

Hypothesenprüfungen III.

Zwei Stichproben t -Test, F-Test,
Varianzanalyse
Stichprobenauswahl, verzerrende
Parameter, Trugschlüsse vermeiden

László Smeller

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.

3

Übersicht der Teste

Verteilung Stichproben	Normalverteilte Daten	Die Verteilung der Daten ist unbekannt
Eine Stichprobe	Einstichproben t -Test ✓	Vorzeichenstest Wilcoxon Test
Zwei Stichproben	Zweistichproben t -Test und Welch Test	Mann-Whitney U-Test
Mehrere Stichproben	ANOVA (Varianzanalyse)	Kruskal-Wallis Test

2

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: $\mu_1 = \mu_2$

4

Zweistichproben t-Test: Beispiel

Ist eine Schlankmittel wirksam?

Zwei Gruppen:

Behandlungsgruppe:

bekommt das neue „Wunderschlankmittel“

Kontrollgruppe: bekommt Placebo.

Nullhypothese:

- Das „Wunderschlankmittel“ 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.

5

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 Durchschnittswerte unterschiedlich sein:

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

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

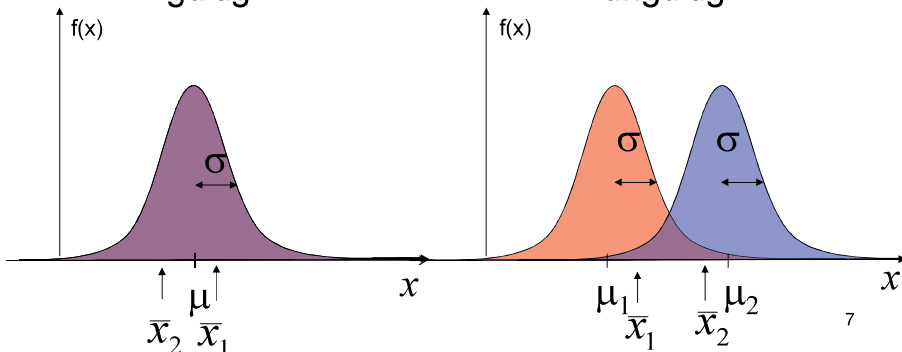
6

Nehmen wir an:

- Die Daten in beiden Gruppen sind normalverteilt

(Bedingung des Zweistichproben-t-Testes)

Nullhypothese ist gültig

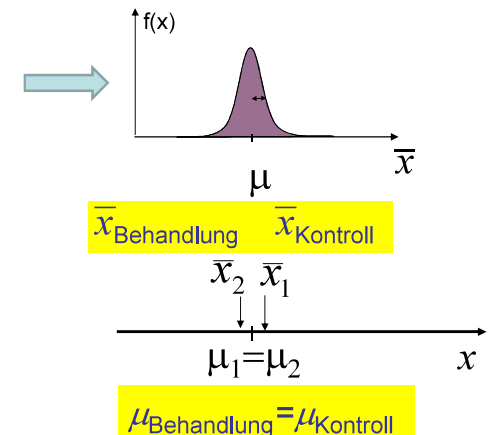


7

Angenommen dass die Nullhypothese gültig ist

Ist $\bar{x}_1 - \bar{x}_2$ 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)



8

Bedingungen:

1. Die Daten in beiden Gruppen sind **normalverteilt**.

Soll erfüllt werden für den Zweistichproben-t-Test und auch für den Welch Test!

2. Die (theoretische) Streuungen in beiden Gruppen sind gleich.

gültig



Zweistichproben-t-Test

T.TEST(Matrix1;Matrix2;Seiten;2)

ungültig



Welch Test

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

Wie entscheidet man zwischen
Zweistichproben-t-Test und Welch Test?

Die Frage ist: Sind die Streuungen Gleich?

Lösung1: Mit F-Test kann man die Streuungen vergleichen

Lösung 2: Man verwendet immer den Welch-Test

Die Berechnung des Parameters t

Der Parameter t wird ähnlich berechnet als beim
Einstichproben-t-Test

Einstichproben-t-Test

$$t = \frac{\bar{x}}{s_{\bar{x}}} = \frac{\bar{x}}{\frac{s}{\sqrt{n}}} \quad FG=n-1$$

Zweistichproben-t-Test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\text{Gesamt}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad FG=n_1+n_2-1$$

Welch-t-Test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad FG=\dots\dots \text{(kompliziert)}$$

