** (z.B. Gewinnwahrscheinlichkeit π bei Binomialverteilung).
- Der „Pfeil“ ist der **Schätzer** dafür: $\hat{\pi}$
- Die **Genauigkeit** kennen wir, wenn wir wissen, wie sich die Differenz $\hat{\pi} - \pi$ (= Abstand vom Ziel) verhält
- Das bedeutet, dass wir die **Verteilung** von $\underbrace{\hat{\pi} - \pi}_{\boxed{\text{ZV, solange Stichprobe noch nicht realisiert ist}}}$ kennen müssen

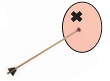
Vertrauensintervalle: Interpretation

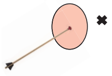
- Für eine konkrete Stichprobe sehen wir nur den „eingeschlagenen Pfeil“ $\hat{\pi}$ (= realisierter Wert des Schätzers basierend auf den beobachteten Daten)
- Da wir für eine andere Stichprobe einen leicht anderen Wert für $\hat{\pi}$ erhalten, wollen wir dem konkreten Wert nicht allzu viel Gewicht geben
- Wir wollen lieber eine Ahnung haben, wo das unbekannte, wahre π in etwa liegt, d.h. wir wollen „den roten Kreis ziehen“
- **Interpretation vom Vertrauensintervall:**
  - Für eine konkrete Realisierung wissen wir leider nicht, ob der Kreis das wahre Ziel „eingefangen“ hat oder nicht
  - Wir wissen aber: Wenn wir diese Strategie verwenden, so „fangen“ wir in 95% der Fälle das Ziel ein und liegen richtig

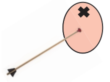
Vertrauensintervalle: Interpretation

Das Ziel ✖ sei fix. Wir schiessen ein paar Mal.

Schuss 1  ✓

Schuss 2  ✓

Schuss 3  —

Schuss 4  ✓

Etc. ➔ Auf langer Sicht werden wir 95% ✓ erhalten

Vertrauensintervall: Definition

Vertrauensintervall

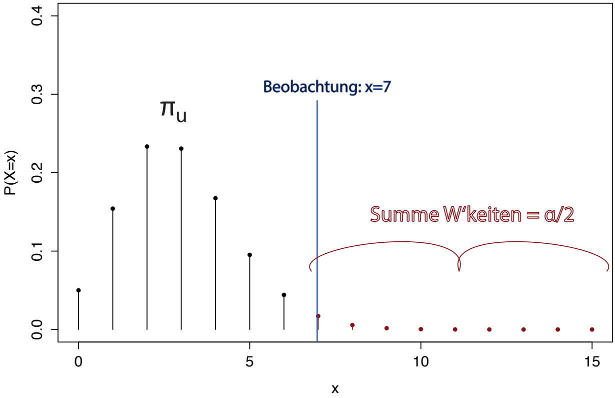
Ein Vertrauensintervall I zum Niveau $1-\alpha$ besteht aus allen Parameterwerten, die im Sinne des statistischen Tests zum Signifikanzniveau α mit der Beobachtung verträglich sind (üblicherweise nimmt man den zweiseitigen Test). Mathematisch heisst dies:

$$I = \{\pi_0;\ \text{Nullhypothese } H_0: \ \pi = \pi_0 \text{ wird belassen}\} = [\pi_u, \pi_o].$$

Das bedeutet also, dass wir sozusagen alle π_0 "durchtesten" und diejenigen "sammeln", bei denen die entsprechende Nullhypothese nicht verworfen wird.

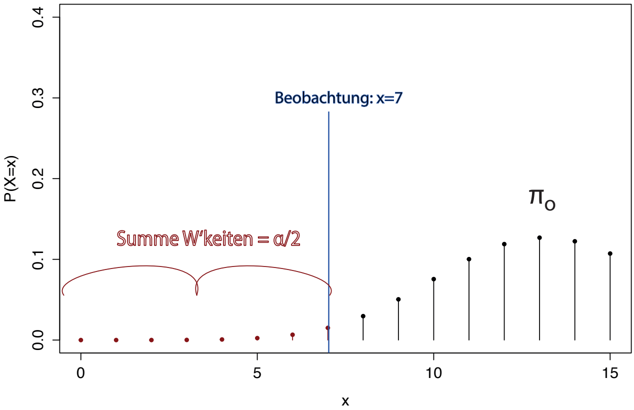
Beispiel: Sie kommen an einer Losbude vorbei und ziehen 50 Lose: darunter sind 7 Gewinne. Es stellt sich also die Frage: welche Werte für die Gewinnwahrscheinlichkeit π sind kompatibel mit Ihrer Beobachtung?

Zweiseitiges Vertrauensintervall



Untere Grenze π_u des 95%-Vertrauensintervalls: wir lassen π_u nach unten "wandern", bis $P_{\pi_u}(X \geq 7) \overset{\approx}{\geq} \alpha/2$ ist.

Zweiseitiges Vertrauensintervall



Obere Grenze π_o des 95%-Vertrauensintervalls: wir lassen π_o nach oben "wandern", bis $P_{\pi_o}(X \leq 7) \overset{\approx}{\geq} \alpha/2$ ist.

