

Leitfaden zur Anwendung empirischer Forschungsmethoden in der nutzerzentrierten Produktentwicklung

Forschung mit und an Menschen in den Ingenieurwissenschaften

Diana Fotler¹, René Germann¹, Barbara Gröbe-Boxdorfer²,
 Werner Engeln² and Sven Matthiesen¹

¹ IPEK - Institute of Product Engineering, Karlsruhe Institute of Technology (KIT),
 Kaiserstraße 10, 76131 Karlsruhe, Germany

² Pforzheim University – School of Engineering,
 Tiefenbronner Straße 65, 75175 Pforzheim, Germany
 {Diana Fotler, René Germann, Sven Matthiesen}@kit.edu
 {Barbara Groebe, Werner Engeln}@hs-pforzheim.de

Keywords: Probandenstudie, User Studies, Forschung mit und an Menschen, Ingenieurwissenschaftliche Forschung, empirische Forschungsmethoden, Mensch-Maschine-Systeme

Der vorliegende Leitfaden hat den Anspruch, Orientierung innerhalb des empirischen Forschungsprozesses zu bieten. Insbesondere beim Aufbau, der Durchführung und der Auswertung in anwendungsorientierten Untersuchungen in der nutzerzentrierten Produktentwicklung, also Untersuchungen mit Probanden (User Studies), soll er unterstützen und dadurch einen Beitrag zur Steigerung der Qualität sowie der Vermeidung von Fehlern leisten. Das Besondere dieses Leitfadens ist die Zusammenführung von Wissen und wissenschaftlichen Methoden aus verschiedenen Fachbereichen mit dem Fokus auf den Menschen in der Forschung und seinen Einfluss auf die Datenerhebung.

An dieser Stelle wird betont, dass kein Anspruch auf Vollständigkeit besteht. Empirische Forschungsmethoden entwickeln sich fortlaufend weiter und sollten in der Anwendung immer überprüft, an die Forschungsfrage und damit einhergehend an die Hypothesen angepasst und diskutiert werden. Für erfolgreiche Forschungsarbeiten soll der Leitfaden ein Grundgerüst darstellen, welches die vielfältigen Forschungsmethoden der verschiedenen wissenschaftlichen Disziplinen sowie den empirischen Forschungsprozess mit Fachbegriffen aufzeigt, Hinweise gibt und für neue Anwendungsmöglichkeiten in der Produktentwicklung inspiriert.

Hinweis zur Nutzung:

Es wird empfohlen, den Leitfaden vor Beginn der Planung eines Forschungsvorhabens zu überfliegen und sich passend zu den persönlichen Herausforderungen im Forschungsfeld wichtige Stellen zu markieren. Anschließend, während der Planungsphase, kann er zur Aufbereitung der einzelnen Forschungsschritte und oder parallel zum Erstellen des eigenen Untersuchungsplans (in der Form: „*Habe ich das beachtet?*“) Hilfestellung leisten. Inhaltlich ist der Leitfaden chronologisch nach dem idealtypischen Ablauf einer empirischen Untersuchung aufgebaut, sodass er bei Bedarf den gesamten Forschungsprozess über als Nachschlagewerk dienen und vor allem als Hinweisgeber genutzt werden kann.

Inhaltsverzeichnis

1.	GRUNDLAGEN – Ethische Richtlinien	5
2.	EMPIRISCHE UNTERSUCHUNG: PROZESS WISSENSCHAFTLICHER ERKENNTNISGEWINNUNG	7
2.1	Planung	7
2.1.1	Klärung Randbedingungen	7
2.1.2	Forschungsfrage auswählen	8
2.1.3	Hypothesen & Theorien generieren, formulieren und einbetten	9
2.1.4	Operationalisierung	12
2.1.5	Forschungsansätze in den empirischen Wissenschaften im Überblick	14
2.1.6	Datenerhebung in den empirischen Wissenschaften im Überblick	25
2.1.7	Gütekriterien empirischer Untersuchungen berücksichtigen	28
2.1.8	Auswahl & Anzahl der Untersuchungsteilnehmer	30
2.1.9	Berücksichtigung von Störvariablen und Störfaktoren	34
2.1.10	Potenzielle Einflussfaktoren / Effekte (sozial)psychologischer Art	39
2.1.11	Untersuchungsplan erstellen	41
2.2	Durchführung	42
2.2.1	Messen und messtheoretische Probleme	42
2.2.2	Skalenniveaus berücksichtigen	43
2.3	Auswertung	45
2.3.1	Datenanalyse & Hypothesenprüfung	45
2.3.2	Interpretation & Diskussion der Ergebnisse	71
2.4	Stichwortartige Zusammenfassung	72
	ANHANG	74
	Anhang A: Checkliste	74
	Anhang B: Untersuchungsplan	75
	Anhang C: Informationspflicht Versuchsteilnehmer	78
	Anhang D:	79
	Literaturverzeichnis	81

Was erwartet mich im Leitfaden?

Der **Gedanke des Leitfadens** ist die Bereitstellung des vorhandenen Methodenwissen anderer Wissenschaften im Bereich Forschung mit und an Menschen. Den Ingenieurwissenschaften steht als interdisziplinäre und empirische Wissenschaft ein breites Spektrum an Methodenwissen anderer Wissenschaften zur Verfügung. Bisher wird hier jedoch ein relativ einseitiges methodisches Feld genutzt. Der Zugang zu Fachliteratur anderer Wissenschaften wird unter anderem durch die (Fach)-Sprache (viele disziplinabhängige Fremdwörter), inkonsistente Begriffsverwendungen und der fehlenden Verfügbarkeit erschwert.

Diesen Zugang soll der Leitfaden bereitstellen. Der Schwerpunkt liegt auf den quantitativen Methoden, welche für Untersuchungen mit und an Probanden eingesetzt werden können. Qualitative Methoden sind sicherlich empfehlenswert, vor allem in explorativen (offene Themenstellung, neues Forschungsfeld) und Voruntersuchungen, auch um quantitative Daten zu erheben. Da diese weniger häufig als die quantitativen Methoden in der Forschung mit und an Menschen in der nutzerzentrierten Produktentwicklung eingesetzt werden, wird nicht vertieft auf sie eingegangen. Zudem sind qualitative Methoden oft mit hohen zeitlichen Aufwänden und vielen Ressourcen verbunden, die in der Forschung in der Produktentwicklung meist nicht gegeben sind.

Der Leitfaden ersetzt keine eigenen themenspezifischen Recherchen.

Wie in Abbildung 1 dargestellt, ist der empirische Forschungsprozess in die Wissenschaftstheorie und Forschungsethik eingebettet. Er erstreckt sich über die vier Phasen der Planung, Durchführung (Erhebung) und Auswertung von Daten sowie der Veröffentlichung. Letzt genannte Phase wird im Leitfaden aufgrund der hohen Komplexität nicht näher betrachtet. (Akremi, Baur & Fromm, 2011, S. 15)

Mit jedem Kapitel soll der/die Forschende für den nächsten Schritt des Forschungsprozesses befähigt werden. Literaturempfehlungen bieten jeweils Vertiefungspotenzial. Unabhängig des entsprechenden Schritts sind Gütekriterien einzuhalten. Zudem sind (spätestens) ab der Operationalisierung mögliche Einflussfaktoren auf die Untersuchung sowie die Skalierungsniveaus der Variablen zu beachten.