10

Zusammenfassung der Excel Funktionen für t- Teste

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

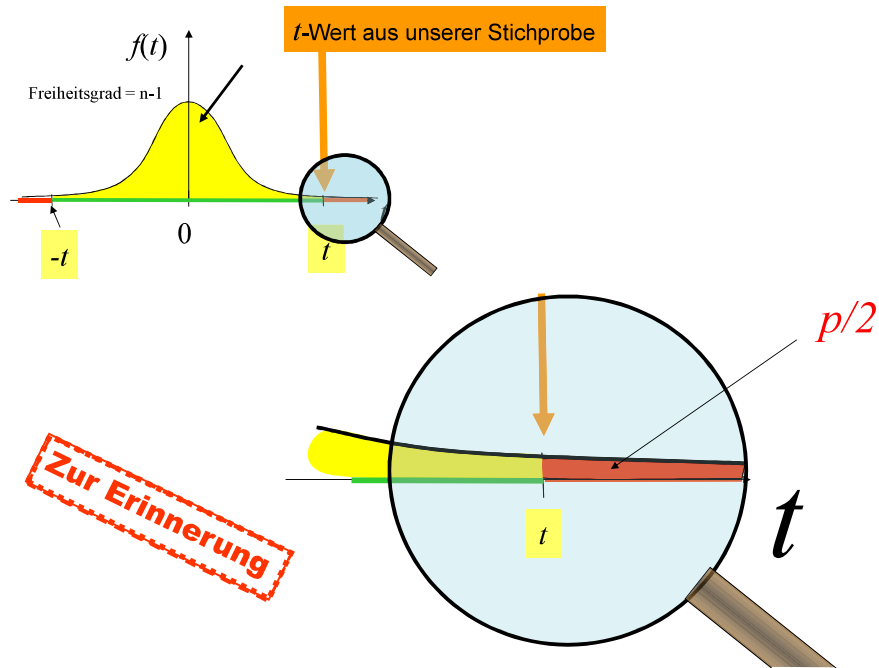
T.TEST(Matrix1; Matrix2; Seiten; Typ)

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

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.)

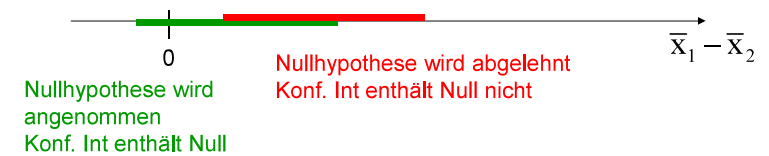
12



Vergleich der Hypothesenprüfungen und Schätzungen bei Zwei Stichproben

Man kann ein Konfidenzintervall für $\bar{x}_1 - \bar{x}_2$ berechnen.

Wenn die Hypothesenprüfung einen signifikanten Effekt zeigt, enthält das entsprechende Konfidenzintervall Null nicht.



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]	Name	Höhe [cm]
Benjamin	189	Anna	175
Christian	175	Eva	155
Daniel	180	Frederike	167
Gabriel	165	Judith	180
Henrik	187		

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



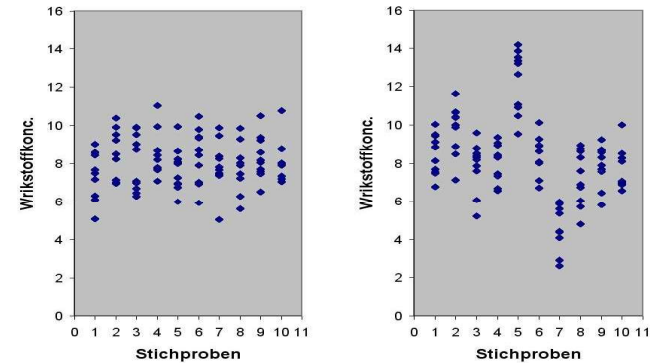
*auch als Güte, Teststärke, Trennscharfe oder Macht genannt

Übersicht der Teste

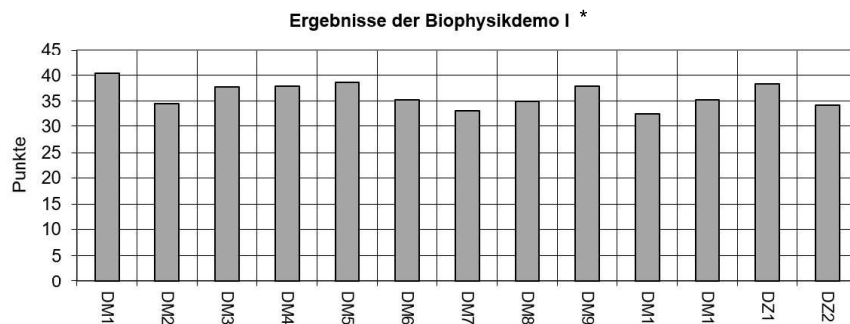
Verteilung Stichproben	Normalverteilte Daten	Die Verteilung der Daten ist unbekannt
Eine Stichprobe	Einstichproben t-Test ✓	Vorzeichenstest Wilcoxon Test
Zwei Stichproben	Zweistichproben t-test ✓	Mann-Whittney U-Test
Mehrere Stichproben	ANOVA (Varianzanalyse)	Kruskal-Wallis Test

17

Vergleich von mehreren Stichproben ANOVA



Vergleich von mehreren Stichproben



*In einem der vorigen Jahren

Bonferroni - Problem

Vergleich von mehreren Stichproben

Paarweise Vergleichen:

