

Mathematik für Biologen

Prof. Dr. Rüdiger W. Braun

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

Heinrich-Heine-Universität Düsseldorf

23. Januar 2015

- 1 **t-Tests für Erwartungswerte**
 - t-Test für unverbundene Stichproben
 - Der p -Wert
 - Indirekte Vergleiche

- 2 **Konfidenzintervalle**
 - Schätzung eines Konfidenzintervalls mit der 3-sigma-Regel
 - Data Snooping

- 3 **Multiple Vergleiche**
 - Problemstellung
 - Bonferroni-Korrektur

t-Tests für Erwartungswerte

Vergleich zweier Erwartungswerte

Zwei Versuchsreihen liefern Messergebnisse. Der Test soll entscheiden, ob sich diese Ergebnisse signifikant unterscheiden.

Unverbundene Stichproben: Die Messergebnisse werden an verschiedenen Populationen gewonnen.

Beispiel: 9 Maisfelder werden mit einem Bodenbakterium behandelt, 10 weitere bleiben unbehandelt. Bei allen wird der Befall mit Maiszünsler bestimmt.

Verbundene Stichproben: Beide Messungen werden an derselben Population unter bis auf den zu testenden Parameter identischen Bedingungen durchgeführt.

Beispiel: Bei 10 Patienten mit Bluthochdruck wird der Blutdruck vor und nach einer Therapie bestimmt.

Man rechnet einen t -Test für die Differenz der beiden Messungen

- Ein Versuch wird n_1 -mal durchgeführt
- Ein Parameter wird geändert
- Der Versuch wird mit dem geänderten Parameter n_2 -mal **mit einem anderen Kollektiv** wiederholt
- Die Messergebnisse werden verglichen
- Da die Stichproben unverbunden sind, ist $n_1 \neq n_2$ möglich

- Gegeben sind Zufallsvariable X_1, \dots, X_{n_1} und Y_1, \dots, Y_{n_2}
- Verteilungsvoraussetzungen sind
 - Alle X_j sind normalverteilt mit unbekanntem Erwartungswert μ_1 und unbekannter Varianz σ^2
 - Alle Y_j sind normalverteilt mit unbekanntem Erwartungswert μ_2 und unbekannter Varianz σ^2
 - Die beiden Varianzen müssen also gleich sein
- Ziel: μ_1 und μ_2 sollen verglichen werden

t-Test für unverbundene Stichproben, Fortsetzung

- x_j und y_j seien Realisierungen
- Bestimme arithmetische Mittelwerte

$$\bar{x} = \frac{1}{n_1} \sum_{j=1}^{n_1} x_j \quad \bar{y} = \frac{1}{n_2} \sum_{j=1}^{n_2} y_j$$

- und Stichprobenstreuungen

$$s_x = \sqrt{\frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (x_j - \bar{x})^2} \quad s_y = \sqrt{\frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (y_j - \bar{y})^2}$$

- $$t = \frac{\bar{x} - \bar{y}}{s_p} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

Unverbundene Stichproben, Fortsetzung

- Das Signifikanzniveau sei α
- Bestimme zugehörige Quantile der t -Verteilung

$$t_{n_1+n_2-2, 1-\alpha/2}$$

beim zweiseitigen Test

$$t_{n_1+n_2-2, 1-\alpha}$$

bei einem einseitigen Test

- Entscheidung

$H_0 = \{\mu_1 = \mu_2\}$: Die Nullhypothese H_0 wird abgelehnt, wenn $|t| > t_{n_1+n_2-2, 1-\alpha/2}$

$H_0 = \{\mu_1 \geq \mu_2\}$: Die Nullhypothese H_0 wird abgelehnt, wenn $t < -t_{n_1+n_2-2, 1-\alpha}$

$H_0 = \{\mu_1 \leq \mu_2\}$: Die Nullhypothese H_0 wird abgelehnt, wenn $t > t_{n_1+n_2-2, 1-\alpha}$

Alternative Formel für die Standardabweichung der gepoolten Stichproben

$$\begin{aligned}
 s_p &= \sqrt{\frac{(n_1 - 1) \cdot s_x^2 + (n_2 - 1) \cdot s_y^2}{n_1 + n_2 - 2}} \\
 &= \sqrt{\frac{1}{n_1 + n_2 - 2} \left(\sum_{j=1}^{n_1} (x_j - \bar{x})^2 + \sum_{j=1}^{n_2} (y_j - \bar{y})^2 \right)}
 \end{aligned}$$

