



Grundlagen

Statistischer Methoden

Dezember 2008

Dr. Guido Strunk

guido.strunk@complexity-research.com

guido.strunk@wu-wien.ac.at

Inhalt

1	Theorie und Hypothese	2
2	Forschungsdesigns	4
3	Hypothetische Konstrukte	6
4	Gütekriterien	7
5	Population und Stichprobe	9
6	Messtheorie	12
7	Skala und Item	15
8	Deskriptive Statistik.....	16
9	Inferenzstatistik – Grundlagen	19
10	Grundsätzliches zur Datenauswertung mit SPSS	22
11	Deskriptive Statistik mit SPSS	26

1 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 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.

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.“

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.

Arten von Alternativ-Hypothesen Übersicht



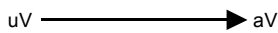
	Unterschieds-Hypothesen		Zusammenhangs-Hypothesen	
	2 Gruppen / Variablen / Objekte	Einfache Gruppenvergleiche		Einfache Korrelation
Mehr als 2 Gruppen / Variablen / Objekte	Multiple Gruppenvergleiche		Multiple Korrelation / Regression	
Hypothesentyp	gerichtet	ungerichtet	gerichtet	ungerichtet

2 Forschungsdesigns

Wichtige Begriffe



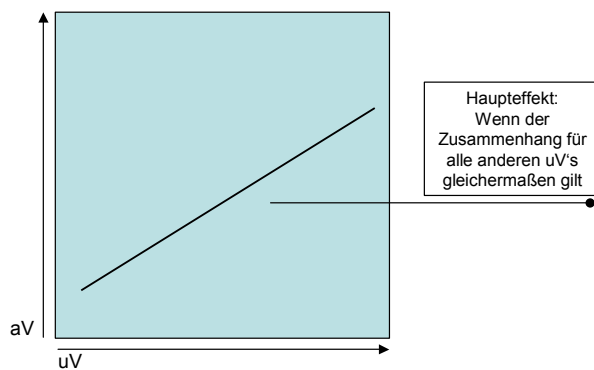
- **Unterscheidung zwischen unabhängigen (uV) und abhängigen Variablen (aV).**
Die uV ist die Variable, deren Auswirkung untersucht werden soll. Die Alternativ-Hypothese geht davon aus, dass die aV von der uV abhängig ist.



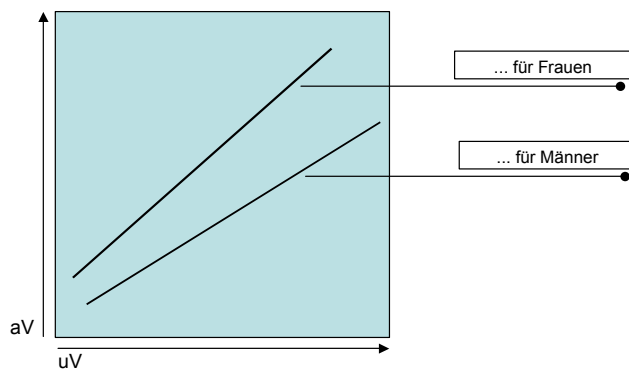
- **Unterscheidung zwischen Prädiktor/en und Kriterium.**
Prädiktoren sind *mehrere* uV's, die in der Regel *eine* aV (Kriterium) vorhersagen helfen.
- **Unterscheidung zwischen Haupteffekten und Interaktion.**
Zeigt eine uV unabhängig von anderen uVs immer die gleiche Wirkung bei der aV, so heißt das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

