

Statistik III

Regressionsanalyse, Varianzanalyse und Verfahren bei Messwiederholung mit SPSS

Verena Hofmann

Dr. phil. des.
Departement für Sonderpädagogik
Universität Freiburg
Petrus-Kanisius-Gasse 21
CH-1700 Freiburg

++41 (0)26 300 77 51
verena.hofmann@unifr.ch



Organisatorisches zur Veranstaltung

- Veranstaltungen HS vom 26.9.16 – 19.12.16
 - > Auftrag anstelle von Veranstaltung: 24.10.16 (Abwesenheit Dozentin)
- Moodle: <https://moodle2.unifr.ch/>:
 - Kursname: **«Statistik III HS 2016»**; Schlüssel: **L051.0479**
 - Zeitplan und Themen
 - Div. Dokumente
 - Datensätze
 - Übungen
 - Links
- Prüfung: 19.12.2016, schriftlich, 45 Min.

Vorwissen

- **Inhalte aus Statistik I und II:**

- Deskriptive Statistik
- 2 Gruppen vergleichen: T-Test
- Zusammenhang zwischen zwei metrischen (kontinuierlichen) Variablen: Korrelation
- Zusammenhang zwischen zwei nominalen Variablen: Chi²-Test

- **Programm SPSS?**

Literatur

Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7th ed.). Berlin: Springer.

Bühl, A. (2016). *SPSS 23: Einführung in die moderne Datenanalyse* (14th ed.). Hallbergmoos: Pearson Studium.

Rasch, B., Frieze, M., Hofmann, W. J., & Naumann, E. (2014). *Quantitative Methoden 1. Einführung in die Statistik für Psychologen und Sozialwissenschaftler*. (4th ed.). Heidelberg: Springer.

-> Deskriptive Statistik

-> T-Test (unabhängige Stichproben)

-> Korrelation

<http://www.lehrbuch-psychologie.de/quantitative-methoden-1>

Rasch, B., Frieze, M., Hofmann, W. J., & Naumann, E. (2014). *Quantitative Methoden 2. Einführung in die Statistik für Psychologen und Sozialwissenschaftler*. Heidelberg: Springer.

-> (Chi²-Test)

<http://www.lehrbuch-psychologie.de/quantitative-methoden-2>

Programm SPSS

- Einführung in das Programm am 3.10.16
- Demonstration jeweils nach theoretischer Einführung eines Verfahrens
- Zusätzliche Übungen als Hausaufgaben
- Download: <http://student.unifr.ch/support/de/software/spss>
- Installation:

Windows: <http://www3.unifr.ch/dit/faq2/article/etudiant-e-s-installation-de-spss-23-sur-windows/?lang=de>

Mac: <http://www3.unifr.ch/dit/faq2/article/studierende-spss-23-auf-mac-os-x-10-10-installieren/?lang=de>

- Datensätze: ALLBUS (allgemeine Bevölkerungsumfrage der Sozialwissenschaften); Beispieldatensätze SPSS; fiktive Datensätze

