Mathematik für Biologen

Prof. Dr. Rüdiger W. Braun

Heinrich-Heine-Universität Düsseldorf

18. Januar 2012

Hypothesentests

- Ein Medikament zur Blutdrucksenkung wird mit einem Placebo verglichen
- Ein Test an 10 Hochdruckpatienten soll klären, ob das Medikament den Blutdruck mehr senkt als das Placebo
- Es muss eine Entscheidung zwischen zwei Optionen gefällt werden. Diese bezeichnet man als Nullhypothese und Alternativhypothese

Nullhypothese und Alternativhypothese

- **Nullhypothese** H_0**:** Das ist diejenige Hypothese, deren fälschliche Ablehnung man nach Möglichkeit vermeiden will

  Im Fall des Medikamentens ist die Nullhypothese die Aussage, dass das Placebo mindestens so gut wirkt wie das Medikament

- **Alternativhypothese** H_1**:** Das ist die Alternative zur Nullhypothese

  Im Falle des Medikaments also die Aussage, dass das Medikament besser wirkt als das Placebo

Nullhypothese und Alternativhypothese, Fortsetzung

- Wissenschaft ist konservativ. Wer mit einer neuen Idee kommt, muss zeigen, dass sie besser ist als die alte
- Daher ist die Nullhypothese in der Regel die Annahme, dass die bestehende Theorie bzw. das vorhandene Medikament mindestens so gut ist wie die Neuerungen

Fehler erster und zweiter Art

- **Der Fehler 1. Art** ist die fälschliche Ablehnung der Nullhypothese.
- **Der Fehler 2. Art** ist die fälschliche Annahme der Nullhypothese

Die Priorität liegt auf der Vermeidung des Fehlers 1. Art. Diese Asymmetrie ist ein entscheidendes Merkmal der Testtheorie.

Gestaltung des Versuchs

- 10 Blutdruckpatienten erhalten eine Woche lang das Medikament und eine Woche lang das Placebo. Der Blutdruck am Ende der jeweiligen Behandlung wird notiert. Zwischen beiden Behandlungen vergehen zwei Wochen mit Standard-Therapie.
- Ob jemand zuerst das Medikament oder zuerst das Placebo bekommt, wird ausgelost.
- Für jeden Patienten wird die folgende Differenz gebildet

$$X_j = \text{Blutdruck unter Medikament} - \text{Blutdruck unter Placebo}$$

- Der mittlere Unterschied ist

$$X = \frac{1}{10}(X_1 + X_2 + \cdots + X_{10})$$

- Positive Werte von X sind ein Zeichen für die Unwirksamkeit des Medikaments. Negative ein Zeichen für seine Wirksamkeit.

Beispiel Blutdrucksenker

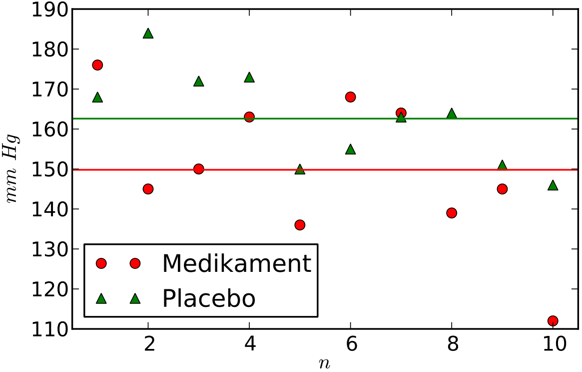
-

Blutdruck [mm hg]	1	2	3	4	5
Medikament	176	145	150	163	136
Placebo	168	184	172	173	150
Differenz X_j	8	-39	-22	-10	-14

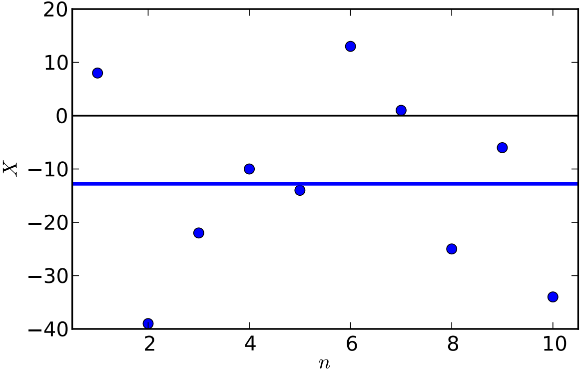
Blutdruck [mm hg]	6	7	8	9	10
Medikament	168	164	139	145	112
Placebo	155	163	164	151	146
Differenz X_j	13	1	-25	-6	-34

- $$X = \frac{1}{10}(8 - 39 - 22 \pm \cdots - 34) = -12.8$$
- Frage: Ist das Zufall?

Blutdrucksenker



Blutdrucksenker, Fortsetzung



Blutdrucksenker, Fortsetzung

- Die Frage

  Ist das Zufall?

  macht keinen Sinn.
- Sinn macht beispielsweise die Frage

  Ist zum Konfidenzniveau 95% sicher, dass die Beobachtung kein Zufall ist?