11

Typen von Effekten 1 Haupteffekt

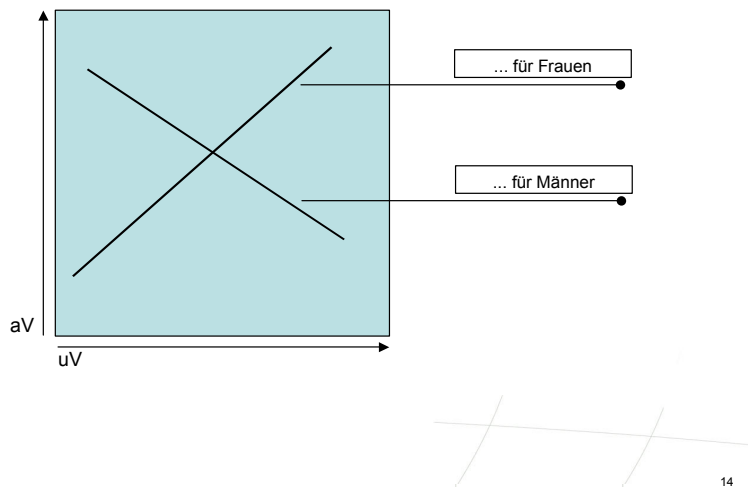


Typen von Effekten 2 Ordinale Interaktion



13

Typen von Effekten 3 Discordale 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
------------------	------------------	---------------------

3 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.
- 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 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.

4 Gütekriterien

Gütekriterien 1: Intersubjektivität

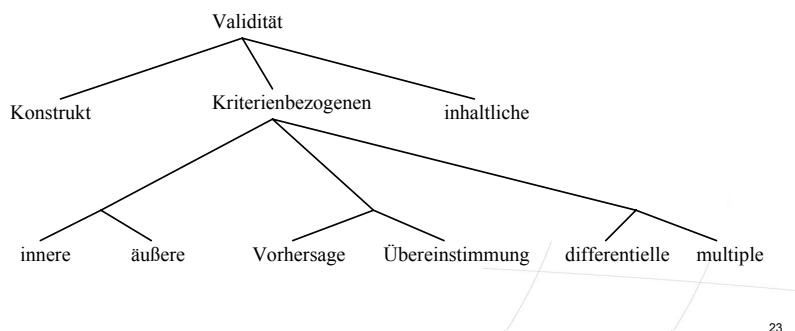


- Die Forderung nach **Intersubjektivität** entspricht im Wesentlichen der historisch älteren Forderung nach **Objektivität**.
- Theoretische Aussagen, Forschungsprozess und Ergebnisse müssen auch von anderen Personen (als den unmittelbar involvierten ForscherInnen) überprüft und nachvollzogen werden 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?



23

- **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**).

- 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. **Erwartungstafeln**, die angeben, wie viel Prozent der Probanden bei einem bestimmten Testwert z.B. eine Ausbildung gut abschließen.
- Die **Übereinstimmungsvalidität** zielt hingegen keine Prognose an, so dass Kriterium und Test gleichzeitig erhoben werden können.
- Die **differenzielle 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



28

5 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.

31

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.

32

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 Nachzügler (die letzten 20%) vergleichen.

33

- **Statistisches Verfahren: Viele statistische Verfahren erfordern die Normalverteilung der Mittelwerte der Datenstichproben. Diese ist ab 30 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.

35

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.

36

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.

37

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.

38

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:
www.psych.uni-duesseldorf.de/aap/projects/gpower/

6 Messtheorie

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

41

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**



42

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**



43

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]

46

Fragebogen 1



Basisdaten	Art der Erhebung (Beispiele)	Skalenniveau
Alter	Alter: _____ Geburtsjahr: _____	Intervall-Skala
	<input type="radio"/> 1 10-15 Jahre <input type="radio"/> 2 16-25 Jahre <input type="radio"/> 3 26-35 Jahre <input type="radio"/> 4 älter als 35 Jahre	Ordinal-Skala

Basisdaten	Art der Erhebung (Beispiele)	Skalenniveau
Geschlecht	<input type="radio"/> 0 männlich <input type="radio"/> 1 weiblich	Nominal-Skala
Beruf	<input type="radio"/> 1 arbeitslos <input type="radio"/> 2 Arbeitsunfähigkeit <input type="radio"/> 3 ArbeiterIn <input type="radio"/> 4 Angestellte/r <input type="radio"/> 5 Selbstständige/r <input type="radio"/> 6 StudentIn <input type="radio"/> 7 Ausbildung <input type="radio"/> 8 Hausfrau / Hausmann <input type="radio"/> 9 RentnerIn <input type="radio"/> 10 Wehr- / Zivildienst Leistender <input type="radio"/> -1 unbekannt	Nominal-Skala