Ziele der Veranstaltung

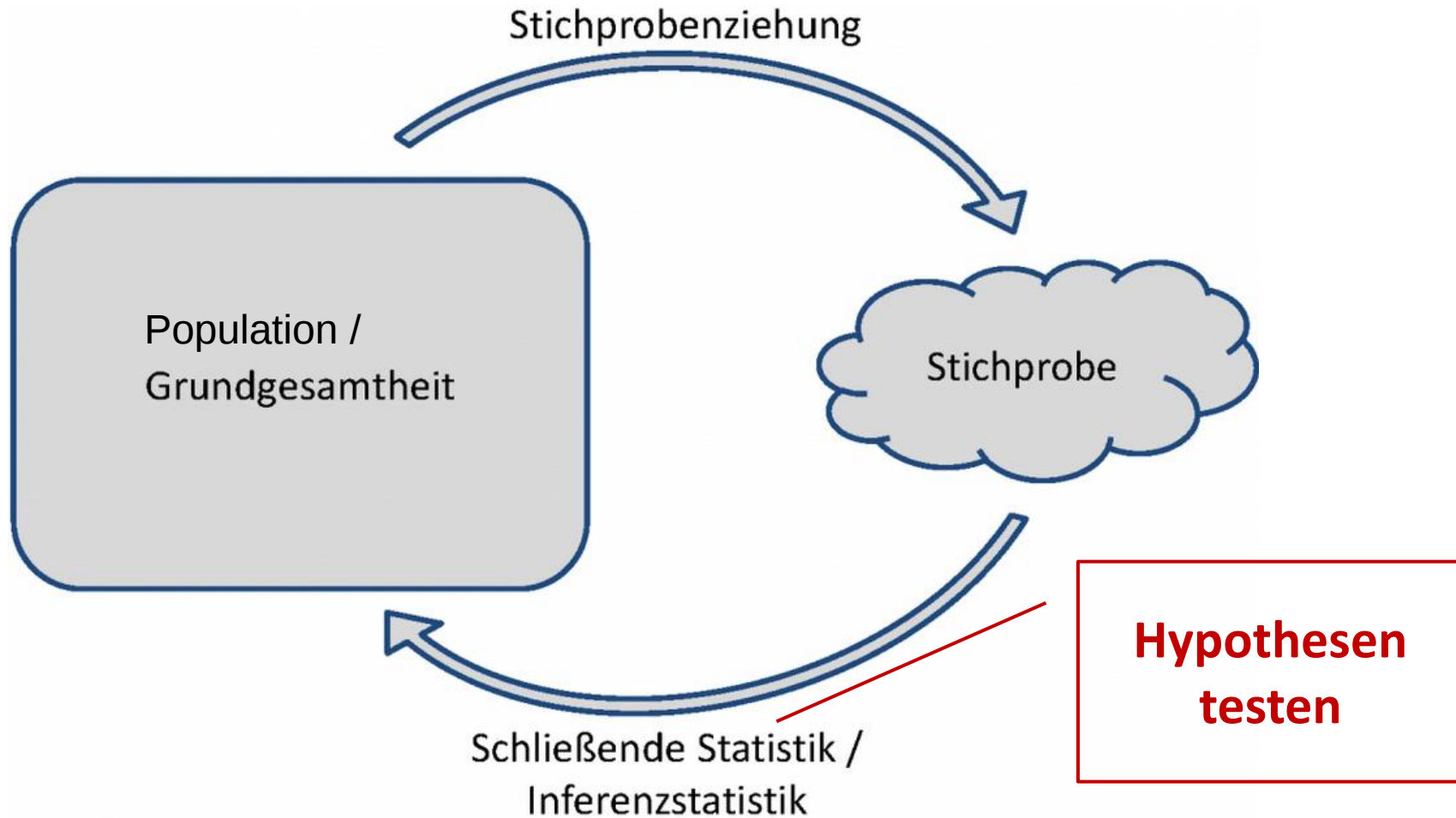
- Wann wende ich welches statistische Verfahren an? (abhängig von Fragestellung und Art der Daten)
- Wie führe ich die Analyse korrekt durch? (mit dem Programm SPSS)
- Wie interpretiere ich die Resultate? (des Programms, aber auch aus der Forschungsliteratur)
- Wie stelle ich die Ergebnisse korrekt dar? (im Text oder tabellarisch)

Inhalte der Veranstaltung

- Repetition zu den Grundprinzipien der Inferenzstatistik
- Die Vorhersage eines Merkmals durch ein (mehrere) andere(s):
Regressionsanalyse
- Vergleich von mehr als zwei Gruppen: Varianzanalyse
- Verfahren bei Messwiederholung

Fragestellung	Skalenniveau AV		Skalenniveau UV	Statistisches Testverfahren
Zusammenhang zwischen zwei Variablen?	Metrisch	↔	Metrisch / Nominal dichotom	Korrelation
Unterschied zwischen zwei Gruppen?	Metrisch		Nominal dichotom	T-Test
Unterschied zwischen zwei oder mehr Gruppen?	Metrisch		Nominal >/= 2 Kategorien	Varianzanalyse (ANOVA)
Zusammenhang zwischen zwei kategorialen Variablen?	Nominal 2 oder mehr Kategorien	↔	Nominal 2 oder mehr Kategorien	Chi ² -Test
Einfluss einer/mehrerer UVs auf eine AV?	Metrisch		Metrisch / Nominal dichotom	Regressionsanalyse
Veränderung eines Merkmals über 2 Messzeitpunkte?	Metrisch		Nominal (2 Zeitpunkte)	T-Test für abhängige Stichproben
Veränderung eines Merkmals über 2 oder mehr Messzeitpunkte? Gruppenspez. Veränderung?	Metrisch		Nominal (>/= 2 Zeitpunkte; >/= 2 Kategorien)	Varianzanalyse (ANOVA) mit Messwiederholung
Diverse	Metrisch		Diverse	Kovarianzanalyse (ANCOVA)

Prinzip der Inferenzstatistik



Prinzip der Inferenzstatistik

- *Populationsparameter* werden auf der Basis von *Stichprobenkennwerten* geschätzt
- Bezogen auf eine Hypothesen heisst das: Kann ein in der Stichprobe gefundener Effekt (Zusammenhangsmass, Differenz) auf die Population generalisiert werden oder ist er rein zufällig zustande gekommen?
- Dies geschieht mittels Signifikanztests

