

Daten- und Entscheidungsanalyse Statistische Methoden

PD Dr. Dr. Dipl.-Psych. Guido Strunk

Hidden Features, Making of, and Additional Resources at ...

<http://www.complexity-research.com/ProjekteLehre.htm>



Inhalt

1.	Übungsaufgaben	4
1.1	Skalenniveaus	4
1.2	Stichprobengröße	5
1.3	Einfache Signifikanztests – GStat	6
1.4	Qualitätskontrolle der Arzt-Patient-Beziehung	7
1.4.1	Theoretische Vorüberlegungen	7
1.4.2	Forschungsvorhaben	7
1.4.3	Aufgabenstellungen	8
2.	Von der Projektarbeit zur Masterarbeit	9
3.	Methodenlehre	10
3.1	Was ist Wissenschaft?	10
3.2	Qualitative Forschung	11
3.3	Quantitative Forschung	14
3.4	Forschungsprozess	15
3.5	Theorie und Hypothese	17
3.6	Forschungsdesigns	22
3.7	Hypothetische Konstrukte	25
3.8	Forschungsmethoden	26
3.9	Gütekriterien	30
3.10	Population und Stichprobe	33
3.11	Messtheorie	36
3.12	Skala und Item	42
4.	Datenverarbeitung	45
5.	Deskriptive Statistik	47
5.1	Maße der zentralen Tendenz	48
5.2	Streuungsmaße	49
6.	Inferenzstatistik – Grundlagen	51
7.	Einfache Testverfahren zum Vergleich von zwei Gruppen	55
7.1	T-Test	55
7.2	T-Test im SPSS	56
7.3	T-Test – GStat	58
7.4	T-Test – Excel	59
8.	Vergleich mehrerer Gruppen	60
8.1	Chi-Quadrat	60
8.2	Einfaktorielle Varianzanalyse	60
9.	Zusammenhangshypothesen	63
10.	Software	67

10.1	Excel.....	67
10.2	GStat.....	67
10.3	SPSS.....	69
10.3.1	Struktur und Aufbau.....	69
10.3.2	Grundsätzliches.....	71
10.3.3	Erstellung von Datenfiles und Kodierung von Fragebögen.....	73
10.3.4	Grundsätzliches zu Datenauswertung mit SPSS.....	75
10.3.5	Deskriptive Statistiken.....	79
10.3.6	Testverfahren im SPSS.....	80
10.3.7	Hilfefunktionen.....	112
10.3.8	Test-Finder.....	113
11.	Glossar für einige wichtige statistische Begriffe.....	115
12.	Darstellung und Abkürzungen.....	120
12.1	Allgemein.....	120
12.2	Korrelationen.....	121
12.3	T-Test.....	122
12.4	Varianzanalyse.....	123
13.	Literatur.....	124

1. Übungsaufgaben

1.1 Skalenniveaus

Phänomen	Skalenniveau	Diskret oder stetig	Kodierung Wie werden Zahlen vergeben?
Geschlecht	<i>Nominal</i>	<i>Diskret</i>	<i>0: männlich 1: weiblich</i>
Temperatur			
Höhe eines Berges			
Reiseziele von Urlaubern			
Güteklassen von Gemüse			
Inflationsraten			
Bruttojahresgehalt			
Anzahl Verkehrstote			
Beruf			
Höchster Bildungsabschluss			
Berufsjahre im Job			
IQ			
Anzahl belegter Betten pro Monat			
Systolischer Blutdruck			
Schulnote in Englisch			
Schmerzskala (VAS)			

1.2 Stichprobengröße

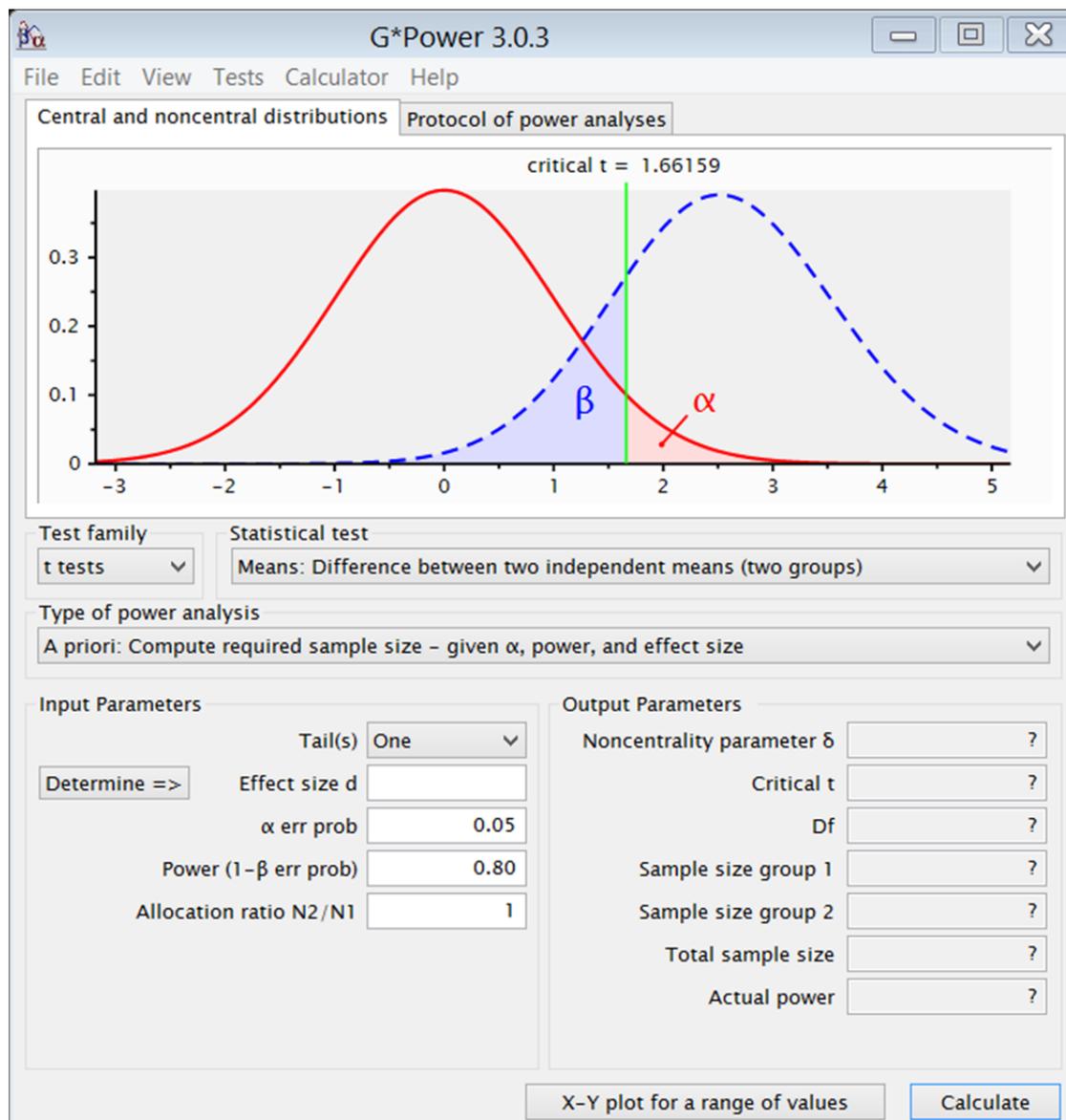
Sie planen eine Untersuchung und möchte vorausberechnen, wie groß Ihre Stichprobe sein muss. Sie planen zwei gleich große Gruppen miteinander zu vergleichen. Verglichen werden soll der mittlere systolische Blutdruck. Vorstudien haben gezeigt in der einen Gruppe liegt RRS bei durchschnittlich 120 ± 25 mmHG und in der anderen bei 135 ± 32 mmHG. Wie groß muss ihre Stichprobe sein bei $\alpha = 0,05$, Power ($1-\beta$) = 0,8.

Welches statistische Testverfahren wird eingesetzt?

Wie groß ist die Effektstärke?

Wie groß sollten die Stichproben sein?

(Software unter: <http://www.gpower.hhu.de/>)



1.3 Einfache Signifikanztests – GStat

Die Berechnungen können durchgeführt werden mit: GStat:

<http://www.complexity-research.com/DownSoft.htm>

- 1) Ist der Unterschied im systolischen Blutdruck aus der vorangehenden Aufgabe statistisch signifikant, wenn jeweils 30 Versuchspersonen teilnahmen?

Testverfahren: _____

1-seitiger oder 2-seitiger Test? 1-seitiger Test 2-seitiger Test

P = _____

Signifikanz? nicht signifikant signifikant sehr signifikant

- 2) In Gruppe A erkrankten 5 von 250 Personen in Gruppe B waren es 10 von 120. Ist die Gruppe A signifikant seltener erkrankt?

Testverfahren: _____

1-seitiger oder 2-seitiger Test? 1-seitiger Test 2-seitiger Test

P = _____

Signifikanz? nicht signifikant signifikant sehr signifikant

- 3) In einer Stichprobe aus den USA fand sich eine Korrelation zwischen Einkommen und Gesundheitsstatus von $r = 0,2$ ($n = 2500$). Eine Replikation der Studie fand in Österreich ein $r = 0,3$ ($n = 500$). Besteht ein Unterschied?

Testverfahren: _____

1-seitiger oder 2-seitiger Test? 1-seitiger Test 2-seitiger Test

P = _____

Signifikanz? nicht signifikant signifikant sehr signifikant

1.4 Qualitätskontrolle der Arzt-Patient-Beziehung

1.4.1 Theoretische Vorüberlegungen

Langewitz, Keller und Denz (1995) gehen auf die Qualitätskontrolle der Arzt-Patienten-Beziehung wie folgt ein (gekürzt aus der Einleitung übernommen):

Dem Aspekt der Qualitätssicherung der ärztlichen Tätigkeit wird auch in Deutschland nicht zuletzt dank entsprechender Vorgaben der Bundesärztekammer und der Kassenärztlichen Vereinigungen zunehmend mehr Beachtung geschenkt. Die Toronto-Konsensus-Konferenz zur Arzt-Patient-Kommunikation hat die Bedeutung effektiven Kommunikationsverhaltens zwischen Arzt und Patient als wesentliches Element ärztlichen Handelns herausgestellt und mit folgenden Argumenten begründet:

„Kommunikationsprobleme in der Praxis sind häufig, sie äußern sich in der unzureichenden Identifikation der Sorgen und Beschwerden von Patienten, in der mangelnden Diagnose psychosozialer und psychiatrischer Probleme und in der fehlenden Übereinstimmung von Arzt und Patient in der Frage, welches Problem für die Konsultation wesentlich ist“.

In der Literatur findet sich eine große Anzahl an Arbeiten, die das Unvermögen traditioneller medizinischer Ausbildung bei der Vermittlung klinischer Kommunikationsfertigkeit beklagen, obwohl die wesentlichen Elemente dieser Fähigkeit bekannt sind und eine Vielzahl von zum Teil hochstrukturierten Weiterbildungsprogrammen publiziert wurde. Wünschenswert wäre es, wenn bereits vorhandene oder noch zu entwickelnde Weiterbildungsprogramme nach einem einheitlichen Muster evaluiert würden. Damit ließe sich beim Vergleich mehrerer Studien für den potentiellen Anwender das Verhältnis aus Aufwand und Ertrag kalkulieren, so dass eine Institution, die z.B. für die in ihr beschäftigten AssistentInnen ein bestimmtes Fortbildungsprogramm etablieren möchte, das ihr gemäße auswählen könnte. Da sich die Interaktion zwischen Arzt und Patient auf vielen Ebenen abspielt und sich zudem in sehr unterschiedlichen situativen Kontexten vollzieht, wird es unmöglich und auch wenig sinnvoll sein, in Analogie zur Bestimmung einzelner Laborwerte einen „Goldstandard“ zu entwickeln, der mit wenigen Messwerten Aussagen über die Qualität einer konkreten Begegnung zwischen Arzt und Patient erlaubt. Ein möglicher Ausweg aus diesem Dilemma könnte die Bestimmung der Zufriedenheit der Patienten mit dem ärztlichen Gespräch sein: ein solches Vorgehen bettet sich ein in die Bemühungen z.B. der EG-Kommission, die Benutzerzufriedenheit der „Kunden“ im Gesundheitssystem zu erfassen. Die Bestimmung der Patientenzufriedenheit als Outcome-Variable ist zusätzlich unter dem Aspekt sinnvoll, dass Patientenzufriedenheit mit Compliance korreliert. Es wäre also prinzipiell wünschenswert und sinnvoll, Interventionsstudien zur Verbesserung der Arzt-Patienten-Kommunikation über ihren Einfluss auf die Patientenzufriedenheit zu evaluieren.

1.4.2 Forschungsvorhaben

Sie werden von einem Spital beauftragt, Qualitätsunterschiede zwischen jungen Assistenzärztinnen mit einem Patientenzufriedenheitsfragebogen zu erheben. Die Spitalsleitung hat durch einzelne Beschwerden von Patientinnen und Patienten den Eindruck gewonnen, dass die jungen Assistenzärztinnen über unterschiedlich hohe Fertigkeiten in der Arzt-Patienten-Kommunikation verfügen. Sie wollen mit der Untersuchung Schwachstellen aufdecken und anschließend ein Kommunikationstraining für diejenigen initiieren, die signifikant schlechter abschneiden als ihre Kolleginnen. Die Untersuchung betrifft fünf Assistenzärztinnen. Es soll der Fragebogen von Langewitz, Keller und Denz (1995) eingesetzt werden. Dieser enthält 14 Fragen zur Messung der Patientenzufriedenheit. Die Antworten auf die Fragen werden durch einen Mittelwert zu einer Gesamtskala zusammengefasst. Der Fragebogen erfasst zudem das

Geschlecht der Patientinnen und Patienten sowie das Alter. Obwohl das Alter intervallskaliert und stetig erhoben wird, sollen nur zwei Altersstufen, „jung“ und „alt“ berücksichtigt werden.

1.4.3 Aufgabenstellungen

1. Welche aVs und uVs werden erhoben? Bitte nennen Sie diese und kennzeichnen Sie sie als aV bzw. uV.
2. Sie weisen die Klinikleitung darauf hin, dass der eingesetzte Fragebogen die üblichen Hauptgütekriterien erfüllen muss. Welche sind das, was bedeuten diese für Ihre Untersuchung und welche können Sie ohne weiteren Erhebungsaufwand nachträglich noch prüfen?
3. Welches Skalenniveau kennzeichnen diese Variablen im allgemeinen, welches Skalenniveau wird in der Untersuchung benutzt?
4. Wie lauten die Hypothesen für den Einfluss der uVs auf die aVs? Bitte nennen Sie alle Alternativhypothesen und die dazu passenden Nullhypothesen.
5. Wie viele Patientinnen und Patienten sollten pro Assistenzärztin untersucht werden, damit Verfahren der Hypothesentestung benutzt werden können, die auf der Normalverteilung beruhen.
6. Auswertung: welche deskriptiven Analysen erscheinen Ihnen sinnvoll?
7. Bitte nennen sie für jede einzelne Hypothese aus 3. ein passendes statistisches Testverfahren.
8. Die grundsätzliche Fragestellung der Spitalsleitung sucht nach einem Unterschied zwischen den Assistenzärztinnen. Alter und Geschlecht der Patientinnen und Patienten können hier einen Einfluss haben. Welches Verfahren muss eingesetzt werden, um den Einfluss aller Variablen gleichzeitig zu berücksichtigen. Tipp: wie viele Faktoren umfasst das Untersuchungsdesign?
9. Die Analyse des Alters soll auf zwei Altersgruppen beschränkt bleiben. Welche Möglichkeiten sehen Sie, diese Gruppen nicht nur willkürlich zu bilden?

Wenn Sie zur Beantwortung der Fragen Annahmen treffen müssen, beschreiben Sie diese in Ihren Antworten. Z. B. wenn man davon ausgeht, dass dieses oder jenes der Fall ist, kann dieses oder jenes Verfahren eingesetzt werden.

Es genügt, wenn Sie die Antworten kurz und knapp formulieren. Dennoch möchte ich Sie bitten so zu schreiben, dass wichtige Annahmen, Überlegungen und Schlussfolgerungen deutlich werden.

2. Von der Projektarbeit zur Masterarbeit

3. Methodenlehre

3.1 Was ist Wissenschaft?

Wissenschaftliches Arbeiten

- Formulieren von Bildern und Modellen über die Realität zur Beschreibung und/oder Erklärung dieser Realität.
- Folgt akzeptierten Regeln und Grundprinzipien, v.a.
 - Transparenz,
 - Nachvollziehbarkeit,
 - Überprüfbarkeit.
- Bezugnahme auf vorhandenes, bereits „gesichertes“ Wissen.
Wissenschaft ist kumulativ.

Grundtypen wissenschaftlicher Fragestellungen

- Beschreibung
 - Was ist der Fall? Wie sieht die „Realität“ aus? (Oder auch: Sieht die Realität wirklich so aus?)
- Erklärung
 - Warum ist etwas der Fall? Warum und unter welchen Bedingungen treten bestimmte Phänomene auf?
- Prognose
 - Was wird zukünftig der Fall sein? Wie wird etwas künftig aussehen? Welche Veränderungen werden eintreten?
- Gestaltung/Technologie
 - Welche Maßnahmen sind geeignet, um ein bestimmtes Ziel zu erreichen?
- Kritik, Bewertung
 - Wie ist ein bestimmter Zustand vor dem Hintergrund explizit genannter Kriterien zu bewerten?

3.2 Qualitative Forschung

Gründe für qualitative Forschung

- Kaum wissenschaftliche Vorarbeiten. Hypothesenprüfendes Vorgehen ist noch nicht möglich, weil Grundlagen fehlen.
- Das Feld ist weitgehend unbekannt.
- Exploration des Unbekannten.
- Gehaltvolle Exploration bietet mehr Informationen als konkrete statistische Kennwerte.
- Man möchte ein Feld inhaltlich besser verstehen, weiß aber nicht was einen erwartet.

Forschungsfragen und Ziele

- Was ist der Fall? (Beschreibung)
- Wie kommt es dazu? (Erklärung suchend)

- Qualitative Forschung steht häufig am Anfang eines Forschungsprozesses.
- Sie liefert Material und Ideen für Theorien (theoriengenerierend).

Grundhaltung

- Unvoreingenommenheit.
- Bewusst eingenommene Unwissenheit.
- Möglichst kein steuernder Einfluss auf die Untersuchungseinheiten.
- Entdeckende unvoreingenommene Analyse.

Mögliches Vorgehen in Interviews

- Kein ausgefeilter Leitfaden, der Fakten abfragt.
- Einstiegsfrage, die zum Erzählen einlädt.
- Danach Gespräch wenig steuern aber zum weitererzählen einladen.

- Das Gespräch wird aufgezeichnet und später wörtlich transkribiert. (Transkriptionsregeln beachten, Zeilennummern benutzen für spätere Auswertung).

Critical Incident

- Eine mögliche Einstiegsfrage orientiert sich an der „*Critical Incident Technique*“, z.B.:

Als Sie damals die XYZ Managementmethode eingeführt haben, gab es da einen Moment, wo Sie dachten, die Einführung könnte scheitern? Was war das für ein Moment? Bitte erzählen Sie ...

Theoretisches Sampling

- Je nach Länge eines Interviews fallen schnell viele Seiten an Transkripten an. N=10 kann schon extrem viel Arbeit sein.
- Auswahl der InterviewpartnerInnen ist daher gut zu begründen.
- Expertinnen, verschiedene Sichtweisen (Extremgruppen).
- Theoretisches Sampling: theoretische Begründung der Auswahl der InterviewpartnerInnen.

Transkription und Analyse

- Mayring, P. (2003) *Qualitative Inhaltsanalyse*. Weinheim und Basel: Beltz Verlag

Transkript

Interview: E
17.05.2013, Dauer: 47:55 Minuten
I: Interviewerin
E: Experte

- 1 I: Danke, dass sie sich bereit erklärt haben für dieses Interview. Das Thema meiner Abschlussarbeit ist Pflegequalität und wie ich es im Vorgespräch schon kurz erklärt habe beleuchte ich dabei die Sicht der Basispflegepersonen. Ich habe mir ein paar Fragen dazu
- 2 überlegt und lasse sie einfach erzählen, ich frage nur gelegentlich nach. Beginnen möchte
- 3 ich gerne mit
- 4
- 5
- 6 I: Wenn sie sich in ihre Tätigkeit denken, sie gehen nach dem Dienst nach Hause, was
- 7 muss für sie sein, dass sie sagen: heute habe ich wirklich gute Pflege geleistet, heute
- 8 habe ich gut gearbeitet?
- 9 E: Für mich ist wichtig, dass der Patient alle Therapien erhalten hat, auch im Zeitrahmen der
- 10 Vorgabe. Dass ich neben dem Stationsalltag auch noch für wichtige Belange des Patienten
- 11 ein Ohr hatte, wenn jetzt wirklich der Patient ein Problem neben den Therapien äußert, dass
- 12 ich das auch wahrnehme und dann dementsprechend darauf auch reagiert habe, das kann

Analyse

- Festlegen der Analyseeinheiten (Worte, Sätze, Sinneinheiten).
- Unvoreingenommenes Lesen und markieren der für die Forschungsfrage relevanten Analyseeinheiten.
- Herauskopieren aller markierten Analyseeinheiten in eine Tabelle. Genauer Verweis darf nicht vergessen werden (Zeile, Seiten, InterviewpartnerIn).

Verweis	Wörtlich zitiert	Paraphrase (Vereinfachung)	Generalisierung (Vereinheitlichung)	Reduktion
---------	------------------	----------------------------	-------------------------------------	-----------

Beispiel

Was sind deine Erfahrungen mit dem Melden von Zwischenfällen, unerwünschten Ereignissen und Fehlern im medizinischen Alltag?

- **Wörtlich:**

Ich bin jetzt seit zehn Jahren im Bereich der Anästhesie, Intensivmedizin und Notfalltherapie tätig und mit zunehmender Erfahrung ist mir aufgefallen, dass über diese Themen unter Kollegen gar nicht gesprochen wird. Viel mehr handelt es sich um ein Tabuthema ...

- **Paraphrase:**

Das Thema Fehler ist ein Tabu, man grenzt sich aus wenn man drüber spricht.

- **Generalisierung:**

Keine Diskussion über Fehler.

- **Reduktion:**

Tabuthema.

3.3 Quantitative Forschung

Gründe für quantitative Forschung

- Konkrete Kennzahlen sind gefragt.
- Gegenstand lässt sich gut in Zahlen abbilden.
- Das Feld ist so bekannt, dass Kennwerte und Konstrukte soweit formuliert sind, dass man weiß, dass man sie messen, zählen kann.
- Komprimierte Darstellung durch statistische Kennwerte.
- Man hat Hypothesen und möchte diese konkret prüfen.

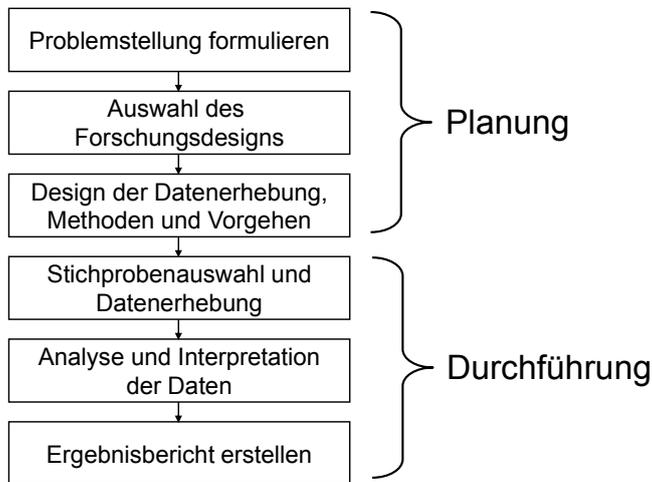
Grundlagen

- Messung: Abbildung empirischer Gegebenheiten im Zahlenraum.
- Deskriptive Statistik: Zusammenfassung von Daten um diese mit wenigen Kennwerten zu beschreiben.
- Prüfende Statistik: Unterschieds- und Zusammenhangshypothesen können statistisch geprüft werden.

Die Art der Messung bestimmt die Methoden der deskriptiven und der prüfenden Statistik.

3.4 Forschungsprozess

Phasen im Forschungsprozess



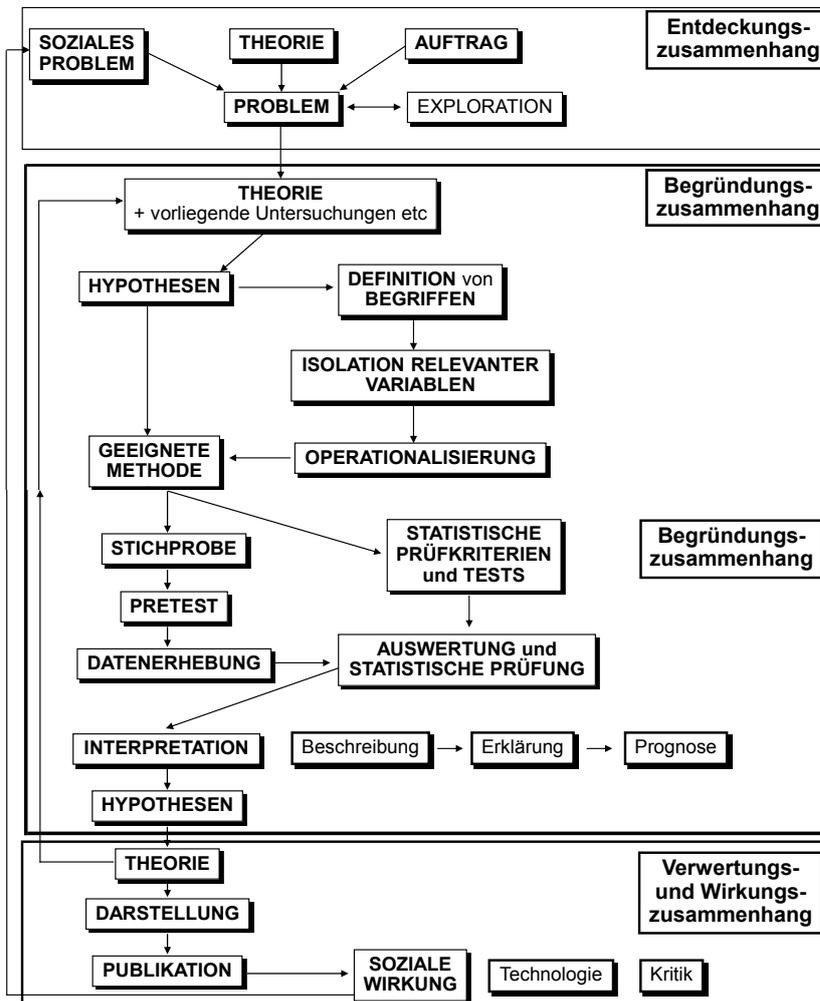
Frei nach: Churchill und Iacobucci (2002, S. 56)

Fragen, für die es Antworten geben sollte I

Phase des Prozesses	Typische Fragen
<i>Problemstellung formulieren</i>	<ul style="list-style-type: none"> • Was ist das Ziel der Studie? Ein Problem zu lösen? Möglichkeiten aufzuzeigen? • Welche Informationen sind nötig, um die Ziele der Studie zu erreichen? • Werden zusätzliche, andere Informationen benötigt? • Wie werden die Ergebnisse der Studie wohl Verwendung finden?
<i>Auswahl des Forschungsdesigns</i>	<ul style="list-style-type: none"> • Was ist eigentlich bereits bekannt? • Können konkrete Hypothesen formuliert werden? • Welche Arten von Fragen sollen konkret beantwortet werden? Beschreibung, Erklärung, Prognose, Gestaltung, Kritik, Bewertung? • Welche Art von Studie (qualitativ, quantitativ, Primärstudie, Sekundäranalysen etc.) kann die Forschungsfragen am besten beantworten?
<i>Design der Datenerhebung, Methoden und Vorgehen</i>	<ul style="list-style-type: none"> • Kann bereits existierendes Datenmaterial genutzt werden? • Was soll gemessen werden? Wie? Operationalisierung? • Welche Datenquellen stehen zur Verfügung? • Sind kulturelle, legale, ethische oder andere Faktoren bei der Datenerhebung zu berücksichtigen? • Können durch Befragungen „objektive“ Daten erwartet werden? • Wie sollte man befragen (offene Fragen, Antwortkategorien, Ratings etc.)? • Können Beobachtungen mit technischen Hilfsmitteln erleichtert werden? • Bei einer Beobachtung, was genau soll beobachtet werden? • Experiment oder Quasi-Experiment?

Phase des Prozesses	Typische Fragen
Stichprobenauswahl und Datenerhebung	<ul style="list-style-type: none"> • Was ist die Ziel-Population? • Gibt es ein vollständiges Verzeichnis der Ziel-Population? • Ist es notwendig eine Stichprobe zu untersuchen? • Ist eine Zufallsstichprobe möglich und nötig? • Wie groß sollte die Stichprobe sein? • Wie wird die Stichprobe ausgewählt? • Wer erhebt die Daten? • Wie lange wird die Erhebung dauern? • Wie wird die Qualität der erhobenen Daten sichergestellt?
Analyse und Interpretation	<ul style="list-style-type: none"> • Wie werden die erhobenen Daten kodiert? • Wer kodiert und wie wird die Qualität überwacht? • Welche Software soll genutzt werden? • Welche deskriptiven Darstellungen (Tabellen, Abbildungen, Kennwerte) werden benötigt? • Mit welchen (inferenz-)statistischen Methoden soll gearbeitet werden?
Ergebnisbericht erstellen	<ul style="list-style-type: none"> • Wer wird den Forschungsbericht lesen? • Wie kompliziert darf der Bericht werden? • In welcher Form soll der Bericht erstellt werden? • Soll es auch eine mündliche Präsentation geben? • Wie ist die mündliche Präsentation zu strukturieren?

Prozessmodell 1



(vgl. Friedrichs, 1990, S. 51)

3.5 Theorie und Hypothese

Theorie

- Eine Theorie ist ein System von Begriffen, Definitionen und Hypothesen. Dieses System sollte in sich geordnet und widerspruchsfrei sein.
- Daraus folgt, dass
 - die Konstruktion und Überprüfung von Theorien zunächst auf einer rein formalen Ebene nach den Regeln der Logik erfolgt;
 - je nach erkenntnistheoretischer Richtung bestimmte Spielregeln zur Überprüfung von Theorien einzuhalten sind (z.B. Falsifikationsprinzip des *Kritischen Rationalismus*);
 - Theorien rein logische Gebäude bleiben. Auch dann, wenn sie empirisch fundiert sind, ergeben sich Probleme der Übertragbarkeit in Bezug auf das Verhältnis zwischen der Theoriesprache einerseits und der sogenannten Realität andererseits.

