

Mathematik für Biologen

Prof. Dr. Rüdiger W. Braun

Heinrich-Heine Universität Düsseldorf

13. Januar 2010

Termine

- ▶ Letzte Vorlesung am 28.01.2010
- ▶ Letzte Übung am 27.01.2010, und zwar für alle Anfangsbuchstaben

t -Tests für Erwartungswerte

- ▶ $X_1, \dots, X_n$ 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

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

- ▶ Die Teststatistik ist

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

t -Tests, Fortsetzung

- ▶ Das Signifikanzniveau sei α
- ▶ Im Gegensatz zum Gauß-Test müssen nun die Quantile der t -Verteilung 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
 $t > t_{n-1, 1-\alpha}$

Zurück zum Filterbeispiel

- ▶ Sie haben bei einem Händler für Laborartikel eine Partie Filter bestellt. Sie sind nur dann bereit, die Ware abzunehmen, wenn zum Niveau $\alpha = 0.05$ sicher ist, dass der Mittelwert der Porengröße weniger als $3.6\mu m$ beträgt.
- ▶ Sie messen 30 Filter aus und finden folgende Werte

Porengröße in μm	Anzahl der Beobachtungen
3.0	2
3.2	4
3.4	6
3.6	8
3.7	8
4.0	2

- ▶ Die Varianz wird durch die empirische Varianz geschätzt.

t -Test im Filterbeispiel

- ▶ Empirische Varianz

$$s_x^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2 = 0.0623$$

- ▶ Also $s = 0.2496$
- ▶ Der Wert der Teststatistik ist

$$t = \frac{\bar{x} - \mu_0}{s_x} \cdot \sqrt{30} = -1.756$$

- ▶ Das benötigte Quantil ist $t_{29,0.95} = 1.699$
- ▶ Wegen $t < -t_{29,0.95}$ wird die Nullhypothese auch in diesem Setting abgelehnt.

p -Wert

- ▶ Der p -Wert für den t -Test ist $p = 0.045$
- ▶ Der p -Wert für den Gauß-Test ist $p = 0.040$
- ▶ Der p -Wert für den Gauß-Test ist etwas besser, weil der Gauß-Test mit mehr Informationen startet.
- ▶ Tabellen für die t -Verteilungen der verschiedenen Freiheitsgrade liegen meist nicht vor. Daher können p -Werte für die t -Verteilung nur mit entsprechender Software bestimmt werden. (Ich verwende <http://www.scipy.org> für die Vorlesung.)

Verteilungsannahmen

- ▶ Alle bisherigen Tests verwenden Verteilungsannahmen.
- ▶ Entweder waren alle Zufallsvariablen normalverteilt oder binomialverteilt.
- ▶ In der Praxis ist oft nicht klar, ob diese Voraussetzungen erfüllt sind.
- ▶ Der t -Test ist *konservativ*. Das bedeutet, dass er auch dann noch gute Ergebnisse liefert, wenn die Zufallsvariablen nur annähernd normalverteilt sind.

Q-Q-Plot

- ▶ Mit dem Quantil-Quantil-Plot kann man auf graphischem Wege beurteilen, ob Messwerte Realisierungen einer normalverteilten Zufallsvariablen sind.
- ▶ Man trägt dazu auf der x -Achse die Quantile der Standardnormalverteilung und auf der y -Achse die Quantile der Beobachtungsdaten auf.
- ▶ Wenn diese Punkte annähernd auf einer Geraden liegen, sind die Daten näherungsweise normalverteilt, ansonsten nicht.