Ratingskalen	Art der Erhebung (Beispiele)	Skalenniveau
Gerade Anzahl an Abstufungen	○ ○ ○ ○ gut schlecht	Intervall-Skala/ Ordinal-Skala
Ungerade Anzahl an Abstufungen	○ ○ ○ ○ ○ gut schlecht	Intervall-Skala/ Ordinal-Skala
Keine Abstufungen	gut ————— schlecht	Intervall-Skala
Ungerade Anzahl Abstufungen Beschriftet	○ ○ ○ ○ ○ sehr gut gut mittel schlecht sehr schlecht	Intervall-Skala/ Ordinal-Skala

Gute Beispiele finden sich auch in Bortz und Döring (2002)

49

Nutzen Sie in Fragebögen die Kategorie „unbekannt“ oder „weiß nicht“, wenn ...

- es wirklich möglich ist, dass etwas unbekannt ist (dann auf jeden Fall)
- wenn Sie den Ausfüllenden mehr Freiheit geben wollen (Vorsicht: Fehlende Daten!)

7 Skala und Item

Messung hypothetischer Konstrukte – ein Beispiel

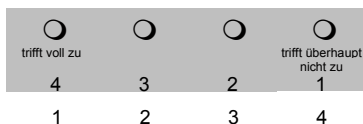


- + Dieser Arzt diese Ärztin hat mich freundlich behandelt.
- Ich habe gewisse Zweifel über die Fähigkeit dieses Arztes / dieser Ärztin.
- Dieser Arzt / diese Ärztin wirkte kühl und unpersönlich.
- + Dieser Arzt/diese Ärztin hat sein/ihr Bestes getan, um mich nicht zu beunruhigen.
- + Dieser Arzt/diese Ärztin hat mich so sorgfältig es notwendig war, untersucht.
- Dieser Art/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.

52

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

- **Skalierung über Ratings**



bei positiven Items
bei negativen Items

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

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

55

8 Deskriptive Statistik

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):**

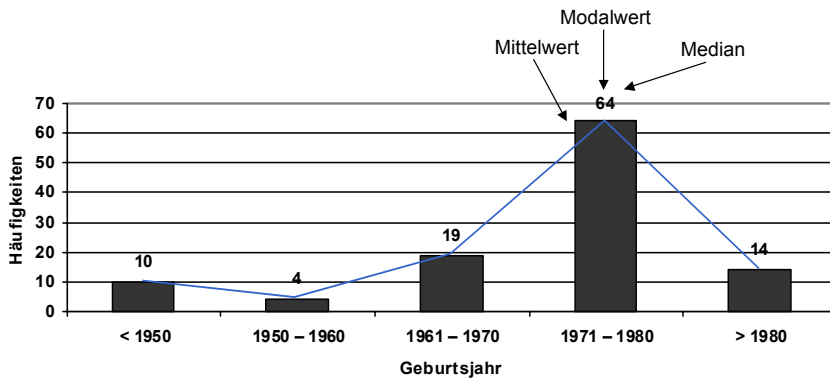
$$GM = \sqrt[n]{\prod_{i=1}^n x_i}$$

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

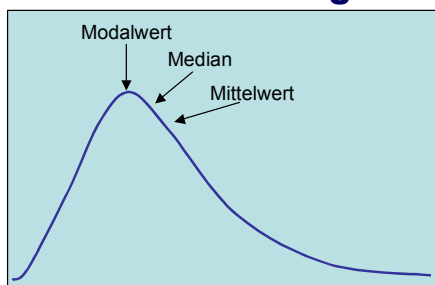
$$HM = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

