Grundlagen der Biostatistik und Informatik

Hypothesenprüfungen III.

Zwei Stichproben *t*-Test, F-Test, Varianzanalyse

Stichprobenauswahl, verzerrende Parameter, Trugschlüsse vermeiden

László Smeller

Übersicht der Teste

Verteilung / Stichproben	Normalverteilte Daten	Die Verteilung der Daten ist unbekannt
Eine Stichprobe	Einstichproben t-Test ✓	Vorzeichnentest Wilcoxon Test
Zwei Stichproben	Zweistichproben t-test	Mann-Whittney U-Test
Mehrere Stichproben	ANOVA (Varianzanalyse)	Kruskal-Wallis Test

Zweistichproben t-Test

Vergleich von zwei Stichproben (zwei Populationen)

Warum?
- zwei wesentlich unterschiedliche Populationen (z.B.: Männer und Frauen)
- Vermeidung des Placeboeffektes mit Anwendung einer Kontrollgruppe. (Doppelblindstudie) (Placebo: Pille ohne Wirkstoff)

Wie?
- Randomisierung ist wichtig! (wenn möglich)
- ethische Hinsicht: kein Patient darf unbehandelt bleiben: Vergleich von alte und neue Medikamente oder Behandlungen.

Zweistichproben t-Test: Frage, Nullhypothese

Frage: Ist der zu vergleichende Parameter unterschiedlich in der zwei Populationen?

Mathematisch: Sind die Erwartungswerte in der zwei Populationen unterschiedlich? (oder stammen die zwei Stichproben aus einer Population?)

$$\mu_1 \neq \mu_2$$

Nullhypothese: Es gibt kein Unterschied, die Erwartungswerte sind gleich: $\mu_1 = \mu_2$

Zweistichproben t-Test: Beispiel

Ist eine Schlankmittel wirksam?

Zwei Gruppen:
  Behandlungsgruppe:
    bekommt das neuen „*Wunderschlankmittel*"
  Kontrollgruppe: bekommt Placebo.

Nullhypothese:

- Das „*Wundersclankmittel*" ist unwirksam.
- Erwartungswert des Gewichtes in beiden Gruppen sind gleich: $\mu_{Behandlung} = \mu_{Kontroll}$
- Die Durchschnitte des Gewichtes in den zwei Gruppen unterscheiden sich voneinander nur zufällig.

Zweistichproben t-Test: Beispiel

Körpermasse (kg)	
Behandelte Gruppe	Kontrollgruppe
95	95
91	98
92	96
93	96
92	97
99	99
96	98
	103
	102
Durchschnittswerte (kg)	
94,0	98,2

Auch wenn $\mu_{Behandlung} = \mu_{Kontroll}$ können die Durchschnittwerte unterschiedlich sein:

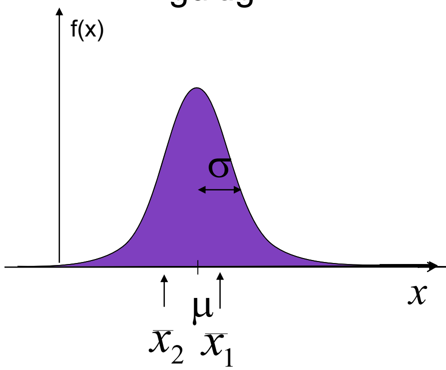
$$\bar{x}_{Behandlung} \neq \bar{x}_{Kontroll}$$

Ist dieser Unterschied zufällig (statistisch), oder ist es die Konsequenz des Unterschiedes zwischen der zwei Populationen (d.h. Konsequenz der Behandlung)?

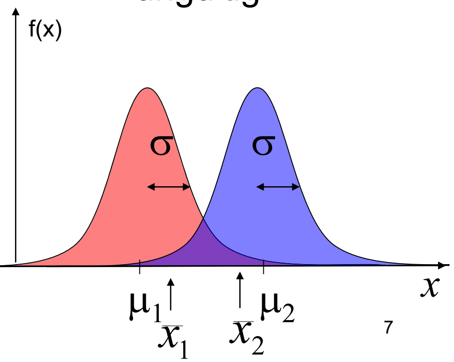
Nehmen wir an:

- Die Daten in beiden Gruppen sind normalverteilt,
- und die Varianzen (Streuungen) sind gleich
  (Bedingungen des Zweistichproben *t*-Testes)

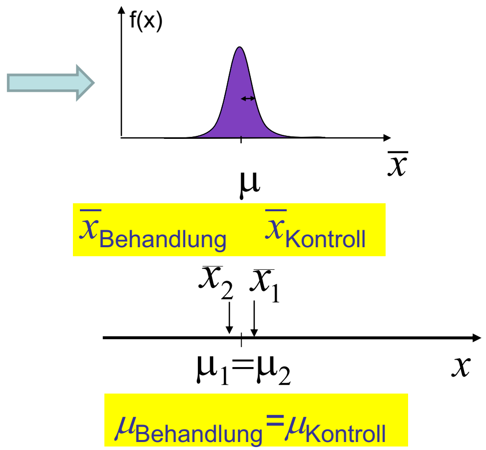
Nullhypothese ist gültig



Nullhypothese ist ungültig



Angenommen dass die Nullhypothese gültig ist



Ist $x_1 - x_2$ zufällig, oder groß genug um die Nullhypothese abzulehnen?

Ein Parameter ist gesucht womit wir die Frage entscheiden können.
(Wie der ***t*** war beim Einstichproben-***t***-Test)

Die Berechnung des Parameters t

Wir brauchen einen Parameter ähnlich zu t beim Einstichprobentest

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{Q_{x1} + Q_{x2}}{n_1 + n_2 - 2}}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$Q_{x1} = \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2$$

$$Q_{x2} = \sum_{i=1}^{n_2} (x_{2i} - \bar{x}_2)^2$$

Ähnlichkeit zum Einstich-probentest: $t = \dfrac{\bar{x}}{s_{\bar{x}}} = \dfrac{\bar{x}}{s_x}\sqrt{n} = \dfrac{\bar{x}}{\sqrt{\dfrac{Q_x}{n-1}}}\sqrt{n}$

Der Parameter t

Bei Gültigkeit der Nullhypothese t folgt eine t–Verteilung mit Freiheitsgrad von n_1+n_2-2.

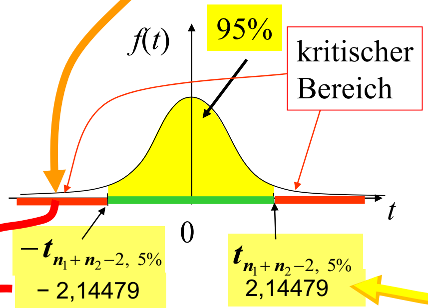
$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{Q_{x1} + Q_{x2}}{n_1 + n_2 - 2}}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

Bedingungen: 1) x muss Normalverteilt sein,
2) die Varianzen in der Gruppen müssen gleich sein.
Entscheidung:
wie bei Einstichproben t-Test

Die Entscheidung

zB: n_1=7, n_2=9 t= -3,02 ⇨ FG: n_1+n_2-2=14

5 %



3,02=|t |> t $_{FG,5\%}$=2,14479

$|t| > t_{n_1+n_2-2,\ 5\%}$ H_0 wird abgelehnt

$|t| < t_{n_1+n_2-2,\ 5\%}$ H_0 wird angenommen

t-Tabelle

FG \ p	0.05	0.02	0.01
2	4.30266	6.96455	9.92499
3	3.18245	4.54071	5.84085
4	2.77645	3.74694	4.60408
5	2.57058	3.36493	4.03212
6	2.44691	3.14267	3.70743
7	2.36462	2.99795	3.49948
8	2.30601	2.89647	3.35538
9	2.26216	2.82143	3.24984
10	2.22814	2.76377	3.16926
11	2.20099	2.71808	3.10582
12	2.17881	2.68099	3.05454
13	2.16037	2.65030	3.01228
14	2.14479	2.62449	2.97685
15	2.13145	2.60248	2.94673
20	2.08596	2.52798	2.84534
50	2.00856	2.40327	2.67779
70	1.99444	2.38080	2.64790
100	1.98397	2.36421	2.62589
unendlich	1.95996	2.32635	2.57583

F-test

Frage:
Sind die Varianzen in zwei Stichproben Gleich?