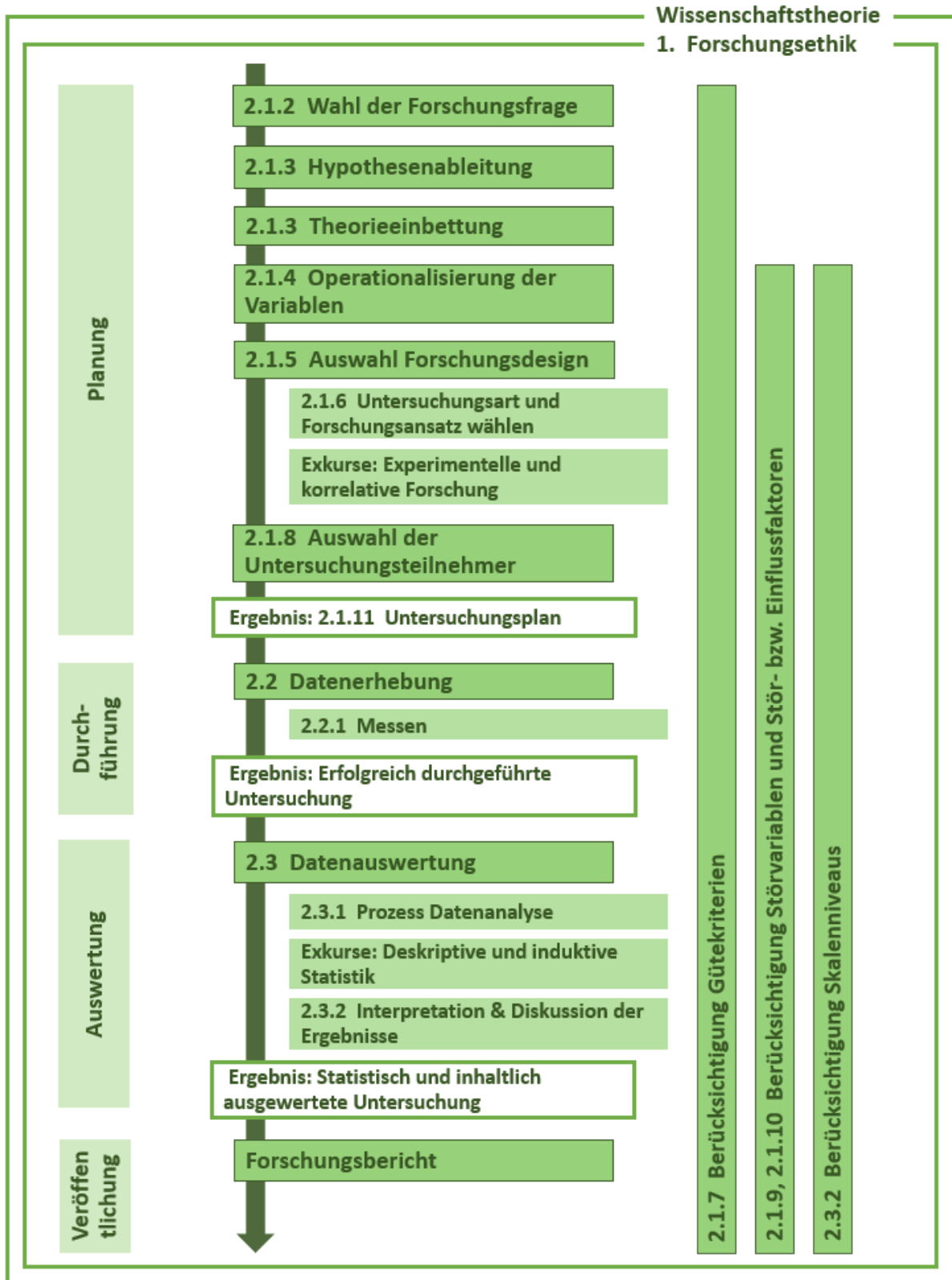


Abbildung 1: Phasen des Forschungsprozesses

1. GRUNDLAGEN – Ethische Richtlinien

Aufgabe des Kapitels: Was erwartet mich hier?

Zum Start eines Forschungsvorhabens wird eine Auseinandersetzung mit ethischen Grundsätzen empfohlen. Passend zur Forschung mit und an Menschen werden folgend Richtlinien erläutert.

Empfehlung: „Leitlinien zur Sicherung guter wissenschaftlicher Praxis – Kodex“ der Deutsche Forschungsgemeinschaft (DFG) von 2019 lesen. Diese grundlegenden Leitlinien, die zentrale Standards guter wissenschaftlicher Praxis zusammenfassen, verankern die verbindliche Kultur wissenschaftlichen Integrität.

🔍 Gute wissenschaftliche Praxis, Grundeinstellungen des Forschenden, wissenschaftstheoretische Einordnung der Ingenieurwissenschaften, Grundlagen empirischer Forschung

Leitlinien für ethische Grundsätze (zusammengefasst)

(DGPs und BDP, 2004; Häder, 2009; Rat für Sozial- und Wirtschaftsdaten [RatSWD], 2017; VDI- Verein deutscher Ingenieure, 2002; WMA Generalversammlung, 2013)

- „Kunst und Wissenschaft, Forschung und Lehre sind frei“ (Art. 5 Abs. 3 GG, Häder, 2009, S. 5)
- Tägliches Handeln ist nach bestem Wissen und Gewissen an ethischen Grundsätzen und den daraus abgeleiteten Handlungsmaximen auszurichten
- Ansprechpartner: Ombudspersonen für die ethischen Grundsätze & Ethikkommission an Forschungseinrichtungen
- Forschungsvorhaben unter Einbeziehung von Probanden: Auseinandersetzung mit sicherheitsrelevanten Risiken der Forschung zur Minimierung von Risiken für Menschenwürde, Leben, Gesundheit, Freiheit, Eigentum und Umwelt sowie des friedlichen Zusammenlebens

Forschungsbezogene ethische Richtlinien:

(Deutsche Vereinigung für Sportwissenschaft e.V. [dvs], 2003; DGPs und BDP, 2004; Eckard Kämper, 2016; Häder, 2009; RatSWD, 2017; WMA Generalversammlung, 2013)

Für die Forschung mit und an Menschen sind forschungsbezogene ethische Richtlinien relevant. Im Folgenden v.a. aus der Medizin, der Psychologie und den Sozialwissenschaften zusammengetragen:

- *Genereller ethischer Grundsatz:* Bei der Teilnahme von Menschen in der Forschung, trägt der/die Versuchsleiter/in die Verantwortung der Versuchsteilnehmer. Die **Würde und Integrität** der teilnehmenden Personen gilt es sicherzustellen, das Selbstbestimmungsrecht, die **Privatsphäre** und **Vertraulichkeit** persönlicher Informationen der Versuchsteilnehmer zu schützen, geeignete Maßnahmen zur Gewährleistung von **Sicherheit** und **Wohl** sowie dem **Ausschluss von Risiken** sind zu treffen. Die Rechte und Interessen der einzelnen Versuchspersonen haben immer Vorrang.

Vor Beginn der Forschung ist eine Aufklärung durchzuführen. Darauf basierend ist die persönliche **Einwilligung** der an der Forschung teilnehmenden Personen einzuholen, die in geeigneter Weise dokumentiert wird. Auf eine informierte Einwilligung kann unter bestimmten Voraussetzungen verzichtet werden. Diese sind zu prüfen, grundsätzlich darf die Teilnahme an der Forschung kein Unbehagen erzeugen, das über alltägliche Erfahrungen hinausgeht.

Die **Aufklärung** beinhaltet: Zweck der Forschung, erwartete Dauer und Vorgehen der Untersuchung, Recht auf Ablehnen / Abbruch der Untersuchung und die dadurch verursachten absehbaren

Konsequenzen, besondere Umstände der Untersuchung (z.B. potenzielle Risiken), voraussichtlicher Erkenntnisgewinn durch die Forschungsarbeit, Gewährleistung von Vertraulichkeit und Anonymität sowie ggf. deren Grenzen, Bonus / Anreize für die Teilnahme, Ansprechpartner bei Fragen (Gelegenheit geben, um Fragen zum Forschungsvorhaben zu beantworten) und eine Nachfrage, ob alle Informationen verstanden wurden.

- Über das *Aufnehmen von Stimmen und Bildern* soll eine auf *Aufklärung* basierende Einwilligung im Rahmen eines Forschungsvorhabens stattfinden
- *Nicht-Teilnahme oder vorzeitige* Beendigung der Teilnahme darf keine nachteiligen Konsequenzen haben
- *Anreize zur Versuchsteilnahme* sind, wenn nötig, verhältnismäßig zu wählen. Werden berufliche Leistungen oder Dienste (z.B. Beratung, Schulung) angeboten, sind die mit ihnen verbundenen Risiken, Verpflichtungen und Grenzen zu erläutern.¹
- *Täuschungen der Versuchsteilnehmer* sind zu vermeiden. Wenn Täuschungen mit psychischen und physischen Belastungen einhergehen, sind diese zu unterlassen. Jede Täuschung ist so früh wie möglich im Experiment aufzuklären, spätestens aber am Ende der Datenerhebung und geht mit der Erlaubnis des Zurückziehens der Daten einher.
- *Ein Debriefing* über das Ziel, die Ergebnisse und Schlussfolgerungen aus der Forschungsarbeit sind sobald wie möglich durchzuführen.
- *Darstellung der Forschungsergebnisse:* Keine Daten werden erfunden oder gefälscht. Bedeutsame Fehler in eigens veröffentlichten Daten gilt es zu korrigieren durch Berichtigung, Zurückziehen, ein Korrekturverzeichnis (Errata) oder andere angemessene Publikationsmittel. Bereits veröffentlichte Daten dürfen nicht als Originaldaten veröffentlicht werden.
- *Forschung muss wissenschaftlichen Grundsätzen entsprechen*, wie bspw. die Vermeidung von Plagiaten, die *Kennzeichnung des Leistungsanteils an einer Forschungsarbeit in Publikationen*, die *Weitergabe von Forschungsdaten (anonym!) nur zum Zweck der Überprüfung*. Als *Gutachter* werden *Vertraulichkeit und Eigentumsrechte von Informationen berücksichtigt*.
- *Mögliche Umweltschäden sollen in der Forschung minimiert werden*
- *Forschung am Menschen soll nur von Personen durchgeführt werden, die angemessen ethisch und wissenschaftlich ausgebildet, geübt und qualifiziert sind. Zudem müssen sie erfahren genug sein, um die Handlungssicherheit zu gewährleisten.*
- *Auswahl gesunder, körperlich und geistig belastbarer Teilnehmer*
- In einem *Studienprotokoll* müssen Planung und Durchführung einer jeden wissenschaftlichen Studie am Menschen klar beschrieben und gerechtfertigt werden. Enthalten soll es ethische Erwägungen, Informationen über Finanzierung, Sponsoren, institutionelle Verbindungen, mögliche Interessenkonflikte, Anreize für Versuchspersonen sowie Informationen bezüglich Vorkehrungen für die Behandlung und/oder Entschädigung von Personen enthalten, die infolge ihrer Teilnahme an der wissenschaftlichen Studie einen Schaden davongetragen haben.

