Stochastik

Statistischer Test

Mirko Birbaumer

Hochschule Luzern Technik & Architektur

Statistischer Test: Idee und Ziel

Können wir basierend auf Daten entscheiden oder nachweisen

- ob ein Grenzwert überschritten wird, z.B. bei Asbestfasern?
- ob ein Hersteller bei einem Produkt die Spezifikationen verletzt?
- ob die Wirkung eines Medikamentes tatsächlich der Behauptung einer Pharmafirma entspricht?

Statistischer Test: Idee und Ziel

- Wir wollen also herausfinden, ob eine bestimmte Annahme oder ein bestimmter Parameter mit unseren beobachteten Daten verträglich ist oder nicht
- Wir müssen eine **objektive, reproduzierbare Entscheidungsregel** verwenden

Beispiel:

- Sie vermuten, dass eine Münze nicht „fair" ist und **zu oft Kopf** zeigt
- In 10 neuen Würfen haben wir 9 mal Kopf erhalten. Passt das mit der Annahme $\pi = 0.5$ (Münze ist fair) zusammen?

Statistischer Test: Vorgehensweise

- Spezifiziere **Nullhypothese** $H_0 : \pi = \pi_0 = 0.5$
  ($\pi =$ Wahrscheinlichkeit für Kopf)
  Dies ist die angezweifelte Behauptung, der „Normalzustand“, was sie verwerfen wollen, „Status Quo“, ...

- Spezifiziere **Alternativhypothese** $H_A : \pi > 0.5$
  (was sie nachweisen wollen, ihre Vermutung)

Was für Möglichkeiten gibt es?

- H_0 stimmt und es war Zufall, dass wir so oft Kopf gesehen haben („kann vorkommen, wenn auch selten“)

- H_A stimmt und deshalb haben wir so oft Kopf gesehen

Ab welcher Grenze H_0 verwerfen?

- Schlussendlich müssen wir eine Grenze finden, ab der wir sagen, dass wir H_0 verwerfen zugunsten von H_A
- In unserem Beispiel macht es Sinn, H_0 zu verwerfen, wenn die Anzahl Würfe mit Kopf sehr hoch ist (da es unter $\pi = 0.5$ sehr unwahrscheinlich ist - **einseitiger Test**).
- X sei die Anzahl Würfe mit Kopf von insgesamt 10 Würfen
- Wir wählen also eine Grenze c derart, dass wir bei $X \geq c$ sagen, dass wir die Nullhypothese verwerfen

Bestimmung Verwerfungsbereich

- Die Wahrscheinlichkeit, dass wir H_0 verwerfen, obwohl H_0 stimmt - soll tief sein, d.h. es soll gelten

$$P_{\pi_0}[X \geq c] \leq \alpha$$

  wobei α das sogenannte **Signifikanzniveau** (auch **Irrtumswahrscheinlichkeit**) ist, typischerweise $\alpha = 0.05,\ 0.01$

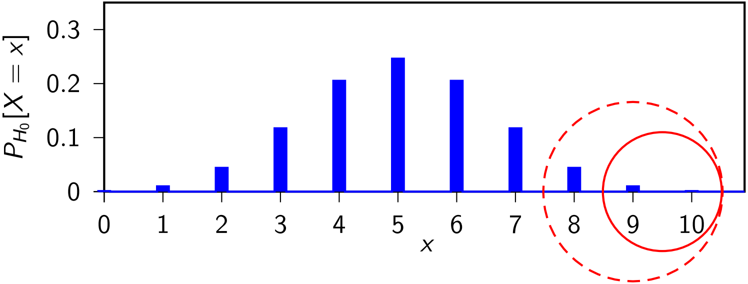
- Falls die Nullhypothese doch stimmen sollte („Münze fair"), dann würde $X \geq c$ nur bei jedem zwanzigsten, resp. jedem hundertsten Spiel passieren

- Wir wählen das **kleinste** c, so dass obige Bedingung erfüllt ist (sonst wird es unnötig schwierig, H_0 zu verwerfen wenn H_A stimmt)

- Ab $X \geq c$ verwerfen wir also die Nullhypothese, d.h. der sogenannte **Verwerfungsbereich** ist $\{c, \ldots, 10\}$

Verteilung unter Nullhypothese

Unter H_0 (d.h. $\pi_0 = 0.5$) ist $X \sim \text{Bin}(10, 0.5)$ verteilt. Wir schreiben dafür auch $P_{\pi_0}[X = x]$.



$$P_{\pi_0}[X = 9] + P_{\pi_0}[X = 10] \leq 0.05$$

$$P_{\pi_0}[X = 8] + P_{\pi_0}[X = 9] + P_{\pi_0}[X = 10] \geq 0.05$$

Berechnung des Verwerfungsbereichs mit R

R-Befehl: dbinom()

```
> sum(dbinom(9:10,size=10,prob=0.5))
[1] 0.01074219
```

oder eleganter

R-Befehl: pbinom()

```
> 1-pbinom(8,10,prob=0.5)
[1] 0.01074219
```

Testentscheid

- Ab $X \geq 9$ verwerfen wir also die Nullhypothese, d.h. der sogenannte **Verwerfungsbereich** ist $\{9, 10\}$
- Wir haben $x = 9$ beobachtet, also können wir die Nullhypothese verwerfen
- Wir haben also „statistisch nachgewiesen", dass die **Münze nicht fair ist und zu oft Kopf zeigt**