Hypothesen 1

Prüfbarkeit. Eine Hypothese blickt nach vorne. Sie ist eine Behauptung, die man prüfen kann. Eine Hypothese mag dem gesunden Menschenverstand widersprechen oder mit ihm übereinstimmen. Sie kann sich als richtig oder falsch erweisen. In jedem Fall führt sie jedoch zu einer empirischen Nachprüfung.

Beantwortbarkeit. Unabhängig von dem Ergebnis ist eine Hypothese eine Frage, die so gestellt ist, dass irgendeine Antwort darauf gegeben werden kann.

Systematische Skepsis. Sie ist ein Beispiel für die systematische Skepsis der Wissenschaft, für ihre Weigerung, irgendeine Behauptung ohne empirische Bestätigung anzuerkennen.

Hypothese und Theorie. Die Verifikation oder Falsifikation von Hypothesen wird in der Regel angestrebt, um bestimmte Elemente oder eine ganze Theorie auf den Prüfstand zu stellen. An einer Hypothese hängt also mehr als nur die Überprüfung einer x-beliebigen Vorhersage.

Hypothesen 2

Hypothesen sind Schlussfolgerungen aus einer Theorie, die die Brücke zur beobachtbaren „*Wirklichkeit*“ schlagen. Aus dem Zutreffen der Hypothese im Rahmen ihrer empirischen Prüfung kann dann die Bewährung der Theorie gefolgert werden.

Hypothesen sind im engeren Sinne nur dann „*wissenschaftlich*“, wenn sie sich in der empirischen Prüfung als *falsch* oder *wahr* herausstellen können.

Wissenschaftstheoretiker, sowohl der *Logischen Empiristen*, als auch der *Kritisch Rationalisten*, nutzen die prinzipielle empirische Überprüfbarkeit als Abgrenzungskriterium, um wissenschaftliches Arbeiten von metaphysischen Spekulationen zu unterscheiden.

Nicht empirisch überprüfbar ist z.B. die Existenz eines sich nicht einmischenden Gottes, die Existenz eines freien Willens, die Existenz des Fremdseelischen.

Hypothesen 3

Verifikation vs. Falsifikation

Die *logischen Empiristen* verlangen die *Verifikation* von Hypothesen und dadurch die Verifikation von Theorien.

Für (*allgemeine*) *Existenzaussagen* ist die Verifikation die einzig logische und die *effektivste* Vorgehensweise. Die Behauptung, dass es kleine grüne Männchen auf dem Mars gibt, kann durch das Vorzeigen nur eines kleinen grünen Männchens vom Mars bewiesen werden.

Die *kritischen Rationalisten* verlangen die Falsifikation von Hypothesen und darüber die Falsifikation von Theorien.

Wissenschaftliche Erkenntnis muss prinzipiell falsifizierbar sein. Widersteht eine Theorie, eine Hypothese in einer Prüfung der Falsifikation, so gilt sie als *vorläufig bewährt*.

K. R. Popper proklamiert: „Wir finden das Richtige, indem wir das Falsche ausschließen.“

Hypothesen 4

Für eine Hypothese, die zur Falsifikation taugt, wird also eine *Schwachstelle der Theorie* gesucht und diese in der Realität geprüft.

Die Logik dahinter geht davon aus, dass (*allgemeine*) *Gesetzesaussagen* zwar beliebig oft verifiziert werden können, dass dies aber kein zwingender Beweis für die Gültigkeit der Gesetzesaussagen ist.

Die Behauptung, dass alle Schwäne weiß seien, ist durch jeden beobachteten weißen Schwan verifizierbar, aber damit nicht vollständig bewiesen. Ein einziger Gegenbeweis, ein einziger schwarzer Schwan, würde jedoch genügen, um die Hypothese als nicht zutreffend festzustellen.

2, 4, 6

Hypothesen 5

Wir können also zu einer Hypothese, die aus einer Theorie folgt zwei verschiedene Aussagen formulieren.

Die Verifikationsannahme führt zur so genannten Alternativ-Hypothese (H_1).

H_1 : „Alle Schwäne sind weiß.“

Die Falsifikationsannahme führt zur so genannten Null-Hypothese (H_0).

H_0 : „Es gibt zu mindest einen schwarzen Schwan.“

Häufig ist die Null-Hypothese das einfache Gegenteil der Alternativ-Hypothese. Das ist aber nicht immer der Fall. Mitunter gibt es viele verschiedene Null-Hypothesen (z.B. grüne Schwäne) oder verschieden spezifische Null-Hypothesen (z.B. nicht-weiß vs. schwarz). Es gilt die Null-Hypothese zu wählen, die bei ihrem Scheitern logisch am stringentesten die Alternativ-Hypothese als bewährt erscheinen lässt.

In einem sauberen Forschungsdesign wird alles getan, um die Falsifikationsaussage, also die H_0 , zu verifizieren.

Arten von Alternativ-Hypothesen Übersicht

	Unterschieds-Hypothesen		Zusammenhangs-Hypothesen	
2 Gruppen / Variablen / Objekte	Einfache Gruppenvergleiche		Einfache Korrelation	
Mehr als 2 Gruppen / Variablen / Objekte	Multiple Gruppenvergleiche		Multiple Korrelation / Regression	
Hypothesentyp	gerichtet	ungerichtet	gerichtet	ungerichtet

Arten von Alternativ-Hypothesen Unterschiede 1

Unterschiedshypothesen

- Zwei Objekte/Gruppen/Treatments
 - H_1 : A ist anders als B (ungerichtete Hypothese / 2-seitig).
 H_0 : A = B.
 - H_1 : A > B (gerichtete Hypothese / 1-seitig).
 $H_{0,1}$: A = B.
 $H_{0,2}$: A < B.
 - H_1 : A < B (gerichtete Hypothese / 1-seitig).
 $H_{0,1}$: A = B.
 $H_{0,2}$: A > B.

Arten von Alternativ-Hypothesen Unterschiede 2

Unterschiedshypothesen

- Mehr als zwei Objekte/Gruppen/Treatments
 - H_1 : Zwischen A, B, C, D bestehen Unterschiede (keine Festlegung darauf, ob sich alle von allen unterscheiden sollen oder ob ein einziger paarweiser Unterschied genügt).
 - H_0 : $A = B = C = D$.

Arten von Alternativ-Hypothesen Unterschiede 3

Unterschiedshypothesen

- Mehr als zwei Objekte/Gruppen/Treatments
 - H_1 : Zwischen A, B, C, D bestehen Unterschiede (alle unterscheiden sich von allen).
 - $H_{0,1}$: $A = B$
 - $H_{0,2}$: $A = C$
 - $H_{0,3}$: $A = D$
 - $H_{0,4}$: $B = C$
 - $H_{0,5}$: $B = D$
 - $H_{0,6}$: $C = D$

Bei mehr als zwei Objekten gilt es zu überlegen, ob A, B, C, D alle in Frage kommenden Objekte sind (*fixed factors*) oder ob es z.B. noch E und F gibt (*fixed factors mit beschränkter Aussage*). Es kann aber auch sehr sehr viele Objekte geben und A, B, C, D sind eventuell nur eine zufällige Auswahl (*random factors*). Das hat eine Bedeutung für die Verallgemeinerbarkeit (Generalisierbarkeit) der Befunde.

Arten von Alternativ-Hypothesen Zusammenhänge 1

Zusammenhangshypothesen

- Zwei Variablen
 - H_1 : x zeigt einen Zusammenhang mit y .
 x korreliert mit y (ungerichtete Hypothese / 2-seitig).
 - H_0 : x zeigt keinen Zusammenhang mit y .
Die Korrelation zwischen x und y ist Null ($r = 0$).
 - H_1 : x korreliert *positiv* mit y . Also: je mehr x , desto mehr y . Und je weniger x , desto weniger y . Und die gleichen Aussagen mit x und y vertauscht (gerichtete Hypothese/ 1-seitig).
 - $H_{0,1}$: Die Korrelation zwischen x und y ist Null ($r = 0$).
 - $H_{0,2}$: Die Korrelation zwischen x und y ist negativ ($r < 0$).

Arten von Alternativ-Hypothesen Zusammenhänge 2

- H_1 : x korreliert *negativ* mit y . Also: je mehr x , desto weniger y .
Und je weniger x , desto mehr y . Und die gleichen Aussagen mit x und y vertauscht (gerichtete Hypothese / 1-seitig).
- $H_{0,1}$: Die Korrelation zwischen x und y ist Null ($r = 0$).
- $H_{0,2}$: Die Korrelation zwischen x und y ist positiv ($r > 0$).

Arten von Alternativ-Hypothesen Zusammenhänge 3

Zusammenhangshypothesen

– Mehr als zwei Variablen

- **Unterscheidung zwischen unabhängigen (uV) und abhängigen Variablen (aV).**
Die uV ist die Variable, deren Auswirkung untersucht werden soll.
Die Alternativ-Hypothese geht davon aus, dass die aV von der uV abhängig ist.
$$uV \longrightarrow aV$$
- **Unterscheidung zwischen Prädiktor/en und Kriterium.**
Prädiktoren sind *mehrere* uV's, die in der Regel *eine* aV vorhersagen helfen.
- H_1 : **Ist in der Regel ungerichtet (2-seitig): uV korreliert mit aV**
- H_0 : **Die Korrelation zwischen aV und uV ist Null ($r = 0$)**

Arten von Alternativ-Hypothesen Zusammenhänge 4

Zusammenhangshypothesen

– Mehr als zwei Variablen

- **Unterscheidung zwischen Haupteffekten und Interaktionseffekten.**
 - Bei einem Haupteffekt besteht die Beziehung einer uV auf eine aV unabhängig von anderen uV's. (Z.B.: je schwerer die Krankheit, desto geringer die Patientenzufriedenheit. Wenn dies unabhängig vom Alter, vom Geschlecht und vom Heilungserfolg etc. gilt, liegt ein Haupteffekt vor.)
 - Bei einem Interaktionseffekt zwischen zwei Variablen (Interaktion 1. Ordnung) ändert sich die Beziehung zwischen einer uV und einer aV je nach Ausprägung einer anderen uV. (Z.B.: mit der Schwere der Krankheit und der Größe des Heilungserfolges wächst die Patientenzufriedenheit.)

3.6 Forschungsdesigns

Wichtige Begriffe

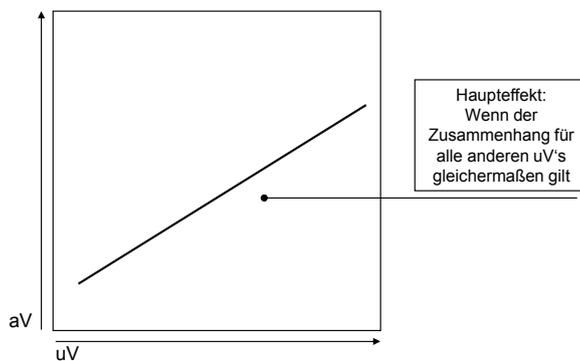
- **Unterscheidung zwischen unabhängigen (uV) und abhängigen Variablen (aV).**

Die uV ist die Variable, deren Auswirkung untersucht werden soll. Die Alternativ-Hypothese geht davon aus, dass die aV von der uV abhängig ist.

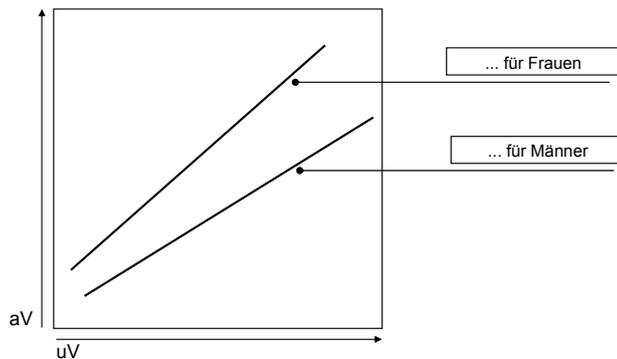
uV \longrightarrow aV

- **Unterscheidung zwischen Prädiktor/en und Kriterium.**
Prädiktoren sind *mehrere* uV's, die in der Regel *eine* aV (Kriterium) vorhersagen helfen.
- **Unterscheidung zwischen Haupteffekten und Interaktion.**
Zeigt eine uV unabhängig von anderen uVs immer die gleiche Wirkung bei der aV, so heißt das Haupteffekt. Eine Interaktion bedeutet unterschiedliche Wirkungen bei Kombination der uVs auf die aV.

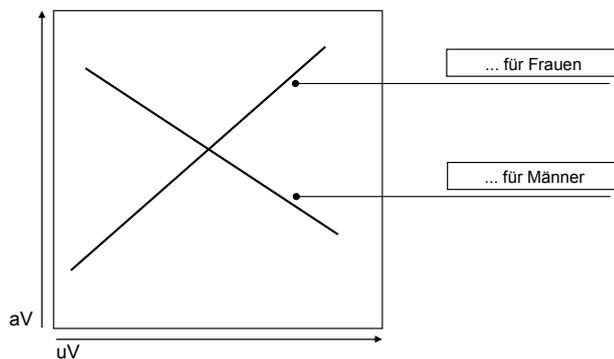
Typen von Effekten 1 Haupteffekt



Typen von Effekten 2 Ordinale Interaktion



Typen von Effekten 3 Disordinale Interaktion



Wichtige Begriffe

- **Unterscheidung zwischen ein- und mehrfaktoriellem Design.**
Liegt nur eine uV vor, die mehrere klare Abstufungen besitzt (z.B. Berufe), so handelt es sich um ein einfaktorielles Design. Mehrere uVs bilden ein mehrfaktorielles Design.
- **Unterscheidung zwischen festen und zufälligen Faktoren.**
Bei mehrstufigen uVs gilt es zu überlegen, ob alle in Frage kommenden Stufen untersucht werden (**fixed factors**) oder ob es noch andere nicht untersuchte Stufen gibt (**fixed factors mit beschränkter Aussage**). Es kann aber auch sehr viele Stufen geben und eventuell wird nur eine zufällige Auswahl (**random factors**) untersucht. Das hat eine Bedeutung für die Verallgemeinerbarkeit (Generalisierbarkeit) der Befunde.

Beispiel Design

	Abteilung 1 (Chirurgie)		Abteilung 2 (Intensiv)		Abteilung 3 (Ambulanz)	
	Frauen	Männer	Frauen	Männer	Frauen	Männer
Spital 2	Kennwert 1	Kennwert 1	Kennwert 1	Kennwert 1	Kennwert 1	Kennwert 1
	Kennwert 2	Kennwert 2	Kennwert 2	Kennwert 2	Kennwert 2	Kennwert 2
	Kennwert 3	Kennwert 3	Kennwert 3	Kennwert 3	Kennwert 3	Kennwert 3
Spital 3	Kennwert 1	Kennwert 1	Kennwert 1	Kennwert 1	Kennwert 1	Kennwert 1
	Kennwert 2	Kennwert 2	Kennwert 2	Kennwert 2	Kennwert 2	Kennwert 2
	Kennwert 3	Kennwert 3	Kennwert 3	Kennwert 3	Kennwert 3	Kennwert 3
Spital 4	Kennwert 1	Kennwert 1	Kennwert 1	Kennwert 1	Kennwert 1	Kennwert 1
	Kennwert 2	Kennwert 2	Kennwert 2	Kennwert 2	Kennwert 2	Kennwert 2
	Kennwert 3	Kennwert 3	Kennwert 3	Kennwert 3	Kennwert 3	Kennwert 3

uV Fester Faktor	uV Zufallsfaktor	uV Kontrollvariable
------------------	------------------	---------------------

Längs-/Querschnitt

- Längsschnitt- / Querschnittuntersuchungen
 - Längsschnittstudien sind angebracht, wenn es um die Messung von Veränderungen geht; z.B.
 - Panelstudien
 - ein- und dieselbe Population (Stichprobe) wird mit demselben Instrument (z.B. Fragebogen) zu verschiedenen Zeitpunkten untersucht
 - Probleme: Panelsterblichkeit, Wahl des Intervalls, Konstanthalten des Instrumentes
 - Folgestudien
 - ein- und dasselbe Instrument zu verschiedenen Zeitpunkten bei verschiedenen Stichproben

3.7 Hypothetische Konstrukte

Echte und operationale Definitionen 1

- Definitionen sind wichtige Elemente wissenschaftlicher Theorien.
- Eine *Realdefinition* ist die Erklärung eines Begriffs, die zum Ziel hat festzustellen, wie der Begriff im Rahmen der Arbeit verwendet wird.
- Eine *Nominaldefinition* ist die explizite Einführung eines Terminus, die zum Ziel hat, für einen Begriff einen Terminus festzusetzen, ihn also durch Kopplung an ein Wort dauerhaft handhabbar zu machen.
- Definitionen sind *Identitäten*. Das *Definiendum* (das, was definiert wird) ist nach der Definition *identisch* mit dem *Definiens* (dem Definierenden). Damit gilt die Forderung der *Eliminierbarkeit*: Das Definiendum muss jederzeit durch das Definiens ersetzbar sein.

Echte und operationale Definitionen 2

- Forderung nach *Nicht-Kreativität*. Es darf nicht erst durch eine Definition eine Wahrheit erzeugt/bewiesen werden, die ohne sie unbeweisbar wäre.
- Für eine empirische Erhebung ist besonders die Definition der zu erhebenden Variablen bedeutsam. Dabei sind zwei Gruppen von Variablen zu unterscheiden:
 - *Echte Definition empirischer Variablen*. Die Definition betrifft Variablen, die direkt wahrnehmbar, zählbar, messbar sind und tatsächlich mit der beobachtbaren Variable „identisch“ sind. Wenn das Gehalt als zu versteuerndes Gehalt aus unselbständiger Arbeit laut Steuerbescheid definiert wird, ist die Definition eindeutig und bezeichnet eine Identität, die bei der Messung des Gehalts 1-zu-1 genutzt werden kann.

Echte und operationale Definitionen 3

- ***Echte operationale Definition***. Die Definition betrifft Variablen, die nicht direkt wahrnehmbar, zählbar, messbar sind. Es muss eine *Operation* durchgeführt werden, um eine interessierende Eigenschaft „hervorzulocken“. Z.B. ist die Wasserlöslichkeit eines Stoffes definiert über das Vorliegen einer vollständigen Lösung eines Stoffes, nachdem man ihn ins Wasser gegeben und lange genug umgerührt hat. Auch diese operationale Definition ist eine Identität.
- ***Unechte operationale Definitionen***. In den Sozialwissenschaften werden häufig *hypothetische Konstrukte* als Elemente einer Theorie benutzt. Z.B. ist die Intelligenz ein solches hypothetisches Konstrukt. „Intelligenz“ ist eine gute „Erklärung“ für bestimmte psychische Phänomene. Wie aber kann Intelligenz definiert werden? Die Zahl der Operationen, um die Intelligenz „hervorzulocken“, ist unbegrenzt. Damit umfasst das hypothetische Konstrukt aber immer *mehr*, als die praktisch begrenzte operationale Definition. Intelligenz als hypothetisches Konstrukt ist immer mehr als der Intelligenztest misst.

Man erzählt, dass drei Schiedsrichter über die Frage des Pfeifens von Fouls uneins waren.

Der erste sagte: „Ich pfeife sie, wie sie sind.“

Der zweite sagte: „Ich pfeife sie, wie ich sie sehe.“

Der dritte und cleverste Schiedsrichter sagte:

„Es gibt sie überhaupt erst, wenn ich sie pfeife.“

(Simons, 1976, S. 29, zitiert in Weick 1985, S.9)



3.8 Forschungsmethoden

Strategien empirischer Forschung 1

- Explorative Datenanalyse:
 - Pilotstudie, Voruntersuchung; Hypothesengenerierung
- Deskriptive Datenanalyse:
 - Beschreibung, Klassifizierung, Bericht der beobachtbaren Daten; deskriptive Statistik (z.B. Verteilungsparameter, Mittelwerte, Streuungsmaße, Tabellen, Diagramme)
- Konfirmatorische Datenanalyse:
 - Hypothesentest, Suche nach Zusammenhängen, Schlüsse von der Stichprobe auf die Grundgesamtheit, Inferenzstatistik (Tests, Wahrscheinlichkeitsrechnung)
- Diagnostische Studien:
 - Erklärung von Einzelfällen; Rückgriff auf bereits bewährte Hypothesen
- Studien zur Technologieentwicklung:
 - Entwicklung und Test von Erhebungsinstrumenten, z.B. eines Fragebogens zur Patientenzufriedenheit

Strategien empirischer Forschung 2

Strategien können unterschieden werden, ...

- nach der Zahl der berücksichtigten Ebenen:
 - Personen, Gruppen, Organisationen, Gesellschaft; Mikro-/Makroebene
- nach der Zahl der verwendeten Methoden:
 - Verfahren derselben Methode (z.B. Vorbereitung von schriftlichen Befragungen durch Interviews), Verfahren unterschiedlicher Methoden (z.B. Beobachtung und Befragung)
- nach der Zahl der Zeitpunkte:
 - Querschnittuntersuchungen, Längsschnittuntersuchungen
- nach der Originalität
 - Primäranalyse, Sekundäranalyse (zweite Analyse bereits abgeschlossener Untersuchungen), Replikationsstudie

Methoden-Verfahren-Instrumente-Regeln 1

- Methoden:
 - grundlegende Ausrichtung und Wege, wie man Modelle/Abbilder der Realität erzeugt
 - konstruieren ein Stück sozialer Realität (Selektivität)
- Verfahren:
 - Techniken der Datenerhebung, die im Rahmen einer Methode angewendet werden (konkrete „Varianten“ einer Methode)
- Instrumente:
 - Werkzeuge, standardisierte Elemente der Datenerhebung
 - Z.B. Meßinstrumente, Indizes, Skalen, Tests (Fragebögen)
- Regeln:
 - Richtlinien, die bei der konkreten Datenerhebung bzw. der Erstellung von Instrumenten zu beachten sind

Methoden-Verfahren-Instrumente-Regeln 2

<i>Methoden</i>	<i>Verfahren</i>	<i>Instrumente</i>	<i>Regeln</i>
Befragung	mündlich: strukturiert / unstrukturiert / Intensivinterview schriftlich: standardisiert / nicht standardisiert / Tagebuch Delphi-Methode	Fragebogen Interview-Leitfaden Testverfahren ...	angemessene Sprache, keine Suggestivfragen, ...
Beobachtung	teilnehmend / nicht teilnehmend offen / verdeckt standardisiert / nicht standardisiert / reaktiv / non-reaktiv	Beobachtungsleitfaden, Beobachtungsschema, Notationssystem ...	Definition der kleinsten Beobachtungseinheit, ...
Textanalyse	qualitativ / quantitativ interpretativ / messend Inhaltsanalyse Diskursanalyse ...	Kategorien-schemata, Beispiel-sammlung ...	Definition von Bedeutungseinheiten, Abgrenzung relevanter Textstellen, Codierung

Methoden-Verfahren- Instrumente-Regeln 3

<i>Methoden</i>	<i>Verfahren</i>	<i>Instrumente</i>	<i>Regeln</i>
Experiment	Laborexperiment / Feldexperiment / Quasi-Experiment	nutzt Instrumente anderer Methoden bei gezielter Manipulation der Experimentalsituation	Isolierende Variation

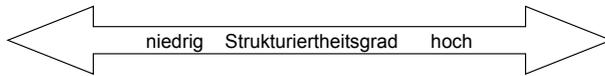
Systematische Beobachtung

- Bei **systematischen Beobachtungsverfahren** ermöglicht ein *Beobachtungsschema* die zielgerichtete Beobachtung und eine systematische Erfassung der Daten in Beobachtungsprotokollen.
- Das Beobachtungsschema umfasst:
 - die Auswahl der Beobachtungsgegenstände (Beobachtungssitems) (Ereignisse, Prozesse, Handlungen), die beobachtet werden sollen;
 - eine Zuordnung der Beobachtungssitems zu den Bedeutungen, die ihnen zugrunde liegen;
 - eine Klassifikation von Ereignissen und Handlungen mit ähnlichen Bedeutungen zu bestimmten Beobachtungskategorien. Diesen Kategorien werden im Beobachtungsprotokoll bestimmte Zeichen oder Zahlen zugeordnet (**Kodierung**).
- Ein Beobachtungsprotokoll beinhaltet Angaben zu Dauer, Ort, anwesenden Personen usw. und die Beobachtungssitems (Ereignisse, Prozesse, Verhaltensweisen), die bestimmten Beobachtungskategorien zugeordnet sind.

Merkmale von Experimenten

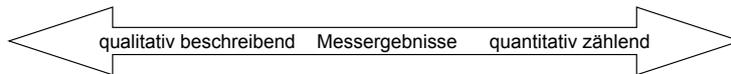
- *Manipulation* der unabhängigen Variablen
- *Messung* der abhängigen Variablen durch Instrumente anderer Methoden
- *Kontrolle* aller anderen eventuell relevanten Bedingungen durch:
 - Elimination von Störvariablen (sofern möglich)
 - Konstanthaltung der Störvariablen
 - Umwandlung von Störvariablen in weitere unabhängige Variablen
 - Parallelisierung von Experimental- und Kontrollgruppe („matching“)
 - Zufallszuweisung („Randomisierung“) der Versuchsteilnehmer zur Experimental- bzw. Kontrollgruppe

Strukturiertheitsgrad



- unstrukturiertes Interview
- nicht standardisierte Befragung
- nicht standardisierte Beobachtung
- qualitative Textanalyse
- Quasi-Experiment
- strukturiertes Interview
- standardisierte Befragung
- standardisierte Beobachtung
- quantitative Textanalyse
- Experiment

Messergebnisse



- unstrukturiertes Interview
- nicht standardisierte Befragung
- nicht standardisierte Beobachtung
- qualitative Textanalyse
- strukturiertes Interview
- standardisierte Befragung
- standardisierte Beobachtung
- quantitative Textanalyse
- Experiment / Quasi-Experiment

Vorsicht: Die Zuordnung von Verfahren ist nicht immer eindeutig.
Die von Instrumenten schon eher.

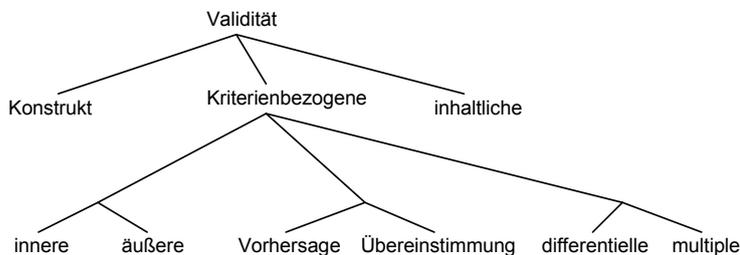
3.9 Gütekriterien

Gütekriterien 1: Intersubjektivität

- Die Forderung nach *Intersubjektivität* entspricht im Wesentlichen der historisch älteren Forderung nach *Objektivität*.
- Theoretische Aussagen, Forschungsprozess und Ergebnisse müssen auch von anderen Personen (als den unmittelbar involvierten ForscherInnen) überprüft und nachvollzogen werden können („*Nachvollziehbarkeit*“).
- Daher sollten in der *Scientific Community* akzeptierte Methoden, Instrumente und Regeln Verwendung finden.
- Wichtig ist eine präzise Definition der zu erhebenden Variablen.

Gütekriterien 2: Validität (Gültigkeit) 1

- Die Validität gibt Antworten auf die folgenden Fragen: Wurde tatsächlich das gemessen, was man messen wollte? Wie groß ist die Übereinstimmung zwischen empirischer Messung und dem zu messenden Konstrukt?



Gütekriterien 2: Validität (Gültigkeit) 2

- **Konstruktvalidität**, fragt danach, wie gut bzw. passend ein Konstrukt tatsächlich erfasst wird (Überprüfung z.B. durch logische Analyse, Extremgruppenvergleiche, Experimente).
- **Inhaltliche Validität**: Beschreibt, ob ein Verfahren nach inhaltlichen Kriterien (Expertenurteile, per Augenschein bzw. face-Validität, auch logische Validität, triviale Validität genannt) erfasst, was er zu erfassen vorgibt. Dies ist bei hypothetischen Konstrukten nur schwer möglich. Daher sollte bei hypothetischen Konstrukten die Konstruktvalidität ermittelt werden. Kennwerte für die inhaltliche Validität gibt es nicht.
- **Kriterienbezogene Validität** wird durch eine Korrelation zu einem Kriterium empirisch bestimmt (empirische Validität).

Gütekriterien 2: Validität (Gültigkeit) 3

- Die **Innere Kriterienbezogene Validität** wählt als Kriterium einen schon bestehenden vergleichbaren Test. Es können aber auch multiple Validitäten aus mehreren verwandten (oder auch divergenten) Tests bestimmt werden.
- Die **Äußere Kriterienbezogene Validität** wählt als Kriterium Expertenurteile oder objektive Maße, wie Fehlzeiten, produzierte Stückzahlen.
- Die **Vorhersagevalidität (prognostische Validität)** muss ermittelt werden, wenn mit dem Verfahren Prognosen über zukünftiges Verhalten angestellt werden sollen (z.B. Eignungstests). In einigen Testmanualen finden sich sog. Erwartungstabellen, die angeben, wie viel Prozent der Probanden bei einem bestimmten Testwert z.B. eine Ausbildung gut abschließen.

Gütekriterien 2: Validität (Gültigkeit) 4

- Die **Übereinstimmungsvalidität** zielt hingegen keine Prognose an, so dass Kriterium und Test gleichzeitig erhoben werden können.
- Die **differentielle Validität** gibt verschiedene Validitätskoeffizienten für (a) verschiedene Kriterien oder (b) verschiedene Stichproben an.
- Die **Multiple Validität** belässt es nicht bei der Aufzählung einzelner differentieller Validitätskoeffizienten, sondern vereinigt verschiedene Kriterien und/oder Stichproben mittels multipler Regressionsgleichungen zu einer Validität.
- Die **ökologische Validität** beschreibt zudem die Künstlichkeit bzw. Echtheit der mit einem Verfahren gewonnenen Ergebnisse und gibt damit an, wie sehr die Ergebnisse generalisiert werden dürfen. Reaktive und experimentelle Verfahren schneiden hier schlechter ab als nicht reaktive, nicht manipulative Verfahren.