Empfehlung:

 Vorlage für Untersuchungen zur Abfrage des Einverständnisses der Probanden:

Anhang C: Informationspflicht Versuchsteilnehmer

¹ Achtung: Finanzielle Anreize können die Studienergebnisse verfälschen, da Versuchspersonen dem Versuchsleiter gefallen möchten (Orne, 1962)

2. EMPIRISCHE UNTERSUCHUNG: PROZESS WISSENSCHAFTLICHER ERKENNTNISGEWINNUNG

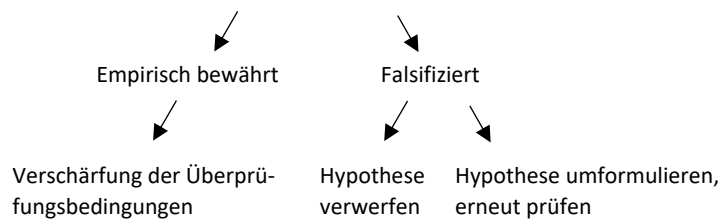
Aufgabe des Kapitels: Was erwartet mich hier?

Dieses Kapitel bildet den Prozess einer empirischen Untersuchung in seiner Gesamtheit ab. Es stellt die wesentlichen Elemente entsprechend der Forschung mit und an Menschen in der nutzerzentrierten Produktentwicklung bereit. Somit besteht die Möglichkeit, sich bei der Erstellung einer eigenen Untersuchung als Forschende(r) entlang dieses Kapitels zu hangeln.

Zur effizienten Arbeitsweise wird empfohlen, sich noch **vor** Beginn der Untersuchungsplanung einen Überblick über das gesamte Kapitel zu verschaffen.

Grundlegendes Schema der empirischen Untersuchung:

1. Hypothesenaufstellung
2. Datenerhebung (empirisch)
3. Datenauswertung → Beobachtungsaussagen



2.1 Planung

2.1.1 Klärung Randbedingungen


Aufgabe des Kapitels: Was erwartet mich hier?

Zu Beginn des Forschungsvorhabens sollen die gegebenen Randbedingungen überblickt werden. So kann Konflikten direkt vorgebeugt und der Bedarf einer Untersuchung abgesichert werden.

In welcher Form Randbedingungen festgehalten werden, ist zweitrangig, Hauptsache man führt sich diese vor Augen. Angeboten wird hierzu eine Checkliste.

Die abgefragten Randbedingungen (z.B. Ziele, Anforderungen, Abhängigkeiten, Aufwände und Ressourcen) dienen zur Orientierung während der Untersuchung und sollen als Vorarbeit für einen Zeitplan, einen Arbeitsplan, etc. genutzt werden.

Empfehlung:

 Vorlage Checkliste für Untersuchungen zur Abfrage der Randbedingungen:

Anhang A: Checkliste

2.1.2 Forschungsfrage auswählen

(Bortz & Döring, 2006, 59 ff.)

Aufgabe des Kapitels: Was erwartet mich hier?

Hier werden notwendige Vorarbeiten für das Formulieren einer Forschungsfrage benannt. Ein Grundverständnis für diesen wichtigen Schritt in der Forschung wird aufgebaut. Man lernt, die Forschungsfrage zu formulieren, in Haupt- und Zusatzfragen einzuteilen und wie daraus anschließend Hypothesen abgeleitet werden.

Forschungsfrage:

Die Fragestellungen, die in einem Forschungsvorhaben untersucht wird, wird Forschungsfrage genannt. Ziel der Untersuchung ist es, die Forschungsfrage am Ende der Arbeit zu beantworten. Falls dies nicht möglich ist, sollte die Untersuchung kritisch hinterfragt werden.

Die Forschungsfrage setzt sich gewöhnlich aus einer oder mehreren Hauptfragen und ggf. mehreren Zusatzfragen zusammen. Die Fragen werden möglichst einfach gehalten. Aus den einzelnen Fragen lassen sich Hypothesen ableiten, die in der/den Untersuchung(en) getestet werden.

Vorarbeit:

- Recherche zu Stand der Forschung (Kenntnisse über theoretische Ansätze, empirische Studien, bisher eingesetzte Methoden) durchführen
- Vorstellungen über die Art der Untersuchung festlegen
- Randbedingungen der Untersuchung, wie Motivation, Verfügbarkeit Forschungsressourcen, Möglichkeit der Publikation klären

Nach der Vorarbeit wird das Untersuchungsthema vorläufig festgelegt.

Möglichkeiten, an ein Thema heranzutreten sind: die Überprüfung von Hypothesen und Forschungsfragen, die Replikation wichtiger Untersuchungen, die Klärung widersprüchlicher Untersuchungen oder Theorien, die Überprüfung neuer methodischer oder untersuchungstechnischer Varianten, die Überprüfung des Erklärungswertes bisher nicht beachteter Theorien oder die Erkundung von Hypothesen in neuen Forschungsfeldern.

Nach der Gesamtplanung der Untersuchung wird das Untersuchungsthema endgültig festgelegt und der Arbeitstitel formuliert. Wichtig ist eine möglichst präzise Formulierung im Kontext des bereits vorhandenen Wissens. Je nach Neuheitsgrad des Forschungsgebiets wird eine allgemeine oder eindeutig, scharf abgegrenzte Formulierung des Titels verwendet.

Beispiele Arbeitstitel:

- Neues Forschungsgebiet: „Zur Frage des Einflusses verschiedener Baumaterialien von Häusern aus das Wohlbefinden ihrer Bewohner – Eine Erkundungsstudie“
- Spezieller Beitrag zu Forschungsgebiet mit langer Forschungstradition: „Vergleichende Analyse exosomatischer und endosomatischer Messungen der elektrodermalen Aktivität in einer Vigilanzsituation – Befunde einer Laboruntersuchung“

2.1.3 Hypothesen & Theorien generieren, formulieren und einbetten

(Bortz & Döring, 2006, 491 ff, 11f)

Aufgabe des Kapitels: Was erwartet mich hier?

Das Verständnis, wie Hypothesen mit Theorien und Beobachtungen zusammenhängen, worum es sich hierbei handelt und wie diese jeweils formuliert und bewertet werden, wird nachfolgend mit anwendungsnahen Beispielen aufgebaut.

Hypothese: (Bortz & Döring, 2006, S. 729)

Eine Hypothese ist eine Annahme über einen Sachverhalt, die in einem Konditionalsatz (Wenn-dann- / Je-desto-Satz) ausgedrückt wird. Mit der Formulierung einer Hypothese findet eine empirische ‚Übersetzung‘ des Forschungsproblems statt. Wissenschaftliche Hypothesen müssen generalisierbar und anhand von Beobachtungsdaten falsifizierbar sein.

Vermutete Zusammenhänge, Unterschiede und Veränderungen werden in den verschiedenen Hypothesenarten ausgedrückt. Des Weiteren sind Hypothesen in ihrer Gültigkeit: universell, beschränkt universell und quasiuniversell sowie in ihrem Abstraktionsniveau zu unterscheiden: Forschungs- (*allg.*), operationale und statistische (*spezifisch*) Hypothesen.