Statistischer Test: Schematisch

1. **Modell:**
   X: Anzahl Würfe, die Kopf zeigen, wenn man 100 mal wirft.

$$X \sim \text{Binomial}(10, \pi)$$

2. **Nullhypothese:**

$$H_0 : \quad \pi_0 = 0.5$$

   Alternative:

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

3. **Teststatistik:**
   T: Anzahl Würfe, die Kopf zeigen, wenn man 100 mal wirft
   Verteilung der Teststatistik unter H_0:

$$T \sim \text{Binomial}(10, 0.5)$$

4. **Signifikanzniveau:**

$$\alpha = 0.05$$

Statistischer Test: Schematisch

5. **Verwerfungsbereich:**

c	$c = 5$	$c = 6$	$c = 7$	$c = 8$	$c = 9$	$c = 10$
$P(X \geq c)$	0.623	0.377	0.172	0.055	0.010	0.001

Aus Tabelle $P[T \geq 9]$ kleiner als 5 % ist

Also ist der Verwerfungsbereich

$$K = \{9, 10\}$$

6. **Testentscheid:**
   - Der beobachtete Wert der Teststatistik $t = 9$ liegt im Verwerfungsbereich $K = \{9, 10\}$
   - Daher kann die Nullhypothese auf dem Signifikanzniveau 5 % verworfen werden
   - D.h.: es gibt nicht genügend statistische Evidenz (auf dem Signifikanzniveau $\alpha = 0.05$) dafür, dass die Münze zu gezinkt ist

Bemerkungen

- Wir müssen Null- und Alternativhypothese vor der Datenerhebung festlegen
- Bzw. der Test muss auf neuen Daten durchgeführt werden
- Man kann zwar basierend auf einer Datenanalyse Hypothesen bilden, **um diese zu verifizieren werden aber neue Daten benötigt**

Binomialtest: Zauberwürfel

Binomialtest

1. Modell: X: Anzahl 6er bei 50 Würfen; $X \sim \text{Bin}(n = 50, \pi)$

2. Nullhypothese: $H_0: \ \pi_0 = \frac{1}{6}$ (Würfel ist fair: status quo)
   Alternative: $H_A: \ \pi > \frac{1}{6}$ (wir glauben, Würfel ist gezinkt, einseitiger Test)

3. Teststatistik T: Anz. 6er bei 50 Würfen
   Verteilung der Teststatistik, wenn Nullhypothese stimmt:

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

4. Signifikanzniveau: $\alpha = 0.05$ (Konvention)

Binomialtest

5. Verwerfungsbereich der Teststatistik:

$$P[T = t] = \binom{n}{t} \pi_0^t (1 - \pi_0)^{n-t}; \quad \text{berechne } P[T \geq t]$$

Verwerfungsbereich

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

Einseitige vs. zweiseitige Tests

- Die Entscheidung für eine zweiseitige oder eine einseitige Alternative H_A hängt von der **Fragestellung** ab
- Eine einseitige Alternative ist dann angebracht, wenn nur ein Unterschied in eine **bestimmte Richtung** von Bedeutung ist (Bsp. Überschreitung Grenzwert)
- Der einseitige Test ist auf der einen Seite „blind“, dafür verwirft er auf der anderen Seite früher als der zweiseitige Test (da der Verwerfungsbereich früher beginnt)
- Man sagt auch, dass er eine **grössere Macht** hat in diesem Bereich (siehe später)

Verwerfungsbereich für Teststatistik

- Form vom Verwerfungsbereich:
  - $K = [0, c_u] \cup [c_o, n]$ falls $H_A:\ \pi \neq \pi_0$ — zweiseitiger Test
  - $K = [c_>, n]$ falls $H_A:\ \pi > \pi_0$
  - $K = [0, c_<]$ falls $H_A:\ \pi < \pi_0$

  $\Big\}$ einseitiger Test

- Grenzen (c's) werden bestimmt, so dass folgendes gilt:
  - $P(T \leq c_u) \approx \frac{\alpha}{2}$
  - $P(T \geq c_o) \approx \frac{\alpha}{2}$