Nullhypothese: Die Varianzen sind gleich

Parameter: $F = \dfrac{s_1^2}{s_2^2}$ $s_1 > s_2$

Bei der Gültigkeit der Nullhypothese F folgt eine F-Verteilung mit n_1-1 und n_2-1 Freiheitsgrade

F-test

$$F = \frac{s_1^2}{s_2^2}$$

Mit Excel: ***F.Test(Matrix1;Matrix2)***
Gibt einen *p* wert an!
(Wie die T.TEST funktion beim *t*-Test.)

Entscheidung:

$p > \alpha$ Nullhypothese wird mit α Irrtumswahrsch. angenommen, d.h. die Varianzen sind gleich

Erfüllt die Bedingungen des Zweistichproben t-Testes

$p < \alpha$ Nullhypothese wird mit α Irrtumsw. abgelehnt d.h. die Varianzen sind nicht gleich

Wenn die Varianzen (und die Streuungen) ungleich sind

Die Daten können transformiert werden so dass der Zweistichprobentest durchgeführt werden kann.
Auch als Welch-Test bekannt.

(Excel kann diese Transformation ausrechnen).

T.TEST(Matrix1;Matrix2;Seiten;3)

Typ

Zusammenfassug der Excel Funtionen für *t*- und *F*-Teste

Excel Funktion für t-Teste:
(Ein- u. Zweistichproben t-Teste)

T.TEST(Matrix1; Matrix2; Seiten; Typ)

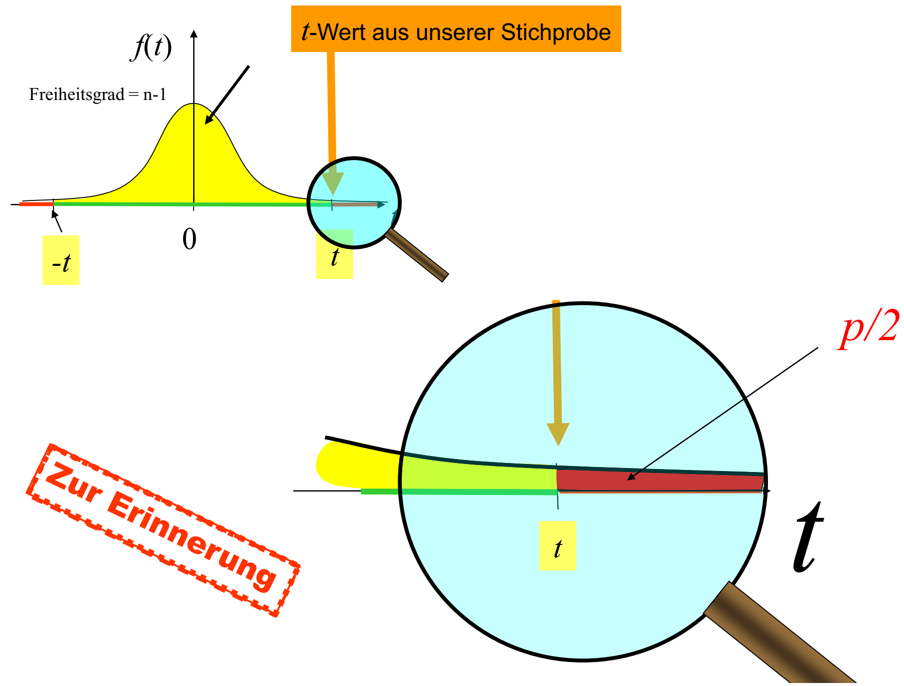
Typ:
1 - gepaart (Eine Stichprobe)
2 - Zwei Stichproben, gleiche Varianz
3 - Zwei Stichproben, ungleiche Varianz

F.TEST(Matrix1; Matrix2)

Diese Funktionen geben *p* an

Entscheidung: p<5% H_0 wird mit 5% Sing.N abgelehnt
p>5% H_0 wird nicht abgelehnt (5% S.N.)

Bemerkung: die F.test() Funktion im Excel gibt p des zweiseitigen Tests



Zusammenfassung: Zweistichproben *t*-Test

Vergleich von zwei Stichproben

Bedingung: Normalverteilung

Prüfung der Varianzen: ***F***-Test

Die Varianzen sind: gleich ungleich

Transformation (oder Typ 3 in Excel)

Berechnung des *t*-Wertes oder des p-Wertes

Ist $t > t_{n-2\,,\,5\%}$ oder $p<\alpha$?

ja → H_0 mit α oder p Irrtumswahrsch. ablehnen

nein → H_0 kann nicht abgelehnt werden (mit α bzw. p Irrtumswahrsch.)

