Grundlagen der Biostatistik und Informatik

Hypothesenprüfungen II.

Fehlern von erste u. zweite Art,

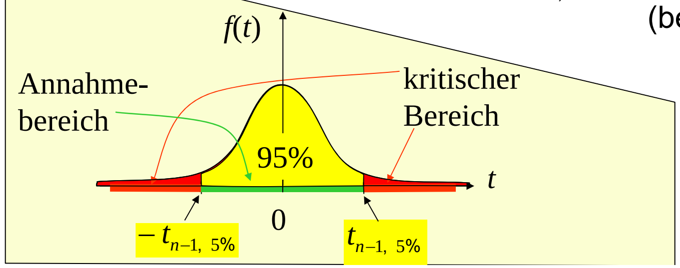
Zwei Stichproben *t*-Test, F-Test,
Bedingungen der Anwendung der *t*-Teste
Varianzanalyse

László Smeller

Widerholung: Grundprinzip der Hypothesenprüfungen

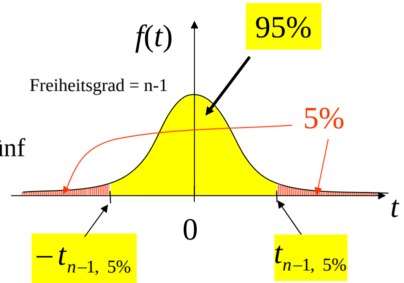
Zur Erinnerung

Zu entscheidende Frage
Indirekter Beweis
Nullhypothese (H_0): nur zufällige Änderungen
mathematisch behandelbar
Ein geeigneter Parameter (zB. *t*)
Bei Gültigkeit der H_0 *t* folgt einer gut bestimmten Verteilung
Zu 95% $|t|<t_{FG,5\%}$ ⇨ Wenn $|t|> t_{FG,5\%}$ ⇨ H_0 wird abgelehnt
bei 5% Irrtumswahrsch.
wenn $|t|< t_{FG,5\%}$ ⇨ H_0 wird beibehalten
(bei 5% Irrtumsw.).



Die Bedeutung des Signifikanzniveaus

Bei einem unwirksamen Medikament
beträgt die Wahrscheinlichkeit dafür,
dass $|t| > t_{n-1,p}$ ist, 5%.
(⇨Bei der Untersuchung von hundert
unwirksamen Pillen werden zufällig fünf
als wirksam gefunden!)
⇩
Fehler erster Art



Fehler von 1. und 2. Art

Fehler erster Art:

Die Nullhypothese wird zufällig abgelehnt werden, obwohl sie richtig ist!
Die Wahrscheinlichkeit des Fehlers erster Art = Signifikanzniveau
zB: Unwirksame Pille als wirksam gefunden
Auch als α-Fehler genannt.

Fehler zweiter Art:

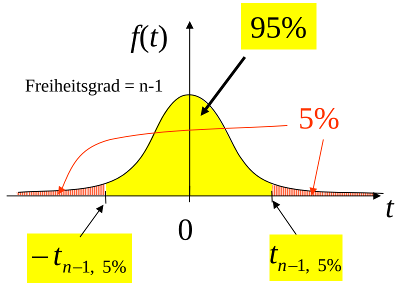
Die Nullhypothese wird nicht abgelehnt, obwohl sie nicht richtig ist.
Wahrscheinlichkeit =?
zB: Die Wirkung einer Pille ist so klein, dass man es aus der Messung nicht beweisen kann. ⇨ *Man braucht noch mehrere Messungen.*
⇨ *So kleine Wirkung ist oft uninteressant*
Auch als β-Fehler genannt.

Einseitige/zweiseitige Teste

Ist es interessant wenn das Medikament die Körpertemperatur erhöht?

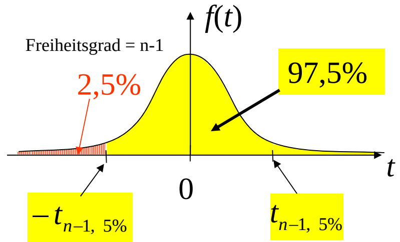
Zweiseitiger Test:

Nullhyp: das Medikament ändert die Körpertemperatur nicht.



Einseitiger Test

Nullhyp: das Medikament erniedrigt die Körpertemperatur nicht.