Gütekriterien 3: Reliabilität (Zuverlässigkeit)

- Das Konzept der **Reliabilität** geht davon aus, dass ein Merkmal „in Wirklichkeit“ eine bestimmte Ausprägung besitzt und dass eine Messung dieser Ausprägung verschieden genau durchgeführt werden kann. Die Genauigkeit ist die Reliabilität.
- Überprüfung der Reliabilität:
 - **Re-Test**: wiederholte Messung, gleiches Instrument, gleiche Objekte, *verschiedene Zeitpunkte. Erinnerungseffekte.*
 - **Parallel-Test**: wiederholte Messung, gleiche Objekte, *ähnliche Instrumente. Eventuell verschiedene Zeitpunkte.*
 - **Split-Half-Verfahren**: Instrument wird in zwei Hälften geteilt (z.B. Items eines Fragebogens) und die Ergebnisse der beiden Hälften werden verglichen.
 - **Innere Konsistenz (Cronbach-Alpha)**: Nicht nur Hälften werden verglichen. Jedes Item wird mit jedem Item verglichen.

Gütekriterien 4: Weitere Kriterien

- Repräsentativität (Generalisierbarkeit)
- Bedeutsamkeit / Relevanz
- Ethischen Kriterien
 - Schutz der Menschenwürde der UntersuchungsteilnehmerInnen
 - Informationspflicht gegenüber den untersuchten Personen
 - Verantwortung der ForscherIn für alle Vorkommnisse während der Untersuchung
 - Freiwillige Teilnahme und Recht auf jederzeitigen Abbruch der Teilnahme
 - Vermeiden psychischer und körperlicher Beeinträchtigungen
 - Anonymität und Datenschutz

3.10 Population und Stichprobe

Population und Grundgesamtheit

- Eine Studie soll Aussagen über eine bestimmte Gruppe von Personen treffen. Welche Gruppe ist gemeint? Wie kann die Gruppe definiert und abgegrenzt werden?
- Die abgegrenzte und definierte Gruppe ist die **Population** bzw. **Grundgesamtheit** der Studie.
- Ist keine **Vollerhebung** möglich, muss eine Stichprobe aus der Population zusammengestellt werden, die für diese möglichst **repräsentativ** ist.

Sampling – Stichprobenauswahl 1

- Eine **Zufallsstichprobe** liegt dann vor, wenn...
... für jedes Element in der Grundgesamtheit die selbe Wahrscheinlichkeit besteht, in die Stichprobe aufgenommen zu werden.
... die Entnahme der einzelnen Elemente unabhängig voneinander erfolgt.
- **Schichtung:** Bei einer proportional geschichteten Stichprobe wird die Grundgesamtheit zunächst in Schichten mit homogenen Merkmalen unterteilt, aus denen dann Zufallsstichproben gezogen werden, deren Größenverhältnis untereinander dem Verhältnis der Teilgesamtheiten in der Grundgesamtheit entspricht.
- **Klumpenstichprobe:** Eine Klumpenstichprobe liegt dann vor, wenn mehrere zufällig ausgewählte Klumpen (natürliche Gruppen, z.B. Schulklassen) vollständig untersucht werden.
- Bei einer **mehrstufigen Auswahl** werden nach einer Klumpenauswahl, in einem zweiten Schritt, die UntersuchungsteilnehmerInnen nach einem anderen Verfahren gewählt.
- Eine bewusste, **gezielte Auswahl** der UntersuchungsteilnehmerInnen ist bei explorativen Studien sinnvoll (z.B. Fallstudien). Rückschlüsse auf eine Grundgesamtheit sind dann aber nur mit Einschränkungen möglich.

Stichprobenauswahl - nicht zufällig

Verfahren	Regel
Bequemlichkeit	Auswahl derjenigen, die für die Studie erreicht werden können.
sehr ähnliche, sehr unterschiedliche	Auswahl von Fällen, die in sich besonders ähnlich sind; oder als Alternative: Auswahl von Fällen, die möglichst unterschiedlich sind.
typische Fälle	Auswahl von Fällen, bei denen man im Vorhinein weiß, dass sie typisch sind und nicht extrem aus dem Rahmen fallen.
kritische Fälle	Auswahl von Fällen, die kritisch sind oder Schlüsselfunktionen haben bei der späteren Anwendung der Studienergebnisse.
Schneeball	Die UntersuchungsteilnehmerInnen verteilen die Fragebögen weiter.
Quotierung	Gezielte Auswahl von Personen, die zur Grundgesamtheit in Hinblick auf bestimmte Merkmale passen.

Stichprobengröße

- Bedeutsam für die Festlegung der Stichprobengröße sind eine Reihe von Faktoren:
 - Forschungsansatz: Qualitative Untersuchungen beruhen auf Daten von einigen wenigen Personen. Das Abbruchkriterium rät dazu mit der Erhebung aufzuhören sobald sich keine neuen Informationen mehr ergeben.
 - Größe der Grundgesamtheit: Wenn es weltweit nur 10 Personen mit der oder der Krankheit gibt, können auch nicht mehr untersucht werden. Ist die Verfügbarkeit gegeben, sollte bei einer kleinen Grundgesamtheit grundsätzlich die gesamte Grundgesamtheit untersucht werden.
 - Verfügbarkeit: Die finanziellen, zeitlichen oder sonstigen Beschränkungen in der Verfügbarkeit einer Stichprobe spielt eine Rolle. Es macht wenig Sinn ein Stichprobe von 100 Personen zu fordern, wenn jede Untersuchung Unsummen kostet.

- Repräsentativität: Die Repräsentativität wächst bei einer echten Zufallsstichprobe mit der Größe der Stichprobe. Bei einer Zufallsstichprobe entscheidet der blinde Zufall über die Repräsentativität. Das ist gut, weil man damit keine systematischen Verzerrungen vornimmt. Das ist schlecht, weil die Stichprobe groß sein muss, damit man „alles mit drin hat“.
Repräsentativ ist eine Stichprobe dann, wenn sie hinsichtlich aller relevanter Merkmale mit der Grundgesamtheit übereinstimmt. Man könnte daher diese relevanten Merkmale in der Stichprobe mit denen in der Grundgesamtheit vergleichen (z.B. über amtliche Statistiken).
Probleme bereiten häufig Untersuchungsverweigerer (echte Zufallsstichproben kann es eigentlich nicht geben). Hier kann man die schnell Antwortenden (die ersten 20%) mit den Nachzüglern (die letzten 20%) vergleichen.

- Statistisches Verfahren: Viele statistische Verfahren erfordern die Normalverteilung der Mittelwerte der Datenstichproben. Diese ist ab 25 bis 50 Befragten Personen pro Untersuchungsgruppe ungefähr erreicht.
- Zu erwartende Effektgröße: In der Regel gilt, dass mit der Größe der Stichprobe auch die Chance wächst tatsächlich vorhandene Unterschiede als signifikant nachzuweisen. Störgrößen mitteln sich bei großen Stichproben heraus und der eigentliche Effekt wird klarer sichtbar. Daher gilt, dass die Stichprobe so groß wie möglich sein sollte.
 - Aus finanziellen oder anderen Gründen muss die Stichprobe aber dennoch meistens begrenzt bleiben. Die Frage danach, wie groß die Stichprobe denn mindestens sein muss, um einen vermuteten Effekt auch zu zeigen, wird durch die Power-Analyse beantwortet.

Power-Analyse Kleine Effekte

Alpha = 0,05 1-Beta = 0,80	Unterschieds- Hypothesen	Zusammenhangs- Hypothesen
2 Gruppen / Variablen / Objekte	Einfache Gruppenvergleiche 310 Fälle pro Gruppe	Einfache Korrelation 614 Fälle Die Angabe meint die Gesamtgröße.
Mehr als 2 Gruppen / Variablen / Objekte	Multiple Gruppenvergleiche 3 Gruppen: 969 Fälle 4 Gruppen: 1096 Fälle 5 Gruppen: 1200 Fälle Die Angabe meint jeweils die Gesamtgröße.	Multiple Korrelation 2 Prädiktoren: 485 3 Prädiktoren: 550 4 Prädiktoren: 602 Die Angabe meint jeweils die Gesamtgröße.

Power-Analyse Mittelgroße Effekte

Alpha = 0,05 1-Beta = 0,80	Unterschieds- Hypothesen	Zusammenhangs- Hypothesen
2 Gruppen / Variablen / Objekte	Einfache Gruppenvergleiche 50 Fälle pro Gruppe	Einfache Korrelation 64 Fälle Die Angabe meint die Gesamtgröße.
Mehr als 2 Gruppen / Variablen / Objekte	Multiple Gruppenvergleiche 3 Gruppen: 159 Fälle 4 Gruppen: 180 Fälle 5 Gruppen: 200 Fälle Die Angabe meint jeweils die Gesamtgröße.	Multiple Korrelation 2 Prädiktoren: 68 3 Prädiktoren: 77 4 Prädiktoren: 85 Die Angabe meint jeweils die Gesamtgröße.

Power-Analyse Große Effekte

Alpha = 0,05 1-Beta = 0,80	Unterschieds- Hypothesen	Zusammenhangs- Hypothesen
2 Gruppen / Variablen / Objekte	Einfache Gruppenvergleiche 20 Fälle pro Gruppe	Einfache Korrelation 22 Fälle Die Angabe meint die Gesamtgröße.
Mehr als 2 Gruppen / Variablen / Objekte	Multiple Gruppenvergleiche 3 Gruppen: 66 Fälle 4 Gruppen: 76 Fälle 5 Gruppen: 80 Fälle Die Angabe meint jeweils die Gesamtgröße.	Multiple Korrelation 2 Prädiktoren: 31 3 Prädiktoren: 36 4 Prädiktoren: 40 Die Angabe meint jeweils die Gesamtgröße.

Power-Analyse G*Power

- Das Programm G*Power wird seit 1992 programmiert und kann als Freeware aus dem Internet bezogen werden.
- Das Programm erlaubt die genau Abschätzung der mindestens nötigen Stichproben-Größe für verschiedene Testverfahren, Alpha- und Beta-Werte.
- Es kann z.B. bezogen werden unter:
<http://www.gpower.hhu.de/>

3.11 Messtheorie

Was ist messen?

- Messen ist die homomorphe Zuordnung eines empirischen Relativs zu einem numerischen Relativ.
- Homomorphie: Strukturerhaltene Abbildung.
- Struktur?
 - Empirisches Relativ: Relationen zwischen Objekten der erfahrbaren/erschließbaren Welt.
 - Numerisches Relativ: Relationen zwischen Zahlen.

Skalenniveaus und Transformation

Skalenniveau	Das darf eine Transformation nicht verändern ...	Zulässige Interpretation
Nominal	Ein-eindeutig Zuordnung	Code, Bezeichnung, Beispiel: Berufe
Ordinal	Reihenfolge	Rangordnung Beispiel: Schulbildung
Intervall	Intervalle zwischen den Zahlen (erlaubt ist die Addition/Subtraktion von Konstanten, sowie die Multiplikation/Division mit Konstanten)	Abstände (Intervalle) zwischen den Zahlen Beispiel: Alter
Verhältnis	Verhältnisse zwischen den Zahlen (erlaubt ist die Multiplikation/Division mit Konstanten)	Verhältnisse zwischen den Zahlen Beispiel: Gehalt
Absolut	Nichts darf verändert werden.	Verhältnisse zwischen den Zahlen, Kardinalzahl Beispiel: Häufigkeiten

Skalenniveaus und Transformation

Skalenniveau	Das darf eine Transformation nicht verändern ...	Zulässige Interpretation
Nominal	Ein-eindeutig Zuordnung	Code, Bezeichnung, Beispiel: Berufe
Ordinal	Reihenfolge	Rangordnung Beispiel: Schulbildung

Metrisch

Intervall	Intervalle zwischen den Zahlen (erlaubt ist die Addition/Subtraktion von Konstanten, sowie die Multiplikation/Division mit Konstanten)	Abstände (Intervalle) zwischen den Zahlen Beispiel: Alter
Verhältnis	Verhältnisse zwischen den Zahlen (erlaubt ist die Multiplikation/Division mit Konstanten)	Verhältnisse zwischen den Zahlen Beispiel: Gehalt
Absolut	Nichts darf verändert werden.	Verhältnisse zwischen den Zahlen, Kardinalzahl Beispiel: Häufigkeiten

Skalenniveaus und Stetigkeit

- **Skalenniveaus (extrem wichtig!)**
 - Nominal
 - Ordinal
 - Metrisch
 - Intervall
 - Verhältnis
 - Absolut
- **Stetigkeit**
 - Diskret
 - Mehrstufig
 - Dichotom
 - Stetig

Skalen-Niveaus I:

Nominal-Skala

Beliebige Zahlen stehen für beliebige Begriffe (z.B. Hausnummern, Geschlecht, Berufe)

Was man darf, was man nicht darf...

- ✓ **Zählen (45 Frauen und 38 Männer nahmen an der Befragung teil)**
- ✓ **Prozentangaben (55% der Befragten waren Frauen)**
- ✓ **Modalwert (die meisten Antwortenden waren Frauen)**
- ⊗ **Kein Median**
- ⊗ **Kein Mittelwert**
- ⊗ **Keine Streuung (Standardabweichung), Varianz**
- ✓ **Häufigkeitstabellen**
- ✓ **Balkendiagramme**
- ✓ **Kreisdiagramme**

Skalen-Niveaus II: Ordinal-Skala

Die Zahlen stehen für Relationen wie größer oder kleiner. Die Ordnung der Zahlen entspricht inhaltlich der Ordnung der Dinge für die sie stehen.

z.B.

- 1 10-15 Jahre
- 2 16-25 Jahre
- 3 26-35 Jahre
- 4 älter als 35 Jahre

Was man darf, was man nicht darf...

- ✓ Zählen (45 Personen fallen in die Kategorie 3)
- ✓ Prozentangaben (55% der Befragten fallen in die Kategorie 3)
- ✓ Modalwert (die meisten Antwortenden gehören zur Kategorie 3)
- ✓ Median (in der Mitte der Antworthäufigkeiten lag die Kategorie 2)
- ⊗ Kein Mittelwert
- ⊗ Keine Streuung (Standardabweichung), Varianz

- ✓ Häufigkeitstabellen
- ✓ Balkendiagramme
- ✓ Kreisdiagramme

Skalen-Niveaus III: Intervall-Skala

Die Zahlen stehen nicht nur für Relationen wie größer oder kleiner, sondern entsprechen in ihren Abständen (Intervallen) inhaltlich und logisch den Abständen von dem was sie bezeichnen.

z.B.

Alter (wenn jemand 2 Jahre älter ist als jemand anders, stimmt der Abstand 2 unabhängig davon ob

- a) die eine Person 10 und die andere 12 Jahre alt ist oder
- b) die eine Person 80 und die andere 82 Jahre alt ist.

Was man darf, was man nicht darf...

- ✓ Zählen (45 Personen waren 20 Jahre alt)
[wenig sinnvoll!]
- ✓ Prozentangaben (55% der Befragten waren 20 Jahre alt)
[wenig sinnvoll!]

- ✓ Modalwert (die meisten Antwortenden waren 25 Jahre alt)
- ✓ Median (die Mitte der Altersverteilung liegt bei 24 Jahren)
- ✓ Mittelwert (im Durchschnitt betrug das Alter 24 Jahre)
- ✓ Streuung (Standardabweichung), Varianz (die Standardabweichung beträgt ± 2 Jahre)

- ✓ Häufigkeitstabellen
[nur sinnvoll, wenn man Gruppen bildet]
- ✓ Balkendiagramme
[auch Mittelwert und Streuung angeben und eventuell einzeichnen]
- ✓ Kreisdiagramme
[nur sinnvoll, wenn man Gruppen bildet]

Übungsaufgaben

Beobachtungen

- Nominal und Ordinal sind immer diskret.
- Metrische Messungen können stetig oder diskret sein.
- Können metrische Merkmale auch dichotom sein?

- Die Art der Kodierung kann aus einem stetigen Merkmal ein diskretes machen.

- Eigentlich gibt es nach der Messung keine stetigen Merkmale mehr. Es ist wichtig sich zu fragen, ob das „echte“ Merkmal hinter der Messung stetig ist.

Skalenniveaus und Stetigkeit

- **Skalenniveaus (extrem wichtig!)**

- Nominal
- Ordinal
- Metrisch
 - Intervall
 - Verhältnis
 - Absolut



Verbesserungsmethoden:
z.B. Nominale Objekte zählen führt zu einer Häufigkeit, die metrisch ist.

- **Stetigkeit**

- Diskret
 - Mehrstufig
 - Dichotom
- Stetig

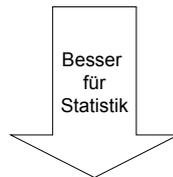


Verbesserungsmethoden:
z.B. Zusammenfassung vieler diskreter Messungen nähert sich einer stetigen Messung an.

Skalenniveaus und Stetigkeit

- **Skalenniveaus (extrem wichtig!)**

- Nominal
- Ordinal
- Metrisch
 - Intervall
 - Verhältnis
 - Absolut

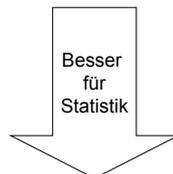


Verschlechterungsmethoden:
z.B. Metrisches Merkmal mit Kategorien abfragen.

Kann helfen:
In Bezug auf Anonymität.

- **Stetigkeit**

- Diskret
 - Mehrstufig
 - Dichotom
- Stetig



Verschlechterungsmethoden:
z.B. Gruppenbildung.

Kann helfen:
Extremgruppenvergleich.

Wofür der ganze Aufwand?

- Maße der zentralen Tendenz ...
- Dispersionsmaße ...
- Häufigkeiten / relative Häufigkeiten ...

3.12 Skala und Item

Messung hypothetischer Konstrukte – ein Beispiel

- + Dieser Arzt/diese Ärztin hat mich freundlich behandelt.
 - Ich habe gewisse Zweifel über die Fähigkeit dieses Arztes / dieser Ärztin.
 - Dieser Arzt / diese Ärztin wirkte kühl und unpersönlich.
 - + Dieser Arzt/diese Ärztin hat sein/ihr Bestes getan, um mich nicht zu beunruhigen.
 - + Dieser Arzt/diese Ärztin hat mich so sorgfältig es notwendig war, untersucht.
 - Dieser Arzt/diese Ärztin hätte mich rücksichtsvoller behandeln sollen.
 - Ich habe gewisse Zweifel über die von diesem Arzt/dieser Ärztin empfohlene Behandlung.
 - + Dieser Arzt / diese Ärztin wirkte sehr kompetent und erfahren.
 - + Dieser Arzt/diese Ärztin schien ein echtes und persönliches Interesse an mir zu haben.
 - Dieser Arzt/diese Ärztin hat mich mit vielen unbeantworteten Fragen über meinen Zustand und die notwendige Behandlung zurückgelassen.
 - Dieser Arzt/diese Ärztin verwendete Fachausdrücke, die ich nicht verstanden habe.
 - + Ich habe großes Vertrauen in diesen Arzt/diese Ärztin.
 - + Ich habe das Gefühl, dass ich diesem Arzt/dieser Ärztin sehr persönliche Probleme hätte anvertrauen können.
 - Ich habe mich nicht getraut, diesem Arzt / dieser Ärztin Fragen zu stellen.
- (Quelle: Langewitz, Keller & Denz, 1995)

Messung hypothetischer Konstrukte

- **Skalierung: z.B. ja / nein**
 – Zählen der Zustimmungen zum Konstrukt (+ und – beachten)
- **Skalierung über Ratings**



**Skalenwert gleich Summe oder
 Skalenwert gleich Mittelwert (Umgang mit fehlenden Werten ist beim Mittelwert besser)**

Herleitung der klassischen Testtheorie

Klassische Testtheorie 1

- Klassische Testtheorie

$$x = \mu + e$$

Der Messwert x entspricht dem wahren Wert μ plus einem Fehler e .
 Der Fehler kann minimiert werden, wenn viele x erhoben und gemittelt werden.

Ein Mittelwert ist gegeben durch:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

n : Anzahl der Messungen
 x : Messwert
 i : Laufvariable für den 1., den 2., 3. ... n-ten Messwert.

Die Varianz s^2 (durchschnittliche quadrierte Abweichung vom Mittelwert) ist gegeben durch:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$S_x^2 = S_\mu^2 + S_e^2$ Varianz der Messwerte ist gleich Varianz der wahren Werte plus Fehlervarianz

$r = \frac{S_\mu^2}{S_x^2}$ Die Reliabilität r ist das Verhältnis der Varianz der wahren Werte zur Varianz der gemessenen Werte. Man spricht hier auch von Varianzaufklärung.

$r = 1 - \frac{S_e^2}{S_x^2}$ Die Reliabilität r ist auch Eins minus das Verhältnis der Varianz der Fehler zur Varianz der gemessenen Werte. Ist die Fehlervarianz groß, so ist die Reliabilität klein.

$\alpha = r = \frac{c}{c-1} \left[1 - \frac{\sum_j^c s_j^2}{S_x^2} \right]$ Die Varianzen jedes einzelnen Items ($j = 1$ bis c) werden aufsummiert und durch die Varianz des Skalenwertes x geteilt. Eins minus diese Zahl wird mit der Zahl der Items – dividiert durch die Zahl der Items minus Eins – multipliziert. Diese Form der Reliabilität heißt Cronbach Alpha

4. Datenverarbeitung

Goldene Regeln Dateneingabe

- Fragebögen vor der Eingabe mit fortlaufenden eindeutigen Nummern versehen!
- Möglichst alle Fragebogendaten als Zahlen kodieren!
- Immer genau die Daten eingeben, die im Fragebogen stehen! Zusammenfassungen, Gruppenbildung etc. wird erst später mit der Software vorgenommen.
- Für dichotome Variablen (ja / nein; männlich / weiblich) 1 und 0 vergeben und nicht 1 und 2!

Vom Fragebogen zur Datendatei

Kodierung des Fragebogens in einer leeren Papierversion

Definition der Variablen in Excel oder SPSS

Dateneingabe mit laufender Nummerierung. Fehlende Daten auslassen

Fragebogenkodierung 1

Laufnummer **Wie zufrieden sind Sie mit Ihrer Physiotherapie?**
Patientenbefragung zur Qualität von physikalischen Instituten

Bitte beantworten Sie folgende Fragen, die sich auf das physikalische Institut beziehen, in dem Sie kürzlich eine oder mehrere Therapien in Anspruch genommen haben.

Bitte zutreffendes Kästchen ankreuzen

1	1. Wie oft haben Sie bereits eine physikalische Therapie in einem physikalischen Institut in Anspruch genommen?
2	<input type="checkbox"/> noch nie, das war meine erste Physiotherapie <input type="checkbox"/> bereits 1 mal zuvor Anzahl mehr als 1 mal zuvor und zwar ungefähr _____ mal
2	2. Über welchen Zeitraum wurde die physikalische Therapie durchgeführt?
2	<input type="checkbox"/> bis zu 2 Wochen Wochen mehr als zwei Wochen und zwar ungefähr _____ Wochen
3	3. Welches physikalische Institut haben Sie kürzlich in Anspruch genommen?
3	Codeliste _____

Fragebogenkodierung 2

4. Auf wessen Empfehlung hin wählen Sie das oben genannte Institut?

0/1 Arzt **0 nein / 1 ja**

0/1 Krankenhaus

0/1 Empfehlungen durch Bekannte

0/1 Selbst gewählt

0/1 Sonstiges _____

Texteingabe bzw. Codeliste

5. Wie klar und vollständig waren die Informationen, die Sie vor der Therapie bekommen haben?

Bitte beurteilen Sie nach dem Schulnotensystem:

1 **2** **3** **4** **5**

sehr gut gut befriedigend genügend nicht genügend

-1 lässt sich nicht sagen / weiß ich nicht

6. Wurden die Aufnahmeformalitäten (Anmeldung, Therapieplan etc.) schnell und unkompliziert durchgeführt?

Bitte beurteilen Sie nach dem Schulnotensystem:

sehr gut gut befriedigend genügend nicht genügend

lässt sich nicht sagen / weiß ich nicht

Fragebogenkodierung 3

15. Hat sich Ihr gesundheitlicher Zustand aufgrund der durchgeführten Therapie(n) verbessert?

0 **1**

nein ja

0 nein / 1 ja

16. Insgesamt bin ich mit Ihrer Therapie.

1 **2** **3** **4** **5**

sehr zufrieden zufrieden wenig zufrieden unzufrieden sehr unzufrieden

17. Angaben zu Ihrer Person:

0 **1**

männlich weiblich

Alter: Zahl Jahre

Dateneingabe im Excel

Jede Zeile ist ein Fall

Jede Spalte ist eine Variable

Fehlende Angaben bleiben leer

Nur Zahlen eingeben und nicht m / w

5. Deskriptive Statistik

Häufigkeitstabellen

Excel: Daten->PivotTabelle

GEBO	GEBJ	GESCHL	Gesamtergebnis
1923	0	1	1
1941		2	2
1945		1	1
1946		3	3
1948		3	3
1950	1	1	1
1952		1	1
1954	1	1	1
1955		1	1
1956		1	1
1957	1	1	1
1958	1	1	1
1960		1	1
1961	1	1	1
1962		1	1
1963	1	4	5
1964	4	2	6
1965	3	3	6
1966		5	5
1967	2	2	4
1968	2	8	10
1969	1	2	3
1970	1	4	5
1971	1	5	6
1972	4	6	10
1973	4	5	9
1974	8	16	24
1975	6	6	12
1976	8	5	13
1977	5	9	14
1978	5	7	12
1979	4	1	5
1980	1	4	5
1981	2	3	5
1982	1	4	5
1983	1	4	5
1984	1	1	1
1985		2	2
1987		1	1
Gesamtergebnis	70	121	191

Tabelle 1: Altersverteilung und Geschlecht (absolute Häufigkeiten)

Geburtsjahr	Frauen	Männer	Summe
< 1950	1	10	11
1950 – 1960	3	4	7
1961 – 1970	15	19	44
1971 – 1980	46	64	110
> 1980	5	14	19
Summe	70	121	191

Tabelle 1: Altersverteilung und Geschlecht (relative Häufigkeiten)

Geburtsjahr	Frauen-Anteil	N (%)
< 1950	9,1%	11 (5,7%)
1950 – 1960	42,9%	7 (3,7%)
1961 – 1970	34,1%	44 (23,0%)
1971 – 1980	41,8%	110 (57,6%)
> 1980	26,3%	19 (10,0%)
Insgesamt	36,6%	191 (100,0%)

Diagramme: Häufigkeiten, Balkendiagramm

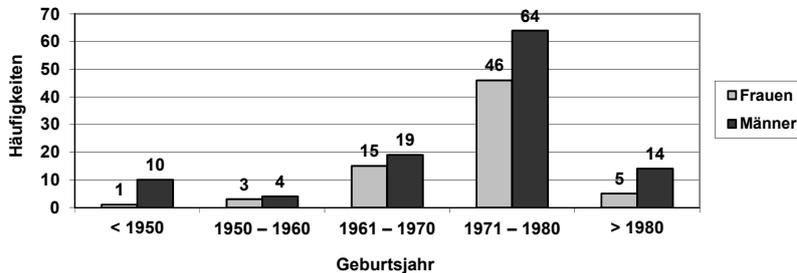


Abbildung 1: Altersverteilung und Geschlecht (absolute Häufigkeiten)

Diagramme: Häufigkeiten, Kreisdiagramm

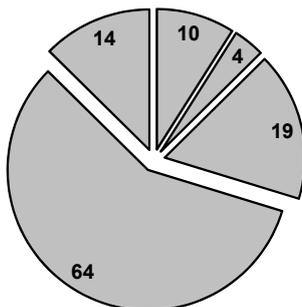


Abbildung 2: Altersverteilung der Männer (absolute Häufigkeiten)

5.1 Maße der zentralen Tendenz

- **Modalwert:** häufigster Wert
- **Median:** Ist der Wert, über dem genau so viele Fälle liegen wie unter ihm. Der Median ist die Mitte der sortierten Häufigkeits-Verteilung.
- **Arithmetisches Mittel:** Summe aller Werte, dividiert durch die Anzahl der Werte.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

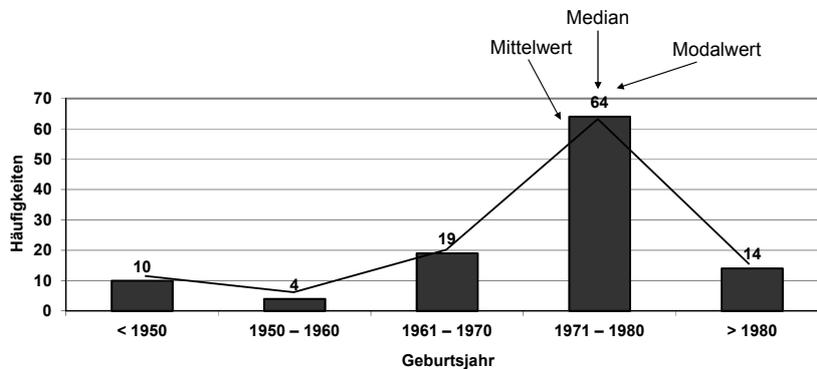
- **Geometrisches Mittel (Verhältnisskala – Mittelwert eines multiplikativen Faktors):**

$$GM = \sqrt[n]{\prod_{i=1}^n x_i}$$

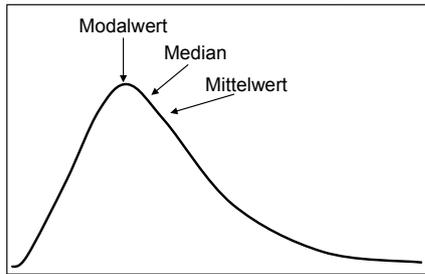
- **Harmonisches Mittel (Verhältnisskala – Mittelwert von Brüchen mit konstanten Zähler):**

$$HM = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Modalwert, Median, Mittelwert

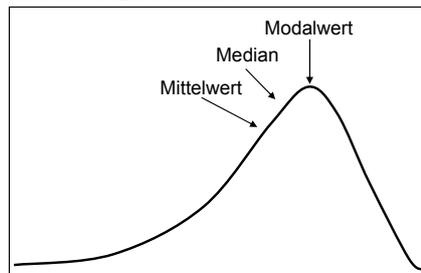


Schiefe Verteilungen



Linkssteil bzw. rechtsschief

Rechtssteil bzw. linksschief



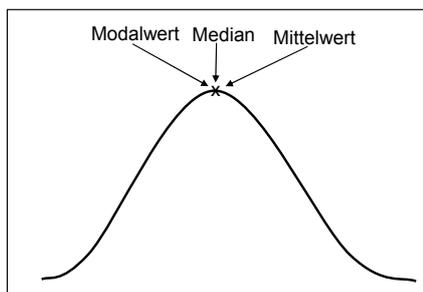
➔ Schiefe

$$Sch = \frac{\bar{x} - \text{Modalwert}}{s}$$

Sch < 0: Rechtssteil bzw. linksschief

Sch > 0: Linkssteil bzw. rechtsschief

Symmetrische Verteilungen



5.2 Streuungsmaße

- **Variationsbreite (range):** größter und kleinster Wert
- **Quartilsabstand:** Differenz zwischen dem Perzentil 25 und dem Perzentil 75. **Perzentil:** x-tes Perzentil ist diejenige Merkmalsausprägung, die x Prozent der Verteilungsfläche abschneidet (Median ist Perzentil 50)

- **Varianz**

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

- **Streuung bzw. Standardabweichung**

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Streuungsparameter Stichprobe oder Population

- **Stichproben-Varianz**

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

- **Stichproben-Streuung bzw. Stichproben-Standardabweichung**

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Ist n groß, dann gleichen sich die Ergebnisse zunehmend an und es wird unbedeutend, ob mit n oder mit n-1 gerechnet wird.

