

Mathematik für Biologen

Prof. Dr. Rüdiger W. Braun

<http://blog.ruediger-braun.net>

Heinrich-Heine-Universität Düsseldorf

29. Januar 2014

oooooooooooooooooooooooooooo

- 1 χ^2 -Tests
 - Verteilungstest
 - χ^2 -Verteilung
 - χ^2 -Vergleichstest
 - χ^2 -Unabhängigkeitstest

- 2 Übungspunkte

oooooooooooooooooooooooooooo

 χ^2 -Tests



Test für ein Merkmal mit nur zwei Ausprägungen

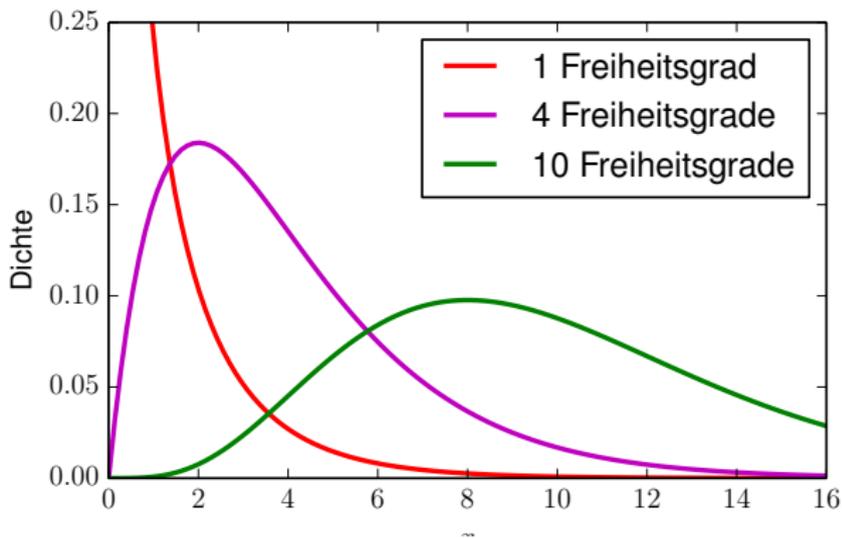
- Beispielaufgabe:

An der HHU sind 60.9% aller Erstsemester weiblich. Im BSc-Studiengang Biologie sind 266 von 432 Erstsemestern weiblich. Das sind 61.6%. Ist der Unterschied beim Anteil weiblicher Erstsemester signifikant zum Signifikanzniveau $\alpha = 0.05$?

- Für solche Fragestellungen verwendet man einen Chi-Quadrat-Anpassungstest. Diese Tests dienen zur Überprüfung der Gleichheit zweier Verteilungen.
- Man könnte diese Beispielaufgabe mit einem Binomialtest rechnen, benötigt dann aber eine Tabelle für $n = 432$