57

Modalwert, Median, Mittelwert



Schiefe Verteilungen



Linkssteil bzw. rechtsschief

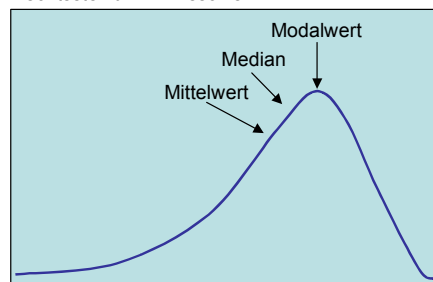
Rechtssteil bzw. linksschief

➔ Schiefe

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

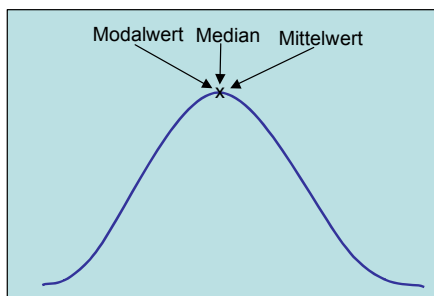
Sch < 0: Rechtssteil bzw. linksschief

Sch > 0: Linkssteil bzw. rechtsschief



59

Symmetrische Verteilungen



Streuungsparameter



- **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**

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

- **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 $n > 30$, dann gleichen sich die Ergebnisse zunehmend an und es wird unbedeutend, ob mit n oder mit $n-1$ gerechnet wird.

62

Tabellarische Darstellung



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

Immer wenn ein Maß der zentralen Tendenz angegeben wird, muss dazu auch das passende Streuungsmaß angegeben werden.

63

9 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.

65

Grundidee – ein Beispiel



- **Der Münzwurf ist mathematisch gesehen binomial verteilt:**

$$P_{(X=k)} = n! / (k! (n-k)!) * P^k * (1-P)^{(n-k)}$$

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 $k=1$, $k=2$, bis $k=15$ (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.**

66

Sind die Frauen im Beispiel jünger?



Tabelle 1: Geburtsjahr nach Geschlecht

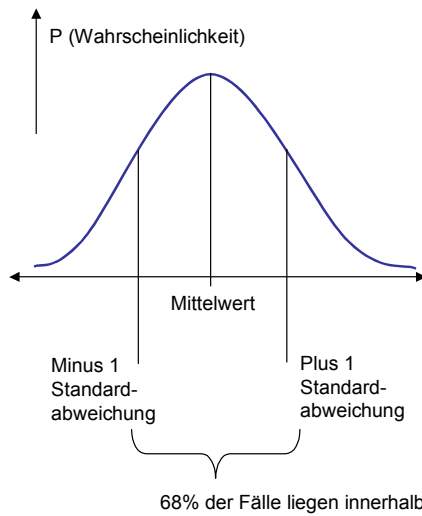
Geschlecht	Mittelwert	Streuung	N
Frauen	1972,73	6,70	121
Männer	1970,62	10,59	70
Gesamt	1971,38	9,40	191

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.

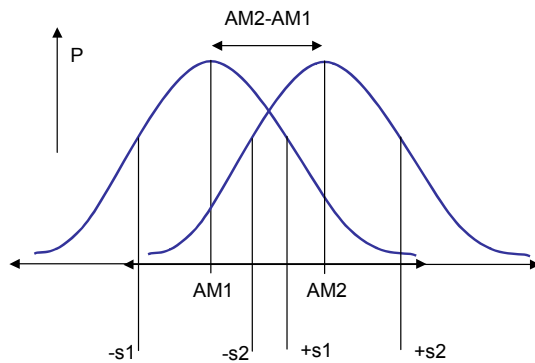
Ob der Unterschied statistisch bedeutsam ist (also signifikant ist) kann mit Hilfe der Wahrscheinlichkeitsrechnung entschieden werden.