Prinzip der Inferenzstatistik: Zusammenhänge und Unterschiede

- Effekt = Zusammenhang zwischen zwei Merkmalen / Unterschied zwischen Gruppen
- Unterschiedliche Masse, aber vom Prinzip her dasselbe: auch Gruppenunterschiede sind Zusammenhänge zwischen zwei Merkmalen

Beispiel:

«Es gibt einen Unterschied zwischen Jungen und Mädchen bezüglich der Leseleistung»

oder:

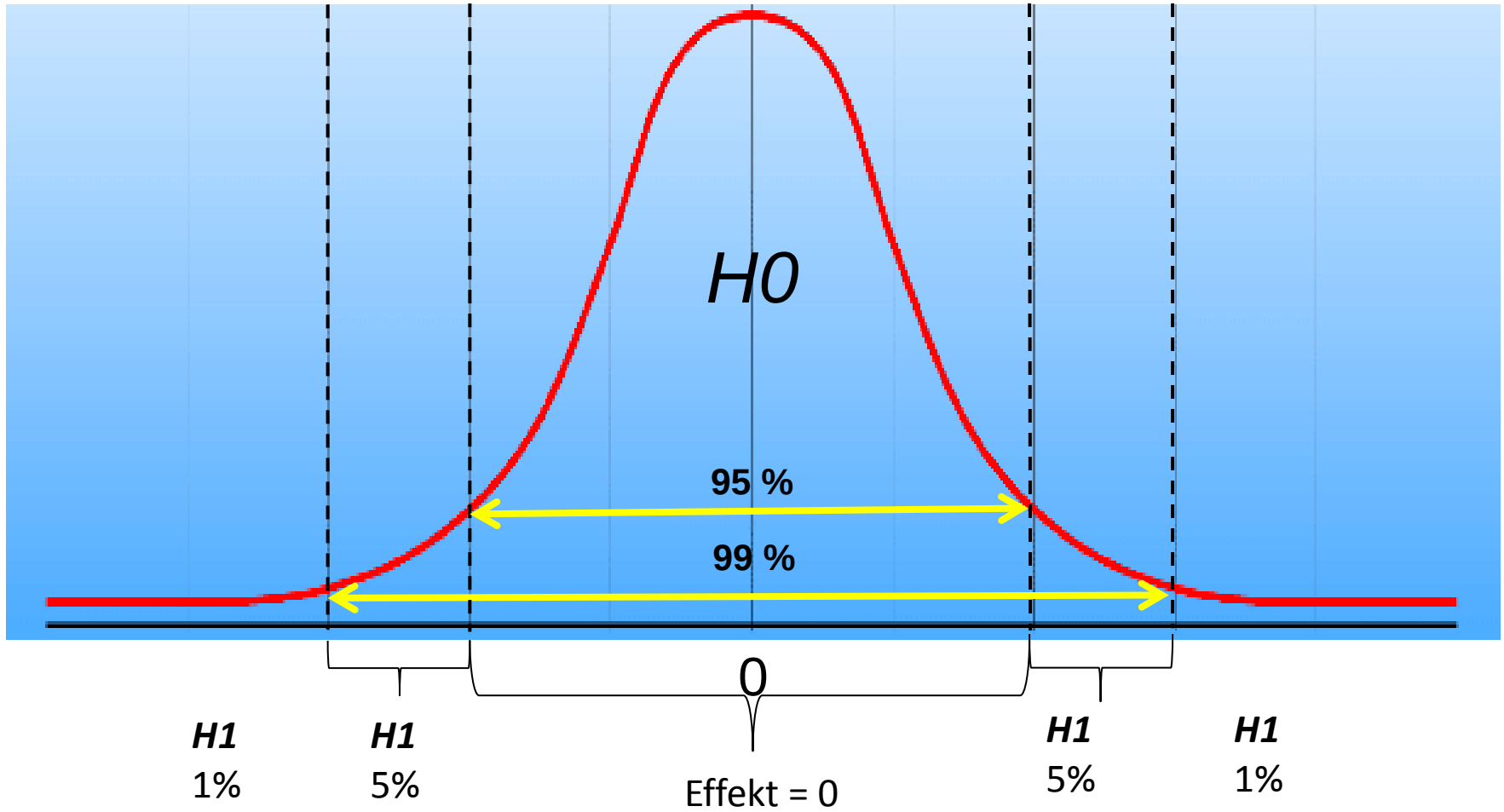
«Es gibt einen Zusammenhang zwischen dem Geschlecht und der Leseleistung»

Signifikanztests

- Es wird immer von der Nullhypothese (H_0) ausgegangen: Wenn die Nullhypothese wahr ist, wie wahrscheinlich ist es, einen solchen Effekt zu finden, wie wir ihn in der Stichprobe gefunden haben?
- Entscheidende Masse: p-Wert und Signifikanzniveau (Alpha, α)
- p-Wert: Die Wahrscheinlichkeit, durch blossen Zufall ein Ergebnis zu erhalten, welches dem Stichprobenergebnis entspricht oder grösser ist
- Signifikanzniveau: Grenzwert für einen signifikanten (überzufälligen) Effekt. Entweder 5% oder 1%.