Gepaarte –ungepaarte Teste

Einstichprobentest

Name	T_{vor}	T_{nach}
Anna	39,7	39,2
Benjamin	38,8	38,4
Christina	37,9	38,7
Daniel	39,2	38,7

Gepaarte Daten

Zweistichprobentest

Name	Höhe [cm]
Benjamin	189
Christian	175
Daniel	180
Gabriel	165
Henrik	187

Name	Höhe [cm]
Anna	175
Eva	155
Frederike	167
Judith	180

Ungepaarte Daten

Diese Daten können nicht in Paare geordnet werden

Vergleich der Effektivität* der gepaarten-ungepaarten Teste

Ungepaarte Test
Zweistichproben t-Test

Kein signifikanter Unterschied

Gepaarte Test:
Einstichproben t-Test

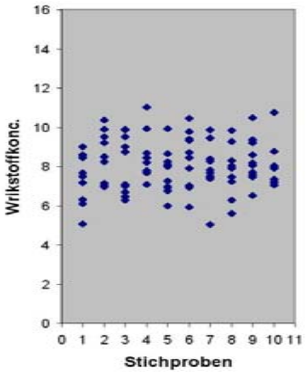
Signifikanter Unterschied

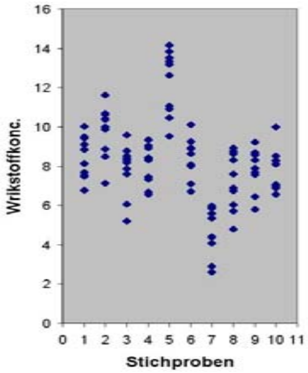
*auch als Güte, Teststärke, Trennscharfe oder Macht genannt

Übersicht der Teste

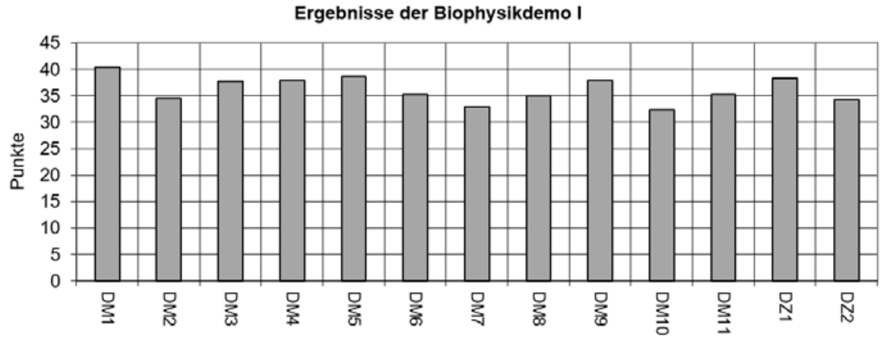
Stichproben \ Verteilung	Normalverteilte Daten	Die Verteilung der Daten ist unbekannt
Eine Stichprobe	Einstichproben t-Test ✓	Vorzeichentest Wilcoxon Test
Zwei Stichproben	Zweistichproben t-test ✓	Mann-Whittney U-Test
Mehrere Stichproben	ANOVA (Varianzanalyse)	Kruskal-Wallis Test

Vergleich von mehreren Stichproben
ANOVA





Vergleich von mehreren Stichproben



Bonferroni - Problem

Vergleich von mehreren Stichproben

Paarweise Vergleichungen:

- Hohe Wahrscheinlichkeit des Fehlers von 1. Art
- z.B.: 10 Stichproben, 45 Vergleichungen
  alle mit 5% Irrtumswahrscheinlichkeit

  Gesamtirrtumsw.:→ $1-(1-0{,}05)^{45}=90{,}0\%$

Lösung (für normalverteilte Daten): ANOVA
(**AN**alysis **O**f **VA**riance)