Herleitungen

6. Inferenzstatistik – Grundlagen

Grundidee – ein Beispiel

- Behauptung: „Ich kann zaubern, die Münze fällt immer auf Kopf!“
 - Ab wann wäre diese Behauptung empirisch gesehen statistisch signifikant abgesichert?
 - Wenn bei 20 mal Werfen 10 mal Kopf kommt?
 - Wenn bei 20 mal Werfen 11, oder 12, oder 13 mal Kopf kommt?
 - Null-Hypothese: „Kein Zaubern, sondern Zufall!“
 - Die Null-Hypothese geht von Zufall aus. Die höchste mathematische Wahrscheinlichkeit für Zufall wäre bei 10 mal Kopf (bei 20 mal Werfen) gegeben.
 - Es lässt sich ausrechnen, wie groß die Wahrscheinlichkeit für 11, 12, 13 usw. mal Kopf ist. Dabei wird davon ausgegangen, dass 10 mal normal wäre und jeder Wurf mit 50%-iger Wahrscheinlichkeit Kopf ergibt.
 - Die Berechneten Wahrscheinlichkeiten gehen also von der Null-Hypothese (50% für Kopf) aus.

- Der Münzwurf ist mathematisch gesehen binomial verteilt:

$$P(X=k) = \frac{n!}{k! (n-k)!} * p^k * (1-p)^{(n-k)}$$

P: Wahrscheinlichkeit für das Ereignis Kopf = 0,5

$P(x=k)$ Wahrscheinlichkeit dafür, dass das Ereignis X (Kopf) bei n Versuchen exakt k mal auftritt.

Im konkreten Beispiel ist jedoch eigentlich $P(X \leq k)$, also die Summe aller

Einzelwahrscheinlichkeiten für $k=1, k=2, \dots, k=15$ (oder bis $k=12$ oder $k=13$, je nach dem) gefragt, um die „Zauberfrage“ beantworten zu können.

- Für $n=20$ Würfe und $k=15$ mal Kopf beträgt die Wahrscheinlichkeit nur mehr 2%. Das dieses Ereignis (15 Treffer) mit einer 50-50-Münze im Einklang steht ist also sehr unwahrscheinlich.
- Weil die Nullhypothese sehr unwahrscheinlich ist, wird sie fallen gelassen. Das heißt, das Ereignis ist signifikant.

Sind die Frauen im Beispiel jünger?

Tabelle 1: Geburtsjahr nach Geschlecht

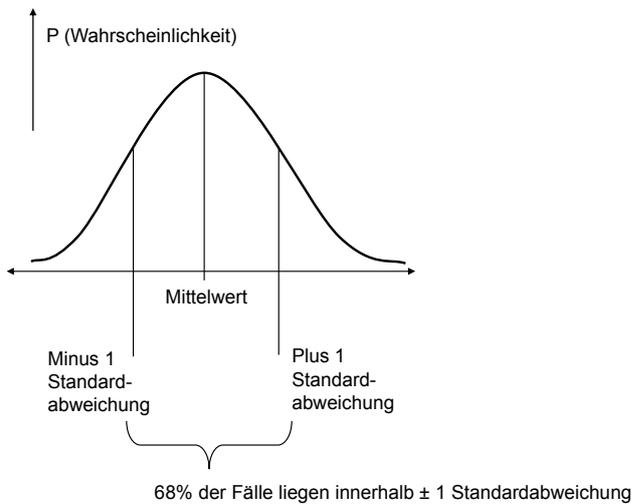
Geschlecht	Mittelwert	Streuung	N
Frauen	1972,73	6,70	121
Männer	1970,62	10,59	70
Gesamt	1971,38	9,40	191

Der Unterschied zwischen den beiden Mittelwerten beträgt rund 2 Jahre (exakt sind es 2,11 Jahre).

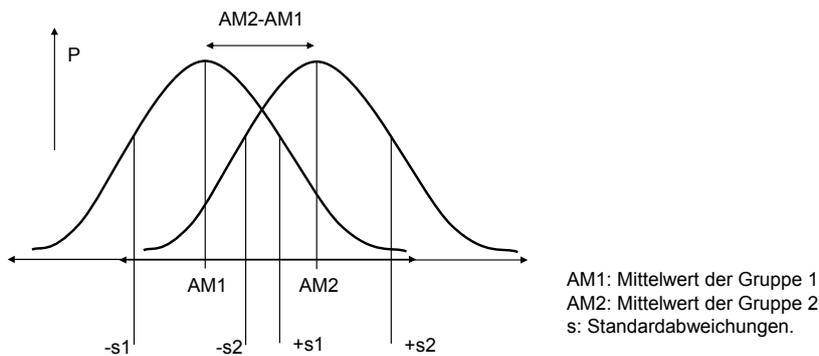
Aber es gibt Frauen, die älter sind als der Durchschnitt der Männer und Männer, die jünger sind als der Durchschnitt der Frauen.

Ob der Unterschied statistisch bedeutsam ist (also signifikant ist) kann mit Hilfe der Wahrscheinlichkeitsrechnung entschieden werden.

Normalverteilung

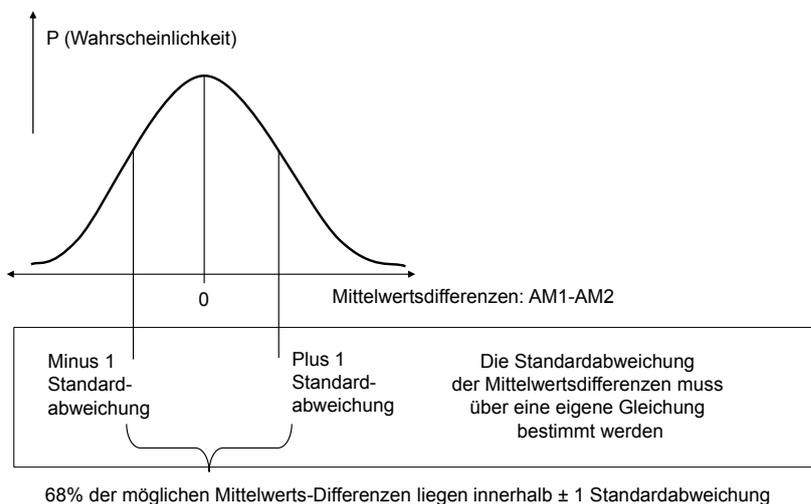


Vergleich zweier Mittelwerte

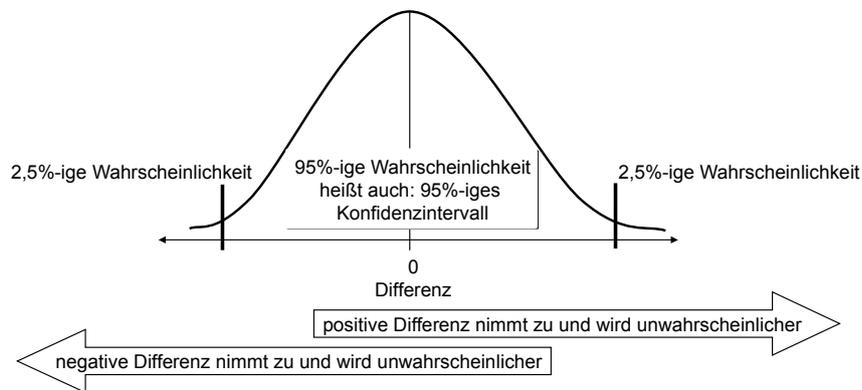


- Ziel wäre es die Wahrscheinlichkeit für den Unterschied zwischen AM1 und AM2 zu kennen.
- Die Verteilung müsste die Nullhypothese (kein Unterschied zwischen AM1 und AM2) als wahrscheinlichsten Fall ansehen.
- Tatsächlich sind Mittelwerts-Differenzen normalverteilt mit einer erwarteten Differenz von Null und einer Standardabweichung, die sich aus den gegebenen Standardabweichungen berechnen lässt.

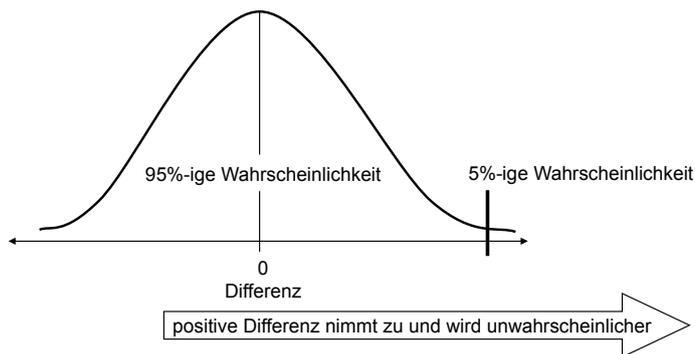
Normalverteilung von Mittelwerts-Differenzen



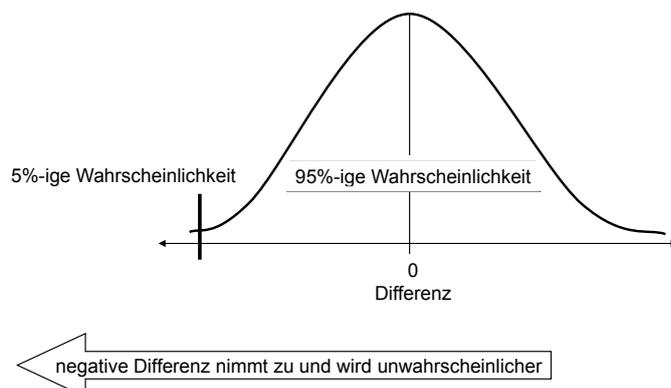
Vergleich zweier Mittelwerte 2-seitig



Vergleich zweier Mittelwerte 1-seitig



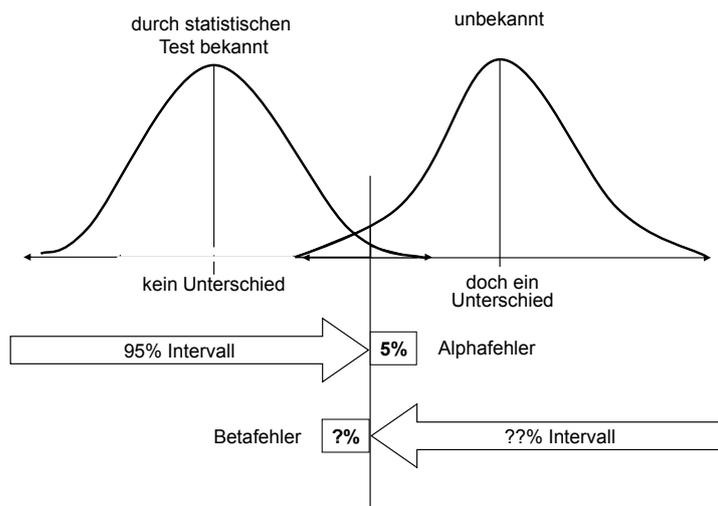
Vergleich zweier Mittelwerte 1-seitig



Zusammenfassung

- Der Vergleich zweier Mittelwerte geht davon aus, dass die **Differenz der Mittelwerte** normalverteilt ist.
- Die Nulldifferenz ist die Mitte der Verteilung. Eine Nulldifferenz ist also am wahrscheinlichsten. **Jeder statistische Test geht von der Null-Hypothese aus.**
- Weicht die tatsächliche Differenz von der Nullhypothese ab, so kann eventuell ein signifikanter Unterschied vorliegen.
- Beträgt die Wahrscheinlichkeit dafür, dass die tatsächliche Differenz Null ist nur mehr 5%, so gilt die Differenz als signifikant. Die Null-Hypothese wird verworfen und die Alternativ-Hypothese wird akzeptiert.
- Der **Alpha-Fehler** beträgt hier $\alpha = 5\%$ (Man schreibt statt $\alpha = 5\%$ auch gerne $\alpha = 0,05$).
- Bei einem Alpha-Fehler von nur mehr 1% spricht man von **sehr signifikanten** Differenzen ($\alpha = 1\%$ bzw. $\alpha = 0,01$).

Alpha- und Betafehler



7. Einfache Testverfahren zum Vergleich von zwei Gruppen

Je nach Datenlage und Fragestellung sind verschiedene Tests zu wählen. Hier folgen nur einfache Testverfahren. Eine Übersicht über weitere Verfahren liefert Kapitel 10.3.6 (S. 80), Kapitel 10.3.8 (S. 113).

7.1 T-Test

Voraussetzungen

- **t- bzw. Normalverteilung der Differenzen.** Ist bei n_1 und n_2 jeweils größer 25 (oder 30 oder 50) automatisch gegeben. Bei kleineren n müssen die Messwerte normalverteilt sein. Dies kann mittels **Kolmogorov-Smirnov-Anpassungstest** geprüft werden. Liegt keine Normalverteilung vor muss ein anderer Test gewählt werden, z.B. der **U-Test** von Mann-Whitney.
- **Varianzhomogenität.** Die Varianzen müssen in beiden Stichproben in etwa gleich groß sein. Dies kann geprüft werden mit dem **F-Test**. Ist das nicht der Fall gibt es Korrekturformeln.
- **Unabhängigkeit der Messwerte.** Die Daten müssen aus zwei voneinander unabhängigen Stichproben stammen. Zwei Messungen an der selben Stichprobe zu verschiedenen Zeiten, verletzen diese Forderung. Hierfür gibt es einen **t-Test für abhängige Stichproben**.
- Sind die Daten abhängig und ist die Normalverteilung ebenfalls verletzt, ist der **Wilcoxon-Rangsummen-Test** zu rechnen.

t-Test

- Vergleich zweier Stichprobenmittelwerte aus unabhängigen Stichproben:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}} \quad \hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$FG = n_1 + n_2 - 2$$

t ist eine Zufallsvariable, die für kleine Stichproben mit FG t -verteilt und für größere Stichproben ($n_1 + n_2 \geq 50$) annähernd normalverteilt ist.

Rechnet man den Test von Hand, so muss in einer Tabelle nachgeschlagen werden, wie wahrscheinlich der errechnete Wert für t bei den gegebenen Freiheitsgraden ist.

Die Wahrscheinlichkeit sinkt mit wachsendem t . Ist die Wahrscheinlichkeit kleiner 0,05 so liegt ein signifikanter Unterschied vor.

Beispiel

Tabelle 1: Geburtsjahr nach Geschlecht

Geschlecht	Mittelwert	Streuung	N
Frauen	1972,73	6,70	121
Männer	1970,62	10,59	70
Gesamt	1971,38	9,40	191

$$\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = 1,251$$

$$t = 2,11 / 1,251 = 1,687$$

$$FG = 189$$

Tabelle der t-Verteilung (Ausschnitt)

FG	t (für $\alpha = 5\%$, 1-seitig)
80	1,664
90	1,662
100	1,660
150	1,655
200	1,653

Ist der positive Wert von $t \geq t_{\text{Tabelle}}$, dann ist der Unterschied signifikant. t_{Tabelle} wird auch als kritische Grenze bzw. t_{krit} bezeichnet.

7.2 T-Test im SPSS

T-Test – unabhängige Stichproben

Teil Anmerkungen Gruppenstatistiken Test bei unabhängigen Stichp

T-Test

Gruppenstatistiken				
	GSCHL	N	Mittelwert	Standardfehler des Mittelwertes
ITEM16	männlich	123	1,67	,70
Gesamtbefriedenheit	weiblich	248	1,75	,74

Test bei unabhängigen Stichproben								
	Levene-Test der Varianzhomogenität	T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehl der Differenz
ITEM16	Varianzen sind gleich	,206	,650	-1,037	369	,300	-8,33E-02	8,03E-1
Gesamtbefriedenheit	Varianzen sind nicht gleich			-1,080	257,815	,290	-8,33E-02	7,86E-1

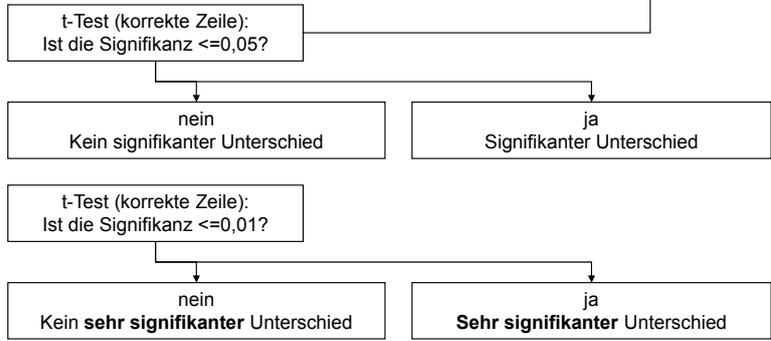
Levene-Test: Ist die Signifikanz $\geq 0,20$?

nein: Varianzen sind nicht gleich: untere Zeile

ja: Varianzen sind gleich: obere Zeile

T-Test – unabhängige Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwerte			
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz
ITEM16 Gesamtzufriedenheit	Varianzen sind gleich	,206	,650	-1,037	369	,300	-8,331
	Varianzen sind nicht gleich			-1,060	257,815	,290	-8,331

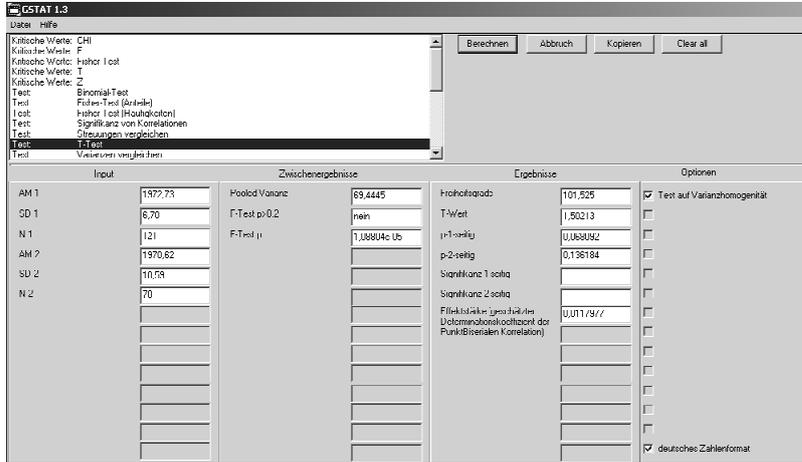


1- oder 2-seitig

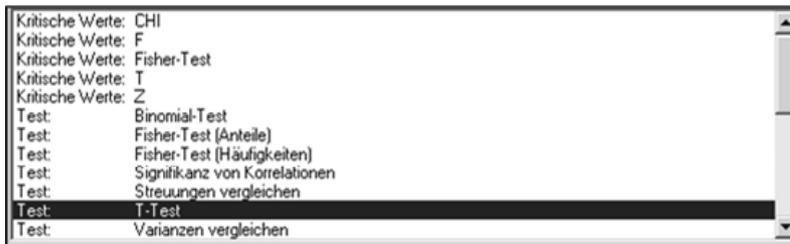
- **Ungerichtete Hypothese:** Die von SPSS angegebene Wahrscheinlichkeit ist 2-seitig zu interpretieren.
- **Gerichtete Hypothese:** Für den 1-seitigen Wert gilt, dass er die Hälfte des 2-seitigen Wertes beträgt.

7.3 T-Test – GStat

<http://www.complexity-research.com/DownSoft.htm>



Verfahren auswählen



Werte eintragen (vorher ausrechnen)

Input	
AM 1	1972,73
SD 1	6,70
N 1	121
AM 2	1970,62
SD 2	10,59
N 2	70

Tabelle 1: Geburtsjahr nach Geschlecht

Geschlecht	Mittelwert	Streuung	N
Frauen	1972,73	6,70	121
Männer	1970,62	10,59	70
Gesamt	1971,38	9,40	191

8. Vergleich mehrerer Gruppen

8.1 Chi-Quadrat

Vergleich erwarteter und beobachteter Häufigkeiten:

$$\text{Testgröße: } \chi^2 = \sum_{j=1}^k \frac{(f_{b(j)} - f_{e(j)})^2}{f_{e(j)}}$$

χ^2 -Methoden basieren auf einen Vergleich von beobachteten f_b und erwarteten Häufigkeiten f_e .

Die Freiheitsgrade FG eines χ^2 -Wertes entsprechen der Anzahl der Summanden k abzüglich der Bestimmungsstücke für die Berechnung der erwarteten Häufigkeiten, die aus den beobachteten Daten abgeleitet werden (in der Regel $FG=k-1$).

Der Test darf nur angewendet werden, wenn nicht mehr als 20% der erwarteten Häufigkeiten Werte kleiner 5 aufweisen und keine der erwarteten Häufigkeiten Werte kleiner 1 aufweisen.

8.2 Einfaktorielle Varianzanalyse

(ANOVA, Analysis Of Variance) 1

- Die einfaktorielle Varianzanalyse untersucht den Zusammenhang einer mehrkategorialen Gruppierungsvariablen (unabhängigen Variablen) mit einer metrischen abhängigen Variablen.
- Die Alternativ-Hypothese behauptet, dass signifikante Mittelwertsunterschiede zwischen den p Populationsgruppen bestehen. Dagegen sind nach der Nullhypothese die Mittelwerte der abhängigen Variablen in den Populationsgruppen gleich.
- Prinzipiell könnten hier auch mehrere T-Tests gerechnet werden. Dabei ist aber zu beachten, dass die Wahrscheinlichkeit für ein signifikantes Ergebnis dadurch ansteigt und entsprechend korrigiert werden muss. (**Alpha-Fehler-Korrektur**)
- **Die Gesamtvarianz wird aufgrund der Varianz der Stichprobe(n) geschätzt:**

$$\hat{\sigma}_{tot}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

x_i meint hier alle gemessenen Werte unabhängig vom Treatment (Gruppe, Kategorie)

- **Bestimmung der Treatmentvarianz:**

Hat das Treatment einen Einfluss, so sollten sich die Mittelwerte der einzelnen Faktorstufen vom Gesamtmittelwert unterscheiden. Man ersetzt daher bei allen Faktorstufen die Messwerte x_i durch den jeweiligen Mittelwert der Faktorstufe um die Abweichungen der Faktorstufen vom Gesamtmittelwert zu bestimmen:

$$\hat{\sigma}_{treat}^2 = \frac{\sum_{j=1}^p n_j \cdot (\bar{x}_j - \bar{x})^2}{p-1}$$

j steht hier für die Faktorstufen.

• **Bestimmung der Fehlervarianz:**

Wenn die abhängige Variable ausschließlich von der unabhängigen Variablen und von keinen anderen Faktoren (Störgrößen) beeinflusst würde, müssten alle Messwerte innerhalb einer Faktorstufe gleich sein, also dem Mittelwert der entsprechenden Stichprobengruppe entsprechen. Der Einfluss der Störvariablen wird also gerade durch das Ausmaß charakterisiert, in dem Messwerte vom Mittelwert der jeweiligen Faktorstufe abweichen. Hieraus folgt für die geschätzte Fehlervarianz aller Faktorstufen:

$$\hat{\sigma}^2_{Fehler} = \frac{\sum_{j=1}^p \sum_m^{n_j} (x_m - \bar{x}_j)^2}{n - p}$$

j steht hier für die Faktorstufen.
 m steht für die Messwerte in den Faktorstufen.

	Gesamt	Treatment	Fehler
Varianz	$\hat{\sigma}^2_{tot} = \frac{\sum_{j=1}^p (x_j - \bar{x})^2}{n - 1}$	$\hat{\sigma}^2_{treat} = \frac{\sum_{j=1}^p n_j \cdot (\bar{x}_j - \bar{x})^2}{p - 1}$	$\hat{\sigma}^2_{Fehler} = \frac{\sum_{j=1}^p \sum_m^{n_j} (x_m - \bar{x}_j)^2}{n - p}$
Quadratsumme	$QS_{tot} = \sum_{j=1}^p (x_j - \bar{x})^2$	$QS_{treat} = \sum_{j=1}^p n_j \cdot (\bar{x}_j - \bar{x})^2$	$QS_{Fehler} = \sum_{j=1}^p \sum_m^{n_j} (x_m - \bar{x}_j)^2$
Freiheitsgrade	$n - 1$	$p - 1$	$n - p$

Der prozentuale Anteil an der Aufklärung der Gesamtvarianz, die auf die Varianz der unabhängigen Variable zurückzuführen ist (Varianzaufklärung), beträgt:

$$\frac{QS_{treat}}{QS_{tot}} \cdot 100\%$$

• **Der eigentliche Signifikanztest:**

Im Signifikanztest wird das Verhältnis von Treatmentvarianz zur Fehlervarianz getestet. Als Prüfgröße wird ein F-Wert ermittelt. Eine relativ große Varianzaufklärung (und damit eine große Treatmentvarianz) führt zu einem hohen F-Wert. Wenn es dagegen keine Unterschiede zwischen den Mittelwerten der einzelnen Faktorstufen und damit keinen Zusammenhang zwischen der unabhängigen und abhängigen Variablen gibt, so ist die Treatmentvarianz im Vergleich zur Fehlervarianz relativ klein und wir erhalten einen relativ kleinen F-Wert:

$$F = \frac{\hat{\sigma}^2_{treat}}{\hat{\sigma}^2_{Fehler}}$$

$FG_{Zähler} = p - 1, FG_{Nenner} = n - p$

Vergleich mehrerer Mittelwerte/Gruppen/Treatments 2

- Für mehrfaktorielle Designs gibt es mehrfaktorielle Varianzanalysen.
- Spezielle Varianzanalysen sind für Messwiederholungsdesigns entwickelt worden.
- Der Unterschied zwischen *fixed* und *random factors* spielt bei der mehrfaktoriellen Varianzanalyse eine wichtige Rolle.
- Bei der mehrfaktoriellen Varianzanalyse können Haupteffekte und Interaktionseffekte einzeln geprüft werden.
- Voraussetzungen sind Intervallskalenniveau, Normalverteilung (bei n pro Stufe > 25 immer gegeben und kein Problem), Varianzhomogenität zwischen den Gruppen. Diese kann getestet werden mit einem F-Test, bei dem $F = s^2_{\max} / s^2_{\min}$; FG sind jeweils n der Gruppe minus Eins.
- Bei schlechterer Datenqualität oder anderen Verletzungen der Voraussetzungen kann der **Kruskal-Wallis-Test** gerechnet werden.
- Eine Varianzanalyse sagt nur, dass Unterschiede zwischen den Gruppen bestehen, nicht jedoch wo. Dafür gibt es so genannte **Kontraste**.

9. Zusammenhangshypothesen

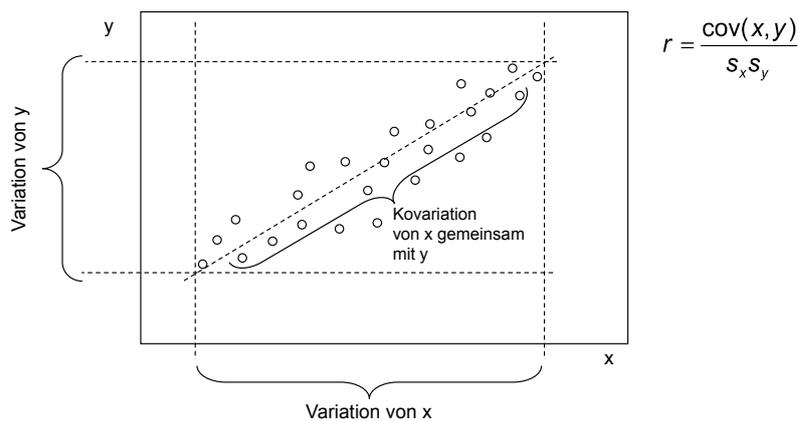
Korrelationsrechnung

- Für die Überprüfung des Zusammenhanges zwischen zwei intervallskalierten Datensätzen wird die Produkt-Moment-Korrelation (auch Pearson- oder Bravais-Pearson-Korrelation genannt) berechnet.

$$r = \frac{\text{cov}(x,y)}{s_x s_y} \quad \text{cov}(x,y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

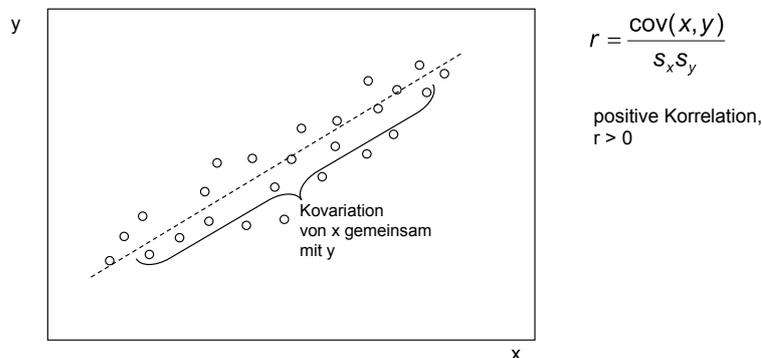
r steht für die Korrelation,
 s sind die Standardabweichungen der beiden Variablen x und y
 Cov(x,y) ist die Kovarianz, also die gemeinsame Varianz der beiden Variablen.

- Ist $r < 0$ liegt eine negative Korrelation vor. Ist $r > 0$ eine positive. Ist $r = 0$ liegt keine Korrelation vor.
- r kann maximal -1 bzw. 1 betragen.