Beispiel: Maiszünsler

- Der Maiszünsler soll mit einem Bodenbakterium bekämpft werden
- Die folgenden Befallraten (in Larven pro Quadratmeter) wurden beobachtet:

Unbehandelte Felder

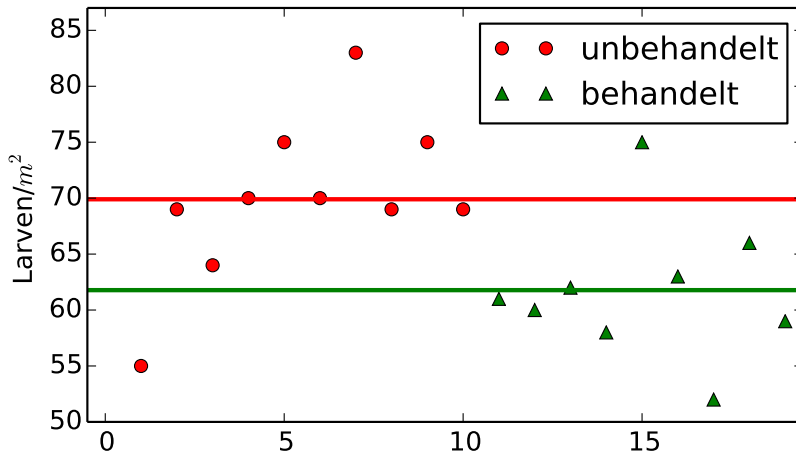
55 | 69 | 64 | 70 | 75 | 70 | 83 | 69 | 75 | 69

Behandelte Felder

61 | 60 | 62 | 58 | 75 | 63 | 52 | 66 | 59

- $n_1 = 10$, $n_2 = 9$
- Da die Stichproben unverbunden sind, sind unterschiedliche Umfänge der beiden Stichproben zulässig

Datensatz des Beispiels



Maiszünsler, Fortsetzung

- Zum Signifikanzniveau $\alpha = 0.05$ soll nachgewiesen werden, dass die Behandlung mit einem Bodenbakterium den Befall mit Maiszünsler verringert
- X_1, \dots, X_{10} sind die Werte für die unbehandelten Felder, Y_1, \dots, Y_9 die Werte für die behandelten
- μ_1 ist der Erwartungswert der X_i , μ_2 der Erwartungswert der Y_i
- Die Nullhypothese ist, dass die Behandlung keinen Nutzen bringt, also

$$H_0 = \{\mu_1 \leq \mu_2\}$$

Maiszünsler, Fortsetzung

- Benötigtes Quantil der t -Verteilung

$$t_{17, 0.95} = 1.740$$

- Die arithmetischen Mittel und die Stichprobenstreuungen betragen

$$\bar{x} = 69.00$$

$$s_x = 7.972$$

$$\bar{y} = 61.78$$

$$s_y = 6.280$$

- Die Streuungen der X_i und der Y_i müssen gleich sein. Wir gehen mal davon aus, dass diese Voraussetzung erfüllt ist. (Auch dafür gibt es einen Test.)
- Die Standardabweichung der gepoolten Stichproben beträgt

$$s_p = 7.226$$

Maiszünsler, Fortsetzung

- Damit kann man die Teststatistik ausrechnen

$$t = 2.175$$

- Das ist größer als $t_{17,0.95} = 1.740$, also wird die Nullhypothese abgelehnt
- Die Behandlung mit dem Bodenbakterium verringert den Befall mit Maiszünsler

Quantile der t -Verteilung

f	90%	95%	97.5%	99%	99.5%	99.9%
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
∞	1.282	1.645	1.960	2.326	2.576	3.090

Der p -Wert

- Der p -Wert beantwortet die Frage

Wie knapp wurde das vorgeschriebene Signifikanzniveau eingehalten bzw. verfehlt?