ANOVA

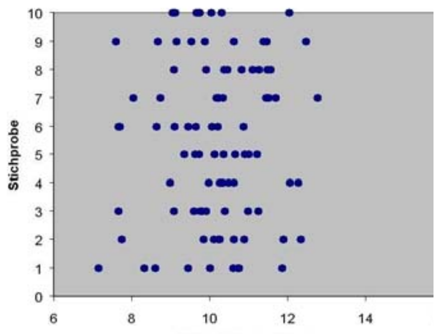
Vorbedingungen:

- Unabhängigkeit der Stichproben
- Normalverteilung
- Gleiche Streuungen

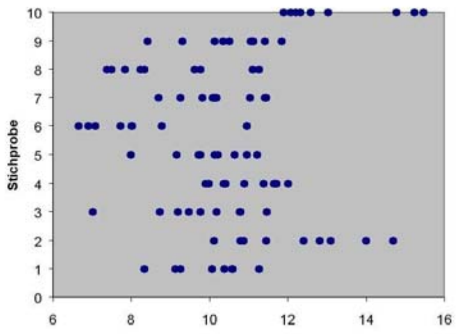
H_0: Alle Stichproben stammen aus der selben Grundgesamtheit

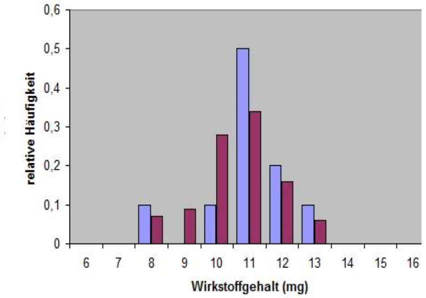
H_1: Mindestens *eine* Stichprobe stammt aus einer anderen Grundgesamtheit

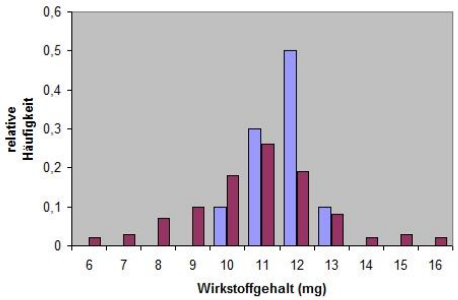
Beispiel 1



Beispiel 2





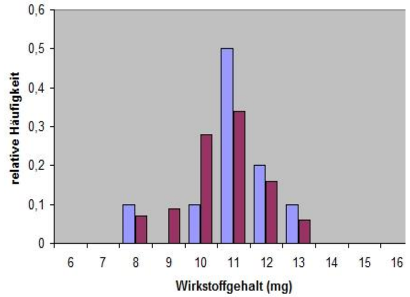


ANOVA

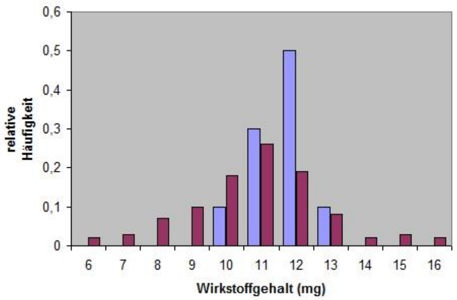
Wenn H_0 gültig ist, sollen die Streuungen *zwischen* den Stichproben und *innerhalb* der Stichproben dieselbe sein.

⇨ Es wird auf einen *F*-Test zurückgeführt.

Beispiel 1



Beispiel 2



ANOVA

$$F = \frac{S_g^2}{S_i^2}$$

↓

F – Test

FG: h-1; N-h

Nicht Signifikant,

keine Unterschiede

H_0: OK

Signifikant,

Es gibt Unterschiede

$\cancel{H_0}$

↓

Post-hoc Test
(Paarweise Vergleichungen
der Stichproben, z.B. Min-Max)

(Paarweise Vergleichungen,
z.B: gegen Kontrolle)

Grundlagen der Biostatistik und Informatik

Woran muss man achten bei Hypothesenprüfungen?

Wie kann man Trugschlüsse vermeiden?