Zweiseitiges Vertrauensintervall mit **R**

R-Befehl: binom.test()

```
> binom.test(7,50)
Exact binomial test
data: 7 and 50 number of successes = 7, number of trials = 50,
p-value = 2.099e-07
alternative hypothesis: true probability
of success is not equal to 0.5
95 percent confidence interval:
0.0581917 0.2673960
sample estimates:
probability of success
0.14
```

Das zweiseitige 95%-Vertrauensintervall ist also

$$I = [0.058, 0.27]$$

Einseitiges Vertrauensintervall

- **Einseitiges nach unten gerichtetes** Vertrauensintervall auf dem Signifikanzniveau α für Beobachtung x hat die Form

$$[0, \pi_o]\,,$$

wobei π_o die folgende Bedingung erfüllt

$$P_{\pi_o}(X \leq x) \overset{\approx}{\geq} \alpha$$

- **Einseitige nach oben gerichtetes** Vertrauensintervall auf dem Signifikanzniveau α für Beobachtung x hat die Form

$$[\pi_u, 1]\,,$$

wobei π_u die folgende Bedingung erfüllt:

$$P_{\pi_u}(X \geq x) \overset{\approx}{\geq} \alpha$$

Einseitiges Vertrauensintervall mit **R**

R-Befehl: binom.test()

```
binom.test(7,50,alternative="less")
Exact binomial test
data: 7 and 50
number of successes = 7, number of trials = 50,
p-value = 1.049e-07
alternative hypothesis: true probability of success is less than
0.5
95 percent confidence interval:
0.0000000 0.2469352
sample estimates:
probability of success
0.14
```

- Das einseitige nach unten gerichtete 95%-Vertrauensintervall ist also $I = [0, 0.25]$.
- Das einseitige nach oben gerichtete 95%-Vertrauensintervall erhält man, indem man `alternative="greater"` wählt

Übersicht Fragestellungen in der Statistik

Wir sind nun in der Lage, folgende Fragestellungen zu beantworten:

- Welcher ist der **plausibelste Wert** eines unbekannten Parameters?
  $\Rightarrow$ $\boxed{\text{Parameterschätzung}}$

- Ist ein **bestimmter vorgegebener Parameterwert** (z.B. ein Sollwert μ_0), mit den beobachteten Daten **verträglich**?
  $\Rightarrow$ $\boxed{\text{Statistischer Test}}$

Im folgenden werden wir uns beschäftigen mit der Fragestellung:

- Was ist der **Bereich** von plausiblen Parameterwerten?
  $\Rightarrow$ $\boxed{\text{Vertrauensintervall}}$

3 Grundfragestellungen der Statistik

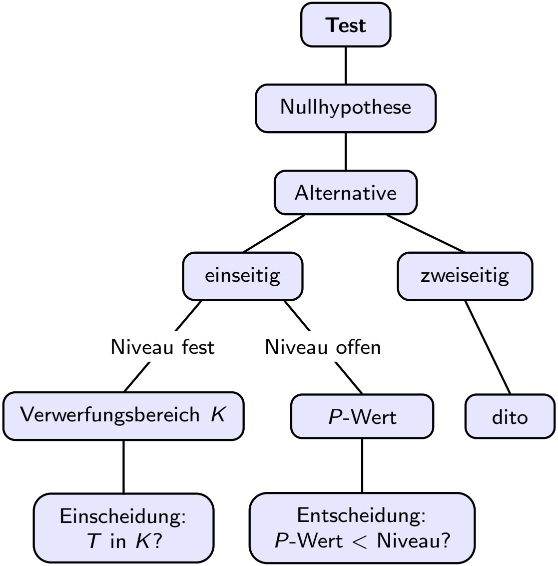
Sei $x = 6$ die effektive Anzahl fehlerhaft übertragener Bits bei der Übertragung von 100 digitalen Signalen. Wir fassen $x = 6$ als **Realisierung** einer Zufallsvariablen X auf, und nehmen an, dass $X \sim Bin(100, \pi)$, also binomialverteilt ist.

- Welches ist der plausibelste Wert π (zur Beobachtung $x = 6$)?
  Antwort: $\hat{\pi} = \frac{6}{100} = 0.06$ (Maximum-Likelihood- oder/Momenten-Methode)

- Ist die Beobachtung $x = 6$ kompatibel mit $\pi_0 = 0.1$ (üblicher Übertragungsfehler von Übertragungskanälen) oder mit $\pi < 0.1$?
  Antwort: P-Wert: $P_{\pi_0}(T \leq t) = 0.12$ (einseitiger Test)

- Welcher Bereich (Intervall) für den Parameter π ist mit der Beobachtung $x = 6$ kompatibel?
  Antwort: 95%-Vertrauensintervall für wahren Parameter $I = [0, 0.11]$

Übersicht : Statistischer Test



Vertrauensintervall:
(meist zweiseitig)
Alle Parameterwerte, bei denen der Test nicht verwirft

Übersicht Statistische Tests: Testentscheid

- Entscheid anhand **Teststatistik**
  Teststatistik $\notin K$: Belasse H_0
  Teststatistik $\in K$: Verwerfe H_0

- Entscheid anhand p-**Wert**
  p-Wert $> \alpha$: Belasse H_0
  p-Wert $< \alpha$: Verwerfe H_0

- Entscheid anhand **Vertrauensintervall** (bei zweiseitigen Tests)
  $\theta_0 \in VI$: Belasse H_0
  $\theta_0 \notin VI$: Verwerfe H_0

Absence of evidence is not evidence of absence

Wird H_0 nicht verworfen (d.h. belassen), so bedeutet dies nicht, dass H_0 damit statistisch bewiesen ist.