Arten von (Forschungs-)Hypothesen in Auswahl

(Bortz & Döring, 2006, 491, 505f; Bortz & Schuster, 2010; Bunge, 2012a; Hussy, Schreier & Echterhoff, 2013, 115f)

Zusammenhangshypothesen: Zwischen zwei/mehr Merkmalen besteht ein Zusammenhang

- **Unterschiedshypothesen:** Zwei/mehrere Populationen unterscheiden sich bzgl. einer/mehrerer abhängiger Variablen
- **Veränderungshypothesen:** Die Ausprägungen einer Variablen verändern sich im Zeitverlauf
- **Universelle Hypothesen:**
Aussage ohne jede Einschränkung; es ist keine vollständige Überprüfung möglich.
z.B.: „Intelligenz bedingt Ängstlichkeit.“
- **Beschränkt universelle Hypothesen:**
Einschränkungen vorhanden; eine Falsifizierung ist möglich, jedoch keine Verifizierung.
z.B.: „Intelligenz bedingt in Gefahrensituationen die Ängstlichkeit.“
- **Quasiuniverselle Hypothesen:** Einschränkungen beziehen sich nur auf die Wahrscheinlichkeit des Zutreffens (nicht auf situative, zeitliche oder personelle Kontexte). z.B.: „Die Intelligenz bedingt zu meist die Ängstlichkeit.“
- **Forschungshypothese:** Aus Voruntersuchungen, eigenen Beobachtungen, Überlegungen und wissenschaftlichen Theorien werden Vermutungen abgeleitet. z.B. „Stress am Arbeitsplatz erhöht die Fehlzeiten.“
- **Operationale Hypothesen:** Entsteht aus der Forschungshypothese durch die Operationalisierung der unabhängigen und abhängigen Variablen. Eine präzise Formulierung ist wichtig und soll helfen zu überprüfen, ob die geplante Untersuchung dazu beiträgt, die zuvor aufgestellte Forschungshypothese zu klären. z.B. Beispiel: „Bei 100 zufällig ausgewählten Mitarbeitern eines bestimmten Betriebes besteht zwischen der Punktzahl in einem Fragebogen zur Erfassung von Stress am Arbeitsplatz und der Anzahl der im vergangenen Jahr registrierten Fehltagen ein positiver Zusammenhang.“

- **Statistische Hypothese:**

Resultiert aus operationale Hypothese und wird als Alternativhypothese (H_1) formuliert. Komplementär zur Alternativhypothese wird eine Nullhypothese (H_0) aufgestellt, um das Hypothesenpaar zu vervollständigen. Statistische Hypothesen können gerichtet/ungerichtet, spezifisch/unspezifisch sein.

- **Einseitige / gerichtete Hypothesen:** In die Hypothese geht die Richtung eines Unterschieds / Zusammenhangs ein.

Beispiele gerichteter Hypothesen:

- Zusammenhangshypothese: Die Korrelation ρ zwischen den Merkmalen „Fehlzeiten“ und „Stress am Arbeitsplatz“ ist positiv bzw. $H_1: \rho > 0$; Daraus folgt: Die Korrelation zwischen den untersuchten Merkmalen ist in der Population, aus der die Stichprobe entnommen wurde, Null oder sogar negativ bzw. $H_0: \rho \leq 0$
- Unterschiedshypothese: Das politische Engagement von Studenten der Politikwissenschaften (μ_1) ist größer als das politische Engagement von Studenten des Maschinenbaus (μ_2). Kurzform: $H_1: \mu_1 > \mu_2$ und $H_0: \mu_1 \leq \mu_2$
- Veränderungshypothese: Die Kaufbereitschaft für ein Produkt durch den Anteil aller Käufer ist vor der Werbung (π_1) kleiner als nach der Werbung (π_2). Kurzform: $H_1: \pi_1 < \pi_2$ und $H_0: \pi_1 \geq \pi_2$

- **Zweiseitige / ungerichtete Hypothesen:** In Alternativhypothesen wird ein Unterschied / Zusammenhang behauptet, die Richtung bleibt offen.

Beispiele ungerichteter Hypothesen:

- Es besteht ein Zusammenhang zwischen den Merkmalen „Fehlzeiten“ und „Stress am Arbeitsplatz“. Kurzform: $H_1: \rho \neq 0$ und $H_0: \rho = 0$
- Das politische Engagement unterscheidet sich zwischen Studenten der Politikwissenschaften (μ_1) und des Maschinenbaus (μ_2). Kurzform: $H_1: \mu_1 \neq \mu_2$ und $H_0: \mu_1 = \mu_2$

- **Unspezifische Hypothesen:** Größe des Unterschieds, der Veränderung oder des Zusammenhangs wird nicht angegeben. Bei großen Stichproben sind die Ergebnisse dadurch fast immer signifikant. Statt unspezifischer Alternativhypothesen lieber spezifische Hypothese mit klar definierter Effektgröße nutzen! Alle bisherigen Beispiele sind unspezifische Hypothesen.
- **Spezifische Hypothesen:** Größe des Unterschieds, der Veränderung oder des Zusammenhangs wird angegeben. Geeignet, wenn viel Erfahrung mit der Untersuchungsthematik besteht. Form: $H_1: \mu_1 \geq \mu_2 + a$ (aus Betrag a wird Effektgröße ermittelt); Stichprobenumfang und Mindestgröße des Effekts werden vorab festgelegt (Good-enough-Prinzip²)

² Good-enough-Prinzip (Bortz und Döring (2006, 28f, 729)): Prinzip, das besagt, dass vor der Untersuchung festgelegt wird, welche Untersuchungsergebnisse „gut genug“ sind, um die Alternativhypothese zu bestätigen.

Beispiele spezifische Hypothesen:

- Zusammenhangshypothese: Der Zusammenhang zwischen den Merkmalen „Fehlzeiten“ und „Stress am Arbeitsplatz“ wird durch eine Korrelation beschrieben, die nicht unter $\rho = 0,3$ liegt. Kurzform: $H_1: \rho \geq 0,3$ und $H_0: \rho < 0,3$
- Unterschiedshypothese: Studenten der Politikwissenschaften sind um mindestens 5 Punkte einer entsprechenden Testskala politisch engagierter als Studenten des Maschinenbaus. Kurzform: $H_1: \mu_1 \geq \mu_2 + 5$ und $H_0: \mu_1 < \mu_2 + 5$
- Veränderungshypothese: Die Werbung erhöht den Käuferanteil um mindestens 4 Prozentpunkte. Kurzform: $H_1: \pi_2 \geq \pi_1 + 0,04$ und $H_0: \pi_2 < \pi_1 + 0,04$

Merkmale wissenschaftlicher Hypothesen:

(Bortz & Döring, 2006, S. 490; Hussy et al., 2013, 27, 31)

Um eine Hypothesen testen zu können, muss sie präzise und widerspruchsfrei formuliert (Angabe von Wirkrichtungen und Effektgrößen), begründbar, prinzipiell widerlegbar und die verwendeten Variablen operationalisiert sein. Ob eine Hypothese angenommen oder zurückgewiesen wird, wird anhand der Ergebnisse der Datenanalyse entschieden. Um eine eindeutige Bestätigung oder Widerlegung einer Hypothese zu erhalten, muss die Datenerhebung kontrolliert (angemessene Operationalisierungen und Untersuchungsdesigns) sowie die Daten korrekt inferenzstatistisch ausgewertet werden (adäquate Wahl und Durchführung von Signifikanztests).

Merkmale wissenschaftlicher Theorien:

Hypothesen, welche durch Beobachtungen gestützt und untersucht werden, sind in Theorien eingebettet. Alternativ bilden ein System aus gut bewährten Hypothesen oder anerkannten empirischen Gesetzmäßigkeiten eine Theorie.

Theorien haben die Funktion, Sachverhalte zu beschreiben, zu erklären und vorherzusagen.

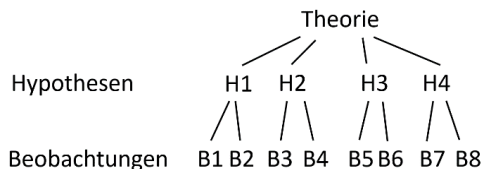


Abbildung 2: Schema Theorie, Hypothesen und Beobachtungen

Kriterien einer „guten“ Theorie:

(Bortz & Döring, 2006, S. 15; Hussy & Jain, 2002, 278f)

- Logisch konsistent (Widerspruchsfreiheit)
- Informativ (potenziell falsifizierbar)
- Sparsam (möglichst viele Befunde durch wenig Annahmen erklären)
- Bewährt (viele verschiedene, strenge Tests bestanden)


Wissenschaftliche Begriffsbildung – Formale Struktur einer Theorie

Für die Erstellung von wissenschaftlichen Theorien ist es wichtig, eine formale Struktur einzuhalten. Eine Theorie enthält Begriffe (Explikations- und Definitionsmethoden), die in Aussagen eingebettet sind, die wiederum nach Kriterien der Logik miteinander verknüpft sind.