- Wenn α das Signifikanzniveau ist

$p \leq \alpha$: Nullhypothese wird abgelehnt

$p > \alpha$: Nullhypothese wird beibehalten

- Da der Test in der Regel durchgeführt wird, um die Nullhypothese abzulehnen, ist ein kleiner p -Wert günstig

Der p -Wert

- Der p -Wert ist das kleinste Signifikanzniveau, zu dem die Daten den Test noch bestehen würden.
- Softwarepakete geben den p -Wert aus. Ist er kleiner als das geforderte Signifikanzniveau, so wird H_0 abgelehnt, ansonsten beibehalten
- Wenn man für einen t -Test einen p -Wert bestimmen will, benötigt man eine Tabelle der Werte der t -Verteilung für die entsprechende Anzahl an Freiheitsgraden
- Freies Software-Paket: <http://www.r-project.org/>

Tabelle der t -Verteilung mit 17 Freiheitsgraden, links

u	0.00	0.01	0.02	0.03	0.04
0.0	0,500000	0,503931	0,507862	0,511792	0,515720
0.1	,539243	,543151	,547055	,550954	,554847
0.2	,578073	,581918	,585754	,589582	,593400
0.3	,616093	,619836	,623566	,627284	,630989
0.4	,652931	,656537	,660127	,663702	,667260
0.5	,688257	,691696	,695115	,698517	,701900
0.6	,721790	,725035	,728260	,731464	,734648
0.7	,753302	,756335	,759346	,762335	,765301
0.8	,782626	,785433	,788218	,790979	,793717
0.9	,809654	,812228	,814778	,817305	,819808
1.0	,834334	,836672	,838987	,841279	,843547
1.1	,856667	,858773	,860856	,862916	,864953
1.2	,876702	,878582	,880440	,882276	,884090
1.3	,894524	,896189	,897833	,899456	,901059
1.4	,910251	,911713	,913156	,914580	,915985
1.5	,924021	,925297	,926554	,927794	,929017
1.6	,935992	,937096	,938184	,939256	,940312
1.7	,946324	,947273	,948208	,949129	,950035
1.8	,955184	,955995	,956794	,957579	,958352
1.9	,962734	,963423	,964101	,964767	,965423
2.0	,969131	,969713	,970285	,970847	,971399
2.1	,974520	,975009	,975489	,975961	,976425

Tabelle der t -Verteilung mit 17 Freiheitsgraden, rechts

u	0.05	0.06	0.07	0.08	0.09
0.0	0,519647	0,523572	0,527495	0,531414	0,535330
0.1	,558735	,562617	,566492	,570360	,574220
0.2	,597208	,601007	,604795	,608572	,612338
0.3	,634681	,638360	,642024	,645674	,649310
0.4	,670802	,674328	,677836	,681327	,684801
0.5	,705263	,708608	,711933	,715239	,718524
0.6	,737810	,740952	,744072	,747170	,750247
0.7	,768246	,771167	,774066	,776942	,779796
0.8	,796431	,799123	,801791	,804435	,807056
0.9	,822288	,824744	,827177	,829586	,831972
1.0	,845792	,848013	,850212	,852387	,854538
1.1	,866967	,868959	,870928	,872875	,874800
1.2	,885883	,887654	,889403	,891131	,892838
1.3	,902641	,904203	,905745	,907267	,908768
1.4	,917371	,918738	,920087	,921416	,922728
1.5	,930222	,931409	,932580	,933734	,934871
1.6	,941352	,942377	,943386	,944380	,945360
1.7	,950928	,951806	,952671	,953522	,954360
1.8	,959113	,959861	,960597	,961321	,962034
1.9	,966067	,966701	,967324	,967937	,968539
2.0	,971943	,972476	,973001	,973516	,974023
2.1	,976880	,977327	,977766	,978198	,978622

Bestimmung des p -Werts

- Bestimmung des p -Werts am Beispiel eines einseitigen oberen Tests $H_0 : \mu \leq \mu_0$ zum Signifikanzniveau α
- Man bestimmt den Wert t der Teststatistik
- H_0 wird abgelehnt, wenn $t \geq t_{n-1, 1-\alpha}$
- Der p -Wert ist das beste α , zu dem noch abgelehnt werden kann
- Also $t = t_{n-1, 1-p}$
- Wir bezeichnen mit F die Verteilungsfunktion der t -Verteilung mit $n - 1$ Freiheitsgraden
- $q \mapsto t_{n-1, q}$ ist die Umkehrfunktion zu F
- Daher $F(t_{n-1, 1-p}) = 1 - p$
- Also $F(t) = 1 - p$ und