Q-Q-Plot: Beispiel

- ▶ Wir legen Daten aus dem Beispiel “Blutdrucksenker” zu Grunde

168 184 172 173 150 155 163 164 151 146

- ▶ Zur Bestimmung der Quantile ordnen wir sie an

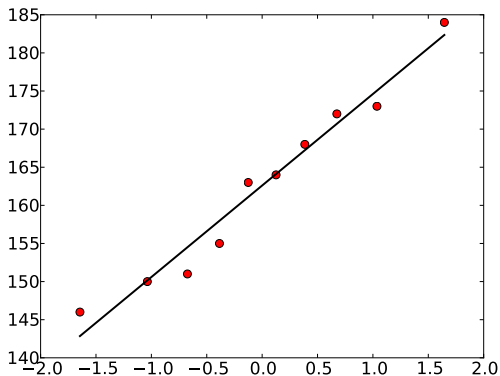
146 150 151 155 163 164 168 172 173 184

- ▶ Wir verteilen $n = 10$ Punkte gleichmäßig über $[0, 1]$

0.05 0.15 0.25 0.35 0.45 0.55 0.65 0.75 0.85 0.95

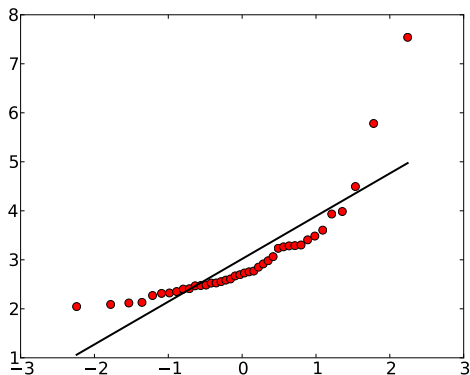
- ▶ Für allgemeines n wählt man in $[0, 1]$ die Punkte $\frac{j-0.5}{n}$,
 $j = 1, \dots, n$
- ▶ Der erste Datenpunkt hat dann die Koordinaten
 $(q_{0.05}, 146) = (-1.645, 146)$, der zweite $(q_{0.15}, 150)$ usw.

Q-Q-Plot



Der Q-Q-Plot der Blutdruckdaten zeigt, dass die Normalverteilungsannahme gerechtfertigt war.

noch ein Q-Q-Plot



Q-Q-Plot von Daten, bei denen die Normalverteilungsannahme nicht gerechtfertigt ist

Abschnitt 3.4

Parameterfreie Tests

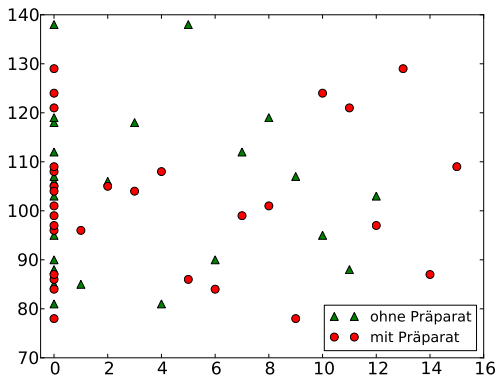
Beispiel: Reaktionszeit unter Alkoholeinfluss

- ▶ Überprüfe, ob die Beeinträchtigung der Reaktionszeit unter Alkoholeinfluss durch Präparat A aufgehoben werden kann.
- ▶ Bestimme dazu die Reaktionszeit von 27 Personen unter Alkoholeinfluss, von denen 15 zusätzlich noch das Präparat genommen haben.
- ▶ Reaktionszeiten sind nicht normalverteilt. Wir benutzen ein Verfahren, das auf Rängen basiert.
- ▶ Seien $x_1, \dots, x_{12}$ die Reaktionszeiten ohne und $y_1, \dots, y_{15}$ mit Präparat A. Wir ordnen alle 27 Daten der Größe nach an, mit dem kleinsten auf Platz 1. Das sind die Ränge.

Beispieldaten

ohne Präparat A		mit Präparat A	
Zeit in [ms]	Rang	Zeit in [ms]	Rang
85	4	96	10
106	17	105	16
118	22	104	15
81	2	108	19
138	27	86	5
90	8	84	3
112	21	99	12
119	23	101	13
107	18	78	1
95	9	124	25
88	7	121	24
103	14	97	11
		129	26
		87	6
		109	20

Grafik



Daten des Beispiels “Reaktionszeit”. Auf der y-Achse kann man die Ränge ablesen.

Bindungen

- ▶ Stimmen mehrere Messwerte überein, so spricht man von *Bindungen*.
- ▶ In diesem Fall wird jedem dieser Messwerte als Rangplatz das arithmetische Mittel der fortlaufend vergebenen Rangplätze zugewiesen.
- ▶ Wir gehen vorerst von dem Fall aus, dass keine Bindungen auftreten.