Wissenschaftliche Kommunikation soll eindeutig sein und eine temporär gültige Kernintension beinhalten. Das heißt, Mehrdeutigkeit & Vagheit müssen reduziert werden. (Achtung bei Übersetzungen!)

Kriterien zur Bewertung von Theorien (Dennis & Kintsch, 2007, S. 143–159)

- **Deskriptive Angemessenheit** (Descriptive adequacy): Stimmt die Theorie mit vorliegenden verhaltensbezogenen, physiologischen, neurophysiologischen und anderen empirischen Daten überein?
- **Präzision und Interpretierbarkeit** (Precision and interpretability): Ist die Theorie ausreichend präzise beschrieben, sodass sie leicht und eindeutig verstanden und interpretiert werden kann?
- **Kohärenz und Konsistenz** (Coherence and consistency): Enthält die Theorie logische Fehlschlüsse? Passen die unterschiedlichen Komponenten einer Theorie zu einem kohärenten Ganzen zusammen? Ist die Theorie mit anderen Theorien anderer Bereiche übereinstimmend (z.B. physikalische Gesetzmäßigkeiten)?
- **Vorhersage und Falsifikation** (Prediction and falsifiability): Ist die Theorie so formuliert, dass empirische Tests zu einer Widerlegung der Theorie führen können? (Falsifikation entspricht wichtigster Anforderung des kritischen Rationalismus)
- **Retrognose und Erklärung** (Postdiction and explanation): Erklärt die Theorie bereits vorliegende empirische Befunde?
- **Einfachheit** (Parsimony): Ist die Theorie so einfach wie möglich?
- **Originalität** (Originality): Ist die Theorie neu oder im Grunde eine Reformulierung bestehender Theorien?
- **Breite** (Breadth): Betrifft die Theorie einen breiten Anwendungsbereich oder lediglich ein Phänomen?
- **Verwendbarkeit** (Usability): Bestehen Auswirkungen durch die Anwendung der Theorie?
- **Rationalität** (Rationality): Macht die Theorie Annahmen über die Architektur des psychischen Systems, die unter Berücksichtigung evolutionärer Bedingungen Sinn machen?

 Anhang D: Generierung von Theorien und Hypothesen als Inspiration

2.1.4 Operationalisierung

Aufgabe des Kapitels: Was erwartet mich hier?

Wir lernen die Typisierung von Variablen und wie diese in einer Operationalisierung eingebunden werden, kennen. Eine Operationalisierung macht Variablen und damit Hypothesen erst messbar.

Arbeitstitel und Untersuchungsart legen in einer hypothesenprüfenden Untersuchung unabhängige und abhängige Variablen fest, die für die anstehende Untersuchung definiert werden müssen.

Variable: (Bortz & Döring, 2006, S. 734)

Eine Variable ist ein Symbol für eine Menge von Merkmalsausprägungen.

Variablen können nach folgenden Kriterien typisiert werden:

- Stellenwert in der Untersuchung: unabhängig/abhängig, Moderator-/Kontroll-/Störvariable
- Merkmalsausprägung: diskret³/stetig⁴/quasi-stetig⁵

³ Diskrete Variable: Merkmale der Variable haben endlich viele Ausprägungen; z.B. Kinderzahl

⁴ Stetige Variable: Merkmale der Variable haben unendlich viele Ausprägungen; z.B. Körpergröße

⁵ Quasi-stetige Variable: nur diskret gemessene Variable, die jedoch immer feiner Abstufungen, wie stetige Merkmale, hat; z.B. Alter in Jahren

- Skalenniveau: nominal-/ordinal-/intervall-/ratioskalierte Variable, dichotome⁶/ polytome⁷ nominalskalierte Variable (s. Skalenniveaus berücksichtigen)
- Empirischer Zugänglichkeit: manifeste/latente Variable

Notwendige Definitionen für eine Operationalisierung: (Bortz & Döring, 2006, Abs. 2.3.6)

- **Analytische Definition:** Begriffsklärung durch Analyse der Semantik und des Gebrauchs; diese ist empirisch überprüfbar, um den Untersuchungsgegenstand transparent zu machen
- **Operationale Definitionen:** Standardisierung eines Begriffs durch die Angabe von Operationen, die zur Erfassung des durch den Begriff bezeichneten Sachverhalt notwendig sind oder durch Angabe von messbaren Ereignissen (Indikatoren); Brauchbarkeit hängt von der Bewährung in der Praxis ab

Bei unabhängigen Variablen ist es häufig ausreichend, nur eine Ausprägung experimentell herzustellen. Bei abhängigen Variablen soll eine Messung in möglichst differenzierten Abstufungen stattfinden. Begriffsoperationalisierungen sind meist **frei gestaltbar** (Sixtl, 1993, S. 24). Bestehen widersprüchliche Operationalisierungen, ist das ein Hinweis für verschiedene Begriffe und eine nicht präzise Bedeutungsanalyse.

Operationalisierungsvarianten für abhängige Variablen:

(Bortz & Döring, 2006, S. 64; Conrad & Maul, 1981, S. 151)

Je nach Messgegenstand muss mindestens eine Operationalisierungsvariante zur Messung des Einflusses der unabhängigen Variable oder eines Treatments auf die abhängigen Variable ausgewählt werden. Beispiele:

- Häufigkeit Bei der Untersuchung zur Rechtschreibschwäche von Kindern:
z.B. Anzahl Fehler in einem Diktat zählen
- Zeit z.B. Reaktionslatenz nach Auftreten eines unerwarteten Verkehrshindernis
- Dauer z.B. Lösungszeit einer Mathematikaufgabe
- Stärke z.B. Stärke der Muskelanspannung als Indikator für Aggressivität
- Auswahl z.B. Präferenzurteile

Die Art der Operationalisierung entscheidet über das **Skalierungsniveau** der abhängigen Variable und damit auch über die statistische Auswertung des Merkmals. Dieses bedingt den Signifikanztest zur Hypothesenprüfung. Dieser sollte mindestens Intervallskalenniveau anstreben. (s. Tabelle 11)

⁶ Dichotome Variable: es bestehen nur 2 Ausprägungen des Merkmals der Variable; z.B. Haustierbesitz: ja/nein

⁷ Polytome Variable: es bestehen mehr als 2 Ausprägungen; z.B. Religionsangehörigkeit: ev./kath./musl./...

Hinweise:

- Es sind immer verschiedene Möglichkeiten der Operationalisierung möglich. Ursachen:
 - verschiedene Datenquellen
 - verschiedene Messinstrumente (unbestimmte und offene Operationalisierung eines Konstrukts)
 - verschiedene Möglichkeiten der Hypothesenuntersuchung
- **Standardisierungen** sind zu bevorzugen (Brähler & Brickenkamp, 2002) (Vorteile: Vergleiche möglich, Gütekriterien bereits nachgewiesen, Aufwand einer Konstruktion neuer diagnostischer Instrumente ersparen), falls nicht möglich: Adaptierung des Vorhandenen
- Unumgänglich: Bestimmung psychometrischer **Gütekriterien** bei Entwicklung von **Forschungsinstrumenten**, idealerweise in Vorstudien (DGPs und BDP, 2004, S. 59)

2.1.5 Forschungsansätze in den empirischen Wissenschaften im Überblick

(Bortz & Döring, 2006, 49ff, Kap 6-9; Flick, 2016; Hussy et al., 2013, 9f, 26f)

Aufgabe des Kapitels: Was erwartet mich hier?

Welche empirischen Forschungsansätze passend zu den operationalisierten Hypothesen für die anstehende Untersuchung in Frage kommen, wird hier genauer betrachtet. Vertieft werden in Form eines Exkurses die experimentelle und korrelative Forschung, da hier großes Potential für die nutzerzentrierte Produktentwicklung besteht.

Grundlage für die Entscheidung der Untersuchungsart(en) sind: die Forschungsfrage, das Vorwissen, das Erkenntnisinteresse einer Untersuchung (Vor-/Hauptstudie) sowie der Stand der Forschung.

Folgende Untersuchungsarten werden unterschieden:

- **Explorative Untersuchung:** wird bei wissenschaftlichem Neuland zur Grundlagenforschung eingesetzt, um sich zu orientieren und neue Hypothesen zu finden.
- **Deskriptive Untersuchung:** dient der Beschreibung einer Population. Sie ist von der Auswahl der Stichprobe abhängig.
- **Explanative Untersuchung:** dienen der Ableitung und Überprüfung gut begründeter Hypothesen.