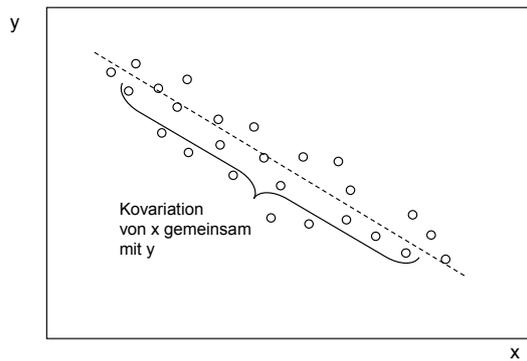


Die eingezeichnete Variationsbreite ist nicht identisch mit der Streuung oder der Varianz, weil zu groß. Dadurch wird die Grafik jedoch klarer.

Positive Korrelation



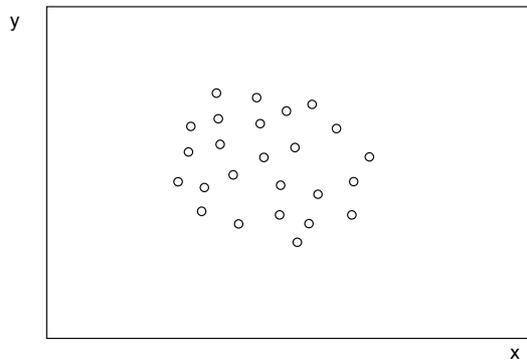
Negative Korrelation



$$r = \frac{\text{cov}(x,y)}{s_x s_y}$$

negative Korrelation,
 $r < 0$

Null-Korrelation



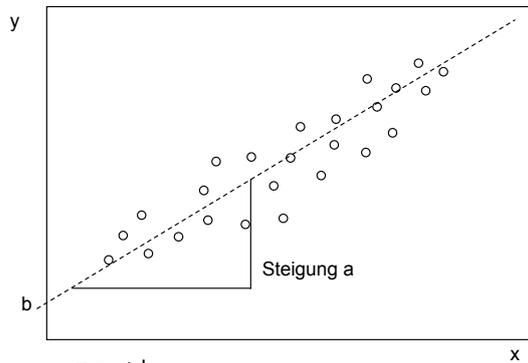
$$r = \frac{\text{cov}(x,y)}{s_x s_y}$$

keine Korrelation,
 $r = 0$

Beurteilung von Korrelationen

- Korrelationen müssen auf Signifikanz geprüft werden.
- Der Determinationskoeffizient $D = r^2$ gibt zudem die Varianzaufklärung an.
- In der Psychologie liegen viele Korrelationen nicht höher als 0,3.
- In den Naturwissenschaften sind 0,9 keine Seltenheit.
- Auch Reliabilität und Validität (siehe oben) sind Korrelationen. Eine Reliabilität ist z.B. erst ab 0,8 ausreichend.

Regressionsrechnung



$$y = ax + b$$

$$a = \frac{\text{cov}(x,y)}{s_x^2} = \frac{r_{s_x s_y}}{s_x^2}$$

$$b = \bar{y} - a\bar{x}$$

Korrelationsverfahren

	x	Intervall	Dichotom	Ordinal
y				
Intervall		Produkt-Moment-Korrelation (Pearson)	Punktbiseriale Korrelation (Alternativ: t-Test)	Rangkorrelation (Spearman)
Dichotom			Phi-Koeffizient (über Chi-Quadrat). Bei 1/0 Kodierung ist die Produkt-Moment-Korrelation identisch mit Phi.	Biseriale Rangkorrelation
Ordinal				Rangkorrelation

Die Produkt-Moment-Korrelation kann auch im Excel bestimmt werden.

Korrelationsverfahren 2

- Produkt-Moment-Korrelation**

$$r = \frac{\text{cov}(x,y)}{s_x s_y} \quad \text{cov}(x,y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- Punktbiseriale Korrelation**

$$r_{pb} = \frac{\bar{y}_1 - \bar{y}_0}{s_y} \cdot \sqrt{\frac{n_0 \cdot n_1}{n^2}}$$

Die Indizes 0 und 1 bezeichnen die beiden Ausprägungen der x-Variable.
 $n = n_0 + n_1$

- Spearman's Rangkorrelation**

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{n \cdot (n^2 - 1)}$$

d_i bezeichnet die Differenz zwischen x_i und y_i .
 x und y sind dabei Ränge.
 Liefert gleiche Ergebnisse wie die Produkt-Moment-Korrelation, wenn die Ränge mit 1, 2, 3, 4... bezeichnet wurden.
 Probleme bei Rangbindungen (gleichen Rängen sog. **ties**).
 n = Zahl der Differenzen.

Kontingenztafeln

	männlich	weiblich	
mit Brille	25 a	10 b	35
ohne Brille	c 25	d 40	65
	50	50	100

$$\chi^2 = \frac{n \cdot (ad - bc)^2}{(a + b) \cdot (c + d) \cdot (a + c) \cdot (b + d)} \quad \text{FG} = 1$$

$$\text{Phi} = \Phi = \frac{ad - bc}{\sqrt{(a + b) \cdot (c + d) \cdot (a + c) \cdot (b + d)}}$$

Phi ist in etwa mit einem r vergleichbar, liegt aber nicht immer zwischen -1 und 1.

Multiple Korrelation / Regression

- Soll eine Kriteriumsvariable y durch viele verschiedene Prädiktorvariablen x_1, x_2, \dots, x_p beschrieben werden, so ergibt sich eine Gleichung der folgenden Form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p$$

- Die β -Gewichte werden mit dem Verfahren der multiplen Regression berechnet.
- Für jedes β -Gewicht muss die Signifikanz bestimmt werden.
- Es ergibt sich zudem eine multiple Korrelation, da jede Prädiktorvariable einen Teil zur Varianzaufklärung beiträgt ist die Gesamtkorrelation R größer als die einzelnen Korrelationen zwischen Prädiktor und Kriterium.

10. Software

10.1 Excel

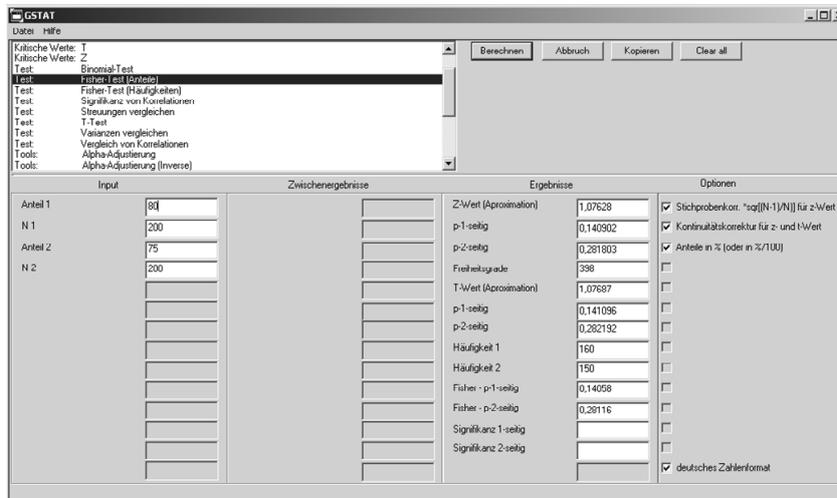
Was geht im Excel?

- Im Excel können Mittelwert, Median, Standardabweichung und Varianz (jeweils für Stichprobe oder Population) berechnet werden.
- Als Testverfahren sind der t-Test und der F-Test vorhanden.
 - Beim t-Test muss angegeben werden, ob von Varianzhomogenität ausgegangen werden kann.
 - Dazu muss zunächst der F-Test für die Varianzen gerechnet werden. Ist das Ergebnis (Wahrscheinlichkeit p) größer 0,2, so kann von Varianzhomogenität ausgegangen werden. Ist $p \leq 0,2$, dann sollte der t-Test für ungleiche Varianzen gerechnet werden.
- Auch die Chi-Verteilung ist im Excel vorhanden. Damit kann die Wahrscheinlichkeit $(1-\text{Alpha})$ für einen berechneten χ^2 -Wert bestimmt werden.
- Korrelationen: Eine einfache Produktmoment-Korrelation kann bestimmt werden.

10.2 GStat

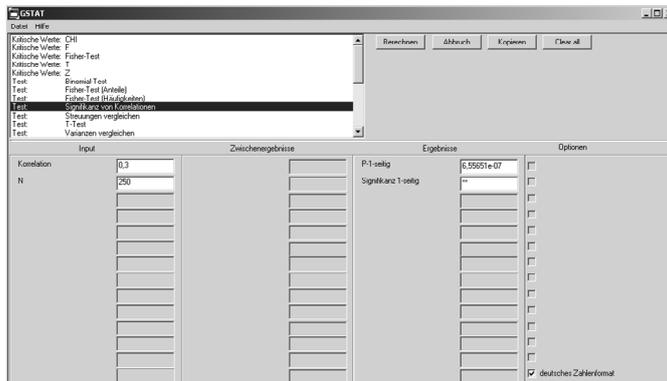
- **Unter www.complexity-research.com, Menü: Downloads, Software kann Gstat heruntergeladen werden.**
 - Sind Mittelwerte oder andere Kennwerte bekannt, kann mit Gstat die Signifikanz bestimmt werden.
 - Beispiel: Fisher-Test für den Vergleich von zwei Prozentwerten.
 - Wenn in zwei Gruppen jeweils 200 Personen befragt wurden und in der einen Gruppe 80% die Küche des Spitals für miserabel halten und in der anderen Stichprobe 75% stellt sich die Frage, ob der Unterschied hier signifikant ist.

Gstat



Korrelationsrechnung Signifikanztestung in GStat

- Die Null-Hypothese geht davon aus, dass eine Null-Korrelation vorliegt.
 $H_0: r = 0$.
- Es kann 1-seitig (Vorhersage der Korrelationsrichtung) oder 2-seitig geprüft werden..



10.3 SPSS

Themenübersicht

- Struktur und Aufbau des Programmpaketes SPSS.
- Grundsätzliches zum Arbeiten mit SPSS.
- Erstellung von Datenfiles und Kodierung von Fragebögen.
- Grundsätzliches zur Datenauswertung mit SPSS.
- Deskriptive Statistiken in SPSS.
- T-Test, 1-faktorielle Varianzanalyse.
- Korrelationen.
- Nichtparametrische Verfahren.
- Hilfsfunktionen, Anleitungen und Literatur.

10.3.1 Struktur und Aufbau

Daten-Editor

Jede Zeile ist ein Fall (Case)

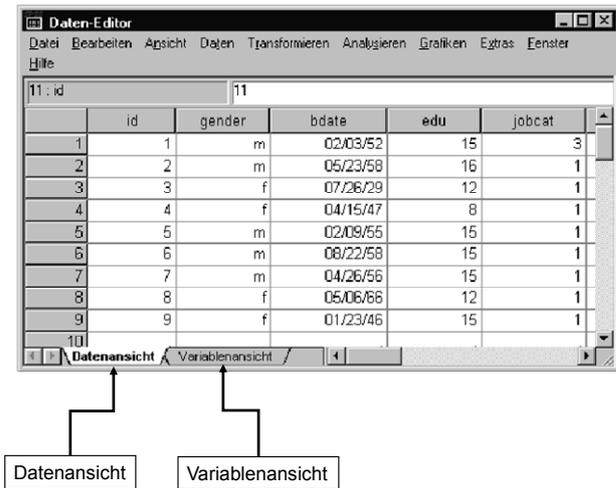
Jede Spalte ist eine Variable

Fehlende Angaben bleiben leer

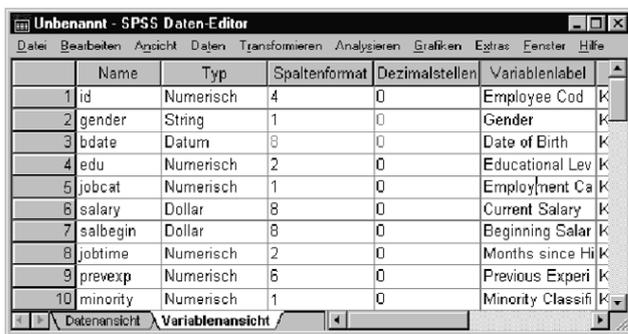
1	auf	dname	pre_4	pre_5	pre_6	pre_7	pre_8	pre_9	pre_10	pre_11	pre_12	pre_13
1	.00	1000000011003447800	.00395	.00278	.00231	.00608	.00642	.00542	.00483	.00262	.00258	.
2	1.00	1000000001000000000	.00967	.01620	.00000	.00914	.07805	.05770	.01507	.01984	.03887	.
3	1.00	1000000001000124000	.01789	.01038	.00438	.02484	.04874	.02944	.02852	.01980	.02480	.
4	1.00	1000000001000124000	.01369	.01284	.01698	.02862	.04874	.01488	.03848	.02200	.01909	.
5	1.00	1000000001000020800	.00484	.05909	.05229	.00691	.01409	.05772	.00718	.09120	.01414	.03828
6	.00	100000000100021801	.00395	.00278	.00231	.00608	.00642	.00542	.00483	.00262	.00258	.
7	.00	1000000001000340260	.00380	.03268	.04314	.00198	.04574	.05803	.05543	.03130	.08271	.04516
8	1.00	1000000001000340000	.00258	.05478	.00004	.05110	.00688	.12428	.02102	.02119	.01414	.02168
9	1.00	1000000001000340000	.01742	.00340	.05959	.01417	.01120	.01488	.01671	.01112	.01415	.04015
10	1.00	1000000001000400000	.04577	.15021	.07516	.00917	.02483	.02451	.11855	.02825	.04888	.12288
11	.00	1000000001000160000	.00734	.01714	.02083	.03883	.04488	.05772	.00003	.02748	.01777	.
12	.00	1000000001000250000	.02409	.02404	.04004	.01254	.00591	.02421	.01558	.01169	.01415	.
13	1.00	1000000001000250000	.06286	.01119	.00080	.01159	.01409	.02451	.04191	.02119	.01414	.
14	.00	1000000001000252801
15	.00	1000000001001699800	.00395	.00278	.00231	.00608	.00642	.00542	.00483	.00262	.00258	.
16	1.00	1000000001000160000	.06876	.11082	.03937	.06420	.01324	.02451	.02053	.04801	.04888	.06669
17	.00	1000000001000390000	.00342	.00700	.00241	.01177	.01480	.01400	.01710	.00789	.04332	.00574
18	1.00	1000000001000480000
19	.00	1000000001004900004	.00482	.00790	.00784	.00920	.01429	.01372	.00953	.01389	.01389	.
20	1.00	1000000001000500000	.06171	.09247	.07249	.04495	.01429	.02451	.02040	.06719	.03478	.
21	.00	1000000001001350800	.00395	.00278	.00231	.00608	.00642	.00542	.00483	.00262	.00258	.

Eine Zelle enthält Werte (Values)

In der Regel wird mit Zahlen gearbeitet

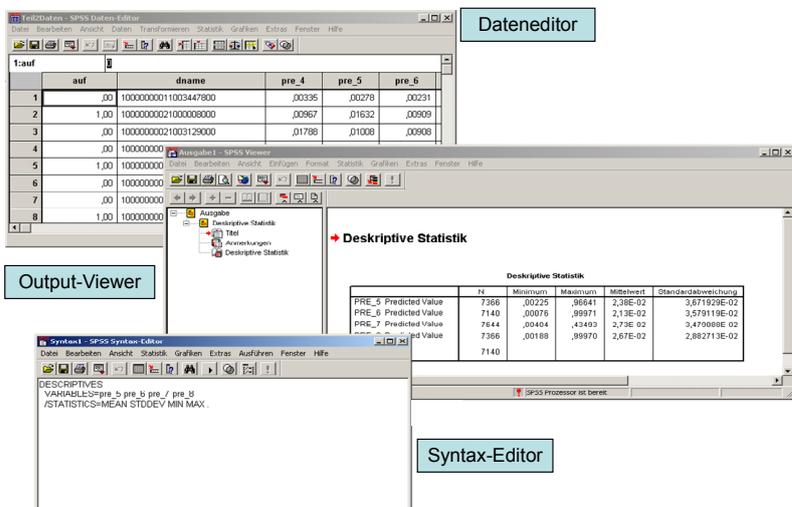


Variablenansicht



- Variablenname
- Datentyp
- Anzahl Ziffern oder Zeichen
- Anzahl Dezimalstellen
- Beschreibende Variablen- und Wertelabels
- Benutzerdefinierte fehlende Werte
- Spaltenbreite
- Messniveau

Programm-Struktur



Datei-Endungen

The screenshot displays the SPSS software interface with three main windows:

- Dateneditor: .sav**: Shows a data table with columns 'auf', 'dname', 'pre_4', 'pre_5', and 'pre_6'. The data is as follows:

auf	dname	pre_4	pre_5	pre_6
.00	1000000011003447800	.00335	.00278	.00231
1.00	10000000021000008000	.00967	.01632	.00909
.00	10000000021003129000	.01788	.01008	.00908
.00	1000000000			
1.00	1000000000			
.00	1000000000			
.00	1000000000			
1.00	1000000000			
- Output-Viewer: .spo**: Shows a 'Deskriptive Statistik' (Descriptive Statistics) table:

	N	Minimum	Maximum	Mittelwert	Standardabweichung
PRE_5 Predicted Value	7366	.00225	.86641	2.38E-02	3.871939E-02
PRE_6 Predicted Value	7140	.00076	.99971	2.13E-02	3.278119E-02
PRE_7 Predicted Value	7644	.00404	.43493	2.73E-02	3.479000E-02
PRE_8 Predicted Value	7366	.00188	.99970	2.67E-02	2.882713E-02
	7140				
- Syntax-Editor: .sps**: Shows the following SPSS syntax code:


```
DESCRIPTIVES
  VARIABLES=pre_5 pre_6 pre_7 pre_8
  /STATISTICS=MEAN STDDEV MIN MAX.
```

10.3.2 Grundsätzliches

Statistisches Wissen

- **SPSS ist ein Statistik-Programm für Profis.**
- **Es umfasst eine Fülle von Funktionen, statistischen Methoden, Algorithmen und Varianten von Methoden.**
- **Sie müssen wissen was Sie tun. Das Programm entscheidet nicht über die Sinnhaftigkeit einer Berechnung: Es rechnet auch den Mittelwert von Hausnummern aus.**
- **Die Ergebnisse der Berechnungen müssen von Ihnen interpretiert werden.**
- **Nicht immer benutzt SPSS die gleichen Methoden und Methodenbezeichnungen, wie gängige Statistik-Bücher.**

Unterschiede zu Excel

- **SPSS ist bewusst unflexibel bei der Dateneingabe.**
- **SPSS ist für die Analyse großer Datensätze geeignet. Auch einfache Fragebögen führen schnell zu über 255 Variablen. Internationale Datensätze können schnell auch mal über 100.000 Cases enthalten.**
- **Berechnungen liefern im SPSS eine Fülle von Detail- und Zwischenergebnissen, die automatisch in Tabellen und Grafiken dargestellt werden. Im Excel wird nur eine Zahl berechnet.**
- **SPSS verfügt über keine automatische Aktualisierung von Ergebnissen oder von Zahlen im Dateneditor.**
- **(in älteren Versionen ist SPSS in den Variablennamen beschränkt auf 8 Zeichen – keine Umlaute – und kann nur jeweils ein Tabellenblatt geöffnet haben.)**

Sonstige Besonderheiten

- **Man kann nur rechnen, wenn die Daten in einem Datenblatt vorhanden sind. Der Vergleich zweier Datenblätter oder eines Datenblattes mit Werten aus der Literatur ist nur schwer oder gar nicht möglich.**
- **Tabellen und Grafiken sind nur selten direkt in wissenschaftlichen Arbeiten nutzbar. Sie enthalten in der Regel mehr Infos als man benötigt oder sind nur schwer zu formatieren. Ein Austausch der Ergebnisse über die Zwischenablage ist aber gut möglich.**
- **SPSS kostet zwischen 1.000 und 5.000 Euro. Studentenversionen und Demoversionen sind jedoch erhältlich.**

10.3.3 Erstellung von Datenfiles und Kodierung von Fragebögen

Definition der Variablen in SPSS

- **Variablenname (8 Zeichen, keine deutschen Sonderzeichen, jeden Namen nur ein mal vergeben), z.B.:**
 - lfnr
 - item1
 - item2
 - item3
 - item4_1
 - item4_2
 - item4_3
 - gschl
 - Alter
- **Datentyp (in der Regel „Numerisch“, aber auch „String“ für Texte und „Datum“ üblich)**

Definition der Variablen in SPSS

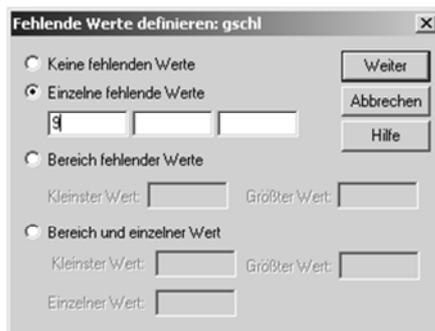
- **Anzahl Ziffern oder Zeichen (bei Text maximal 255)**
- **Anzahl Dezimalstellen (bei ganzen Zahlen sieht es schöner aus, wenn hier 0 eingegeben wird, ist aber egal)**
- **Beschreibende Variablen- und Wertelabels**
- **Benutzerdefinierte fehlende Werte**
- **Spaltenbreite (nur für die Anzeige relevant)**
- **Messniveau (Nominal, Ordinal, Metrisch). Dient als Gedächtnisstütze, hat aber sonst keine Bedeutung.**

Beschreibende Wertelabels



Zur Vergabe von Wertelabels den Wert und den Label eingeben und mit „Hinzufügen“ zur Liste hinzufügen. Am Schluss mit „Weiter“ den Dialog verlassen.

Benutzerdefinierte fehlende Werte



„weiß ich nicht“ ist eine Angabe der befragten Person, die kodiert werden sollte, aber bei Berechnungen ignoriert werden muss. Es handelt sich um einen „Definierten fehlenden Wert“ (Missing).
Fehlen Angaben im Fragebogen, kann bei der Dateneingabe einfach die Zelle im Dateneditor übersprungen werden. Man spricht hier von einem „System Missing“.

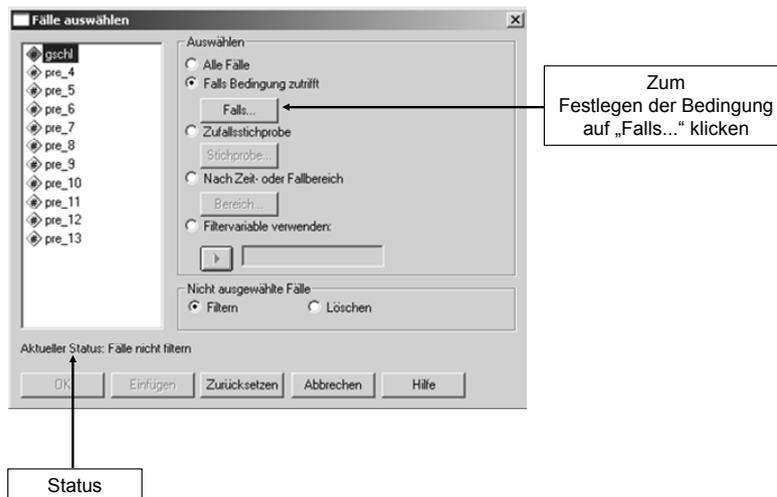
Eine der vier Optionen wählen und geforderte Angaben ausfüllen.
Am Schluss mit „Weiter“ den Dialog verlassen.

10.3.4 Grundsätzliches zu Datenauswertung mit SPSS

Vorbereitende Schritte

- Soll nur mit einem Teil der Daten gerechnet werden?
(Falls ja => Menü: Daten / Fälle auswählen ...)
- Müssen aus den Daten zunächst neue Variablen erzeugt werden?
(Falls ja => Menü: Transformieren / Berechnen ...)

Fälle auswählen ...



Fälle auswählen ... Falls...

Formel für die Bedingung

~ bedeutet „nicht“

Variablenliste

Kopiert eine ausgewählte Variable in den Formel-Editor

Sobald die Formel steht, Kann der Dialog mit „Weiter“ geschlossen werden.

Fälle auswählen ... Falls ... beendet

Bedingung wird angezeigt

Löschen?

„OK“ und „Einfügen“ wird verfügbar

Filter an

1:gschl	gschl	dname	pre_4	pre_5	pre_6	pre_7
1	0	10000000011003447800	,00335	,00278	,00231	,0060
2	1	10000000021000000000	,00967	,01632	,00909	,0291
3	0	10000000021003129000	,01788	,01008	,00908	,0246
4	0	10000000021003129002	,01369	,01284	,01036	,0258
5	1	10000000031000020800	,00484	,03309	,03323	,0063
6	0	10000000031000321801	,00335	,00278	,00231	,0060
7	0	10000000041003404059	,09380	,03256	,04314	,0919
8	1	10000000051000039000	,10255	,05478	,09004	,0371

Auswahl der Fälle

Filterstatus

Transformieren / Berechnen ...

Formel

Variablenliste

neue Zielvariable definieren

Kopiert eine ausgewählte Variable in den Formel-Editor

Transformieren / Berechnen ... Falls...

Gruppe=1

Die neue Variable „gruppe“ bekommt den Wert „1“ falls...

neue Zielvariable definieren

Transformieren / Berechnen ... Falls...

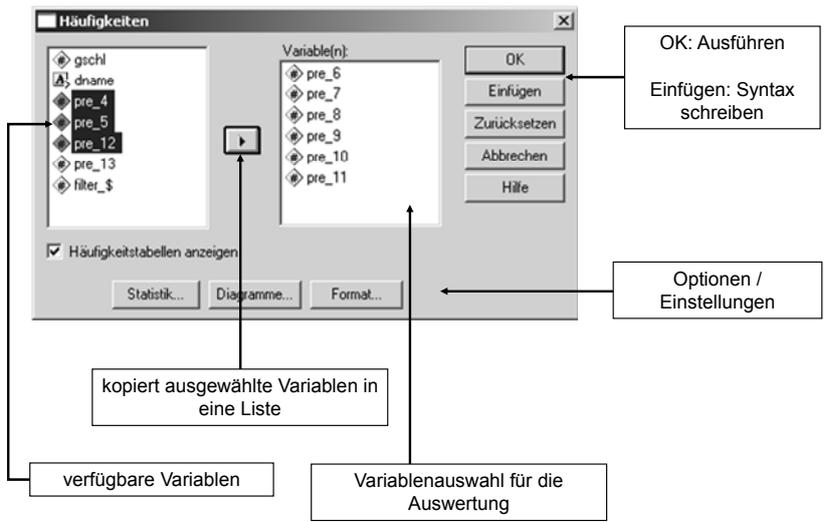
2. Bedingung

3. „Weiter“

1. Fall einschließen, wenn...

**Und-Verknüpfung: AND
Oder-Verknüpfung: OR**

Typischer SPSS-Dialog

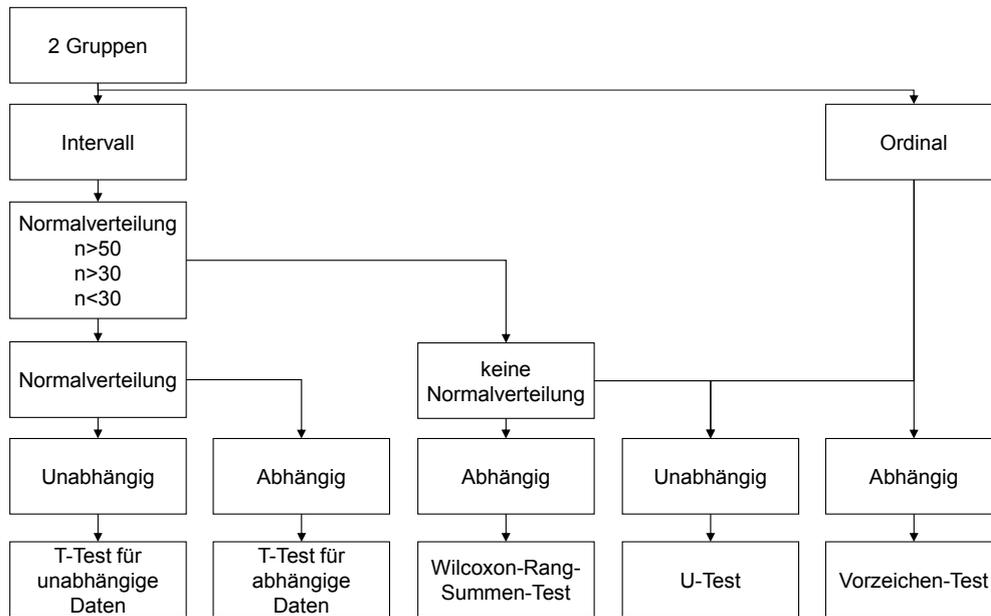


10.3.5 Deskriptive Statistiken

Statistik \ Zusammenfassen \ Häufigkeiten

Statistik \ Zusammenfassen \ Deskriptive Statistiken

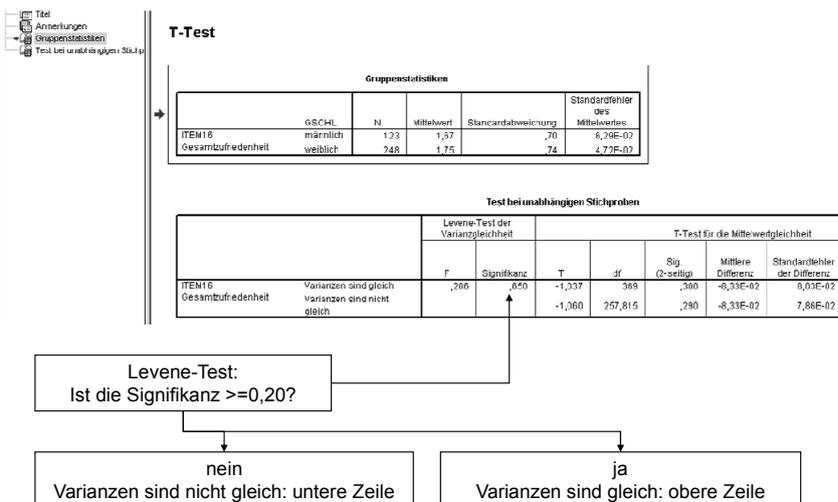
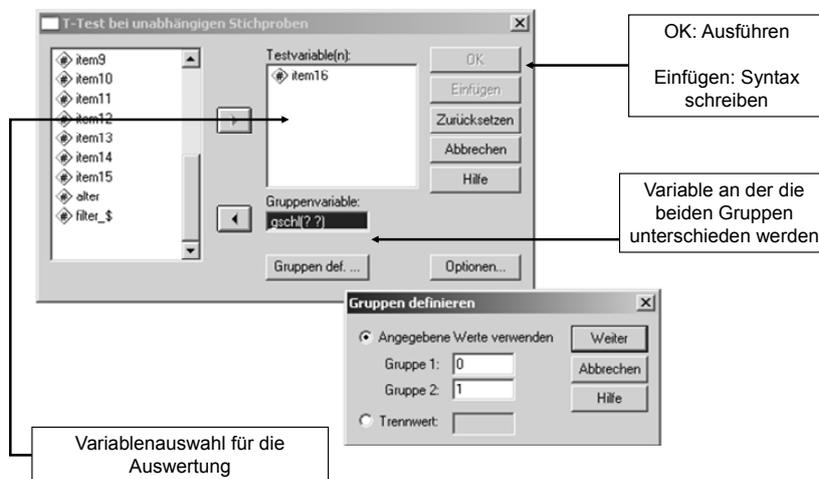
10.3.6 Testverfahren im SPSS



T-Test – unabhängige Daten

- SPSS:
- Statistik \ Mittelwerte vergleichen
T-Test bei unabhängigen ...