Verallgemeinerung: $\mu_0 \neq 0$

Beispiel:

Eine Maschine stellt Pillen mit einem nominalen Wirkstoffsgehalt von 20mg her. Man mißt 10 Tabletten und die Wirkstoffgehalte sind (in mg):

20,1 19,8 19,5 17,9 18,8 19,9 18,6 20,3 19,2 19,3

Durchschnitt 19,34 mg, Standardabweichung 0,74 mg, Standardfehler 0,24 mg

Nullhypothese: $\mu_0 = 20$ mg

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$$

$t = -2{,}80$ → H_0 wird .... (Hausaufgabe)

Übersicht der Teste

Stichproben \ Verteilung	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)
- 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: $\mu_{\text{Behandlung}} = \mu_{\text{Kontroll}}$
- 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 $\mu_{\text{Behandlung}} = \mu_{\text{Kontroll}}$ können die Durchschnittwerte unterschiedlich sein:

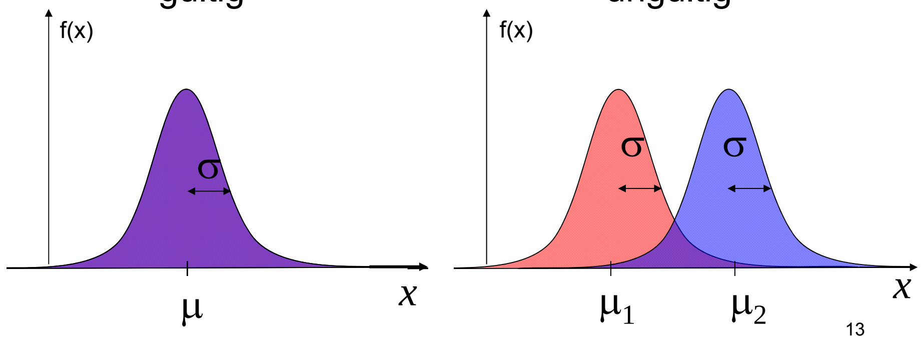
$$\overline{x}_{\text{Behandlung}} \neq \overline{x}_{\text{Kontroll}}$$

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)

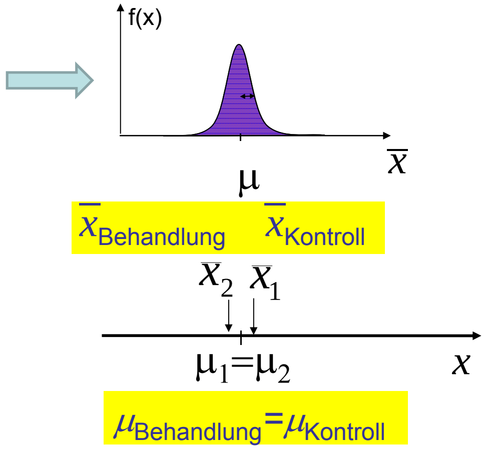
Nullhypothese ist gültig

Nullhypothese ist ungültig



Aus der zwei Populationen nimmt man zwei Stichproben, man kann die zwei Durchschnitts-werte vergleichen.

Angenommen dass die Nullhypothese gültig ist



Ist $\overline{x}_1-\overline{x}_2$ 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

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

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

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

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

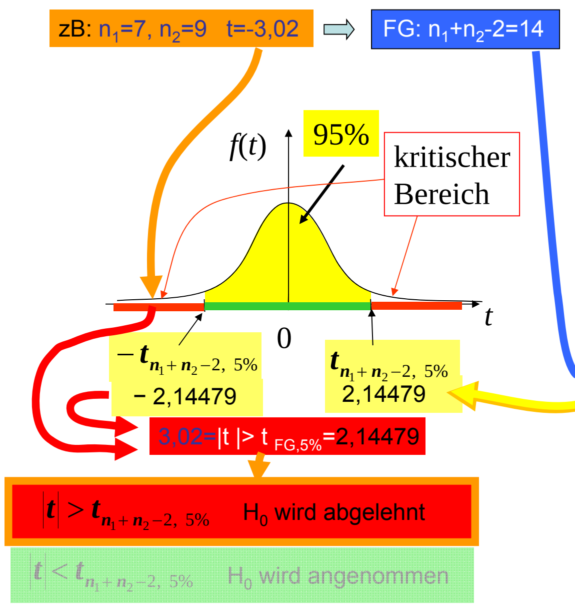
Der Parameter *t*

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

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

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

Die Entscheidung



5 %

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: $\boxed{F = \frac{s_1^2}{s_2^2}}$ $s_1 > s_2$

Bei der Gültigkeit der Nullhypothese F folgt eine F-Verteilung mit n_1-1 und n_2-1 Freiheitsgrade
Bemerkung: Tabelle zum einseitigen Test!
Wir brauchen einen zweiseitigen Test!