Normalverteilung



68

Vergleich zweier Mittelwerte

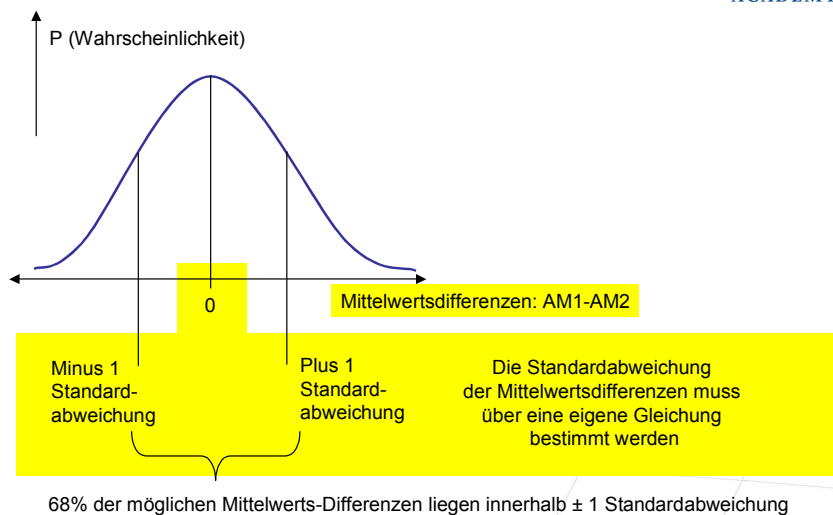


AM1: Mittelwert der Gruppe 1
AM2: Mittelwert der Gruppe 2
s: Standardabweichungen.

- 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.