$$p = 1 - F(t)$$

Der p -Wert, Fortsetzung

$$H_0 = \{\mu = \mu_0\}: p = 2 \cdot (1 - F(|t|))$$

$$H_0 = \{\mu \geq \mu_0\}: p = 1 - F(-t)$$

$$H_0 = \{\mu \leq \mu_0\}: p = 1 - F(t)$$

p -Wert im Beispiel Maiszünsler

- 17 Freiheitsgrade
- F ist die Verteilungsfunktion der t -Verteilung mit 17 Freiheitsgraden
- $t = 2.175$ und $F(2.175) = 0.978$. Wir erhalten

$$1 - p = 0.978$$

- Der p -Wert ist 2.2%

Beispiel: Babyfotos

- Experiment zur Überprüfung, ob Babies ihrem Vater ähnlicher sind als ihrer Mutter
- Foto jeweils eines Babies wurde zusammen mit Fotos von drei Männern, darunter der Vater, und drei Frauen, darunter die Mutter, einer Anzahl von Auswertern vorgelegt, die Vater und Mutter erraten sollen
- Für jedes Baby wurde sowohl der Anteil der korrekten Antworten für die Mutter und als auch für den Vater bestimmt
- Es sei μ_1 die Wahrscheinlichkeit, dass die Mutter richtig bestimmt wird, und μ_2 die Wahrscheinlichkeit, dass der Vater richtig bestimmt wird
- Getestet werden soll $H_0 = \{\mu_1 \geq \mu_2\}$ zum Signifikanzniveau $\alpha = 0.05$
- Der Stichprobenumfang war $n = 24$

Babyfotos: Daten

Die Tabelle zeigt für jedes Baby den Anteil der Auswerter, die Mutter bzw. Vater korrekt bestimmt haben

Baby	1	2	3	4	5	6	7	8
Mutter	0.34	0.16	0.29	0.35	0.33	0.41	0.35	0.29
Vater	0.26	0.55	0.21	0.54	0.46	0.28	0.70	0.46
Baby	9	10	11	12	13	14	14	16
Mutter	0.46	0.54	0.55	0.55	0.13	0.37	0.23	0.35
Vater	0.26	0.42	0.47	0.41	0.42	0.39	0.29	0.47
Baby	17	18	19	20	21	22	23	24
Mutter	0.28	0.42	0.30	0.52	0.34	0.40	0.53	0.33
Vater	0.35	0.48	0.47	0.38	0.15	0.42	0.44	0.30

Babyfotos: Tests auf $\mu \leq 0.33$

- Wir testen zuerst, ob die Wahrscheinlichkeit, die richtigen Eltern zu erkennen, größer als 33% ist
- $H_0 = \{\mu \leq \mu_0\}$ für $\mu_0 = 0.33$
- Zuerst für die Mütter (t ist die Teststatistik)

$$\bar{x} = 0.3674, \quad s_x = 0.1161, \quad t = \frac{\bar{x} - \mu_0}{s_x} \sqrt{n} = 1.441$$

- Das benötigte Quantil ist $t_{23,0.95} = 1.714$
- H_0 kann nicht abgelehnt werden: Es ist nicht gezeigt, dass Babies ihren Müttern überhaupt ähnlich sind

Babyfotos: Tests auf $\mu \leq 0.33$

- Jetzt dasselbe für die Väter



$$\bar{y} = 0.3997, \quad s_y = 0.1224, \quad t = 2.658$$

- H_0 kann abgelehnt werden: Es ist gezeigt, dass Babies ihren Vätern ähnlich sind

Babyfotos: Direkter Vergleich

- Unverbundene Stichproben
- Standardabweichung der gepoolten Stichproben

$$s_p = \sqrt{\frac{(n_1 - 1) \cdot s_x^2 + (n_2 - 1) \cdot s_y^2}{n_1 + n_2 - 2}} = 0.1193$$

- Teststatistik für $n_1 = n_2 = n$

$$t = \frac{\bar{x} - \bar{y}}{s_p} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} = -0.9175$$