F-test

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

Aus der Tabelle

$F < F_{n_1-1,n_2-1;5\%}$ Nullhypothese ist gültig zu 5% Irrtumswahrsch.
d.h. die Varianzen sind gleich

Erfüllt die Bedingungen des Zweistichproben t-Testes

$F > F_{n_1-1,n_2-1;5\%}$ Nullhypothese ist ungültig zu 5% Irrtumsw.
d.h. die Varianzen sind nicht gleich

Wenn die Streuungen sind ungleich

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

Hypothesenprüfungen mit Excel

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

t.test(Matrix1; Matrix2; Seiten; Typ)

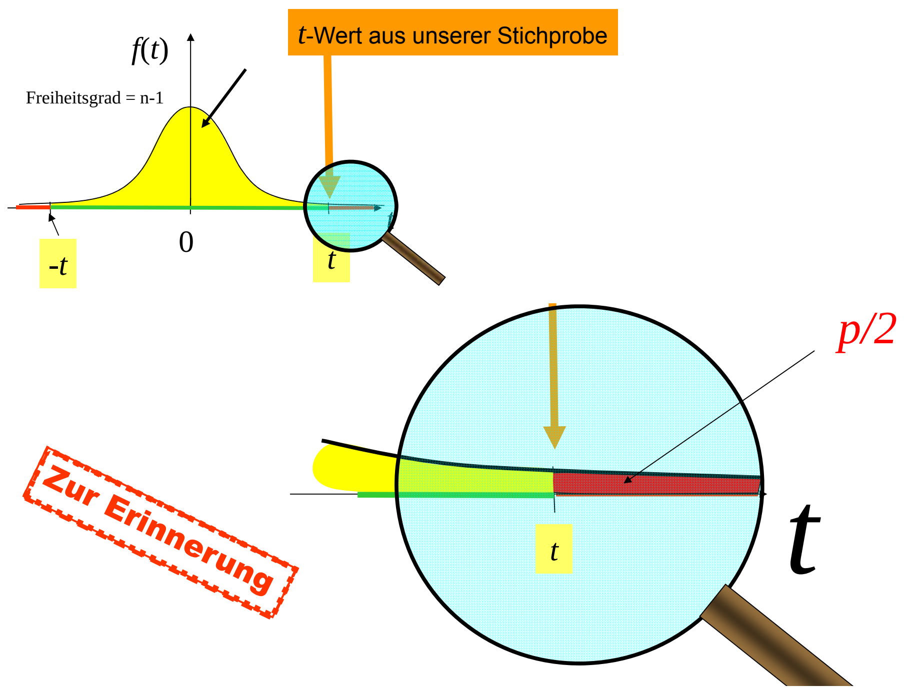
Typ: 1 - gepaart (Eine Stichprobe)
2 - Zwei Stichproben, gleiche Varianz
3 - Zwei Stichproben, ungleiche Varianz

F.test(Reihe1; Reihe2)

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 Populationen durch zwei Stichproben

Bedingung: Normalverteilung mit derselben Varianz

Prüfung der Varianzen: $\boldsymbol{F}$-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<5% ?

ja — H_0 mit 5% oder p Irrtumswahrsch. ablehnen

nein — H_0 kann nicht abgelehnt werden (mit 5% 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