Die Wahl der Untersuchungsart hilft bei der Auswahl des Forschungsansatzes und der konkreten Forschungsmethode. Der Forschungsansatz bestimmt die übergreifende Vorgehensweise und hat Auswirkungen auf die Stichprobenziehung sowie die Methoden der Datenerhebung und -auswertung. Eine geeignete Kombination der Untersuchungsart(en) und Forschungsansätzen ist auszuwählen (s. Abbildung 5). Zur Untersuchungsart muss die Forschungsmethode passen. Die Ressourcen, die Verfügbarkeit, die Stichprobengröße und das Untersuchungsinteresse geben die Forschungsmethode vor (s. Tabelle 1). Folgende methodische Ansätze sind zu unterscheiden:

- **Quantitativen Methoden,**
- **Qualitativen Methoden** und
- **Mixed Methods** (Mix aus quanti- und qualitativen Methoden).

Methodenklasse	Quantitativer Ansatz	Qualitativer Ansatz
Forschungsansatz (Forschungsdesign)	Experimentelle Ansätze <ul style="list-style-type: none"> • Experimentelle Forschung <ul style="list-style-type: none"> - (Labor-)Experiment - Feldexperiment - Quasiexperiment - Einzelfallforschung Nichtexperimentelle Ansätze <ul style="list-style-type: none"> • Korrelationsforschung <ul style="list-style-type: none"> - Prognosestudie - Metaanalyse 	<ul style="list-style-type: none"> • Deskriptive Feldforschung • Handlungsforschung • Biografieforschung • Gegenstandsbezogene Theoriebildung • Fallstudie
(Daten-) Erhebungsmethoden	Beobachten Zählen Urteilen Testen	Interview Struktur-lege-Verfahren Gruppendiskussion Teilnehmendes Beobachten
(Daten-) Analysemethoden	Beschreibende Methoden (deskriptiv) Schlussfolgernde Methoden (induktiv) Multivariate Methoden Modelltests	Inhaltsanalyse Hermeneutik Semiotik Diskursanalyse

Tabelle 1: Auswahl: Methodische Ansätze empirischer Wissenschaften im Überblick (in Anlehnung an (Hussy et al., 2013, 26f))

In Tabelle 2 werden quanti- und qualitative Untersuchungsansätze gegenübergestellt.

Quantitative Forschung findet in einem eher linearen Prozess, qualitative Forschung in einem zirkulären Prozess statt (s. Abbildung 3). Zusammengefasst ist festzustellen, dass quantitative Forschung innerhalb kurzer Zeit einen guten Überblick zu sehr vieler Untersuchungsobjekte und Verallgemeinerungen auf die Grundgesamtheit⁸ zulässt. Hierbei wird der Kontext nicht berücksichtigt, eine vergleichsweise oberflächliche und standardisierte Betrachtung wird durchgeführt. Es kann passieren, dass Untersuchungen die relevanten Aspekte und Zusammenhänge nicht erfassen und diese unentdeckt bleiben. Qualitativer Forschung wird durch eine intensive und offene Beschäftigung mit wenigen Untersuchungsobjekten ein größerer Spielraum zur Verfügung gestellt, um diese Aspekte zu identifizieren und zu messen. Wenige Fälle reichen für tiefgehende Untersuchungen aus. Dabei ist der höherer (Zeit)Aufwand und die geringe Übertragbarkeit auf die Grundgesamtheit zu bemängeln.

Beste interne und externe Validität bei hypothesenprüfender Forschung:

Experimentelle Felduntersuchung (Shadish et al., 2002)

In der **Wissenschaft** wird die **Hypothesenprüfung präferiert**; in der Praxis findet sich oft eine **anwendungsbezogenen Fragen orientierte Deskription**

Datenerhebungs- und damit verknüpfte Datenanalysemethoden sind individuell zu bewerten:

Es gibt keine ideale Erfassungsmethode. Die Empfehlung lautet: **multi-methodal zu erfassen** (Eid & Diener, 2006)

⁸ Grundgesamtheit (GG): Als GG wird die Menge aller potenziellen Untersuchungsobjekte für eine gegebene Fragestellung verstanden; Synonym: Population (Hussy, Schreier und Echterhoff, 2013, S. 118).)

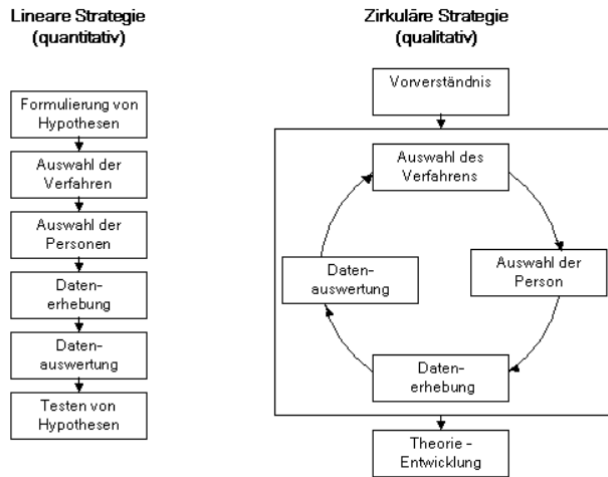


Abbildung 3: Schematische Darstellung der Forschungsstrategien (Witt, 2001)

Quantitative und Qualitative Untersuchungsansätze im Vergleich	
Quantitativ	Qualitativ
Vorbild ist der naturwissenschaftliche Ansatz	Ein typisch geisteswissenschaftlicher Ansatz liegt vor
Die Untersuchungen finden unter standardisierten (auch Labor-) Bedingungen statt	Die Untersuchungen werden möglichst unter unverfälschten Feldbedingungen durchgeführt
Ein deduktives Vorgehen (Allgemeines → Besonderes) wird eingesetzt, dieses gilt als wahrheitsbewahrend	Ein induktives Vorgehen (Besonderes → Allgemeines) findet statt, es gilt als wahrheitserweiternd
Die Überprüfung von aufgestellten Hypothesen ist das Ziel	Es geht um Entdeckungen, das Prinzip der Offenheit gilt
Die Sachverhalte sollen erklärt werden, dazu müssen die äußeren Ursachen (U) herangezogen werden	Die Sachverhalte sollen verstanden werden, dafür sind die inneren Gründe zu betrachten
Zur Informationsgewinnung werden Stichprobenuntersuchungen genutzt, eine Irrtumswahrscheinlichkeit kann angegeben werden	Einzelfälle werden analysiert, ein bewusstes theoretisches Sampling findet statt
Die Arbeit geschieht mit großen Fallzahlen	Nur relativ wenige Fälle werden untersucht
Harte, das heißt voll standardisierte Methoden kommen zum Einsatz	Weiche, kaum standardisierte Methoden werden benutzt
Das Prinzip des Messens und der Operationalisierung wird praktiziert	Es geht um das Beschreiben von Fällen und um die Sensibilisierung
Erkenntnisziele sind Aussagen über Aggregate und statistische Zusammenhänge	Die Aussagen erfolgen fallbezogen, diese werden rekonstruiert
Verallgemeinerungen sind angestrebt (Häufigkeiten)	Typenbildungen (innere Logik) werden vorgenommen
Gütekriterien sind Objektivität, Reliabilität und Validität	Gütekriterium ist die Gegenstandsbezogenheit der Methoden

Tabelle 2: Quantitative und Qualitative Untersuchungsansätze im Vergleich (in Anlehnung an (Häder, 2019, S. 67))

Exkurs: Vertiefung ausgewählter quantitativer Forschungsansätze: Experimentelle & korrelative Forschung

Folgend werden die zwei wichtigsten Forschungsansätze empirischer Wissenschaften näher erläutert: die experimentelle sowie die korrelative Forschung (s. Tabelle 3).

Die Korrelationsforschung wird vor allem eingesetzt, wenn Experimente aus prinzipiellen, ökonomischen und ethischen Gründen, keinen Sinn ergeben.

Experimentelle Forschung	Korrelative Forschung
Variationen der abhängigen Variablen durch Manipulation der unabhängigen Variablen selbst herstellen	Zusammenhänge bereits existent, Variationen zwischen Merkmalen

Tabelle 3: Gegenüberstellung experimentelle Forschung vs. Korrelationsforschung

Experimentelle Forschung (Campbell & Stanley, 1967; Huber, 2019; Hussy et al., 2013, 141f; Sedlmeier & Renkewitz, 2018) (Hussy & Jain, 2002; Sarris & Reiß, 2005)

Das Experiment dient als Standardmethode zur Untersuchung kausaler Zusammenhänge und wird als Königsweg zur Überprüfung von Ursache-Wirkungs-Zusammenhängen bezeichnet. Es wird definiert als eine objektive Beobachtung von Phänomenen, die in streng kontrollierten Situationen hervorgerufen werden. Dabei werden 1 / mehrere (unabhängige) Variablen systematisch variiert und die Effekte beobachtet bzw. gemessen, alle anderen (Störvariablen) werden konstant gehalten bzw. ausgeschaltet. (Huber, 2019)(Huber, 2019)(Huber, 2019)(Huber, 2019)

Hierfür sind Techniken zur Kontrolle von Störvariablen (Drittvariablen) notwendig

Eine grobe Übersicht über die experimentelle Forschung wurde in Tabelle 3 gegeben, eine detaillierte Ansicht bietet die Systematisierung experimenteller Designs in Abbildung 4.