- Das benötigte Quantil ist

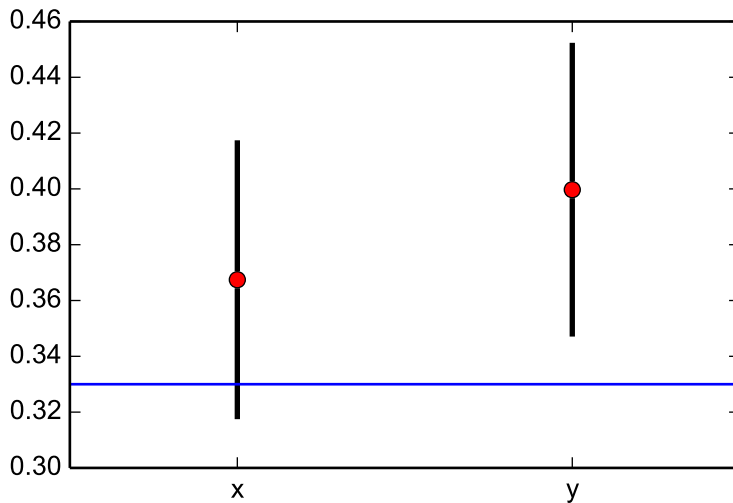
$$t_{46, 0.95} = 1.679$$

- H_0 kann nicht abgelehnt werden
- Wie kann das sein?

Babyfotos: Erklärung des scheinbaren Widerspruchs

- Nicht-Ablehnung von H_0 ist kein Argument für die Gültigkeit von H_0
- Das Bild auf der folgenden Folie zeigt die (zweiseitigen) Konfidenzintervalle für die Erkennungsraten bei Vätern und Müttern
- Diese überlappen: Keine Trennung der Erkennungsraten

Babyfotos: Konfidenzintervalle



Schätzung eines Konfidenzintervalls mittels der 3σ -Regel

- X_1, \dots, X_n seien normalverteilt mit unbekanntem μ
- Für μ soll ein Konfidenzintervall zum Konfidenzniveau 95% geschätzt werden.
- Das ist ein Intervall, welches den wahren Wert mit 95%-tiger Wahrscheinlichkeit enthält
- Den Mittelpunkt des Konfidenzintervalls bildet das arithmetische Mittel

$$\bar{Y} = \frac{1}{n}(X_1 + \dots + X_n)$$

Die 3σ -Regel

- X_1, \dots, X_n unabhängig, alle mit derselben Verteilung
- $\mu = E(X_1) = \dots = E(X_n)$ und $\sigma^2 = \text{Var}(X_1) = \dots = \text{Var}(X_n)$
- Arithmetisches Mittel der X_j

$$Y = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

- n ausreichend groß, dann näherungsweise

$$P\left(\left|Y - \mu\right| \leq \frac{\sigma}{\sqrt{n}}\right) \geq \frac{2}{3}$$

$$P\left(\left|Y - \mu\right| \leq \frac{2\sigma}{\sqrt{n}}\right) \geq 0.95$$

$$P\left(\left|Y - \mu\right| \leq \frac{3\sigma}{\sqrt{n}}\right) \geq 0.99$$

Schätzung eines Konfidenzintervalls mittels der 3σ -Regel, Fortsetzung

- 3σ -Regel

$$P\left(\mu - \frac{2\sigma}{\sqrt{n}} \leq \bar{Y} \leq \mu + \frac{2\sigma}{\sqrt{n}}\right) = 0.95$$

- Beachte die Gleichwertigkeit der Ungleichungen

$$\bar{Y} \leq \mu + \frac{2\sigma}{\sqrt{n}} \quad \text{und} \quad \mu \geq \bar{Y} - \frac{2\sigma}{\sqrt{n}}$$

- Damit sieht die 3σ -Regel so aus

$$P\left(\bar{Y} - \frac{2\sigma}{\sqrt{n}} \leq \mu \leq \bar{Y} + \frac{2\sigma}{\sqrt{n}}\right) = 0.95$$

- Also ist folgendes Intervall ein Konfidenzintervall zum Konfidenzniveau 95%

$$\left[\bar{y} - \frac{2s}{\sqrt{n}}, \quad \bar{y} + \frac{2s}{\sqrt{n}}\right]$$

wobei

$$s = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2}$$

Schätzung eines Konfidenzintervalls mittels der 3σ -Regel, Fortsetzung

- Da s geschätzt wurde, muss bei genauer Rechnung anstelle der 2 das 0.975-Quantil der t -Verteilung mit $(n - 1)$ -Freiheitsgraden verwendet werden

Beispiel: Messungenaugigkeiten

- Eine Schadstoffkonzentration in einem Gewässer wird bestimmt
- Die folgenden Werte (in [ppm]) wurden gemessen