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

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

Lösung (für normalverteilte Daten): **ANOVA**
(ANalysis Of VAriance)

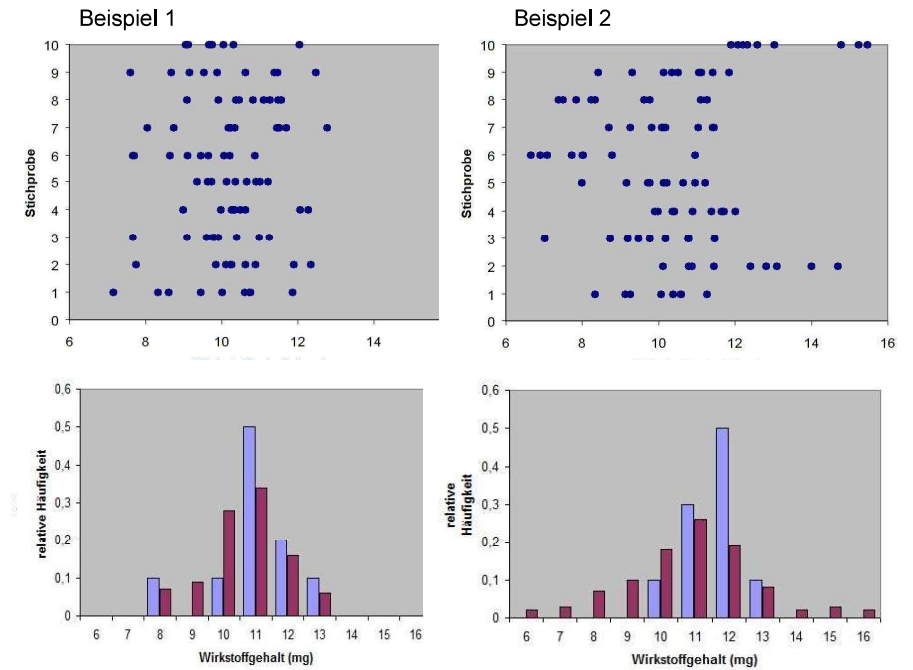
ANOVA

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

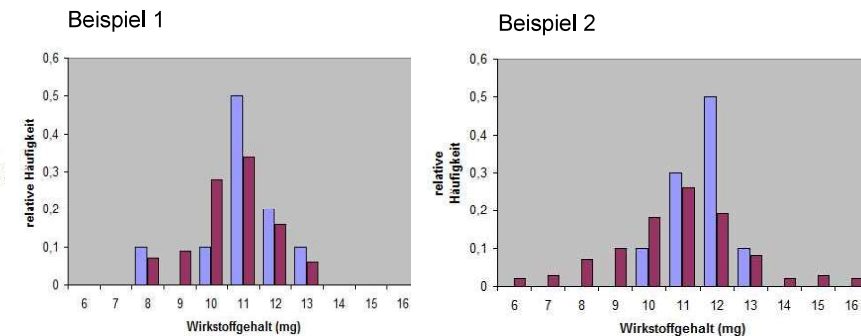


ANOVA

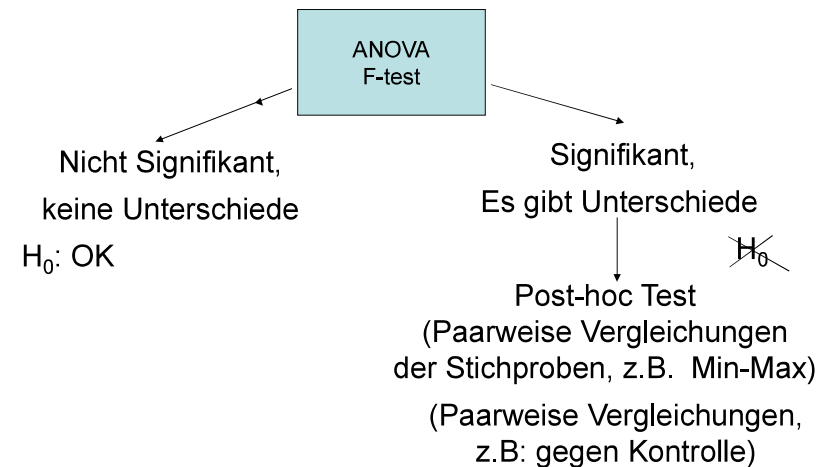
Wenn H_0 gültig ist, sollen die Streuungen *zwischen* den Stichproben und *innerhalb* der Stichproben dieselbe sein.



Die Streuungen werden mit einem F -Test verglichen.



ANOVA




Woran muss man achten bei Hypothesenprüfungen?

Wie kann man Trugschlüsse vermeiden?


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).


• Planung

-  – 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

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 verworfen 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 Diagnosemethoden 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

- 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 die objektivste Möglichkeit um den Effekt eines neuen Medikaments zu prüfen.
- Nicht immer möglich (z.B. Chirurgie).

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.

Zusammenfassung der Untersuchungsarten