- Dazu berechnen wir das Konfidenzintervall von X
- Wenn 0 im Konfidenzintervall liegt, ist das Ergbnis zufällig, sonst nicht

Blutdrucksenker, Berechnung des Konfidenzintervalls

- Daten

$$8, -39, -22, -10, -14, 13, 1, -25, -6, -34$$

- Arithmetisches Mittel $\overline{x} = -12.8$
- Stichprobenstreuung $s = 17.36$
- Quantil $t_{9,\,0.975} = 2.262$
- Obere Vertrauensgrenze

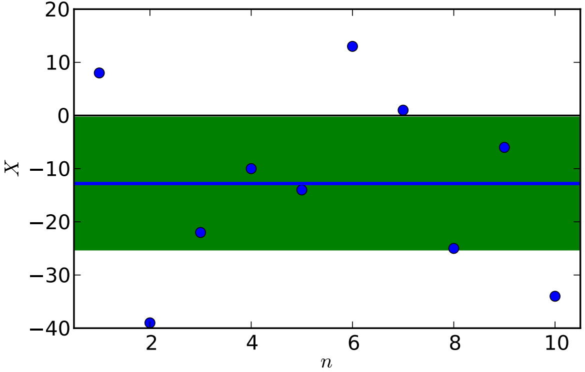
$$g_o = \overline{x} + \frac{s \cdot t_{9,\,0.975}}{\sqrt{n}} = -12.8 + \frac{17.36 \cdot 2.262}{\sqrt{10}} = -0.3785$$

- Untere Vertrauensgrenze ist $g_u = -25.22$

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

Blutdrucksenker, Konfidenzintervall



Der Wert 0 liegt nicht im Konfidenzintervall: Wir können zur Irrtumswahrscheinlichkeit 5% sagen, dass der beobachtete Unterschied in der Wirksamkeit nicht auf Zufall beruht

Ein- und zweiseitige Tests

- Wenn das Medikament deutlich schlechter als das Placebo gewesen wäre, dann wäre 0 ebenfalls nicht im Konfidenzintervall gewesen.
- Man sagt: "Der Test ist zweiseitig."
- Das bedeutet, es wird nur überprüft, ob ein Unterschied vorliegt.
- Will man testen, ob ein Datensatz größere (kleinere) Werte aufweist als der andere, so macht man einen einseitigen Test.
- Da im Beispiel der zweiseitige Test die Wirksamkeit des Medikaments bereits gezeigt hat, verzichten wir auf den einseitigen.

Testverfahren

- Es sei Θ eine Menge von Parametern. Zu jedem $\theta \in \Theta$ gebe es eine Verteilung P_θ
- Die Zufallsvariablen $X_1, \ldots, X_n$ seien unabhängig und alle nach demselben P_θ verteilt. Dieses θ sei unbekannt
- Der Parameterraum sei in zwei Mengen H_0 und H_1 zerlegt. Dabei ist H_0 die Nullhypothese und H_1 die Alternative

Testverfahren, Fortsetzung

Ein *Test* besteht aus einer Vorschrift, die zu jedem möglichen Versuchsausgang festlegt, ob die Nullhypothese H_0 angenommen oder abgelehnt wird.

	H_0 wird angenommen	H_0 wird abgelehnt
$\theta \in H_0$	richtige Entscheidung	Fehler 1. Art
$\theta \in H_1$	Fehler 2. Art	richtige Entscheidung

Interpretation im Beispiel

- Im Beispiel bezeichnen die Zufallsvariablen $X_1, \ldots, X_{10}$ die Unterschiede zwischen Medikament und Placebo für die einzelnen Patienten
- Die X_j sind $N(\mu, \sigma^2)$-verteilt für unbekannte μ und σ.
- Der Parameterraum ist $\Theta = \mathbb{R}$
- Die Nullhypothese ist $H_0 = \{\mu = 0\}$, die Alternative $H_1 = \{\mu \neq 0\}$

Interpretation im Beispiel

- Ich habe den folgenden Test vorgestellt:

  H_0 wird genau dann angenommen, wenn

$$\overline{x} - \frac{s \cdot t_{9,\,0.975}}{\sqrt{10}} \leq 0 \leq \overline{x} + \frac{s \cdot t_{9,\,0.975}}{\sqrt{10}}$$

- Das formuliert man um

  H_0 wird genau dann angenommen, wenn

$$-t_{9,\,0.975} \leq \frac{\overline{x}\sqrt{10}}{s} \leq t_{9,\,0.975}$$

- Die Zahl $\frac{\overline{x}\sqrt{10}}{s}$ heißt *Teststatistik.* Allgemein ist die Teststatistik die Zahl, die mit dem Quantil verglichen werden muss.