  $\Big\}$ zweiseitiger Test

  - $P(T \geq c_>) \approx \alpha$
  - $P(T \leq c_<) \approx \alpha$

  $\Big\}$ einseitiger Test

Verwerfungsbereich für Teststatistik: Grenzen exakt

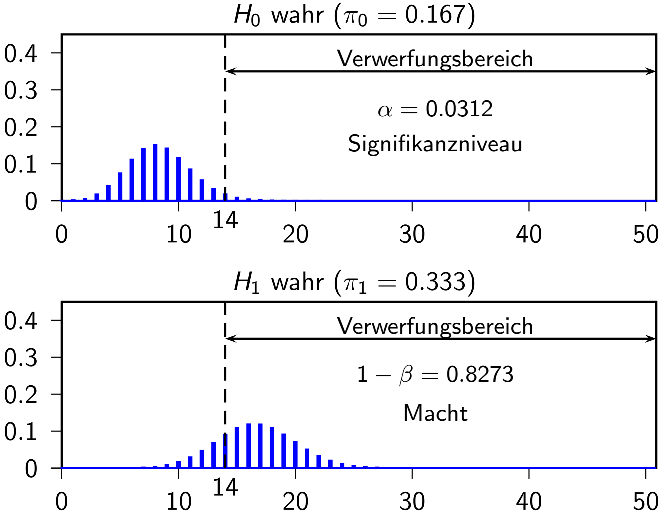
- Beispiel: $T \sim \text{Bin}(10, 0.4)$ mit $\alpha = 0.05$

t	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$P[T = t]$	0.006	0.04	0.12	0.21	0.25	0.20	0.11	0.04	0.01	0.002	0.0001

- Aus Tabelle:
  - $c_u = 0$
  - $c_o = 8$
  - $c_> = 8$
  - $c_< = 1$

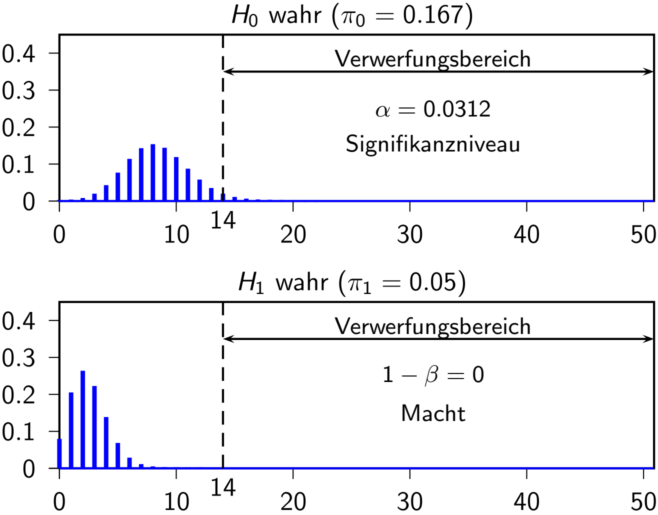

Casino

Einseitiger Test - bei zu vielen 6er



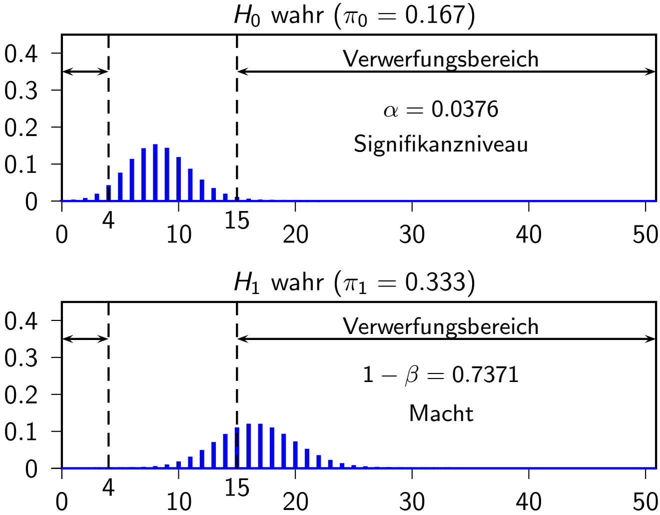
$c = 14$

Einseitiger Test - bei zu wenigen 6er



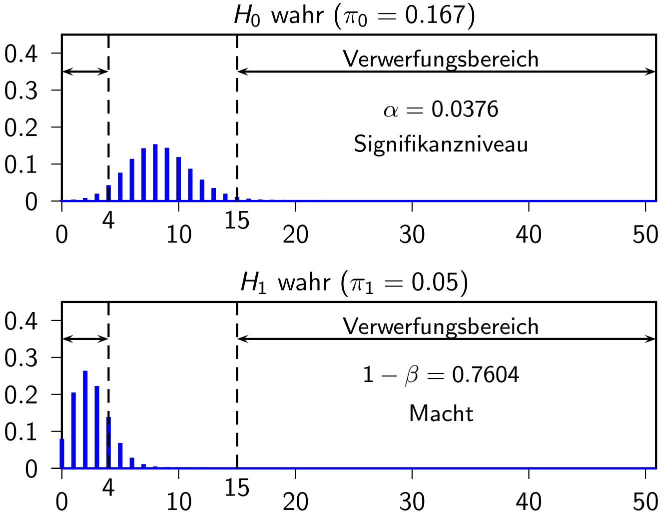
$c = 14$

Zweiseitiger Test - bei zu vielen 6er



$c = 4, 15$

Zweiseitiger Test - bei zu wenigen 6er



$c = 4, 15$

Zweiseitiger versus Einseitiger Test

- **Einseitig**:
  - Auf einer Seite blind
  - Auf anderer Seite sehr grosse Sehschärfe (grosse Macht)

- **Zweiseitig**:
  - Sieht auf beide Seiten
  - Sieht auf keiner Seite besonders gut (kleine Macht)