

Mathematik II für Biologen

Schließende Statistik: Hypothesentests

Stefan Keppeler

2. Mai 2014



Nullhypothese und Alternativhypothese

Verteilung der Teststatistik

Simulation der Teststatistik mit `MATLAB`

Signifikanzniveau und Verwerfungsbereich

Testentscheidung

Testablauf



- ▶ **Beispiel:** Population von 1000 Individuen heißt “gesund”, falls für die Anzahl k “kranker” Individuen gilt: $k \leq 80$.
- ▶ Annahme: Population ist gesund, d.h. $k \leq 80$, also schlimmstenfalls $k = 80$.

Nullhypothese $H_0 : k = 80$

- ▶ Jemand behauptet: Die Population ist krank (d.h. $k > 80$).

Alternativhypothese $H_A : k > 80$.

- ▶ Alle Individuen zu untersuchen ist zu teuer. Sie planen, $n = 10$ Individuen zu untersuchen.

$$\begin{aligned} X &= \#\{\text{krankte Individuen unter den } n \text{ untersuchten}\} \\ &= \text{“Teststatistik”} \end{aligned}$$

X ist zufällig!

Ziel: Entscheide aufgrund von X , ob H_0 verworfen wird.



Annahme: H_0 stimmt. Welche Werte kann dann X annehmen?

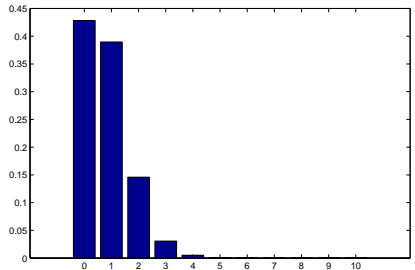
- ▶ Theoretisch: Alle Werte $X = 0, 1, 2, \dots, 10$ sind möglich
- ▶ Praktisch: Einige Werte von X sind unwahrscheinlicher als andere; z.B. ist es “praktisch unmöglich”, $X = 10$ zu beobachten.

Bestimme Verteilung der Teststatistik X

- ▶ mittels Simulation
 - ▶ nehme Urne mit 1000 Kugeln, 80 schwarze, 920 weiße
 - ▶ mische, ziehe $n = 10$ Kugeln $X = \#\{\text{scharze Kugeln}\}$
 - ▶ wiederhole oft, z.B. 10^6 mal
 - ▶ Histogramm für X -Werte
- ▶ mittels Wahrscheinlichkeitstheorie



```
N=1000; % Größe der Urne
k=80;   % Anzahl schwarze Kugeln
X=[];
for wdh=1:10^4 % besser 10^5
    urne=[ones(1,k),zeros(1,N-k)];
    n=10; % Stichprobenumfang
    stichprobe=[];
    for j=1:n
        kugel=unidrnd(length(urne));
        stichprobe=[stichprobe,urne(kugel)];
        urne(kugel)=[];
    end;
    X=[X,sum(stichprobe)];
end
verteilung=hist(X,0:10)./wdh;
bar(0:10,verteilung)
```



X	0	1	2	3	4	5	6 – 10
#	4283	3896	1458	306	51	6	0
	≈ 99,4%			≈ 0,6%			
	≈ 96,4%				≈ 3,6%		



- ▶ Lege **Signifikanzniveau** α fest, z.B. $\alpha = 5\%$.
(auch üblich: 10%, 1%, 0,1% oder 0,01%)
- ▶ Erkläre diejenigen theoretisch möglichen Werte von X
 - ▶ die zusammen höchstens α ausmachen
 - ▶ und am stärksten für H_A sprechen,für “praktisch unmöglich” (falls H_0 gilt).

Verwerfungsbereich $K = \{\text{“praktisch unmögliche” Werte}\}$

Hier¹: $K = \{3, 4, 5, \dots, 10\}$ (für $\alpha = 5\%$).

¹Bemerkung: $K^C = \{0, 1, 2\}$ heißt “Annahmebereich”



- ▶ **Erst jetzt:** Untersuche n Individuen, bestimme X ,
z.B. $X = 3$ beobachtet.
- ▶ **Testentscheidung**
 - ▶ Falls $X \in K$ beobachtet wird:
 H_0 wird verworfen zugunsten von H_A .
Man sagt: *Es ist "statistisch bewiesen" auf $\alpha = 5\%$, dass H_0 nicht gelten kann.* Oder: H_0 ist mit Beobachtung nicht vereinbar.
 - ▶ Falls $X \notin K$ beobachtet wird:
 H_0 wird nicht verworfen.
 H_0 kann gelten, muss aber nicht. Daten und H_0 scheinen sich nicht zu widersprechen, aber "nichts bewiesen".



	Allgemein	im Beispiel
1	Nullhypothese H_0	$k = 80$
2	Alternativhypothese H_A	$k > 80$
3	Wähle Teststatistik X	$\# \left\{ \begin{array}{l} \text{kranke Individuen unter} \\ 10 \text{ zufällig gewählten} \end{array} \right\}$
4	Verteilung von X falls H_0 wahr	Simulation
5	Wähle Signifikanzniveau α	$\alpha = 5\%$
6	Verwerfungsbereich K (aus 4 & 5)	$K = \{3, 4, 5, \dots, 10\}$
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 $K = \{4, 5, \dots, 10\}$ gewesen und damit $X = 3 \notin K$...