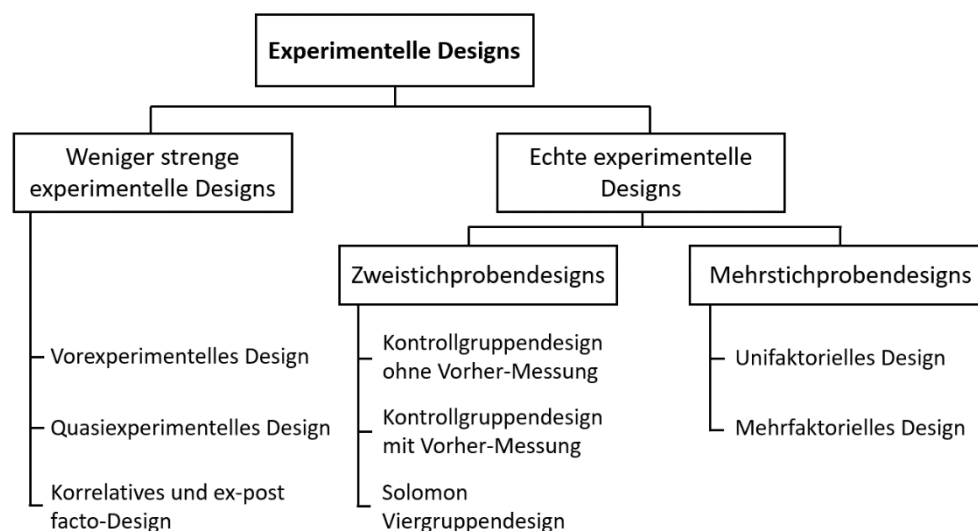


Abbildung 4: Systematisierung experimenteller Designs (in Anlehnung an (Garnefeld, 2009, S. 111))

Die beste Untersuchung von Kausalzusammenhängen ist durch die Nutzung echter experimenteller Designs gewährleistet. Hierbei unterscheiden sich die Zweistichprobendesigns zu den Mehrstichprobendesigns in der Anzahl der experimentell manipulierbarer Stufen der unabhängigen Variable (Zweistichprobendesigns: nur zwei) und der Anzahl berücksichtigter unabhängiger Variablen (Mehrstichprobendesigns mehr als eine). (Steinhoff, 2014, 129ff)

Die häufigsten Studiendesigns experimenteller und quasiexperimenteller Untersuchungen werden nachfolgend aufgeführt. Legende: T = Treatment; M = Messung (Cook & Campbell, 1979; Shadish, Cook & Campbell, 2002)

- Quasiexperimentelle Untersuchung:
 - Eingruppen-Posttest-Design (One-Shot Case Study): nur Experimentalgruppe, keine Vortests; kausal nicht interpretierbar: $T \rightarrow M$
 - Ein-Gruppen-Pretest-Posttest-Design (One-Group Pretest-Posttest Design): Messungen vor und nach dem Treatment, um zwischenzeitliche Veränderungen zu berücksichtigen: $M_1 \rightarrow T \rightarrow M_2$
- Experimentelle Untersuchung:
 - Ex-post-facto-Studie: Vergleich der „behandelten“ mit einer nichtbehandelten, nichtäquivalenten Kontrollgruppe: $\frac{T \rightarrow M_1}{M_2}$
 - Kontrollgruppenplan mit Pre- und Posttest: wiederholte Messungen bei zwei Gruppen: $M_{11} \rightarrow T \rightarrow M_{12}$
 $M_{21} \rightarrow M_{22}$
 - Variationen Designtypen (s. Sedlmeier & Renkewitz, 2018)
 - Between – subjects – design: Ein Proband wird nur einer Stufe der unabhängigen Variable zugeordnet
 - Within – subjects – design: Ein Proband durchläuft alle experimentellen Bedingungen
- Zu untersuchende Hypothesen: Veränderungs⁹- oder Unterschiedshypothesen

Zentrale Anforderungen / Voraussetzungen für ein Experiment: (Bortz & Döring, 2006, S. 58)

- Rahmenbedingungen identisch in Experimentalgruppe und Kontrollgruppe (Ceteris paribus-Kriterium)
- Versuchspersonen aus derselben Grundgesamtheit (GG) und exakt vergleichbar → Randomisierte Versuchsgruppen
- Gruppen unterscheiden sich nur durch das Treatment¹⁰
- Strikte Kontrolle von Störgrößen
- Kriterien
 - Planmäßigkeit der Untersuchungsdurchführung (Willkürlichkeit)
 - Wiederholbarkeit der Untersuchung
 - Variierbarkeit der Untersuchungsbedingungen

⁹ Veränderungshypothese: Abhängige Variable min. 1x vor und 1x danach (Prä- und Posttest) messen, z.B. Fragebogen und Multiple Choice

¹⁰ Treatment: entspricht Stufen der unabhängigen Variable; entweder Variation bestimmter situativer Bedingungen oder Interventionen (Eingriff/Veränderung/Behandlung/Aufgabe) → Prüfung von Veränderungshypothesen, z.B. ob Lernprogramme schulische Leistungen verbessert (Bortz und Döring, 2006, S. 730))

Bewertungskriterien zur Auswahl experimenteller Forschungsmethoden: (Campbell & Stanley, 1967)

- Innere Gültigkeit (interne Validität mit statistischer Validität) & äußere Gültigkeit (externe Validität mit Konstruktvalidität) (s. Abbildung 5, Tabelle 6)
- Felduntersuchungen vs. Laboruntersuchungen (s. Tabelle 9)

	Experimentell	Quasiexperimentell
Feld	Interne Validität +	Interne Validität –
	Externe Validität +	Externe Validität +
Labor	Interne Validität +	Interne Validität –
	Externe Validität –	Externe Validität –

Abbildung 5: Kombinationen der Untersuchungsvarianten (Bortz & Döring, 2006, S. 58)

Achtung: Validität nicht nur nach diesen Kriterien zu bewerten; Feld und Labor stellen Extreme eines Kontinuums von Untersuchungen mit unterschiedlicher Kontrolle untersuchungsbedingter Störvariablen dar.

Beispiele experimenteller Forschung: (Bortz & Döring, 2006, 57f)

- Quasiexperimenteller Felduntersuchung: Zusammenlegung von Schulen mit weißen und People of Color-Schülern auf das akademische Selbstbild der Schüler
- Experimentelle Felduntersuchung: Veränderungen politischer Einstellungen durch das Lesen zweier überregionaler Tageszeitungen
- Quasiexperimentelle Laboruntersuchung: Unterschiede in der Fingerfertigkeit männlicher und weiblicher Untersuchungsteilnehmer
- Experimentelle Laboruntersuchung = klassisches Experiment (Wundt, 1898): Einfluss eines Verbalisierungstrainings auf die Denkleistung von Kindern

Einzelfallforschung als experimentelle Untersuchungsform: (Hussy et al., 2013, 146f)

In der quantitativen Einzelfallforschung (auch N=1-Forschung genannt) wird eine einzelne Untersuchungseinheit (z.B. Personen, Gruppen, Kulturen) bezüglich ein oder mehreren abhängigen Variablen unter möglichst kontrollierten Bedingungen wiederholt betrachtet (unabhängig von Erhebungstechnik), um die Wirkung einer unabhängigen Variable messen und bewerten zu können. Dies wird häufig bei explorativen Untersuchungen eingesetzt.

Einzelfalldesigns konzentrieren sich auf Entwicklungs- und Veränderungsprozesse, die durch wiederholte Untersuchungen an einer oder mehreren Personen durchgeführt werden. Besonderheit: Im Unterschied zu korrelativen Designs können Interventionen bzw. Treatments erfolgen, wie typisch für within-subject-designs ist.