Signifikanztests

- Für jedes $\theta \in H_0$ bezeichnet man die Wahrscheinlichkeit $P_\theta($"H_0 wird abgelehnt"$)$ als eine *Fehlerwahrscheinlichkeit erster Art*
- Ein Test heißt *Signifikanztest* zum Niveau α, wenn alle Fehlerwahrscheinlichkeiten erster Art $\leq \alpha$ sind
- Im Beispiel hatte ich einen Signifikanztest zum Niveau $\alpha = 0.05$ angegeben. Übliche Niveaus sind 0.1, 0.05 und 0.01

t-Tests für Erwartungswerte

- $X_1, \ldots, X_n$ bezeichnen unabhängig erhobene, gleichartige Messwerte.
- Verteilungsvoraussetzungen: Alle X_j sind normalverteilt mit unbekanntem Erwartungswert μ und unbekannter Varianz σ^2
- Ziel: μ soll mit einem festen Referenzwert μ_0 verglichen werden.
- x_j seien Realisierungen der X_j
- Bestimme arithmetisches Mittel und Stichprobenstreuung

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \qquad \text{und} \qquad s_x = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(x_j - \overline{x})^2}$$

- Die Teststatistik ist

$$t = \frac{\overline{x} - \mu_0}{s_x}\sqrt{n}$$

t-Tests, Fortsetzung

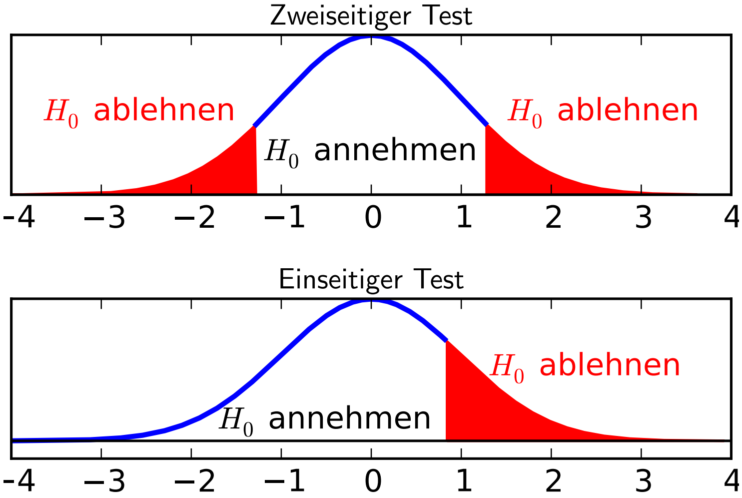
- Das Signifikanzniveau sei α
- Die Quantile der t-Verteilung müssen verwendet werden

$t_{n-1,\,1-\alpha/2}$	beim zweiseitigen Test
$t_{n-1,\,1-\alpha}$	bei einem einseitigen Test

- Entscheidung:

  $H_0 = \{\mu = \mu_0\}$: Die Nullhypothese H_0 wird abgelehnt, wenn
  $$|t| > t_{n-1,\,1-\alpha/2}$$
  $H_0 = \{\mu \geq \mu_0\}$: Die Nullhypothese H_0 wird abgelehnt, wenn
  $$t < -t_{n-1,\,1-\alpha}$$
  $H_0 = \{\mu \leq \mu_0\}$: Die Nullhypothese H_0 wird abgelehnt, wenn
  $$t > t_{n-1,\,1-\alpha}$$

Ein- und zweiseitige Tests



In beiden Fällen ist die rote Fläche gleich groß. Sie bezeichnet den Fehler 1. Art.

Beispiel: Baumschule

- Bauer S. Claus besitzt eine Baumschule
- Gemeinsam mit seinem Mitarbeiter K. Ruprecht beschließt er, die Christbäume auf Feld 13 zu verkaufen, falls ihre mittlere Höhe $1.88m$ übersteigt
- Das soll zum Signifikanzniveau 5% festgestellt werden
- Die Höhe von Christbäumen wird als normalverteilt angenommen; man einigt sich daher auf einen t-Test
- Der Test ist einseitig
- Die Nullhypothese is $H_0 : \mu \leq 1.88$

Baumschule: Fortsetzung

- 10 Bäume werden sorgfältig vermessen

Baum	1	2	3	4	5
Höhe	2.05	2.02	1.86	1.81	1.87

Baum	6	7	8	9	10
Höhe	1.93	1.81	2.00	2.01	1.88

- Dann $\overline{x} = 1.924$ und $s = 0.09021$
- Damit berechnen die beiden die Teststatistik

$$t = \frac{\overline{x} - \mu_0}{s} \cdot \sqrt{n} = \frac{1.924 - 1.88}{0.09021} \cdot \sqrt{10} = 1.542$$

- Benötigt wird das Quantil $t_{9,\,0.95} = 1.833$
- Die Bäume bleiben stehen

Andere Tests

Gaußtest: Wie t-Test, aber bei bekannter Streuung. In der Praxis unrealistisch

U-Test: Wie t-Test, aber ohne Verteilungsannahme

Binomialtests: Zur Überprüfung einer Erfolgswahrscheinlichkeit bei binomialverteilten Zufallsvariablen

Chi-Quadrat-Anpassungstest: Zum Vergleich zweier Verteilungsannahmen

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)