9.80 9.75 9.81 9.80 9.77 9.76 9.77 9.83 9.80 9.75

- Arithmetisches Mittel der Daten

$$\bar{x} = 9.784$$

- Stichprobenstreuung

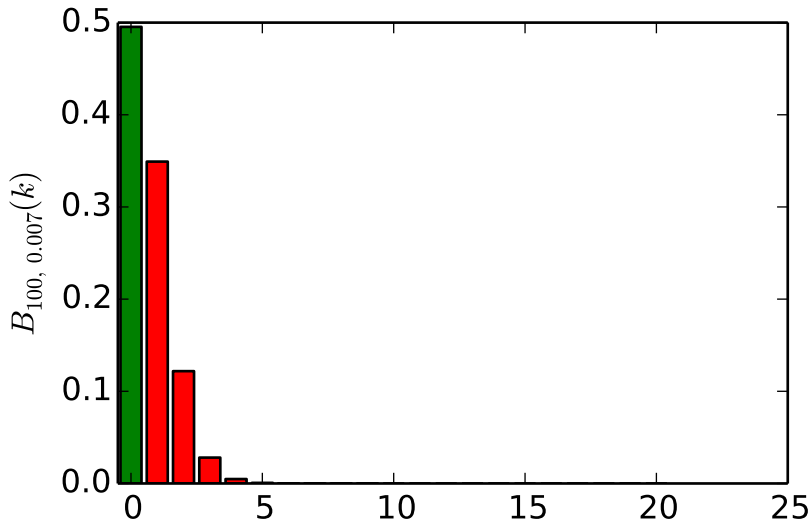
$$s = 0.02757$$

- Konfindenzintervall

$$\left[\bar{x} - \frac{2s}{\sqrt{n}}, \bar{x} + \frac{2s}{\sqrt{n}} \right] = [9.767, 9.801]$$

Data Snooping

- “Snooping” = “Schnüffeln”
- Data Snooping bedeutet, dass man den Test für dieselben Daten rechnet, die man auch für die Formulierung der Hypothese benutzt hat
- Die nächste Folie stammt aus einem schlechten Buch (und wird daher am Netz nicht gezeigt)

Binomialverteilung $B_{100,0.007}$ 

t-Tests für Erwartungswerte

oooooooooooooooooooooooooooo

Konfidenzintervalle

ooooooo

Multiple Vergleiche

oooo

Multiple Vergleiche

Jelly Beans

<http://xkcd.com/882>

Multiple Vergleiche

Möglichkeiten für korrektes Vorgehen

- Man testet die Nullhypothese

$$H_0 = \{ \text{"Grüne Gummibärchen verursachen keine Akne"} \}$$

in einem neuen Experiment zum Signifikanzniveau 5%

- Man rechnet des multiple Experiment mit *Bonferroni-Korrektur*

Bonferroni-Korrektur

- Wenn man simultan n Vergleiche durchführt, dann schreibt die Bonferroni-Korrektur vor, dass man jeden einzelnen Vergleich zum Signifikanzniveau $\frac{\alpha}{n}$ durchführt.
- Im Beispiel der Gummibärchen hätten die Einzelversuche zum Signifikanzniveau $\alpha = 0.0025$ durchgeführt werden müssen.
- Im Beispiel der Schwarzstörche hätte für jeden Einzeltest das Signifikanzniveau $\alpha = \frac{0.05}{100} = 0.0005$ gewählt werden müssen.

False Discovery Rate

- Beispiel Bilddaten: 20 MNR-Aufnahmen von gesunden Gehirnen und 20 MNR-Aufnahmen von erkrankten Gehirnen, wobei die Gehirne auf einen Standardatlas normalisiert werden
- Im einem Bild werden alle Voxel (3D-Pixel) markiert, bei denen der Eisengehalt der Gruppe der Erkrankten signifikant über dem der Gruppe der Gesunden liegt
- Millionen von parallelen Experimenten: Bonferroni-Korrektur sehr schwierig
- Alternative: False Discovery Rate
- FDR von 5% sagt: im Schnitt sind nur 5% aller markierten Voxel falsch
- Zum Vergleich: Ein multipler Test zum Signifikanzniveau 5% sagt: Die Wahrscheinlichkeit, dass auch nur ein Voxel zu Unrecht markiert ist, beträgt höchstens 5%