- Normalverteilung:** Die Normalverteilung der Messwerte ist keine Voraussetzung für den T-Test. Vielmehr müssen die Differenzen der Mittelwerte normalverteilt sein. Das ist bei großen Stichproben automatisch der Fall. Je nach Misstrauen gegenüber dem zentralen Grenzwertsatz werden entweder 30 oder 50 Messungen pro Gruppe als Minimum empfohlen. Bei kleinen Stichproben müssen hingegen die Messwerte normalverteilt sein.
- Varianzhomogenität:** Es gibt zwei verschiedene T-Tests, einen für und einen ohne Varianzhomogenität.



T-Test

Gruppenstatistiken					
	USCHL	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
ITEM16	männlich	123	1,67	,70	6,29E-02
Gesamtbefriedenheit	weiblich	249	1,75	,74	4,72E-02

		Levene-Test der Varianzhomogenität		T-Test für die Mittelwertgleichheit				
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz der Differenz	Standardfehler der Differenz
ITEM16	Varianzen sind gleich	,206	,650	-1,037	368	,300	-8,33E-02	8,03E-02
Gesamtbefriedenheit	Varianzen sind nicht gleich			-1,060	257,815	,290	-8,33E-02	7,86E-02

t-Test (korrekte Zeile):
Ist die Signifikanz $\leq 0,05$?

nein
Kein signifikanter Unterschied

ja
Signifikanter Unterschied

T-Test

Gruppenstatistiken					
	USCHL	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
ITEM16	männlich	123	1,57	,70	6,29E-02
Gesamtbefriedenheit	weiblich	249	1,75	,74	4,72E-02

		Levene-Test der Varianzhomogenität		T-Test für die Mittelwertgleichheit				
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz der Differenz	Standardfehler der Differenz
ITEM16	Varianzen sind gleich	,206	,650	-1,037	368	,300	-8,33E-02	8,03E-02
Gesamtbefriedenheit	Varianzen sind nicht gleich			-1,060	257,815	,290	-8,33E-02	7,86E-02

t-Test (korrekte Zeile):
Ist die Signifikanz $\leq 0,01$?

nein
Kein **sehr** signifikanter Unterschied

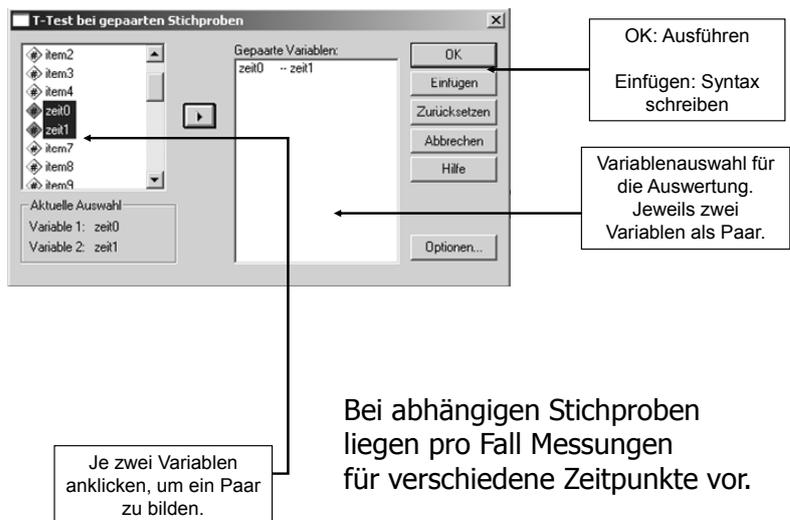
ja
Sehr signifikanter Unterschied

1- oder 2-seitig

- *Ungerichtete Hypothese:* Die von SPSS angegebene Wahrscheinlichkeit ist 2-seitig zu interpretieren.
- *Gerichtete Hypothese:* Für den 1-seitigen Wert gilt, dass er die Hälfte des 2-seitigen Wertes beträgt.

T-Test – abhängige Daten

- SPSS:
- Statistik \ Mittelwerte vergleichen
T-Test bei gepaarten ...
- **Normalverteilung:** Die Normalverteilung der Messwerte ist keine Voraussetzung für den T-Test. Vielmehr müssen die Differenzen der Mittelwerte normalverteilt sein. Das ist bei großen Stichproben automatisch der Fall. Je nach Misstrauen gegenüber dem zentralen Grenzwertsatz werden entweder 30 oder 50 Messungen pro Messzeitpunkt als Minimum empfohlen. Bei kleinen Stichproben müssen hingegen die paarweise berechneten Messwert-Differenzen normalverteilt sein.



Statistik bei gepaarten Stichproben

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren	ZEIT0	1,90	358	,95	5,04E-02
	ZEIT1	1,36	358	,81	3,24E-02

Korrelationen bei gepaarten Stichproben

		N	Korrelation	Signifikanz
Paaren	ZEIT0 & ZEIT1	358	,456	,000

Ist die Korrelation negativ, sollte der Test nicht gerechnet werden. Als Alternative muss dann der Wilcoxon-Test berechnet werden.

Test bei gepaarten Stichproben

		Gepaarte Differenzen				T	df	Sig. (2-seitig)	
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere				Obere
Paaren	ZEIT0 - ZEIT1	,54	,87	4,59E-02	,45	,63	11,752	357	,000

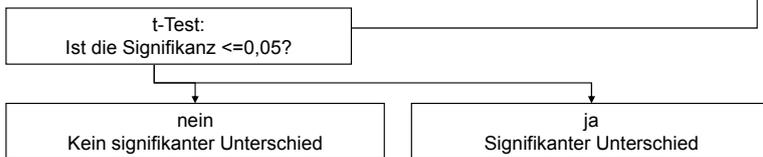
		Mittelwert	N	Standardabweichung	des Mittelwertes
Paaren	ZEIT0	1,90	358	,95	5,04E-02
	ZEIT1	1,36	358	,81	3,24E-02

Korrelationen bei gepaarten Stichproben

		N	Korrelation	Signifikanz
Paaren	ZEIT0 & ZEIT1	358	,456	,000

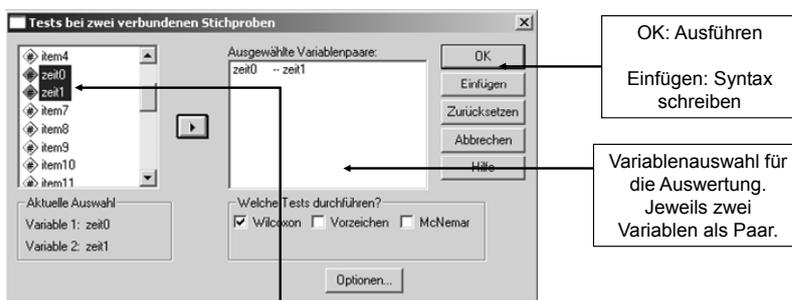
Test bei gepaarten Stichproben

		Gepaarte Differenzen				T	df	Sig. (2-seitig)	
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere				Obere
Paaren	ZEIT0 - ZEIT1	,54	,87	4,59E-02	,45	,63	11,752	357	,000



Wilcoxon-Rangsummen-Test

- SPSS:
- Statistik \ Nichtparametrische Tests
2 verbundene Stichproben ...
- Der Test gilt als Alternative zum T-Test für abhängige Stichproben, falls die Normalverteilungsannahme nicht erfüllt ist.
- Der Test zieht die Messwertpaare voneinander ab, wobei positive, negative und Nulldifferenzen unterschieden werden.
- Der Test berücksichtigt dabei auch die Höhe der Differenzen.



Je zwei Variablen anklicken, um ein Paar zu bilden.

Bei abhängigen Stichproben liegen pro Fall Messungen für verschiedene Zeitpunkte vor.

Wilcoxon-Test

Ränge

		N	Mittlerer Rang	Rangsumme
ZEIT1 - ZEIT0	Negative Ränge	157 ^a	87,46	13731,50
	Positive Ränge	15 ^b	76,43	1146,50
	Bindungen	186 ^c		
	Gesamt	358		

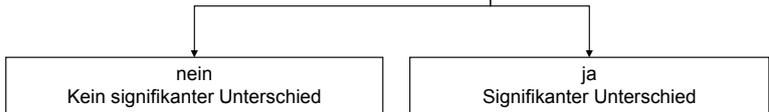
- a. ZEIT1 < ZEIT0
- b. ZEIT1 > ZEIT0
- c. ZEIT0 = ZEIT1

Statistik für Test^b

	ZEIT1 - ZEIT0
Z	-10,192 ^a
Asymptotische Signifikanz (2-seitig)	,000

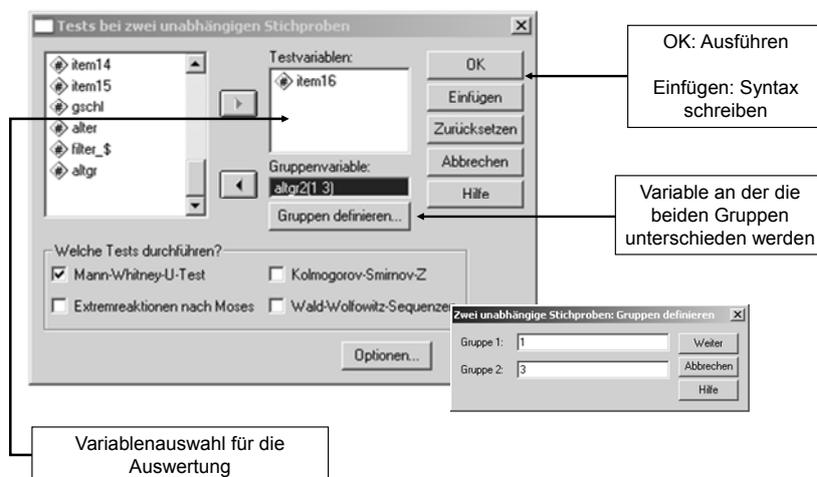
- a. Basiert auf positiven Rängen.
- b. Wilcoxon-Test

z-Test für Wilcoxon-Test:
Ist die Signifikanz $\leq 0,05$?

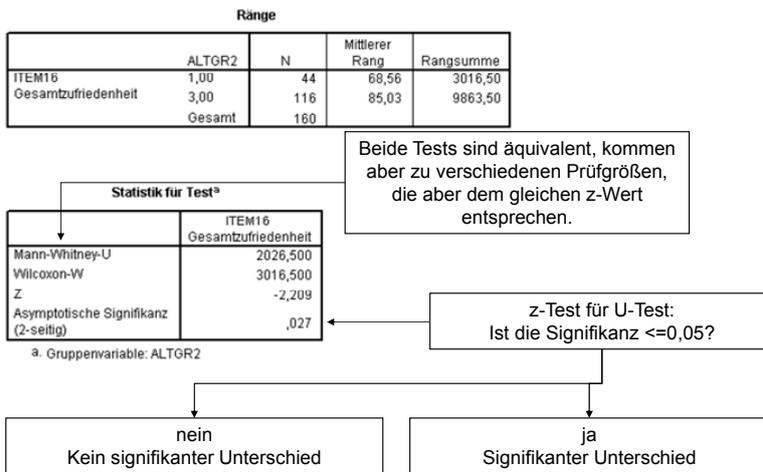


U-Test

- SPSS:
- Statistik \ Nichtparametrische Tests
2 unabhängige Stichproben ...
- Der Test gilt als Alternative zum T-Test für unabhängige Stichproben, falls die Normalverteilungsannahme bzw. die Intervallskalierung der Daten nicht erfüllt sind.
- Der Test sortiert die Messdaten und vergibt Ränge.
- Er zählt, wie häufig Rangplatzunterschiede vorkommen.

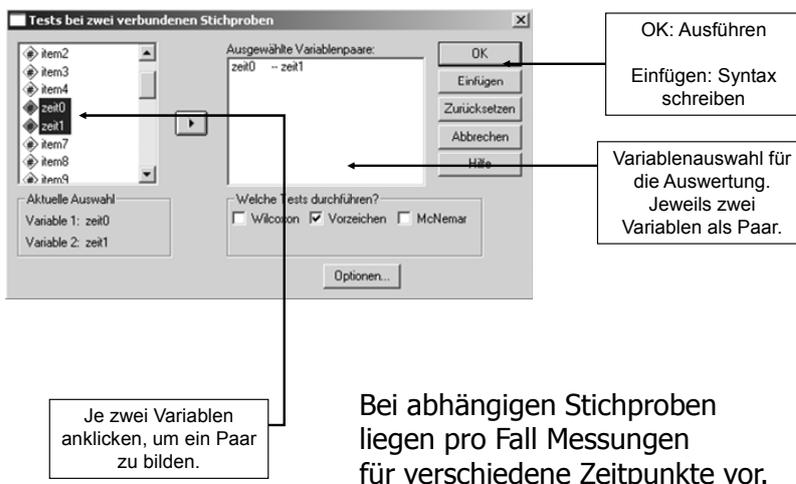


Mann-Whitney-Test



Vorzeichen-Test

- SPSS:
- Statistik \ Nichtparametrische Tests
2 verbundene Stichproben ...
- Der Test zieht die Messwertpaare voneinander ab, wobei positive, negative und Nulldifferenzen unterschieden werden.
- Der Test berücksichtigt nicht die Höhe der Differenzen, sondern zählt diese. Deswegen ist bei intervallskalierten Daten, für die der T-Test nicht gerechnet werden kann, der Wilcoxon-Rang-Summen-Test vorzuziehen.



Vorzeichentest

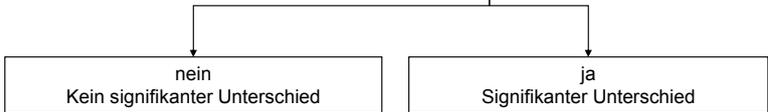
Häufigkeiten		N
ZEIT1 - ZEIT0	Negative Differenzen ^a	157
	Positive Differenzen ^b	15
	Bindungen ^c	186
	Gesamt	358

- a. ZEIT1 < ZEIT0
- b. ZEIT1 > ZEIT0
- c. ZEIT0 = ZEIT1

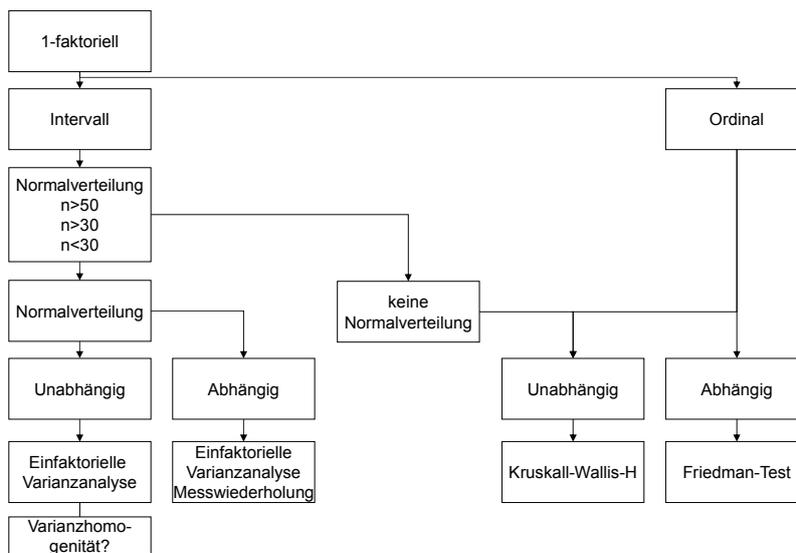
Statistik für Test ^a	
	ZEIT1 - ZEIT0
Z	-10,751
Asymptotische Signifikanz (2-seitig)	,000

a. Vorzeichentest

z-Test für Vorzeichen-Test:
Ist die Signifikanz $\leq 0,05$?



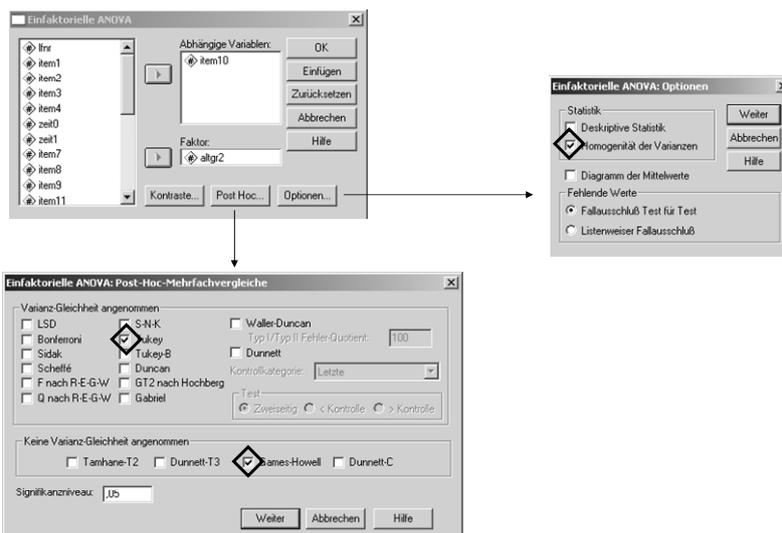
Mehr als zwei Gruppen / 1 Faktor



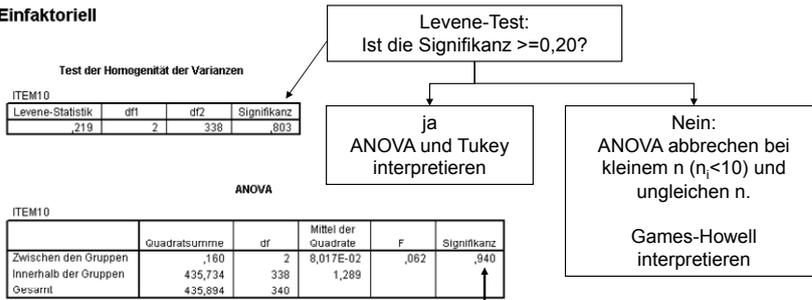
Einfaktorielle Varianzanalyse (ANOVA)

- SPSS:
- Statistik \ Mittelwerte vergleichen
einfaktorielle ANOVA ...
- **Varianzhomogenität:** Bei der Durchführung muss unbedingt die Varianzhomogenität geprüft werden. "Im Fall normalverteilter Populationen mit gleichen Stichprobenumfängen ist die Varianzanalyse relativ robust gegenüber mäßigen Unterschieden zwischen den Populationsvarianzen. Sie reagiert jedoch sehr empfindlich auf Varianzheterogenität, wenn die Stichproben nicht gleich groß sind." (Diehl & Arbinger, 1990, S. 214)
- **Bei fehlender Varianzhomogenität** sollte der Brown-Forsythe-JM- oder der Welch-JM-Test gewählt werden (beide in SPSS nicht enthalten). Im SPSS kann der H-Test von Kruskal-Wallis gerechnet werden, der aber in diesem Fall weniger robust ist, als die beiden zuerst genannten Verfahren.

- **Feste vs. zufällige Effekte:** Die klassische einfaktorielle Varianzanalyse geht davon aus, dass feste Effekte untersucht werden. Es wird also angenommen, dass man genau die Gruppen untersucht, die man auch untersuchen möchte und über die man Aussagen treffen möchte. Zufällige Effekte würden hingegen vorliegen, wenn aus einer Menge möglicher Gruppen, per Zufall eine begrenzte Anzahl gezogen wurde. Das Ziel ist dann in der Regel über diese Gruppen hinweg auf die anderen möglichen aber nicht untersuchten Gruppen zu generalisieren. In diesem Falle wäre eine klassische, einfaktorielle Varianzanalyse nicht angemessen.
- Die Varianzanalyse sagt nichts darüber aus, zwischen welchen Gruppen Unterschiede bestehen, sondern nur dass/ob überhaupt Unterschiede vorliegen:
 - **Tukey-Test:** Prüft auf Einzelunterschiede bei varianzhomogenen Stichproben.
 - **Games-Howell-Test:** Prüft auf Einzelunterschiede bei varianzheterogenen Stichproben.
- Für beide Testverfahren bedeuten Verletzungen der Normalität kein sehr großes Problem. Nur bei sehr schiefen Verteilungen und kleinen Gruppengrößen ist mit erheblichen Problemen zu rechnen.



Einfaktoriell



Post-Hoc-Tests

Die Quadratsumme der Abweichungen innerhalb und zwischen den Gruppen wird verglichen. Ist die Signifikanz $\leq 0,05$ bzw. $0,01$ liegt ein signifikanter Unterschied vor.

Abhängige Variable: ITEM10

	(I) ALTGR2	(J) ALTGR2	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
						Untergrenze	Obergrenze
Tukey-HSD	1,00	2,00	5,77E-02	,195	,953	-,40	,52
		3,00	7,32E-02	,209	,935	-,42	,56
	2,00	1,00	-5,77E-02	,195	,953	-,52	,40
		3,00	1,55E-02	,137	,993	-,31	,34

Post-Hoc-Tests

Wenn der Levene-Test: Signifikanz $\geq 0,20$?

Wenn der Levene-Test: Signifikanz $\leq 0,20$?

Mehrfachvergleiche

Abhängige Variable: ITEM10

	(I) ALTGR2	(J) ALTGR2	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
						Untergrenze	Obergrenze
Tukey-HSD	1,00	2,00	5,77E-02	,195	,953	-,40	,52
		3,00	7,32E-02	,209	,935	-,42	,56
	2,00	1,00	-5,77E-02	,195	,953	-,52	,40
		3,00	1,55E-02	,137	,993	-,31	,34
Games-Howell	1,00	2,00	5,77E-02	,195	,947	-,38	,50
		3,00	7,32E-02	,209	,930	-,41	,55
	2,00	1,00	-5,77E-02	,195	,947	-,50	,38
		3,00	1,55E-02	,137	,993	-,31	,35

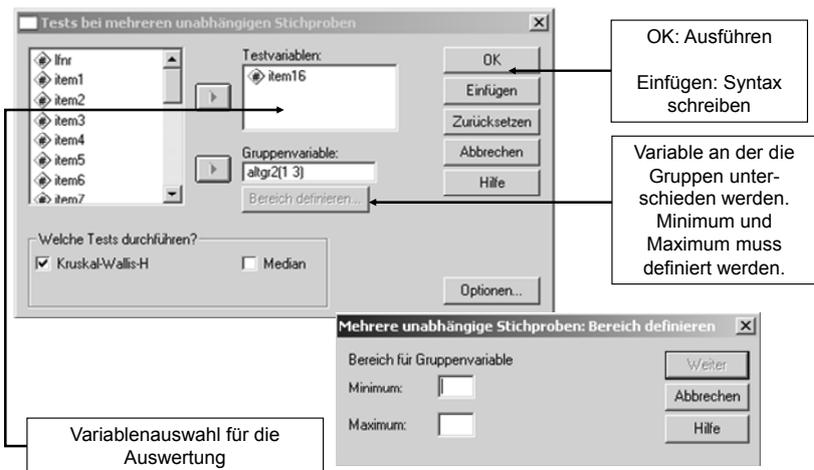
Paarweise wird jede Gruppe mit jeder anderen verglichen. Ist die Signifikanz $\leq 0,05$ bzw. $0,01$ liegt ein signifikanter Unterschied vor.

Varianzanalyse mit Messwiederholung

- SPSS:
- Statistik \ Allgemeines lineares Modell
GLM-Messwiederholung ...

Kruskal-Wallis-H-Test

- SPSS:
- Statistik \ Nichtparametrische Tests
k unabhängige Stichproben ...
- **Varianzhomogenität und Symmetrie der Verteilung.** Der Kruskal-Wallis H-Test gilt als Alternative für die einfaktorielle Varianzanalyse bei Verletzung der Normalverteilung und der Varianzhomogenität. Es gibt jedoch Hinweise darauf, dass auch der Kruskal-Wallis H-Test bei heterogenen Varianzen und schiefen Verteilungen ähnliche Probleme aufweist wie die Varianzanalyse (Diehl & Arbinger, 1990, S. 216).
- **Chi-Quadrat-Verteilung.** Der Test kann exakt berechnet werden und setzt dann keine bestimmte Verteilung voraus. In der Regel wird er jedoch über die Chi-Quadrat-Verteilung approximiert. Hier gilt, dass für 3 Gruppen $n_i > 8$ und für 5 Gruppen $n_i > 3$ die Untergrenze bilden für grobe „Routineentscheidungen“ (vgl. Bortz & Lienert, 2000, S. 225).



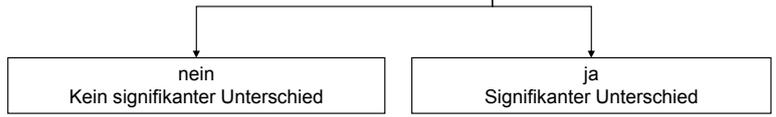
Kruskal-Wallis-Test

Ränge			
	ALTGR2	N	Mittlerer Rang
ITEM16	1,00	44	165,81
Gesamtzufriedenheit	2,00	207	176,80
	3,00	116	203,74
Gesamt		367	

Statistik für Test ^{a,b}	
	ITEM16 Gesamtzufriedenheit
Chi-Quadrat	7,643
df	2
Asymptotische Signifikanz	,022

a. Kruskal-Wallis-Test
b. Gruppenvariable: ALTGR2

Chi-Quadrat-Test für Kruskal-Wallis-Test:
Ist die Signifikanz $\leq 0,05$?



Friedman-Test

- SPSS:
- Statistik \ Nichtparametrische Tests
k verbundene Stichproben ...
- Der Friedman-Test prüft rangskalierte abhängige Daten auf Unterschiede. Ein solcher Fall liegt z.B. vor, wenn n Beurteiler J verschiedene Elemente in eine Rangordnung bringen. Zum Beispiel könnten n=10 Statistiker gebeten werden J=5 gegebene statistische Testverfahren zu bewerten. Der beste Test bekommt den Rangplatz 1 der zweitbeste den Rangplatz 2 und so weiter. Die Bewertungen der Tests sind nun voneinander abhängig und zudem rangskaliert.
- Ein anderes typisches Beispiel für den Einsatz des Friedman-Tests stellt ein Design dar, in dem an n Personen J-mal wiederholt eine Messung einer Variablen vorgenommen wird. Es handelt sich hier um eine klassische Messwiederholung. Werden für jede Person die ermittelten Messwerte in Ränge umgewandelt (der geringste Messwert der Person erhält eine 1, der Nächsthöhere eine 2 und so weiter), so ist der Friedman-Test der passende Test.

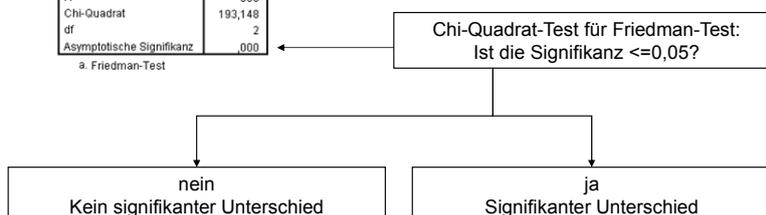


Friedman-Test

Ränge	
	Mittlerer Rang
ZEIT0	2,40
ZEIT1	1,80
ZEIT2	1,80

Statistik für Test ^a	
N	355
Chi-Quadrat	193,148
df	2
Asymptotische Signifikanz	,000

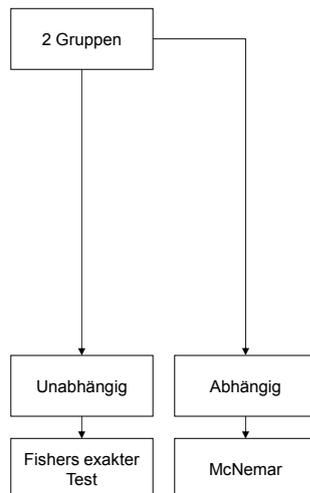
a. Friedman-Test



Mehrfaktorielle Designs

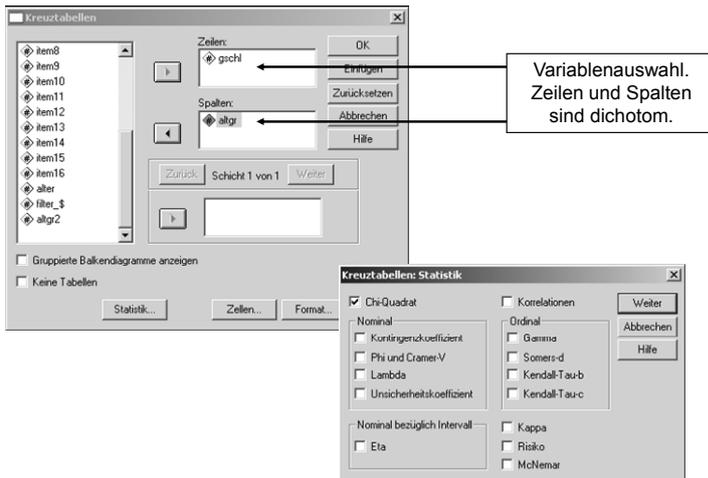
- **Chi-Quadrat:** Man kann versuchen die Daten in Häufigkeiten umzuwandeln (z.B: Median-Split).
- **Mehrfaktorieller Friedman-Test:** Es gibt Verallgemeinerungen des Friedman-Testes für mehr als einen Faktor. Diese sind in SPSS nicht implementiert.
- **GLM:** Mehrfaktorielle Designs finden sich in SPSS unter dem Namen Allgemeines Lineares Modell.
- **Regressions-Modelle:** Mehrfaktorielle Varianzanalysen sind äquivalent zu Regressions-Modellen (Zusammenhangshypothesen).
 - **Lineare Regression:** Metrische abhängige Variable.
 - **Logistische Regression:** Dichotome abhängige Variable.
 - **Multimodale Logistische Regression:** Mehrfach gestufte abhängige Variable.
 - **Cox Regression:** Vorhersage von Wartezeiten bis ein Ereignis eintritt.