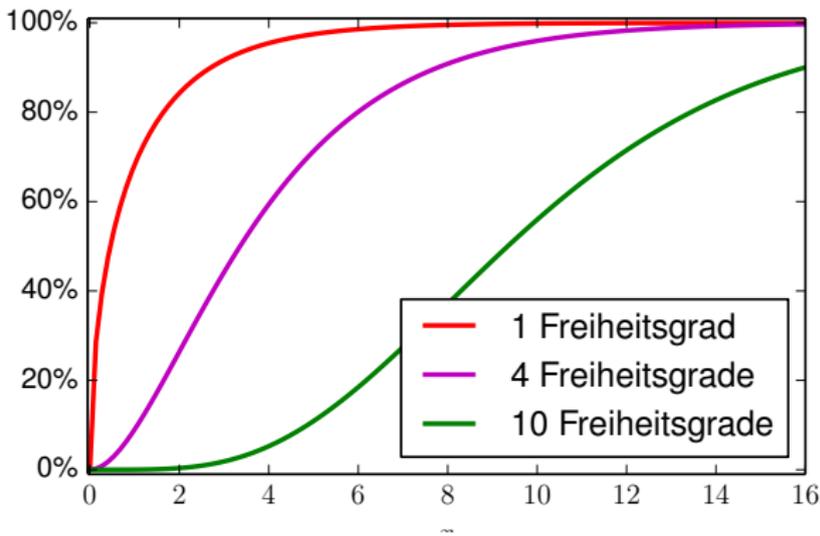


Graphen von Dichten von χ^2 -Verteilungen





Graphen von Verteilungsfunktionen von χ^2 -Verteilungen





Quantile der χ^2 -Verteilung

f	90%	95%	97.5%	99%	99.5%	99.9%	99.95%
1	2.71	3.84	5.02	6.63	7.88	10.83	12.12
2	4.61	5.99	7.38	9.21	10.60	13.82	15.20
3	6.25	7.81	9.35	11.34	12.84	16.27	17.73
4	7.78	9.49	11.14	13.28	14.86	18.47	20.00
5	9.24	11.07	12.83	15.09	16.75	20.52	22.11
6	10.64	12.59	14.45	16.81	18.55	22.46	24.10
7	12.02	14.07	16.01	18.48	20.28	24.32	26.02
8	13.36	15.51	17.53	20.09	21.95	26.12	27.87
9	14.68	16.92	19.02	21.67	23.59	27.88	29.67
10	15.99	18.31	20.48	23.21	25.19	29.59	31.42
11	17.28	19.68	21.92	24.72	26.76	31.26	33.14
12	18.55	21.03	23.34	26.22	28.30	32.91	34.82
13	19.81	22.36	24.74	27.69	29.82	34.53	36.48
14	21.06	23.68	26.12	29.14	31.32	36.12	38.11
15	22.31	25.00	27.49	30.58	32.80	37.70	39.72
16	23.54	26.30	28.85	32.00	34.27	39.25	41.31
17	24.77	27.59	30.19	33.41	35.72	40.79	42.88
18	25.99	28.87	31.53	34.81	37.16	42.31	44.43
19	27.20	30.14	32.85	36.19	38.58	43.82	45.97
20	28.41	31.41	34.17	37.57	40.00	45.31	47.50



χ^2 -Test auf Übereinstimmung mit einer gegebenen Verteilung

- Gegeben ein Signifikanzniveau α
- Berechne Teststatistik

$$t = \sum_{j=1}^s \frac{(y_j - n \cdot \pi_j)^2}{n \cdot \pi_j}$$

- Bestimme Quantil $\chi_{s-1, 1-\alpha}^2$ der χ^2 -Verteilung mit $s - 1$ Freiheitsgraden
- Falls

$$t \geq \chi_{s-1, 1-\alpha}^2$$

dann lehne H_0 ab



Bemerkungen zum χ^2 -Test

- Der χ^2 -Test verwendet eine Approximation
- Er ist daher nur zulässig, wenn

$$n \cdot \pi_1 \geq 5$$

$$n \cdot \pi_2 \geq 5$$

$$\vdots$$

$$n \cdot \pi_s \geq 5$$

- Die Zahl der Freiheitsgrade beträgt $s - 1$



Beispiel Geschlechterverteilung

- Y_1 gibt die Anzahl der weiblichen und Y_2 die Anzahl der männlichen Erstsemester an
- Beobachtet wurden $y_1 = 266$ und $y_2 = 166$
- Unter H_0 hätte man erwartet $E(Y_1) = 432 \cdot 60.9 = 263.1$ und $E(Y_2) = 432 \cdot 0.391 = 168.9$



Beispiel Geschlechterverteilung

j	beobachtet y_j	erwartet $n \cdot \pi_j$	$\frac{(y_j - n \cdot \pi_j)^2}{n \cdot \pi_j}$
1	266	263.1	0.0320
2	166	168.9	0.0498
Σ	432	432.0	0.0818

Der Wert der Teststatistik ist $t = 0.0818$. Das Quantil ist $\chi_{1,0.95} = 3.84$.

H_0 kann nicht abgelehnt werden.



Beispiel Geburten pro Wochentag

In einem amerikanischen Krankenhaus wurden im Jahr 1999 die folgenden Anzahlen an Geburten beobachtet

Wochentag	Tage im Jahr	Anzahl Geburten
Montag	52	41
Dienstag	52	63
Mittwoch	52	63
Donnerstag	52	47
Freitag	53	56
Samstag	52	47
Sonntag	52	33

Sind diese Zahlen mit der Nullhypothese vereinbar, dass Geburten an allen Wochentagen gleich häufig auftreten? Diese Frage soll zum Signifikanzniveau $\alpha = 0.05$ beantwortet werden.