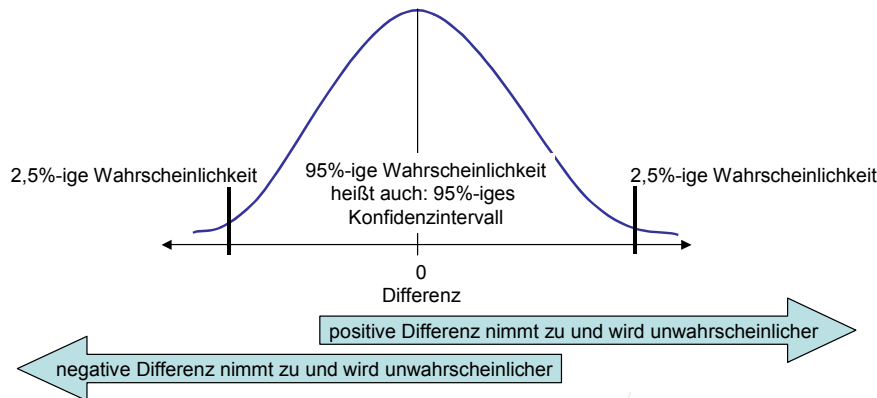
69

Normalverteilung von Mittelwerts-Differenzen



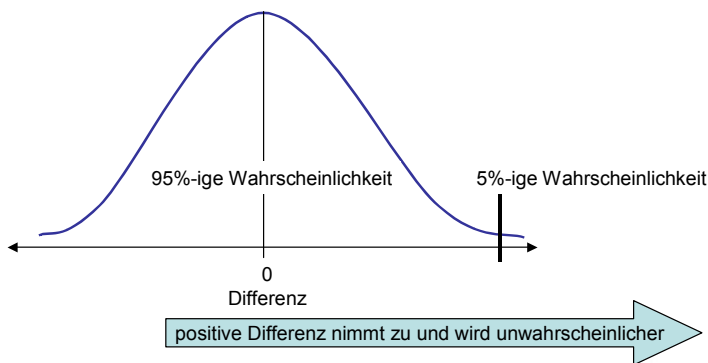
70

Vergleich zweier Mittelwerte 2-seitig



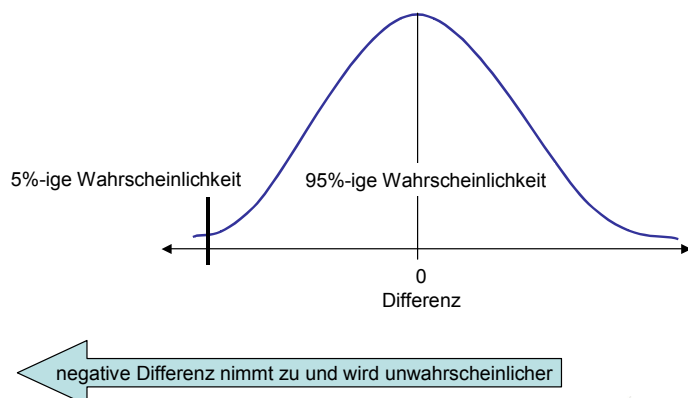
71

Vergleich zweier Mittelwerte 1-seitig



72

Vergleich zweier Mittelwerte 1-seitig



73

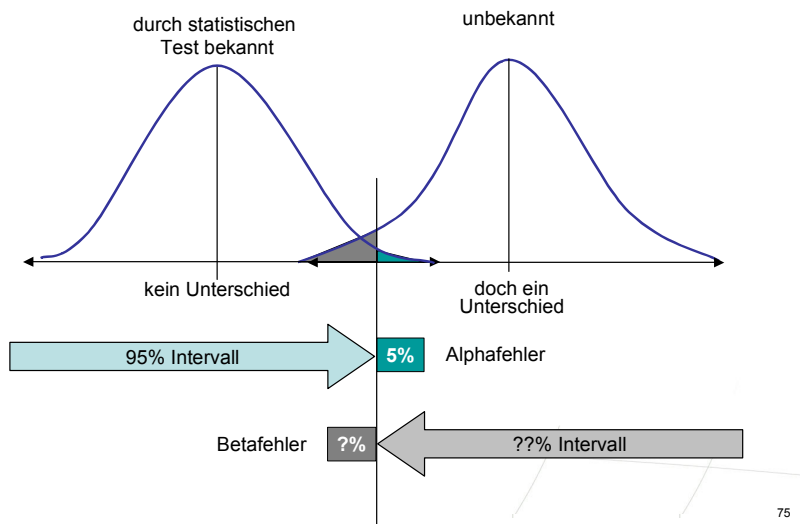
Zusammenfassung



- 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. $\alpha = 0,01$).

74

Alpha- und Betafehler



75

10 Grundsätzliches zur 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 ...



Zum Festlegen der Bedingung auf „Falls...“ klicken

Status

78

Fälle auswählen ... Falls...



Formel für die Bedingung

Variablenliste

~ bedeutet „nicht“

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

79

Fälle auswählen ... Falls ... beendet



Bedingung wird angezeigt

Löschen?

„OK“ und „Einfügen“ wird verfügbar

80

Filter an



Auswahl der Fälle

Filterstatus

81

Transformieren / Berechnen ...



Formel

Variablenliste

Kopiert eine ausgewählte Variable in den Formel-Editor

neue Zielvariable definieren

82

Transformieren / Berechnen ... Falls...



Gruppe=1

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

neue Zielvariable definieren

83

Transformieren / Berechnen ... Falls...



The screenshot shows the 'Variable berechnen' dialog box in SPSS. A secondary window titled 'Variable berechnen: Falls Bedingung erfüllt ist' is open, showing the condition 'alter >= 6 AND alter < 18'. Annotations include:

- 1. Fall einschließen, wenn...**: Points to the 'Fall einschließen, wenn Bedingung erfüllt ist' radio button.
- 2. Bedingung**: Points to the text input field containing the condition.
- 3. „Weiter“**: Points to the 'Weiter' button.

Below the dialog boxes, text reads: **Und-Verknüpfung: AND** and **Oder-Verknüpfung: OR**.

84

Typischer SPSS-Dialog



The screenshot shows the 'Häufigkeiten' dialog box in SPSS. Annotations include:

- verfügbare Variablen**: Points to the list of variables on the left.
- kopiert ausgewählte Variablen in eine Liste**: Points to the arrow button between the variable lists.
- Variablenauswahl für die Auswertung**: Points to the 'Variable(n):' list on the right.
- Optionen / Einstellungen**: Points to the 'Statistik...', 'Diagramme...', and 'Format...' buttons.
- Einfügen: Syntax schreiben**: Points to the 'Einfügen' button.
- OK: Ausführen**: Points to the 'OK' button.

85

11 Deskriptive Statistik mit SPSS

Statistik \ Zusammenfassen \ Häufigkeiten



OK: Ausführen

Einfügen: Syntax schreiben

Optionen / Einstellungen

kopiert ausgewählte Variablen in eine Liste

verfügbare Variablen

Variablenauswahl für die Auswertung

87

Statistik \ Zusammenfassen \ Deskriptive Statistiken



OK: Ausführen

Einfügen: Syntax schreiben

Optionen / Einstellungen

kopiert ausgewählte Variablen in eine Liste

verfügbare Variablen

Variablenauswahl für die Auswertung

88