2 Gruppen / Häufigkeiten



Fishers exakter Test

- SPSS:
 - Statistik \ Zusammenfassen
 - Kreuztabellen
-
- **Idee:** Wenn Häufigkeiten in zwei Gruppen ermittelt werden, so ergibt sich eine 4-Felder-Kreuztabelle. Das Merkmal für das die Häufigkeit bestimmt wird liegt entweder vor oder nicht, was zwei Bedingungen sind. Weitere zwei ergeben sich für die beiden Gruppen, die untersucht und verglichen werden.
 - **Exakter Test:** Der Test bildet eine exakte Wahrscheinlichkeitsverteilung, die in jedem Fall korrekt ist und keinerlei Einschränkungen unterliegt, wie sie z.B. für den Chi-Quadrat-Test gelten. Aber, bei großen Stichprobenumfängen treten bei der Berechnung auch sehr große Zahlen auf. Einige Computer-Programme geben bei einem N größer 1000 auf. Hier sollte dann doch der Chi-Quadrat-Test herangezogen werden, oder eine Aproximation für den Fisher-Test bestimmt werden.
 - **SPSS:** Verfügt SPSS über das Modul „Exakte Tests“ kann der Fisher-Test auch für große Stichproben und für beliebig große Kontingenztafeln bestimmt werden.



GSCHL * ALTGR Kreuztabelle

Anzahl

		ALTGR		Gesamt
		1,00	2,00	
GSCHL	0 männlich	25	45	70
	1 weiblich	72	78	150
Gesamt		97	123	220

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	2,922 ^a	1	,087		
Kontinuitätskorrektur ^a	2,445	1	,118		
Likelihood-Quotient	2,955	1	,086		
Exakter Test nach Fisher				,109	,058
Zusammenhang linear-mit-linear	2,909	1	,088		
Anzahl der gültigen Fälle	220				

a. Wird nur für eine 2x2-Tabelle berechnet
 b. 0 Zellen (.0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 30,86.

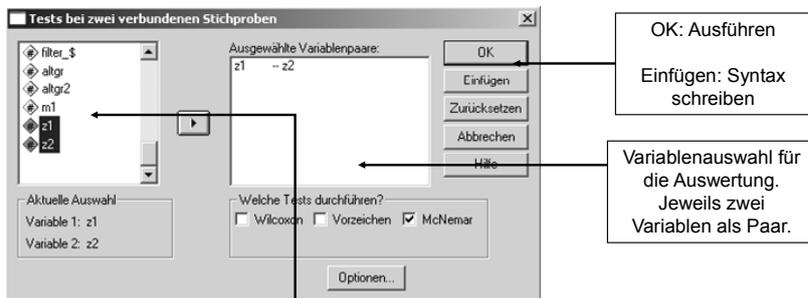
Exakter Wert für Fisher-Test:
 Ist die Signifikanz $\leq 0,05$?

nein
 Kein signifikanter Unterschied

ja
 Signifikanter Unterschied

McNemar

- SPSS:
- Statistik \ Nichtparametrische Tests
Zwei verbundene Stichproben
- Der Test kann für kleine Stichproben exakt berechnet werden. Für große Stichproben erfolgt eine Approximation über die Normalverteilung.
- **SPSS:** Die beiden gepaarten Variablen müssen dichotom kodiert sein und beide müssen gleich kodiert sein.



Je zwei Variablen anklicken, um ein Paar zu bilden.

Bei abhängigen Stichproben liegen pro Fall Messungen für verschiedene Zeitpunkte vor. Der Test erwartet gleich kodierte, dichotome Variablenpaare.

McNemar-Test

Kreuztabellen

1 & 2

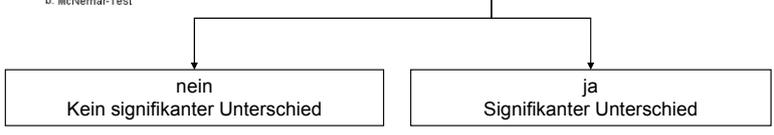
	Z2	
Z1	0	1
0	25	45
1	72	78

Statistik für 1 test^a

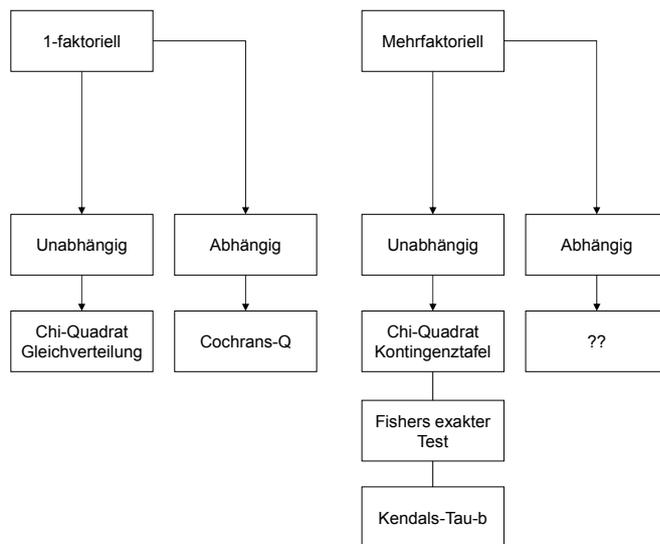
	1 & 2
N	220
Chi-Quadrat ^a	5,778
Asymptotische Signifikanz	,016

a. Kontinuität korrigiert
b. McNemar-Test

Chi-Quadrat-Test für McNemar-Test:
Ist die Signifikanz <= 0,05?



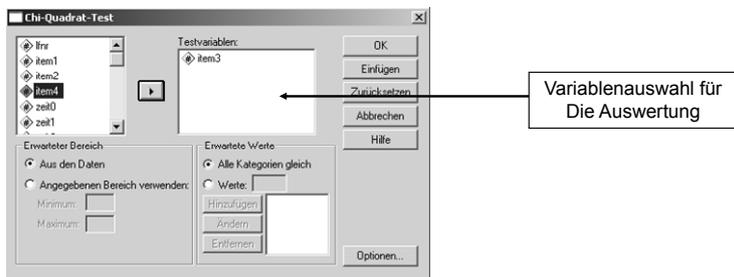
>2 Gruppen / Häufigkeiten



Chi-Quadrat – Gleichverteilung

- SPSS:
- Statistik \ Nichtparametrische Tests\
Chi-Quadrat - Erwartete Werte: Alle
Kategorien gleich
- **Post-Hoc:** Der Chi-Quadrat-Test zeigt nicht, zwischen welchen Gruppen Unterschiede bestehen, sondern nur dass/ob überhaupt Unterschiede vorliegen. Einzelunterschiede können mit mehrfachen Paarvergleichen getestet werden. Hier ist der Fisher-Test ein exaktes Verfahren. Der Alpha-Fehler muss für die Zahl der Paarvergleiche adjustiert werden. Zudem sind Gleichungen für die Bestimmung von Kontrasten verfügbar (vgl. Diehl & Arbinger, 1990, S. 455).

- Chi-Quadrat-Verteilung.** Im Test werden erwartete und beobachtete Häufigkeiten verglichen. Dabei ergibt sich approximativ eine Chi-Quadrat-Verteilung, wenn die erwarteten Häufigkeiten nicht zu klein sind. Als Faustregel gilt heute, dass keine erwartete Häufigkeit kleiner 1 sein darf und höchstens 20% eine erwartete Häufigkeit kleiner 5 aufweisen dürfen. Alternativ können eventuell exakte Tests durchgeführt werden. Auch paarweise Vergleiche mit dem Fisher-Test und eine Alpha-Fehler-Adjustierung sind denkbar. Oder man kann inhaltlich passende Gruppen so zusammen legen, dass die erwarteten Häufigkeiten die Grenze überschreiten.



ITEM3 Institut

	Beobachtetes N	Erwartete Anzahl	Residuum
1 Institut A	3	37,2	-34,2
2 Institut B	18	37,2	-19,2
3 Institut C	83	37,2	45,8
4 Institut D	29	37,2	-8,2
5 Institut E	76	37,2	38,8
6 Institut F	43	37,2	5,8
7 Institut G	5	37,2	-32,2
8 Institut H	58	37,2	20,8
9 Institut I	41	37,2	3,8
10 Institut J	16	37,2	-21,2
Gesamt	372	37,2	

Statistik für Test

	ITEM3 Institut
Chi-Quadrat ^a	192,892
df	9
Asymptotische Signifikanz	,000

^a Bei 0 Zellen (0%) werden weniger als 5 Häufigkeiten erwartet. Die kleinste erwartete Zellenhäufigkeit ist 37,2.

Die kleinste erwartete Häufigkeit muss größer 1 sein. Höchstens 20% dürfen eine erwartete Häufigkeit kleiner 5 aufweisen.
Für Alternativen siehe oben.

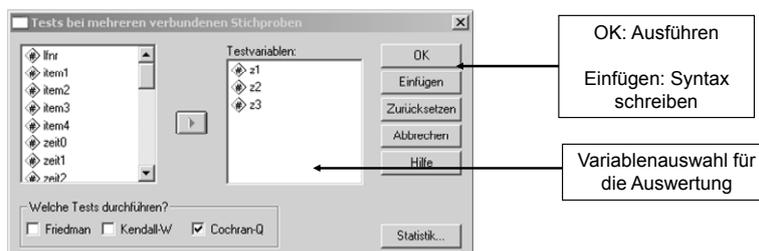
Chi-Quadrat-Test:
Ist die Signifikanz $\leq 0,05$?

nein
Kein signifikanter Unterschied

ja
Signifikanter Unterschied

Cochrans-Q

- SPSS:
 - Statistik \ Nichtparametrische Tests
 - k verbundene Stichproben
-
- **Post-Hoc:** Der Test zeigt nicht, zwischen welchen Gruppen Unterschiede bestehen, sondern nur dass/ob überhaupt Unterschiede vorliegen. Einzelunterschiede können mit mehrfachen Paarvergleichen getestet werden. Hier ist der McNemar-Test ein passendes Verfahren. Der Alpha-Fehler muss für die Zahl der Paarvergleiche adjustiert werden. Zudem sind Gleichungen für die Bestimmung von Kontrasten verfügbar (vgl. Diehl & Arbinger, 1990, S. 482).
 - **Chi-Quadrat-Verteilung.** Der Test kann exakt berechnet werden und setzt dann keine Verteilung voraus. In der Regel wird er jedoch über die Chi-Quadrat-Verteilung approximiert. Hier gilt, dass für 3 Messzeitpunkte $n=10$ und für 5 Messzeitpunkte $n=8$ die Untergrenze bilden für grobe „Routineentscheidungen“.
 - **SPSS:** Die verbundenen Variablen müssen dichotom kodiert sein und müssen dabei gleich kodiert sein.



Bei abhängigen Stichproben liegen pro Fall Messungen für verschiedene Zeitpunkte vor. Der Test erwartet gleich kodierte, dichotome Variablen.

Cochran-Test

Häufigkeiten

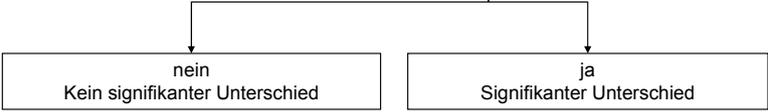
	Wert	
	0	1
Z1	70	150
Z2	97	123
Z3	70	150

Statistik für Test

N	220
Cochrans Q-Test	12,462 ^a
df	2
Asymptotische Signifikanz	,002

a. 1 wird als Erfolg behandelt.

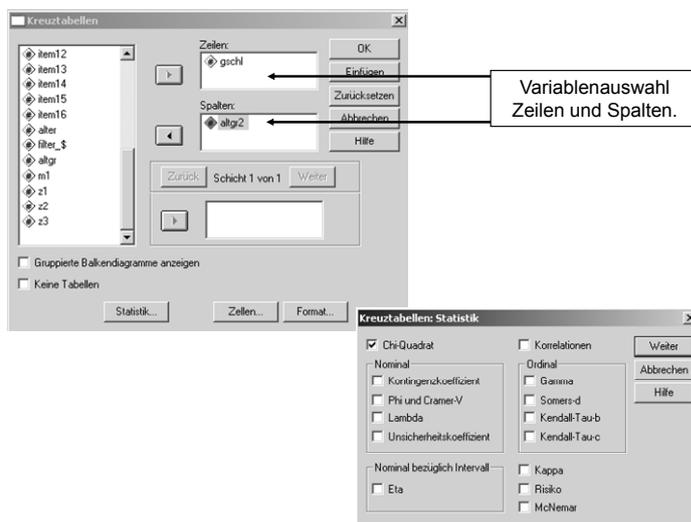
Asymptotische Signifikanz für Cochrans-Q-Test:
Ist die Signifikanz $\leq 0,05$?



Chi-Quadrat – Kontingenztafel

- SPSS:
 - Statistik \ Zusammenfassen
 - Kreuztabellen
-
- **Post-Hoc:** Der Chi-Quadrat-Test zeigt nicht, zwischen welchen Gruppen Unterschiede bestehen, sondern nur dass/ob überhaupt Unterschiede vorliegen. Einzelunterschiede können mit mehrfachen Paarvergleichen getestet werden. Hier ist der Fisher-Test ein exaktes Verfahren. Der Alpha-Fehler muss für die Zahl der Paarvergleiche adjustiert werden. Zudem sind Gleichungen für die Bestimmung von Kontrasten verfügbar (vgl. Diehl & Arbinger, 1990, S. 455).
 - **Chi-Quadrat-Verteilung.** Im Test werden erwartete und beobachtete Häufigkeiten verglichen. Dabei ergibt sich approximativ eine Chi-Quadrat-Verteilung, wenn die erwarteten Häufigkeiten nicht zu klein sind. Als Faustregel gilt heute, dass keine erwartete Häufigkeit kleiner 1 sein darf und höchstens 20% eine erwartete Häufigkeit kleiner 5 aufweisen dürfen. Alternativ können eventuell exakte Tests durchgeführt werden. Auch paarweise Vergleiche mit dem Fisher-Test und eine Alpha-Fehler-Adjustierung sind denkbar. Oder man kann inhaltlich passende Gruppen so zusammen legen, dass die erwarteten Häufigkeiten die Grenze überschreiten.

- Vorsicht bei geordneter Zellenstruktur:** Dem Test ist es egal, was die Zeilen und Spalten der Kontingenztabelle bedeuten. Enthalten diese eine Ordnung, z.B. indem sie die Stufen eines Ratings abbilden, so ist es inhaltlich nicht mehr bedeutungslos, wenn man diese Kategorien vertauscht. Der Test merkt hier aber keinen Unterschied. Beschreiben die Zeilen oder Spalten Ordnungsrelationen, so ist Kendals-Tau (siehe unten) zu rechnen.



GSCHL * ALTGR2 Kreuztabelle

Anzahl		ALTGR2			Gesamt
		1,00	2,00	3,00	
GSCHL	0 männlich	10	69	41	120
	1 weiblich	34	138	75	247
Gesamt		44	207	116	367

Die kleinste erwartete Häufigkeit muss größer 1 sein. Höchstens 20% dürfen eine erwartete Häufigkeit kleiner 5 aufweisen.
Für Alternativen siehe oben.

Chi-Quadrat-Tests

	Wert	df	Asymmetrische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	2,395 ^a	2	,302
Likelihood-Quotient	2,522	2	,283
Zusammenhang linear-mit-linear	1,728	1	,189
Anzahl der gültigen Fälle	367		

a. 0 Zellen (.0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 14,39.

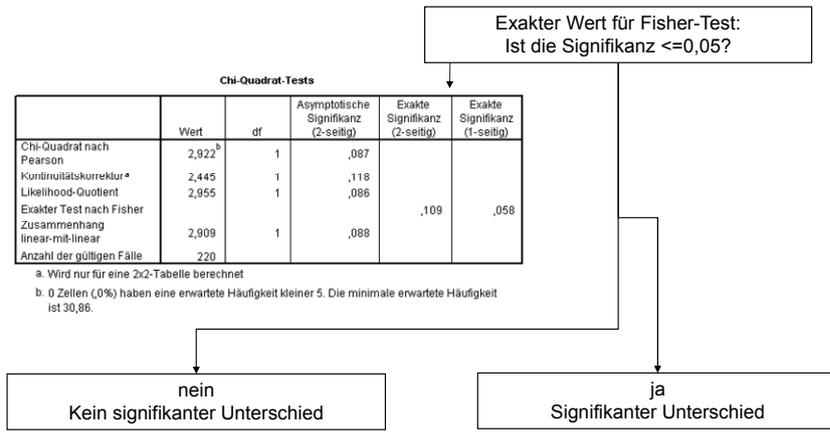
Chi-Quadrat-Test nach Pearson:
Ist die Signifikanz $\leq 0,05$?

nein
Kein signifikanter Unterschied

ja
Signifikanter Unterschied

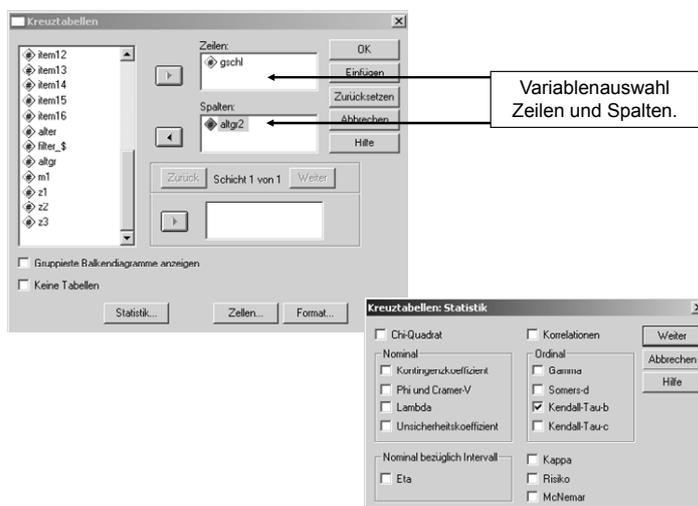
Fishers exakter Test

- SPSS:
 - Statistik \ Zusammenfassen
 - Kreuztabellen
 - Exakte Tests
-
- **Idee:** Wenn Häufigkeiten in zwei Gruppen ermittelt werden, so ergibt sich eine 4-Felder-Kreuztabelle. Das Merkmal für das die Häufigkeit bestimmt wird liegt entweder vor oder nicht, was zwei Bedingungen sind. Weitere zwei ergeben sich für die beiden Gruppen, die untersucht und verglichen werden. Große Kontingenztafeln lassen sich in viele kleine 4-Felder-Tafeln zerlegen.
 - **Exakter Test:** Der Test bildet eine exakte Wahrscheinlichkeitsverteilung, die in jedem Fall korrekt ist und keinerlei Einschränkungen unterliegt, wie sie z.B. für den Chi-Quadrat-Test gelten. Aber, bei großen Stichprobenumfängen treten bei der Berechnung auch sehr große Zahlen auf. Einige Computer-Programme geben bei einem N größer 1000 auf. Hier sollte dann doch der Chi-Quadrat-Test herangezogen werden, oder eine Aproximation für den Fisher-Test bestimmt werden.
 - **SPSS:** Verfügt SPSS über das Modul Exakte Tests kann der Fisher-Test auch für große Stichproben und für beliebig große Kontingenztafeln bestimmt werden.
 - **Vorsicht bei geordneter Zellenstruktur:** Dem Test ist es egal, was die Zeilen und Spalten der Kontingenztafel bedeuten. Enthalten diese eine Ordnung, z.B. indem sie die Stufen eines Ratings abbilden, so ist es inhaltlich nicht mehr bedeutungslos, wenn man diese Kategorien vertauscht. Der Test merkt hier aber keinen Unterschied. Beschreiben die Zeilen oder Spalten Ordnungsrelationen, so ist Kendals-Tau (siehe unten) zu rechnen.



Kendals-Tau-b

- SPSS:
 - Statistik \ Zusammenfassen
 - Kreuztabellen
-
- **Beim Chi-Quadrat-Test bzw. Fisher-Test gilt: Vorsicht bei geordneter Zellenstruktur:** Den Verfahren ist es egal, was die Zeilen und Spalten der Kontingenztafel bedeuten. Enthalten diese eine Ordnung, z.B. indem sie die Stufen eines Ratings abbilden, so ist es inhaltlich nicht mehr bedeutungslos, wenn man diese Kategorien vertauscht. Der Test merkt hier aber keinen Unterschied. Beschreiben die Zeilen oder Spalten Ordnungsrelationen, so ist Kendals-Tau (siehe unten) zu rechnen.
 - Es handelt sich bei Kendals-Tau-b um ein Zusammenhangsmaß zwischen Rangordnungsdaten.
 - **Rangbindungen.** Es wird mit Rängen gerechnet und so kann es vorkommen, dass mehrere Fälle den gleichen Rang aufweisen. Rangbindungen können zwar herausgerechnet werden, verschlechtern aber die Statistik.



GSCHL * ALTGR2 Kreuztabelle

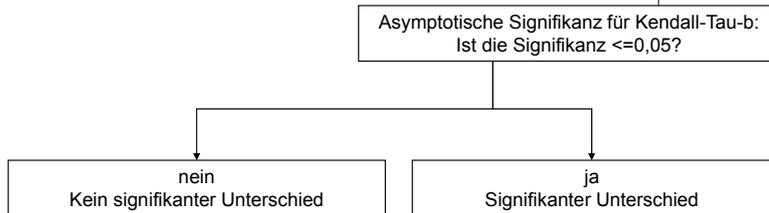
Anzahl		ALTGR2			Gesamt
		1,00	2,00	3,00	
GSCHL	0 männlich	10	68	41	120
	1 weiblich	34	138	75	247
Gesamt		44	207	116	367

Symmetrische Maße

Ordinal- bzgl. Ordinalmaß	Kendall-Tau-b	Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b	Näherungsweise Signifikanz
Anzahl der gültigen Fälle		-,062	,049	-1,263	,207

a. Die Null-Hypothese wird nicht angenommen.

b. Unter Annahme der Null-Hypothese wird der asymptotische Standardfehler verwendet.



10.3.7 Hilfsfunktionen

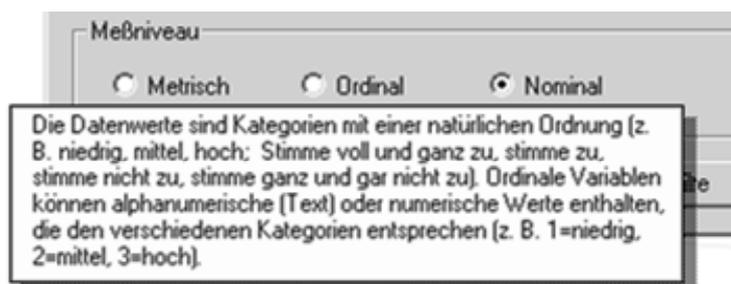
Hilfen...

- Die Hilfe von SPSS ist durchaus komfortabel. Es gibt einen Statistik-Assistenten und für alle Teile des Programmpaketes gibt es Online-PDF-Handbücher.
- Die Pop-Up-Hilfe bietet eine gute Unterstützung während der Arbeit.
- Zahlreiche Handbücher zum Arbeiten mit SPSS sind im Buchhandel zu haben. Auch zahlreiche SPSS-Handbücher zu aufwändigen Verfahren sind erschienen. Da ständig neue SPSS-Versionen erscheinen wird hier keine Empfehlung gegeben.
- Einige Statistikbücher geben auch einen kurze Einführung zu Rechnen in SPSS, so z.B. in: Bortz, J. (1989). Statistik für Sozialwissenschaften. Berlin, Heidelberg, New York, Springer.

Hilfen...

- Viel Hilfe findet sich im Internet, so z.B. auf: <http://www.uni-koeln.de/themen/statistik/software/spss/index.html>
- Besonders hilfreich sind die SPSS-PDF-Dokumente des Universitäts-Rechenzentrum Trier. Suchworte „SPSS Universität Trier“ in google bringt gute Treffer.

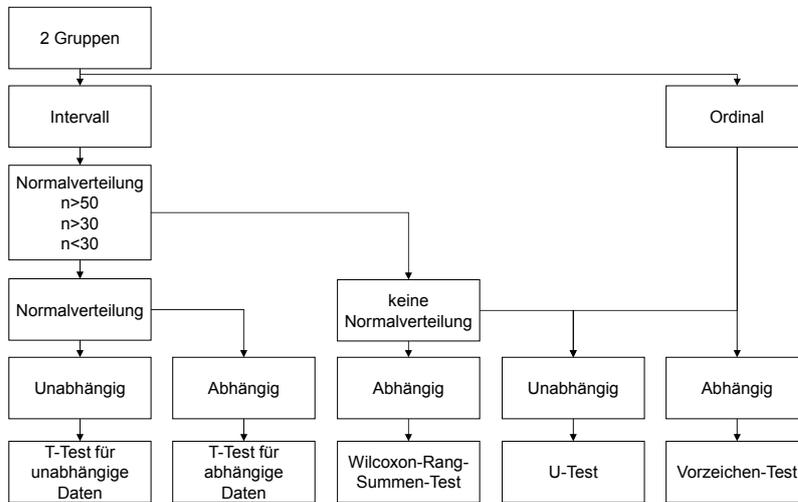
Popup-Hilfe



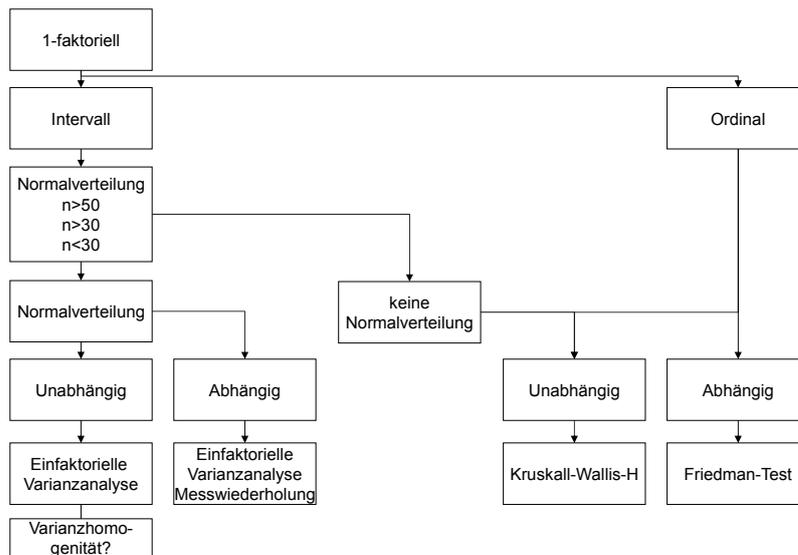
Ein Klick mit der rechten Maustaste auf ein Dialogelement aktiviert die Popup-Hilfe für dieses Dialogelement.

10.3.8 Test-Finder

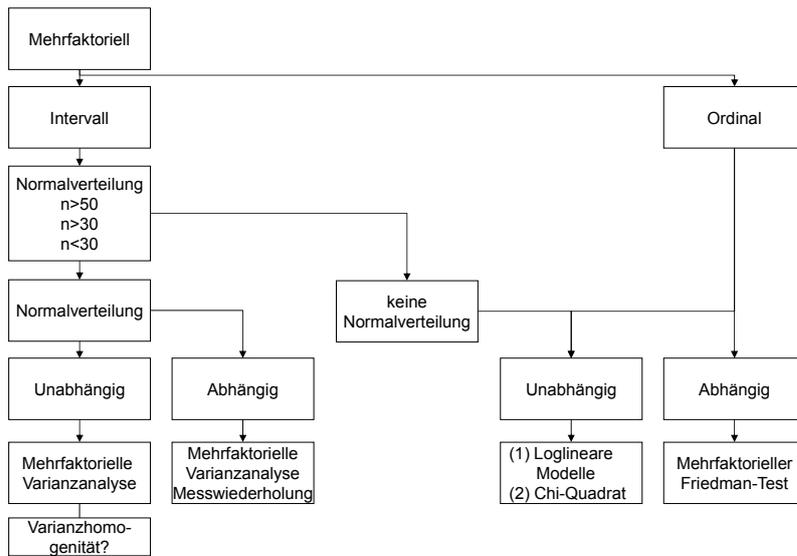
1 Faktor / 2 Gruppen / zentrale Tendenz



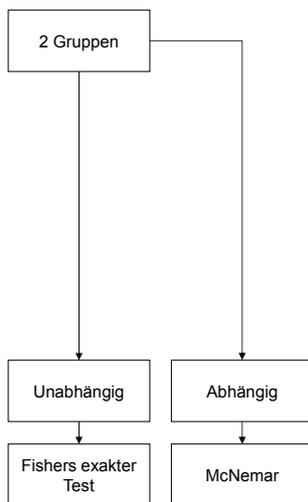
1 Faktor / > 2 Gruppen / zentrale Tendenz



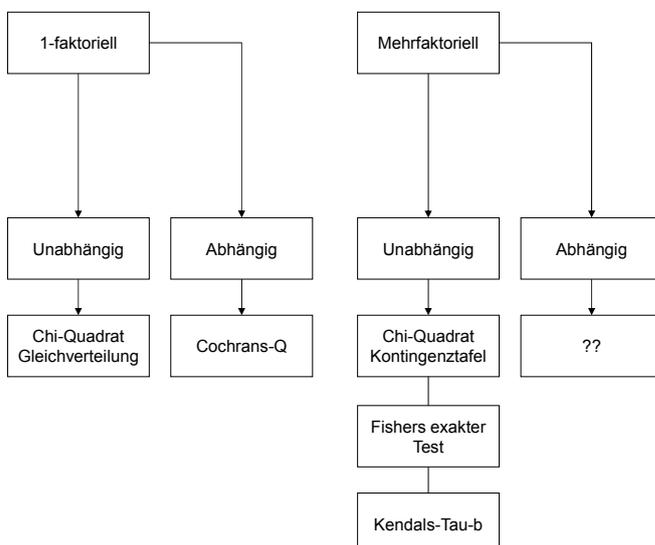
>1 Faktor / > 2 Gruppen / zentrale Tendenz



2 Gruppen / Häufigkeiten



>2 Gruppen / Häufigkeiten



11. Glossar für einige wichtige statistische Begriffe

Alpha-Fehler. Ein Signifikanztest (\nearrow statistische Signifikanz) befindet den Unterschied zwischen zwei Messwerten dann als signifikant, wenn der Unterschied so groß ist, dass es nach den Gesetzen der Wahrscheinlichkeitsrechnung als extrem unwahrscheinlich angesehen werden kann, dass kein Unterschied besteht. Nun ist es jedoch relativ offen, welche Wahrscheinlichkeit als klein genug gelten kann. Es handelt sich daher um eine Übereinkunft, dass gemeinhin bei einer Wahrscheinlichkeit von 5% (und darunter) von Signifikanz gesprochen wird. Nun heißt dies jedoch, dass ein Signifikanztest, der zwei Messwerte nur mit 5% Wahrscheinlichkeit für ähnlich hält, dazu verleitet, die beiden Messwerte eben für unterschiedlich zu halten. Dennoch besteht laut Test aber eine Wahrscheinlichkeit von 5%, dass sie doch ähnlich sind und sich nicht unterscheiden. Wenn man aufgrund des Tests also davon ausgeht, dass sie sich unterscheiden, macht man mit eben jener 5%-tigen Wahrscheinlichkeit einen Fehler. Dieser Fehler wird Alpha-Fehler genannt.