Beispiel Geburten

Tag	beobachtet	erwartet	
	y_j	$n \cdot \pi_j$	$\frac{(y_j - n \cdot \pi_j)^2}{n \cdot \pi_j}$
Montag	41	49.86	1.57
Dienstag	63	49.86	3.46
Mittwoch	63	49.86	3.46
Donnerstag	47	49.86	0.16
Freitag	56	50.82	0.53
Samstag	47	49.86	0.16
Sonntag	33	49.86	5.70
Σ	350	350	15.04

Der Wert der Teststatistik ist $t = 15.04$. Das Quantil ist $\chi_{6,0.95}^2 = 12.59$.

H_0 kann abgelehnt werden. Die Geburtenverteilung ist merkwürdig.
Der p -Wert ist sogar besser als 0.025

χ^2 -Unabhängigkeitstest

- Der Unabhängigkeitstest überprüft, ob zwei Merkmale stochastisch unabhängig sind.
- Im Beispiel wird Merkmal 1 die Erkrankung an einem Parasiten und Merkmal 2 der Tod durch einen Reiher sein
- Die Zufallsvariablen X_1, \dots, X_n sind unabhängig mit gleicher Verteilung
- Die Zufallsvariablen Y_1, \dots, Y_n sind unabhängig mit gleicher, aber möglicherweise anderer Verteilung
- Die Zufallsvariable X_j beschreibt ein Merkmal mit den Ausprägungen w_1, \dots, w_s
- Die Zufallsvariable Y_j beschreibt ein Merkmal mit den Ausprägungen v_1, \dots, v_r
- Die Nullhypothese ist
$$P(X_1 = w_\ell, Y_1 = v_k) = P(X_1 = w_\ell) \cdot P(Y_1 = v_k)$$



Kontingenztafel

- Der Stichprobenumfang sei n
- Wir stellen die Kontingenztafel auf

	v_1	v_2	\dots	v_r	Σ
w_1	$n_{1,1}$	$n_{1,2}$		$n_{1,r}$	M_1
w_2	$n_{2,1}$	$n_{2,2}$		$n_{2,r}$	M_2
\vdots					
w_s	$n_{s,1}$	$n_{s,2}$		$n_{s,r}$	M_s
Σ	N_1	N_2		N_r	n

- Dabei ist $n_{\ell,k}$ die Anzahl der Individuen mit der Merkmalkombination w_ℓ, v_k

χ^2 -Unabhängigkeitstest, Fortsetzung

- Wir vergleichen die Kontingenztafel mit den erwarteten Werten unter H_0

	v_1	v_2	\dots	v_r	Σ
w_1	$e_{1,1}$	$e_{1,2}$		$e_{1,r}$	M_1
w_2	$e_{2,1}$	$e_{2,2}$		$e_{2,r}$	M_2
\vdots					
w_s	$e_{s,1}$	$e_{s,2}$		$e_{s,r}$	M_s
Σ	N_1	N_2		N_r	n

- Hierbei ist

$$e_{\ell,k} = \frac{M_{\ell} \cdot N_k}{n}$$

die erwartete Anzahl an Individuen mit der Merkmalskombination w_{ℓ}, v_k unter H_0

- Wir vergleichen die beiden Tafeln durch einen χ^2 -Test



χ^2 -Unabhängigkeitstest, Teststatistik

- Die Teststatistik des Unabhängigkeitstests ist

$$t = \sum_{\ell=1}^s \sum_{k=1}^r \frac{(n_{\ell,k} - e_{\ell,k})^2}{e_{\ell,k}}$$

- Die Anzahl der Freiheitsgrade ist $(s - 1) \cdot (r - 1)$
- Zum Signifikanzniveau α wird das Quantil $\chi_{(s-1) \cdot (r-1), 1-\alpha}^2$ benötigt

Vierfeldertest

- Ein χ^2 -Unabhängigkeitstest mit $s = r = 2$ heißt Vierfeldertest
- Als Beispiel wollen wir feststellen, ob das Geschlechterverhältnis in der Mathematik sich von dem in der Biologie unterscheidet
- In Biologie sind 432 Erstsemester eingeschrieben, davon 61.6% weiblich. In Mathematik sind 722 Erstsemester eingeschrieben, davon 55.3% weiblich.

