Mathematik II für Biologen

Schließende Statistik: Hypothesentests (Forts.)

Stefan Keppeler

22. Mai 2009

	Allgemein	im Beispiel
1	Nullhypothese H_0	$k = 80$
2	Alternativhypothese H_A	$k > 80$
3	Wähle Teststatistik X	$\#$ { kranke Individuen unter 10 zufällig gewählten }
4	Verteilung von X falls H_0 wahr	Simulation
5	Wähle Signifikanzniveau α	$\alpha = 5\%$
6	Verwerfungsbereich K (aus 4 & 5)	$K = \{3, 4, 5, \ldots, 10\}$
7	Bestimme X aus Daten	$X = 3$
8	Testentscheidung: $X \in K$ oder $X \notin K$?	$X \in K$: H_0 wird verworfen

Merke: Je kleiner α desto schwieriger wird es, H_0 zu verwerfen.

Beispiel: Hätten wir $\alpha = 1\%$ gewählt, so wäre $K = \{4, 5, \ldots, 10\}$ gewesen und damit $X = 3 \notin K$...

- Einseitige Tests haben die Form

$$H_0 : \; k = k_0$$
$$H_A : \; k > k_0 \qquad \text{(oder } H_A : \; k < k_0\text{)}$$

  Behauptung, die gezeigt werden soll:
  k ist größer (kleiner) als erwartet.

- Zweiseitige Tests haben die Form

$$H_0 : \; k = k_0$$
$$H_A : \; k \neq k_0$$

  Behauptung, die gezeigt werden soll:
  k weicht von Erwartung ab (egal in welche Richtung.)

Im Beispiel: Falls $X = 3$ beobachtet wurde, so wird...

...H_0 verworfen für $\alpha = 5\%$

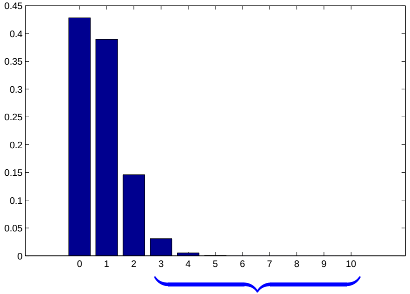
...H_0 nicht verworfen für $\alpha = 1\%$

...H_0 verworfen, falls $\alpha \geq 3{,}6\%$

...H_0 nicht verworfen für $\alpha < 3{,}6\%$

Definition: Das Signifikanzniveau, an dem der Test zwischen *verwerfen* und *nicht verwerfen* schwankt, heißt p-Wert.

Im Beispiel: Für $X = 3$ beträgt der p-Wert $3{,}6\%$



$\approx 3{,}6\%$

Äquivalent: Der p-Wert ist die Wahrscheinlichkeit (unter H_0) dafür, etwas zu beobachten, das H_A mindestens so stark unterstützt wie das, was tatsächlich beobachtet wurde.

Im Beispiel: $X = 3$ beobachtet (in 3% der Fälle); noch mehr als $X = 3$ würde $X = 4, 5, \ldots, 10$ die Alternativhypothese H_A unterstützen.

$X \geq 3$ wird in $3{,}6\%$ der Fälle beobachtet (p-Wert).

Interpretation verschiedener p-Werte:

(mit Vorsicht zu genießen)

p-Wert $> 5\%$	Test nicht einmal signifikant auf 5%-Niveau: kein (statistischer) Widerspruch zwischen Daten und Modell H_0
$5\% \geq$ p-Wert $> 1\%$	Test signifikant auf 5%- aber nicht auf 1%-Niveau: schwacher Widerspruch zwischen Daten und H_0
$1\% \geq$ p-Wert $> 0{,}1\%$	Test signifikant auf 1%- (also auch auf 5%-) aber nicht auf $0{,}1\%$-Niveau
$0{,}1\% \geq$ p-Wert	Test (sogar) signifikant auf $0{,}1\%$-Niveau: sehr starker Widerspruch zwischen Daten und H_0

Angenommen, $X = 3$ wurde beobachtet (und $\alpha = 5\%$) dann...

- wird H_0: $k = 80$ verworfen zugunsten von $H_A : k > 80$
- würde H_0: $k = 79$ erst recht verworfen zugunsten von $H_A : k > 79$, ebenso wie H_0: $k = k_0$ zugunsten von $H_A : k > k_0$ für alle $k_0 \leq 80$.

Was passiert für $k_0 > 80$?