Alpha-Fehler-Adjustierung. In der Regel sind Signifikanztests in der Lage, nur zwei Messwerte miteinander zu vergleichen. Einige Fragestellungen machen daher mehrere Vergleiche zwischen jeweils zwei Messwerten nötig, um die Frage insgesamt beantworten zu können. Beantworten drei Personengruppen einen Fragebogen (Gruppe A, B, C), so kommt man auf insgesamt drei paarweise Vergleiche (A mit B; A mit C und B mit C). Allgemein gilt Anzahl der Vergleiche = [Anzahl der Gruppen mal [Anzahl der Gruppen minus Eins]] geteilt durch 2. So ergeben sich für vier Gruppen bereits: $(4 \times 3)/2 = 6$ Vergleiche. Wenn die Fragestellung relativ offen formuliert ist und generell nach Unterschieden zwischen den Gruppen gefragt wird, so wächst die Wahrscheinlichkeit, einen Unterschied zu finden, je mehr Vergleiche möglich werden. Da man ja bei jedem Paarvergleich einen Alpha-Fehler von 5% begeht, summieren sich die Fehler von Paarvergleich zu Paarvergleich. Bei drei Vergleichen macht man also einen viel höheren Fehler als bei nur einem. Höhere Fehler als 5% sind jedoch nach der oben angesprochenen Vereinbarung nicht signifikant. Um insgesamt nur auf einen Fehler von 5% zu kommen, müssen für jeden Einzelvergleich strengere Alpha-Fehler-Grenzwerte festgelegt werden. Für 3 Vergleiche ergibt sich z.B. ein Wert von 1,7%, bei vier Vergleichen sind es 1,3%, bei 10 Vergleichen 0,5%, usw. Eine Alternative für die Berechnung vieler Signifikanztests, die nur jeweils zwei Messwerte vergleichen können ist die sogenannte Varianzanalyse (\nearrow Varianzanalyse, ANOVA).

Chi-Quadrat-Test. Der Chi-Quadrat-Test ermöglicht den Vergleich von erwarteten Häufigkeiten mit tatsächlich beobachteten Häufigkeiten. Erwartet man aufgrund von Vorerfahrungen oder aus der Literatur zum Beispiel, dass jeder vierte männliche Österreicher Raucher ist, so würde man bei 100 befragten Personen 25 Raucher erwarten. Der Chi-Quadrat-Test vergleicht die erwarteten 25 Raucher dann mit den tatsächlich im Rahmen einer Befragung vorgefundenen Rauchern. Im Rahmen eines Chi-Quadrat-Tests können beliebig viele verschiedene Häufigkeiten miteinander verglichen werden. So ergibt sich beim Chi-Quadrat-Test auf eine Gleichverteilung hin die erwartete Häufigkeit als Mittelwert der beobachteten. Aufgrund geringer Voraussetzungen kann der Chi-Quadrat-Test immer berechnet werden, wenn es um Häufigkeiten geht und eine bestimmte oder mehrere bestimmte Häufigkeiten erwartet werden können. Der Chi-Quadrat-Test ermittelt einen Chi-Quadrat-Wert, für den zusammen mit den sog. Freiheitsgraden (in der Regel Zahl der Messwerte minus eins) die Wahrscheinlichkeit bekannt ist. Die Wahrscheinlichkeit ist das Ergebnis des Tests. Man spricht von einer \nearrow statisti-

schen Signifikanz, wenn diese Wahrscheinlichkeit kleiner als der vorher festgelegte α -Fehler ist.

Fishers exakter Test. Ein besonders *sicherer* Test ist Fishers exakter Test, da er kaum an Voraussetzungen gebunden ist und immer berechnet werden kann, wenn es um den Vergleich zweier Prozentzahlen geht. Eine Berechnung durch einen Computer setzt jedoch meist voraus, dass insgesamt nicht mehr als 1000 Personen befragt wurden, da bei der Berechnung extrem hohe Zahlen als Zwischenergebnisse auftreten. Neben der exakten Variante dieses Tests gibt es für große Stichproben daher auch Näherungsformeln über den α -T-Test, die jedoch mit Vorsicht zu genießen sind. Fishers exakter Test liefert ohne weitere Kennwerte die Wahrscheinlichkeit für die Übereinstimmung der beiden Prozentzahlen. Die Wahrscheinlichkeit ist das Ergebnis des Tests. Man spricht von einer α -statistischen Signifikanz, wenn diese Wahrscheinlichkeit kleiner als der vorher festgelegte α -Fehler ist.

Korrelationen. Eine Korrelation beschreibt den statistischen Zusammenhang zwischen zwei Merkmalen. Beide Merkmale müssen in unterschiedlichen Ausprägungen vorkommen können. Ist das nicht der Fall, so kann keine Korrelation berechnet werden. Wird z.B. die Frage danach gestellt, ob die Zahl der Geburten und die Zahl der Störche einen Zusammenhang (also eine Korrelation) aufweist, so muss sowohl die Zahl der Störche, als auch die Zahl der Geburten variieren können. Es bietet sich hier an, die Zahl der Geburten und die Zahl der Störche pro Monat zu erheben. Dadurch erhält man Zahlenpaare von Geburtenzahl und Storchpopulation für jeden Monat. Es stellt sich nun die Frage, ob sich die Zahl der Störche und die Zahl der Geburten über das Jahr hinweg in die gleiche Richtung entwickelt, also ob mit ansteigender Zahl der Geburten auch die Zahl der Störche wächst und ob mit sinkender Zahl der Geburten auch die Zahl der Störche abnimmt. Ist es so, dass die Zahl der Störche und die Zahl der Geburten sich jeweils in die gleiche Richtung entwickeln, so spricht man von einer positiven Korrelation. Steigt jedoch die Zahl der Geburten, immer wenn die Zahl der Störche abnimmt (und umgekehrt: die Zahl der Geburten sinkt und gleichzeitig nimmt die Zahl der Störche zu), so spricht man von einer negativen Korrelation. Korrelationen können Zahlenwerte zwischen -1 und $+1$ annehmen. Dabei zeigt das Vorzeichen nur an, ob es sich um einen positive oder um eine negative Korrelation handelt. Je näher die Zahlenwerte bei 1 (bzw. -1) liegen, desto „perfekter“ ist der Zusammenhang. Ist eine Korrelation jedoch 0 , dann liegt gar keine Korrelation, also auch kein Zusammenhang vor. Viele Zusammenhänge, die z.B. in der Psychologie beschrieben werden haben relativ kleine Werte um $0,3$ (bzw. $-0,3$), wohingegen z.B. in der Physik nicht selten Korrelationen um $0,9$ (bzw. $-0,9$) gefunden werden können. Ob eine kleine Korrelation nicht eventuell doch auf das Fehlen einer Korrelation (Null-Korrelation) hinweist, kann nur durch einen Signifikanztest (α -Statistische Signifikanz) entschieden werden. Es wird grundsätzlich 2-seitig getestet (α -P-2-seitig). Erst, wenn eine Korrelation sich als signifikant herausstellt, kann sie interpretiert werden. Ist sie nicht signifikant, so kann man nicht davon ausgehen, dass ein Zusammenhang beobachtet wurde. Ist sie jedoch signifikant, so bedeutet das noch nicht, dass der beobachtete Zusammenhang kausal zu interpretieren ist. Es gibt Studien, die zeigen, dass die Zahl der Störche mit der Zahl der Geburten in einigen Gegenden im Verlauf des Jahres tatsächlich korreliert. Das bedeutet jedoch nicht, dass die Störche die Kinder bringen.

Mann-Whitney-U-Test. Besteht der Verdacht, dass die Voraussetzungen für einen α -T-Test verletzt sein könnten, kann am besten der U-Test von Mann und Whitney berechnet werden.

Mittelwert, Median, Modalwert. (Maße der zentralen Tendenz) Ergebnisse einer Untersuchung, wie z.B. einer schriftlichen Befragung werden häufig unter Rückgriff auf das sog. arithmetische Mittel, den Mittelwert, berichtet. Der Mittelwert, als Summe aller Antworten durch die Zahl der Antworter, trägt der Tatsache Rechnung, dass in der Regel von verschiedenen Personen auch verschiedene Antworten gegeben werden. Dennoch kann man als zentrale Tendenz aller Antworten den Mittelwert als gute Näherung für einen Großteil der Antworter ansehen. Allerdings ist der Mittelwert nicht das einzig sinnvolle Maß. So ist der Mittelwert empfindlich gegen extreme Antworten, auch dann, wenn diese nur von wenigen Personen gegeben werden.

Der Median weist eine solche Empfindlichkeit nicht auf. Der Median ist die Mitte der Messwerte, die sich dort finden lässt, wo exakt 50% der befragten Personen einen geringeren Messwert aufweisen und exakt 50% einen höheren (bzw. den gleichen). Der Modalwert ist relativ einfach definiert; er ist der Messwert, der insgesamt am häufigsten vorkommt. Bei einigen Fragestellungen ergibt es sich, dass Mittelwert, Median und Modalwert exakt den gleichen Wert aufweisen. Dies ist jedoch nicht immer der Fall. Aus der Anordnung der drei Werte gewinnen Statistiker wichtige Informationen über das Antwortverhalten der befragten Personen. Bei Merkmalen, die durch extreme Antworten verzerrt sein könnten (z.B. Altersangaben, bei denen einige wenige Personen sehr alt sind), sollte immer der Median angegeben werden. Bei einigen anderen Messwerten kann der Mittelwert zudem nicht sinnvoll berechnet werden.

P-2-seitig. Das Ergebnis eines Signifikanztests (\rightarrow statistische Signifikanz) ist im wesentlichen die Wahrscheinlichkeit dafür, dass sich zwei Messwerte nicht voneinander unterscheiden. Da Wahrscheinlichkeit auf Englisch *Probability* heißt, wird sie mit dem Buchstaben „p“ abgekürzt. p kann jedoch grundsätzlich auf zwei verschiedene Arten berechnet werden. p kann 1-seitig oder auch 2-seitig bestimmt werden. Welche der beiden Berechnungen im Einzelfall anzugeben ist, entscheidet sich durch die Fragestellung, die mit dem Signifikanztest beantwortet werden soll. Eine zweiseitige Fragestellung prüft, ob zwischen zwei Messwerten ein Unterschied besteht, ohne genauer darauf einzugehen, welche Richtung der Unterschied hat (ob der eine Messwert größer als der andere ist oder ob das Umgekehrte zu erwarten ist, wird nicht berücksichtigt). Eine einseitige Fragestellung prüft nicht nur, ob allgemein ein Unterschied besteht, sondern zudem, ob er in die erwartete Richtung geht. Der 2-seitige Wert wird also bei ungerichteten Signifikanztests angegeben. Er ist immer exakt doppelt so hoch wie der entsprechende 1-seitige Wert. Der 1-seitige Wert hat es damit „leichter“ signifikant zu werden, erfordert aber die genauere Vorhersage.

Standardabweichung, Streuung, Varianz. Die Standardabweichung oder auch Streuung genannt, ist ein Wert für die mittlere Abweichung der Messwerte vom Mittelwert. Die Streuung gibt damit einen Eindruck von der Variationsbreite der Antworten und damit zum Teil auch über die Messgenauigkeit. Bei ideal verteilten Messwerten liegt der \rightarrow Mittelwert exakt in der Mitte (zusammen mit dem \rightarrow Median und dem \rightarrow Modalwert) aller Antworten. Insgesamt 68% aller Antworten befinden sich dann in dem Messwertebereich zwischen dem Mittelwert minus der Streuung und dem Mittelwert plus der Streuung. Beispiel: Ein IQ-Test weist in der Regel einen Mittelwert von 100 und eine Streuung von 10 auf. Damit liegen 68% aller Menschen mit ihrem IQ zwischen einem IQ von 90 und 110. Die Varianz ist nichts anderes als das Quadrat der Streuung.

Statistische Signifikanz. (Statistische Bedeutsamkeit) Jeder im Rahmen einer Messung gewonnene Messwert ist mit einer gewissen Fehlertoleranz behaftet. Die Ergebnisse einer Befragung sind daher nie exakt. Die Genauigkeit einer Messung kann in vielen Fällen mit Hilfe der Wahrscheinlichkeitsrechnung angegeben werden. In diesem Sinne bezeichnet z.B. die σ Streuung eine bestimmte Schwankungsbreite für einen Messwert. Wenn nun zwei Messwerte verglichen werden sollen, z.B. die Messwerte Männer mit denen für Frauen, so muss immer auch mitbedacht werden, dass beide Messwerte ungenau sind. Sagen z.B. 40% der Männer auf eine bestimmte Frage "ja" und antworten auf dieselbe Frage nur 30% der Frauen mit "ja", so ist es nicht leicht zu entscheiden, ob dieser Unterschied durch Messungenauigkeiten zustande kommen konnte oder auf Unterschiede zwischen Männern und Frauen zurückgeführt werden kann. Ein statistischer Signifikanztest beantwortet nun die Frage danach, ob ein Unterschied zwischen zwei Messwerten durch Messungenauigkeiten erklärt werden kann. Erst wenn die Wahrscheinlichkeit für den vorgefundenen Unterschied so klein ist, dass nur noch mit einer Wahrscheinlichkeit von 5% (oder weniger) davon ausgegangen werden kann, dass kein Unterschied zwischen den Messwerten vorliegt, sagt man, dass die Unterschiede statistisch signifikant sind. D.h., dass ein statistischer Signifikanztest niemals behaupten würde, dass ein Unterschied zwischen den Männern und den Frauen besteht. Statistisch signifikant heißt nur, dass es unwahrscheinlich (aber nicht unmöglich) ist, dass kein Unterschied besteht.

Je nach erhobenen Daten müssen verschiedene Verfahren für die Signifikanzprüfung angewandt werden. Wichtige Testverfahren sind z.B.: \rightarrow T-Test, \rightarrow Fishers exakter Test, \rightarrow Chi-Quadrat-Test, \rightarrow Mann-Whitney-U-Test, \rightarrow Varianzanalyse. Das wichtigste Ergebnis eines Testes ist die Wahrscheinlichkeit (\rightarrow P-2-seitig) dafür, dass sich die Messwerte nicht unterscheiden. Diese Wahrscheinlichkeit wird mit einem vorher festgelegten Grenzwert, der Signifikanzgrenze (\rightarrow Alpha-Fehler-Adjustierung) verglichen.

T-Test. Ein besonders gebräuchlicher Signifikanztest für den Vergleich von zwei Mittelwerten ist der T-Test (\rightarrow statistische Signifikanz). Der T-Test besitzt jedoch einige Voraussetzungen, die erfüllt sein müssen, damit er berechnet werden kann. Diese Voraussetzungen sind allerdings nicht immer erfüllt. Zu den Grundvoraussetzungen gehört u.a., dass mit gutem Gewissen ein \rightarrow Mittelwert und die dazu gehörige \rightarrow Streuung berechnet werden können. Die Verteilung der Mittelwerte muss einer Normal- bzw. T-Verteilung folgen, was bei kleinen Stichproben Probleme machen kann. Bei Stichproben mit einer Gruppengröße von mindestens 25 bis 50 Personen pro Untersuchungsgruppe, liegt automatisch eine Normalverteilung der Mittelwerte vor (\rightarrow zentraler Grenzwertsatz), so dass dann kein Problem bei der Anwendung des T-Tests besteht. Der T-Test berechnet einen t-Wert, für den zusammen mit den sog. Freiheitsgraden (in der Regel Zahl der Messwerte minus eins) die Wahrscheinlichkeit bekannt ist. Die Wahrscheinlichkeit ist das Ergebnis des Tests. Man spricht von einer \rightarrow statistischen Signifikanz, wenn diese Wahrscheinlichkeit kleiner als der vorher festgelegte \rightarrow Alpha-Fehler ist.

Validität, prognostische. Wenn Verfahren zur Personalauswahl eingesetzt werden, so verspricht man sich davon Hinweise, die es tatsächlich erlauben, die geeignetsten Kandidatinnen bzw. die geeignetsten Kandidaten aus den Bewerberinnen und Bewerbern heraus zu suchen. Die Verfahren sollen also im weitesten Sinne die „Eignung“ feststellen. Ob ein Verfahren tatsächlich das misst, was es zu messen vorgibt, hier die „Eignung“, wird als Validität des Verfahrens bezeichnet. Zur Feststellung der Validität wird in der Regel eine \rightarrow Korrelation zwischen den Ergebnissen des eingesetzten Verfahrens und passender Außenkriterien (z.B. Leistungsbeurteilung durch einen Vorgesetzten) berechnet. Damit ist die Validität

quantifizierbar mit Werten zwischen Null und Eins, wobei hohe Werte einer hohen Validität entsprechen. Da es bei der Personalauswahl darum geht, die Eignung zu prognostizieren und als passende Außenkriterien Merkmale in Frage kommen, die in der Zukunft liegen, spricht man von einer prognostischen Validität, also von der Fähigkeit des eingesetzten Verfahrens, Vorhersagen über die Verwendbarkeit einer Bewerberin eines Bewerbers zu erstellen. Wie hoch die Validität im Idealfall sein soll, hängt vom Einsatzziel (z.B. von der Anzahl der wahrscheinlich ohnehin geeigneten Bewerberinnen und Bewerbern: sind wahrscheinlich ohnehin alle für die Stelle geeignet, kann die Auswahl einfach gehalten werden) und vom Aufwand (Kosten vs. Nutzen) ab. Eine hohe Validität wird Verfahren mit einem Wert über 0,3 zugesprochen. Hierzu gehört z.B. das Assessment Center, wohingegen Bewerbungsunterlagen, Schulnoten und graphologische Gutachten darunter liegen.

Varianzanalyse. (heißt auch ANOVA) In der Regel sind Signifikanztests in der Lage nur zwei Messwerte miteinander zu vergleichen. Einige Fragestellungen machen daher mehrere Vergleiche zwischen jeweils zwei Messwerten nötig, um die Frage insgesamt beantworten zu können. Beantworten drei Personengruppen einen Fragebogen (Gruppe A, B, C), so kommt man auf insgesamt drei paarweise Vergleiche (A mit B; A mit C und B mit C). Obwohl es hier möglich ist, jede Kombination der Gruppen einzeln zu vergleichen und eine Alpha-Fehler-Adjustierung vorzunehmen (\nearrow Alpha-Fehler-Adjustierung), ist eine Varianzanalyse eleganter und weniger aufwändig zu rechnen. Die Varianzanalyse löst das Problem durch einen Trick: Es werden im wesentlichen zwei Varianzen (\nearrow Streuung) ermittelt und diese mit einem F-Test verglichen. Es werden also auch hier nur zwei Werte durch den Test verglichen. Die eine Varianz ist die innerhalb der Gruppen, die andere ist die zwischen den Gruppen. Sind die Unterschiede (also die Varianz) zwischen den Gruppen größer als die Unterschiede innerhalb der Gruppen, so unterscheiden sich die Gruppen. Allerdings ist dann noch nicht bekannt, welche Gruppen sich voneinander unterscheiden. Um dies heraus zu finden werden anschließend doch wieder paarweise Vergleiche durchgeführt.

Zentraler Grenzwertsatz. Ist eine Untersuchungsstichprobe groß, so ergibt sich unabhängig von der Verteilung der Rohdaten für den Mittelwert eine Normalverteilung. Diesen Zusammenhang kann man sich wie folgt vorstellen: Es wird aus einer größeren Stichprobe eine begrenzte Zufallsauswahl getroffen und für diese Zufallsauswahl ein Mittelwert berechnet. Dies wird mehrfach wiederholt. Jeder berechnete Mittelwert beruht dann nur auf einer Zufallsauswahl und stimmt damit mit dem echten Mittelwert nur mehr oder weniger gut überein. Es zeigt sich, dass die Mittelwerte der Zufallsauswahlen um den echten Mittelwert normalverteilt streuen und zwar unabhängig von der Verteilung der eigentlichen Rohwerte. Testverfahren wie der \nearrow T-Test oder die \nearrow Varianzanalyse benötigen solche normalverteilten Mittelwerte. Diese sind nach dem zentralen Grenzwertsatz für große Stichproben immer gegeben. Was groß ist und was nicht hängt vom jeweiligen Lehrbuch ab. Einige sagen, dass 25 Personen pro Untersuchungs-Gruppe genügen, andere fordern 30 und ganz strenge sogar 50 Personen. Da für kleine Stichproben der zentrale Grenzwertsatz nicht gilt, müssen hier dann auch die Rohwerte strikt normalverteilt sein, damit Testverfahren wie der \nearrow T-Test oder die \nearrow Varianzanalyse angewendet werden können.

12. Darstellung und Abkürzungen

12.1 Allgemein

AM oder	<i>Mittelwert (arithmetisches Mittel; Mean)</i>
M oder	
\bar{x}	
SD oder	<i>Standardabweichung (Standard Deviation; Streuung)</i>
s oder	
Std.	
df oder	<i>Freiheitsgrade (degree of freedom)</i>
FG	
N	<i>Größe der erfassten Grundgesamtheit.</i>
n	<i>Größe einer Stichprobe oder Gruppe aus der Grundgesamtheit.</i>
p	<i>Wahrscheinlichkeit (kann Werte zwischen 0 und 1 annehmen. 0,6 bedeutet also eine Wahrscheinlichkeit von 60%).</i>
p-2-seitig	<i>Wahrscheinlichkeit dafür, dass etwas nicht signifikant ist (2-seitig getestet).</i>
p-1-seitig	<i>Wahrscheinlichkeit dafür, dass etwas nicht signifikant ist (1-seitig getestet).</i>
*	<i>Der Unterschied ist signifikant bei einem Alphafehler von 5% ($p \leq 0,05$)</i>
**	<i>Der Unterschied ist hoch signifikant bei einem Alphafehler von 1% ($p \leq 0,01$)</i>

12.2 Korrelationen

r Korrelation.

p-2-seitig Wahrscheinlichkeit dafür, dass eine Korrelation Null ist.

* Die Korrelation ist mit einem Alphafehler (einer Wahrscheinlichkeit) von 5% ($p \leq 0,05$) Null. Die Korrelation ist signifikant.

** Die Korrelation ist mit einem Alphafehler (einer Wahrscheinlichkeit) von 1% ($p \leq 0,01$) Null. Die Korrelation ist signifikant.

Im Text: „...Es besteht also ein hoch signifikanter Zusammenhang zwischen dem Verhalten der beobachteten Frauen und dem Verhalten der beobachteten Männer ($r = 0,52$; p -2-seitig = $0,001$)....“

In Tabellen:

Tabelle 1:

Interkorrelationsmatrix der Skalen

N = 506	Skala 1	Skala 2	Skala 3	Skala 4	Skala 5
Skala 2	0,758 **				
Skala 3	0,924 **	0,723 **			
Skala 4	0,815 **	0,589 **	0,292 *		
Skala 5	0,810 **	0,491 **	0,587 **	0,517 **	
Skala 6	0,849 **	0,599 **	0,062	0,706 **	0,562 **

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Skala 1: Neurotizismus

Skala 2: Emotionale Stabilität

Skala 3: Extraversion

Skala 4: Selbstdarstellung

Skala 5: Führungsmotivation

12.3 T-Test

- t *Prüfgröße für den T-Test.*
- df *Freiheitsgrade (degree of freedom).*
- p-2-seitig *Wahrscheinlichkeit dafür, dass zwei Mittelwerte sich nicht signifikant unterscheiden (2-seitig getestet).*
- p-1-seitig *Wahrscheinlichkeit dafür, dass zwei Mittelwerte sich nicht signifikant unterscheiden (1-seitig getestet).*
- * *Der Unterschied ist signifikant bei einem Alphafehler von 5% ($p \leq 0,05$)*
- ** *Der Unterschied ist hoch signifikant bei einem Alphafehler von 1% ($p \leq 0,01$)*

Im Text: „...Es besteht also ein hoch signifikanter Unterschied zwischen dem Verhalten der beobachteten Frauen und dem Verhalten der beobachteten Männer ($t = 3,52$; $df = 255$; p -2-seitig = $0,003$)...“

In Tabellen:

Tabelle 2:

Ergebnisse der Befragung: Männer vs. Frauen¹

	Männer			Frauen			t	df	p
	AM	SD	n	AM	SD	n			
Skala 1	5,25	1,32	500	6,01	1,12	420	2,57	918	0,004**
Skala 2	4,98	1,04	499	5,98	1,20	420	2,62	917	0,003**

** Die Unterschiede sind auf dem Niveau von 0,01 (2-seitig) signifikant.
 * Die Unterschiede sind auf dem Niveau von 0,05 (2-seitig) signifikant.

Skala 1: Neurotizismus
 Skala 2: Emotionale Stabilität

Tabelle 2 (alternative):

Ergebnisse der Befragung: Männer vs. Frauen²

	Männer		n	Frauen		t	df	p	
	AM	SD		AM	SD				
Skala 1	5,25	(1,32)	500	6,01	1,12	420	2,57	918	0,004**
Skala 2	4,98	(1,04)	499	5,98	1,20	420	2,62	917	0,003**

** Die Unterschiede sind auf dem Niveau von 0,01 (2-seitig) signifikant.
 * Die Unterschiede sind auf dem Niveau von 0,05 (2-seitig) signifikant.

Skala 1: Neurotizismus
 Skala 2: Emotionale Stabilität

¹ Alle Zahlenwerte (inklusive der Ergebnisse des Tests) sind frei erfunden

² Alle Zahlenwerte (inklusive der Ergebnisse des Tests) sind frei erfunden

12.4 Varianzanalyse

- F *Prüfgröße für den F-Test, der in der Varianzanalyse benutzt wird*
- df *Freiheitsgrade (degree of freedom).*
- p *Wahrscheinlichkeit dafür, dass sich die Gruppen nicht signifikant unterscheiden*
- * *Der Unterschied zwischen den Gruppen ist signifikant bei einem Alphafehler von 5% ($p \leq 0,05$)*
- ** *Der Unterschied zwischen den Gruppen ist hoch signifikant bei einem Alphafehler von 1% ($p \leq 0,01$)*

Im Text: „...Es besteht also ein hoch signifikanter Unterschied zwischen den beobachteten Gruppen ($F = 3,52$; $p = 0,003$)....“

In Tabellen:

Tabelle 3:
Ergebnisse der Beobachtung³

Skala 1	AM	SD	n	F	p
Männer	5,25	1,32	500	3,25	0,004**
Frauen	4,98	1,04	499		
Kinder	3,12	0,98	356		
Skala 2					
Männer	4,25	1,32	500	2,56	0,003**
Frauen	4,98	1,04	499		
Kinder	3,12	0,98	356		

** Die Unterschiede sind auf dem Niveau von 0,01 (2-seitig) signifikant.
 * Die Unterschiede sind auf dem Niveau von 0,05 (2-seitig) signifikant.
 Skala 1: Neurotizismus
 Skala 2: Emotionale Stabilität

Tabelle 4:
Signifikante Gruppenunterschiede nach Scheffé-Test⁴

Skala 1	Frauen	Kinder
Männer	*	**
Frauen		**
Skala 2		
Männer	*	
Frauen		**

** Die Unterschiede sind auf dem Niveau von 0,01 (2-seitig) signifikant.
 * Die Unterschiede sind auf dem Niveau von 0,05 (2-seitig) signifikant.
 Skala 1: Neurotizismus
 Skala 2: Emotionale Stabilität

Bei einer Varianzanalyse bieten sich zusätzlich auch Grafiken an.

³ Alle Zahlenwerte (inklusive der Ergebnisse des Tests) sind frei erfunden
⁴ Alle Zahlenwerte (inklusive der Ergebnisse des Tests) sind frei erfunden

13. Literatur

- Bortz, J. (1999) *Statistik für Sozialwissenschaften (5. Auflage)*. Berlin, Heidelberg, New York: Springer
- Bortz, J. & Döring, N. (2002) *Forschungsmethoden und Evaluation*. Berlin, Heidelberg: Springer
- Bühner, M. (2004) *Einführung in die Test- und Fragebogenkonstruktion*. München u. a.: Pearson Studium
- Churchill, G. A. & Iacobucci, D. (2002) *Marketing Research: Methodological Foundations*. Orlando, FL: Harcourt College Publishers
- Friedrichs, J. (1990) *Methoden empirischer Sozialforschung*. Opladen: Westdeutscher Verlag
- Langewitz, W., Keller, A. & Denz, M. (1995) Patientenzufriedenheits-Fragebogen (PZF): Ein taugliches Mittel zur Qualitätskontrolle der Arzt-Patient-Beziehung? *Zeitschrift für Psychotherapie Psychosomatik und medizinische Psychologie*, 45, 351-357
- Lienert, G., A. & Raatz, U. (1994) *Testaufbau und Testanalyse*. Weinheim: Beltz
- Mayring, P. (2003) *Qualitative Inhaltsanalyse*. Weinheim und Basel: Beltz Verlag
- Weick, K. E. (1985) *Der Prozeß des Organisierens*. Frankfurt am Main: Suhrkamp