Kontingenztafel zum Beispiel

Kontingenztafel

	Bio	Mathe	Σ
♀	266	399	665
♂	166	323	498
Σ	432	722	1154

erwartete Werte unter H_0

	Bio	Mathe	Σ
♀	248.9	416.1	665.0
♂	186.4	311.6	498.0
Σ	432.0	722.0	1154.0



Tabelle der Differenzen

- Die Tabelle enthält die Terme $\frac{(n_{\ell,k} - e_{\ell,k})^2}{e_{\ell,k}}$

	Bio	Mathe
♀	1.17	0.70
♂	2.23	0.42

- Die Summe über alle Einträge ist

$$t = 4.52$$

- Das Quantil ist $\chi_{1,0.95}^2 = 3.84$
- Der Unterschied ist signifikant



Beispiel: Parasit

- Ein Parasit aus der Klasse der Saugwürmer lebt als erwachsenes Tier in Vögeln
- Die Eier entwickeln sich in Schnecken
- Wenn eine befallene Schnecke von einem Fisch gefressen wird, besiedelt der Saugwurm das Gehirn des Fisches
- Wenn der Fisch von einem Reiher gefressen wird, beginnt der Zyklus von neuem
- Es fällt auf, dass befallene Fische gerne an der Oberfläche schwimmen, wo sie eine leichte Beute für Reiher sind
- Das soll durch einen χ^2 -Unabhängigkeitstest überprüft werden



Daten des Saugwurm-Experiments

Reiheropfer?	gesund	leichter Befall	starker Befall	Σ
ja	1	10	37	48
nein	49	35	9	93
Σ	50	45	46	141

erwartete Werte unter H_0				
Reiheropfer?	gesund	leichter Befall	starker Befall	Σ
ja	17.0	15.3	15.7	48.0
nein	33.0	29.7	30.3	93.0
Σ	50.0	45.0	46.0	141.0

Tabelle der Werte $\frac{(n_{\ell,k} - e_{\ell,k})^2}{e_{\ell,k}}$

Reiheropfer?	gesund	leichter Befall	starker Befall
ja	15.1	1.8	28.9
nein	7.8	0.9	15.0



Teststatistik für das Saugwurm-Experiment

- Das Signifikanzniveau sei $\alpha = 0.01$
- Die Teststatistik ist $t = 69.5$
- Die Anzahl der Freiheitsgrade ist $(3 - 1) \cdot (2 - 1) = 2$
- Das benötigte Quantil ist

$$\chi_{2,0.99}^2 = 9.21$$

- H_0 wird abgelehnt
- Der Parasit beeinflusst das Verhalten des Fisches

Übungspunkte

- Es gab insgesamt 312 Übungspunkte
- Für jeden Studierenden werden die Übungspunkte zusammengerechnet
- **Dann werden 15 Punkte “Korrekturfehlerpauschale” addiert**
- Beschwerden werden nur akzeptiert, wenn die Punktezahl um mehr als diese 15 Punkte steigt. Dann wird der die 15 Punkte übersteigende Anteil hinzugezählt.
- Beispiel: Herr X hat 180 Punkte. Dann werden 195 Punkte in Klausurpunkte umgerechnet (= 4 Klausurpunkte).
- Drei Aufgaben von Herrn X wurden übersehen. Dadurch entgingen Herrn X 18 Punkte. Nach der Beschwerde werden 198 Punkte in Klausurpunkte umgerechnet (= 4 Klausurpunkte)

Übungspunkte, Fortsetzung

- Die Klausurpunkte werden über das Studierendenportal mitgeteilt. Diese Mitteilung erhält, wer mindestens einen Klausurpunkt erworben hat.
- Nachfragen richten Sie bitte an die Korrektoren

Korrektorenzimmer 25.13.U1.31