Ablauf des statistischen Testes

- Plannung
  - Fragestellung (Nullhypothese, Alternativhyp., Irrtumsw.)
  - Auswahl der Stichproben
  - Messung
- Mathematische Auswertung
  - Ausrechnung des Testparamters (t, U, z…)
  - Entscheidung (Ablehnen oder behalten von H_0)
- Interpretation
  - Beantwortung der Frage
  - Publikation

Auswahl der Stichproben für Vergleich von zwei Stichproben

- Die Stichproben müssen die Populationen **repräsentieren**.
- Alle bekannte Parametern, (Geschlecht, Alter, sozialer Zustand Wohnumgebungen, Beruf, Ausbildungsgrad…) die die Ergebnisse beeinflussen können, müssen **gleiche Verteilung** haben wie in der Population.
  Bei gepaarten Test gleiche Verteilung in beiden Stichproben.
- Vermeidung des Effektes der versteckten Parametern:
  Randomisierung: Zufällige Zuordnung zu Kontrollgruppe und Behandlungsgruppe (nicht immer möglich).

Experiment-Beobachtung

Geplantes Experiment:

Zuerst Fragestellung und die Stichprobe wird danach gewählt. Man kann an die Repräsentativität achten.

Beobachtung:

Mit Hilfe der Daten die zu Verfügung stehen, wollen wir eine Hypothese überprüfen.

Beobachtung:

Wenn die Daten (und damit die Stichprobe) schon vor der Testfragestellung aufgenommen wurden:

Randomisierung: Zufällige Daten müssen verwerfen werden bis die wichtige Faktoren in beiden Gruppen eine identische Verteilung haben.

zB: Körperhöhe: Gibt es einen Unterschied zwischen Körperhöhen von ungarischen u. deutschen Studenten?

Wir haben 158 Daten: 63D, (22M, 41W) und 95 U (40M 55W)

Probleme bei der Auswahl der Stichprobe

Als Kontrollgruppe nimmt man die Patienten an einer anderen Abteilung des Hospitals. Die zwei Stichproben unterscheiden voneinander dann nicht nur in dem untersuchten Parameter (z.B die angewendete Behandlung) sondern in vielen anderen Faktoren.

Als Kontrollgruppe nimmt man die Patienten die früher behandelt wurden. (Historische Kontroll). Die zwei Stichproben unterscheiden voneinander dann auch z.B. weil die Diangnosemethoden inzwischen entwickelt wurden.

Probleme bei der Auswahl der Stichprobe

Falsche Praktiken:

Die Patienten mit gutem Zustand werden operiert, die alte, und die die noch andere Erkrankungen haben oder es ist fraglich ob sie die Operation überleben werden in Kontrollgruppe eingeordnet. Die Operation wird natürlich danach als sehr effektiv in einem Zweistichprobentest gefunden!

In die Behandlungsgruppe sind nur die Patienten die in der Untersuchung teilgenommen wollen (Freiwilligen-Bias), in der Kontrollgruppe aber alle.

Versteckte Faktoren: Verzerrende Störgrössen (confounder)

Alter

Geschlecht

Rauch- und

Trinkgewohnheiten

Soziale Faktoren

Häufigkeit der Erkrankung,

Gesundheitliche Bewusstheit,…

Wenn nur untersuchungswillige Patienten teilnehmen

…

Placeboeffekt

- Placebo: Pille ohne Wirkstoff
- Es ist wissenschaftlich bewiesen dass Placebopillen haben eine Wirkung: Placeboeffekt
- Die auch haben Nebenwirkungen! Noceboeffekt (nocere)

Um Placeboeffekte zu vermeiden braucht man die Verblindung der Studie

Doppelblindstudie (Double-blind experiment)

Die Patienten werden am Anfang **zufällig** in zwei Gruppen geteilt. Weder die Patienten noch die Ärzte (die die Untersuchung der Patienten auswerten) wissen es, welcher Patient zu der Kontrollgruppe und welcher zu der behandelten Gruppe gehört.

Doppelblindstudie

- Wenn alle Daten gemessen wurden, die den Effekt des neuen Medikaments beschreiben, wird das veröffentlicht welcher Patient in welcher Gruppe war.
- Danach wird der mathematische Teil der Hypothesenprüfung durchgeführt.
- Das ist das objektivste Möglichkeit um den Effekt eines neuen Medikamentes zu prüfen.
- Nicht immer möglich (z.B. Chirurgie).

Zusammenfassung der Untersuchungsarten

- Untesuchung
  - Mit Kontrolle
    - Gleichzeitige
      - Experiment
        - Randomisiert
        - Nicht randomisiert
      - Beobachtung
    - Historische
  - Ohne Kontrolle