Neue Rechnung/Simulation (wiederhole Testschritt 4) zeigt

- H_0: $k = k_0$ wird verworfen zugunsten von $H_A : k > k_0$ für alle $k_0 \leq 87$,
- H_0: $k = k_0$ wird nicht verworfen zugunsten von $H_A : k > k_0$ für alle $k_0 > 87$,

denn (für $X = 3$):

k_0	87	88
p-Wert	$4{,}87\% < 5\%$	$5{,}02\% > 5\%$

Definition: Das $(1-\alpha)$-Vertrauensintervall (Konfidenzintervall, -bereich) für k besteht aus denjenigen Zahlen k_0 für die $H_0 : k = k_0$ **nicht verworfen** wird.

Im Beispiel: Falls $X = 3$ beobachtet wurde:
95%-Vertrauensintervall für k: $\{88, 89, \ldots, 1000\}$.

Merke: Das Vertrauensintervall hängt (wie der p-Wert) vom beobachteten Wert von X ab.

	H_0 wird verworfen	H_0 wird nicht verworfen
H_0 ist wahr	☹ Fehler 1. Art	☺ Richtige Entscheidung
H_0 ist falsch	☺ Richtige Entscheidung	☹ Fehler 2. Art

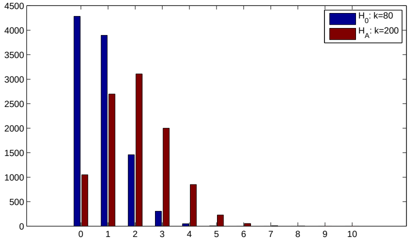
- Fehler 1. Art $\leftrightarrow$ Signifikanzniveau:
  Falls H_0 gilt, so begeht der Test in $\leq \alpha$ der Fälle einen Fehler 1. Art (gemäß Konstruktion des Verwerfungsbereichs K).
- Fehler 2. Art $\leftrightarrow$ Zusätztliche Annahmen, Simulation...

Beispiel: $H_0 : k = 80$, $\alpha = 5\%$, $K = \{3, 4, 5, \ldots, 10\}$
Annahme: In Wirklichkeit gilt nicht H_0 sondern $H_A : k = 200$.

```
N=1000;
k=200;
X=[];
for wdh=1:10^4 % besser 10^5
  urne=[ones(1,k),zeros(1,N-k)];
  n=10;
  stichprobe=[];
  for j=1:n
    kugel=unidrnd(length(urne));
    stichprobe=[stichprobe,urne(kugel)];
    urne(kugel)=[];
  end;
  X=[X,sum(stichprobe)];
end

verteilungH0=[4283 3896 1458 306 51 6 0 0 0 0 0]
Z=[verteilungH0,hist(X,0:10)];
Z=reshape(Z,11,2);
bar(0:10,Z)
legend('H_0: k=80','H_A: k=200')
```

Fehler 2. Art $\beta \approx 69\%$ | Macht des Tests $1-\beta \approx 31\%$



$\approx 3{,}6\%$
Fehler 1. Art

Beispiel (Forts.): (Annahme: H_A gilt)

- In 31% der Fälle wird H_0 verworfen, in 69% der Fälle nicht.
- Wahrscheinlichkeit für einen Fehler 2. Art ist $\beta = 69\%$ (falls H_A gilt).
- $1 - \beta$ (hier $= 31\%$) heißt Macht des Tests.

	einfacher Test	Test mit p-Wert
1	Nullhypothese H_0	
2	Alternativhypothese H_A	
3	Wähle Teststatistik X	
4	Verteilung von X falls H_0 wahr	
5	Wähle Signifikanzniveau α	
6	Verwerfungsbereich K (aus 4 & 5)	
7	Bestimme X aus Daten	
8	Testentscheidung: $X \in K$ oder $X \notin K$?	
9		p-Wert (aus 4 & 7)
10		Testentscheidung: p-Wert $\leq \alpha$ oder $> \alpha$

$X \in K$ bzw. p-Wert $\leq \alpha$: H_0 wird verworfen
(H_A statistisch bewiesen)

	Testschemata	$(1-\alpha)$-Vertrauensintervall	
1	Nullhypothese H_0	enthält einen Parameter	
2	Alternativhypothese H_A		
3	Wähle Teststatistik X		
4	Verteilung von X falls H_0 wahr		
5	Wähle Signifikanzniveau α	bereits festgelegt	
6	Verwerfungsbereich K (aus 4 & 5)		
7	Bestimme X aus Daten		
8	Testentscheidung: $X \in K$ oder $X \notin K$?		
9	p-Wert (aus 4 & 7)		
10	Testentscheidung: p-Wert $\leq \alpha$ oder $> \alpha$		

$X \in K$ bzw. p-Wert $\leq \alpha$: Wert des Parameters liegt nicht im Vertrauensintervall für den Parameter

$X \notin K$ bzw. p-Wert $> \alpha$: Wert des Parameters liegt im Vertrauensintervall