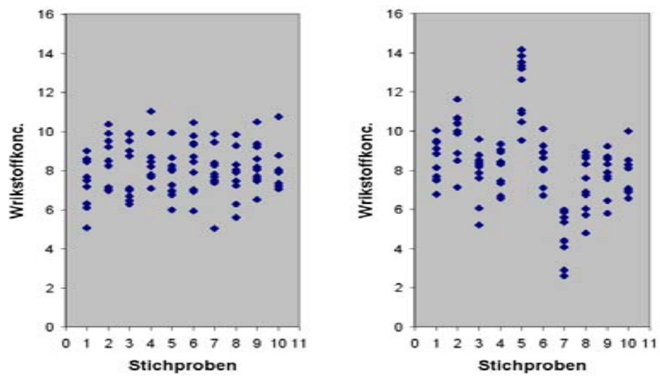
Gepaarte Test:
Einstichproben t-Test

Signifikanter Unterschied

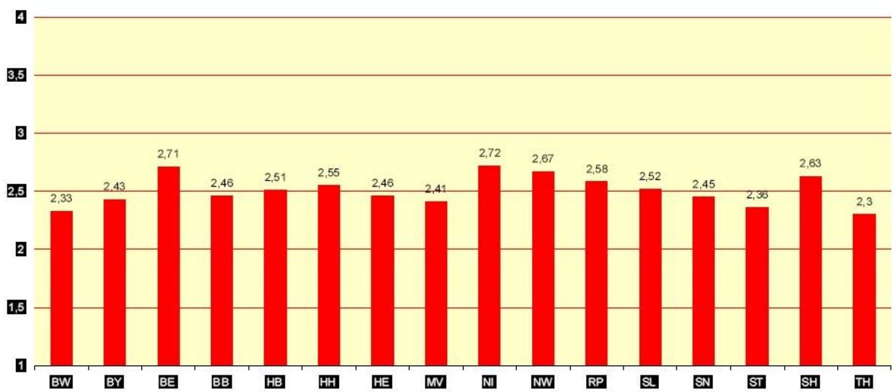
Übersicht der Teste

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

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

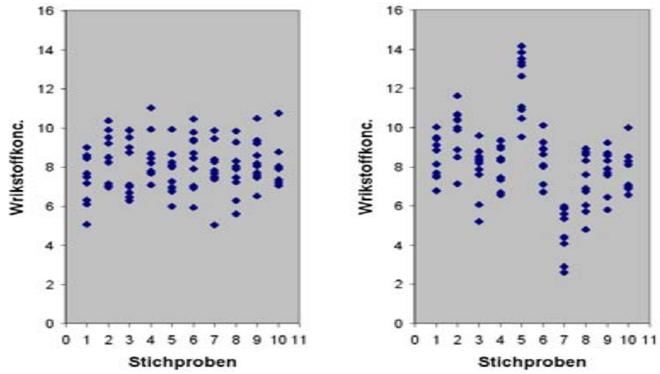
H_0: Alle Stichproben stammen aus der selben Grundgesamtheit

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

ANOVA

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

Beispiel 1.   Beispiel 2.



ANOVA

h: Anzahl der Stichproben

H_0: $\mu_1 = \mu_2 = \ldots \mu_h$

Varianz innerhalb der Stichproben: S_i^2

Varianz zwischen den Stichproben: S_g^2

Wenn $S_i^2 << S_g^2$ → Varianzen sind unterschiedlich → H_0 ablehnen

Wenn $S_i^2 \approx S_g^2$ → Varianzen sind die Schätzungen derselben Varianz → H_0 annehmen

$$F = \frac{S_g^2}{S_i^2}$$ *F* – Test; Einseitig, Freiheitsgrad: *h-1; N-h*

ANOVA

Varianz zwischen den Stichproben:

$$s_g^2 = \frac{\sum_{j=1}^{h} n_j \left(\bar{x}_j - \bar{\bar{x}}\right)^2}{h-1} = \frac{Q_g}{h-1}$$

h: Anzahl der Stichproben

n_j: Anzahl der Elementen in der *j*-ten Stichprobe

$\bar{\bar{x}}$: Durchschnitt von allen Elementen

$\bar{x}_j$: Durchschnitt in der *j*-ten Stichprobe

ANOVA

Varianz innerhalb der Stichproben:

$$s_i^2 = \frac{\sum_{j=1}^{h} Q_j}{N-h} = \frac{\sum_{j=1}^{h}\sum_{i=1}^{n_j}(x_{ij} - x_j)^2}{N-h} = \frac{Q_i}{N-h}$$

h: Anzahl der Stichproben

n_j: Anzahl der Elementen in der *j*-ten Stichprobe

x_{ij}: *i*-ten Element der *j*-ten Stichprobe

$\bar{x}_j$: Durchschnitt in der *j*-ten Stichprobe

N: Gesamte Anzahl der Stichprobenelementen

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

~~H_0~~

Post-hoc Test

(Paarweise Vergleichungen

der Stichproben, z.B. Min-Max)

(Paarweise Vergleichungen,

z.B: gegen Kontrolle)