-> Alpha-Fehler-Wahrscheinlichkeit = 5% (resp. 1%): Die Wahrscheinlichkeit einen signifikanten Effekt zu finden, obwohl dieser gar nicht existiert ist 5% (resp. 1%)
- Ist der p-Wert $< .05$ ist das Ergebnis auf dem 5% Signifikanzniveau signifikant (analog dazu $p < .01$ für das 1% Signifikanzniveau)
- Ist ein Ergebnis signifikant, kann es auf die Population generalisiert werden -> Alternativhypothese (H_1) wird angenommen

Signifikanztests



Signifikanztests

- Signifikanz ist abhängig von der Grösse des Effekts (möglichst weit von null entfernt) und dem Standardfehler
- Standardfehler ist abhängig von der Varianz und der Stichprobengrösse
- Problem: Sehr kleine Effekte können signifikant werden, wenn die Stichprobe genügend gross ist und grosse Effekte werden evtl. nicht signifikant in sehr kleinen Stichproben
- Deshalb Vorsicht bei der Interpretation:
 - Signifikanter Effekt ist statistisch bedeutsam, d.h.: Er ist nicht zufällig zustande gekommen und darf auf die zugrundeliegende Population übertragen werden
 - Aber: Statistische Bedeutsamkeit $\neq$ inhaltliche/praktische Bedeutsamkeit

Effektstärke (Effektgrösse)

- Effektstärke = inhaltliche Bedeutsamkeit (praktische Relevanz) eines Effektes
- Standardisiertes Mass für einen in der Stichprobe gefundenen Effekt (wichtig für die Vergleichbarkeit von Studien)
- Keine inferenzstatistische Aussage (auf Stichprobe bezogen)
- Unabhängig von der Stichprobengrösse
- Verschiedene Arten von Effektstärken (z.B. Korrelationskoeffizient r oder Cohen's d) -> Wird im Rahmen der behandelten Testverfahren noch genauer thematisiert
- Auch hier gibt es Grenzwerte für die Bedeutsamkeit eines Effektes

Effektstärken und Cutoff-Werte

Mass der Effektstärke	Kein Effekt	Kleiner Effekt	Mittlerer Effekt	Grosser Effekt
Korrelationskoeffizient r	< .1	ab .1	ab .3	ab .5
Cohen's d	< .2	ab .2	ab .5	ab .8

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale: Lawrence Erlbaum Associates.

Teststärke (Testpower)

- Wahrscheinlichkeit einen signifikanten Effekt zu finden, falls dieser tatsächlich existiert
- Abhängig von der Effektstärke und der Stichprobengrösse
- Studienumfangsplanung: Wie gross muss das N sein, um bei einer erwarteten Effektstärke genügend Teststärke zu haben, um einen signifikanten Effekt finden zu können?

(Programm G*Power: <http://www.gpower.hhu.de/>)

- Genügend Teststärke heisst mindestens 90% (Rasch et al., 2006)
- Beta-Fehler Wahrscheinlichkeit 10% (Nullhypothese wird fälschlicherweise angenommen)

Zusammenhang Teststärke, Effektstärke und Signifikanz

Zwei fiktive Beispiele von experimentellen Designs:

- Beispiel 1: Eine Intervention zur Förderung der selektiven Aufmerksamkeit soll zu einer besseren Schulleistung führen. Erwartet wird eine mittlere Effektstärke von mindestens $r = 0.3$
- Beispiel 2: Gezielter Nachhilfeunterricht in den problematischen Fächern führt zu einer besseren Schulleistung. Erwartet wird eine hohe Effektstärke von mindestens $r = 0.5$

Zusammenhang Teststärke, Effektstärke und Signifikanz

Beispiele anhand von zwei fiktiven experimentellen Designs

- Beispiel 1:

- $r = 0.3$; $\alpha = 5\%$; Power = 90%; $\beta = 10\%$; zweiseitig

-> **N = 109**

- Beispiel 2:

- $r = 0.5$; $\alpha = 5\%$; Power = 90%; $\beta = 10\%$; zweiseitig

-> **N = 34**