Achtung: In Einzelfallstudien wird der Durchschnitt der Gruppen betrachtet, nicht die einzelnen Probanden. Es kann zu hypotheseninkonformen Effekten kommen, obwohl die Hypothese bestätigt wird. Schlussfolgerungen hierzu sind:

- 1) Die Ergebnisse gruppenbezogener Studien dürfen nicht ohne weitere Prüfung auf die einzelnen Probanden dieser Studien übertragen werden, die sog. „Mittelwertsfalle“. (Asendorpf, 2000)
- 2) Wenn das Erleben und Verhalten (und damit verbundene Veränderungen) einzelner Personen interessiert, sollen einzelne Personen untersucht werden.

Korrelative¹¹ Forschung (Bortz & Döring, 2006, 512f; Häder, 2019, 462ff; Hussy et al., 2013, 174f)
 = Untersuchungsformen (statistische Analyseverfahren) von Fragen/Hypothesen zu Zusammenhängen zwischen Variablen und Konstrukten; Korrelative Zusammenhänge sind nicht deterministisch, lediglich stochastisch (zufallsabhängig) → nur mit gewisser Wahrscheinlichkeit zutreffend

Datenerhebung: alle Variablen (Prädiktor(en) und Kriterium/Kriterien; bestenfalls: intervallskaliert) werden an einer repräsentativen Stichprobe erhoben, Erhebungsmethode individuell zu wählen

Datenauswertung (s. unten): Bravais-Pearson Korrelationskoeffizienten, regressions- und faktorenanalytische Analyseverfahren (ab min. 3 Variablen)

Vorteile: Zusammenhänge vieler Variablen sind untersuchbar (im Experiment nur wenige)

Nachteile: Korrelative Zusammenhänge dürfen nicht kausal interpretiert werden (dann sind inhaltliche Überlegungen / Längsschnittstudien / Cross-laged Panel-designs möglich, um die Anzahl kausaler Erklärungsalternativen einzuschränken bzw. zu falsifizieren; auch abhängig vom zugrundeliegenden Kausalitätskonzept)

Unterscheidung:

- Lineare / Nicht-lineare Zusammenhänge
- Querschnitts- / Längsschnittuntersuchungen

An dieser Stelle werden die für die Psychologie typischen Begriffe des Prädiktors und des Kriteriums eingeführt, die synonym zur unabhängigen und abhängigen Variablen verwendet werden. Die unterschiedlichen Begriffe sollen der Verwechslungsgefahr zwischen experimenteller und korrelativer Forschung vorbeugen. (Bortz & Döring, 2006, 512, 732, 737)

Prädiktor x = Vorhersage-Variable in Korrelations- und Regressionsanalysen (unabhängige/exogene Variable)

Kriterium y = zu vorhersagende Variable in Korrelations- und Regressionsanalysen (abhängige/endo-gene Variable)

Untersuchung bivariater Zusammenhangshypothesen:

Werden zwei Variablen (=bivariat; 1 Prädiktor, 1 Kriterium) in ihrem Zusammenhang untersucht, wird empfohlen, den Bravais-Pearson Korrelationskoeffizienten r (s. *Pearson r*) heranzuziehen. Eine bivariate Regression ist ebenfalls möglich, jedoch sind der standardisierte Betakoeffizient der Regression mit dem Korrelationskoeffizienten identisch. (Hussy et al., 2013, S. 177)

Untersucht werden gerichtete / ungerichtete Zusammenhangshypothesen, betreffend vermuteter Assoziationen zwischen 2 Merkmalen (Richtung der Assoziation: neg. / pos. nur bei gerichteten Zusammenhangshypothesen möglich) sowie die Intensität [-1 bis 1]

Unterschied Korrelation & Regression:

- Korrelation ist symmetrisch → lediglich Zusammenhang wird untersucht, keine kausale Interpretation möglich (Auswahl des Korrelationskoeffizienten nach Skalenniveau, *Skalenniveaus berücksichtigen*)

¹¹ Korrelation: Allgemeine Bezeichnung zur Beschreibung von Zusammenhängen zwischen Variablen. (Bortz und Döring, 2006, S. 732))

- Regression ist asymmetrisch → eine Wirkungsrichtung muss festgelegt werden (UV = Ursache, AV = Wirkung; Ursache geht Wirkung zeitlich voraus und bringt diese hervor); Interpretation in Einheiten möglich (z.B. „Mit jeder Einheit X, steigt Y um 7,2 Einheiten.“); Modell ist leicht erweiterbar

Untersuchung multivariater Zusammenhangshypothesen

Werden mehrere Variablen zur korrelativen Untersuchung von Zusammenhängen herangezogen, spricht man von multivariaten Zusammenhangshypothesen. „Die Überprüfung einer multivariaten Zusammenhangshypothese durch mehrere bivariate Korrelationen führt meistens zu Fehlinterpretationen“ (Bortz & Döring, 2006, S. 516); Simpson Paradox).

Die wichtigsten Untersuchungsdesigns multivariater korrelativer Zusammenhänge werden im Folgenden näher betrachtet. Sie unterscheiden sich in der Anzahl der Kriterien. Die Erfassung der Prädiktoren sollte der des Kriteriums bzw. der Kriterien vorangehen. (Bortz & Döring, 2006, 511f; Hussy et al., 2013, 174f)

1. **Drittvariablenkontrolle: 1 Prädiktor, 1 Kriteriumsvariable, dritte Variable: Z**

Wird eine Drittvariable vermutet oder soll dies ausgeschlossen werden bei der Untersuchung eines Zusammenhangs zwischen Prädiktor und Kriterium, ist eine Drittvariablenkontrolle sinnvoll. [Beim Experiment: Ausschaltung von Drittvariablen durch zufällige bzw. kontrollierte Zusammensetzung von Versuchs- und Kontrollgruppen

Bei Nichtexperimentellen Untersuchungen: Im Nachhinein durch multivariate statistische Kontrolle]

- Teilgruppenanalyse
- Partielle Korrelation
- Trivariate bzw. multivariate Regression

2. **Multiple Zusammenhänge: Mehrere Prädiktoren, 1 Kriteriumsvariable**

(idealerweise linear unabhängige Prädiktoren)

- **Multiple Korrelation**

Ergebnis: Zusammenhang zwischen mehreren Prädiktoren und Kriterium; (er entspricht der Produkt-Moment-Korrelation zwischen dem vorhergesagten und dem tatsächlichen Kriteriumswert bei der Regression mit mehreren Prädiktorvariablen) oder

- **Multiple Regressionsrechnung**

Ergebnis: Gleichung zur Vorhersage der Kriteriumswerte und dem multiplen Korrelationskoeffizient¹² (s. Unterschied Korrelation & Regression:

3. **Mehrere Prädiktoren, mehrere Kriterien (kanonische¹³ Zusammenhänge)**

(Bei nominalskalierten Variablen s (Bortz & Döring, 2006, 512ff))

- Kanonische Korrelation¹⁴ (hat viele Spezialfälle, (Bortz & Döring, 2006, Kap. 19))
 - Multiple Korrelation
 - Produkt-Moment-Korrelation
 - Diskriminanzanalyse bzw. multivariate (Ko-)Varianzanalyse
 - Univariate Varianzanalyse
 - T-Test für unabhängige Stichproben

¹² Der multiple Korrelationskoeffizient quantifiziert den Zusammenhang zwischen den vorhergesagten und den tatsächlichen Kriteriumswerten.

¹³ Kanonische Zusammenhänge werden statistisch mit Hilfe kanonischer Korrelation ermittelt

¹⁴ „Die kanonische Korrelation erfasst den Zusammenhang zwischen mehreren Prädiktorvariablen und mehreren Kriteriumsvariablen.“ (Bortz und Döring, 2006, S. 627))

- $k \times l - \chi^2$ -Test
- $k \times 2 - \chi^2$ -Test
- Vierfelder- χ^2 -Test
- Faktoranalyse¹⁵ (Bortz & Döring, 2006, 516f; Häder, 2019, S. 462)
 - Explorative / konfirmatorische Faktoranalyse zur Prüfung der Hypothese: wechselseitige Zusammenhänge vieler beobachteter bzw. gemessener Variablen lassen sich durch wenige, i.d.R. voneinander unabhängigen (orthogonalen) Faktoren erklären → Viele miteinander korrelierende Variablen werden auf wenige, voneinander unabhängige Faktoren reduziert (mehrere Items werden zu Gesamt-Score kombiniert) [Faktoren = latente Variablen; inhaltlich sinnvolle Faktoren = hypothetisches Konstrukt]
 - Voraussetzung: normalverteilte, intervallskalierte Variablen; ausreichend große Anzahl an Variablen, die jeweils in min. 10 Fällen erhoben wurden
 - Vorteile:
 - Reduktion der Datenmenge
 - Messungen aggregieren; nach dem Aggregationsprinzip liefert die Summe mehrerer (miteinander korrelierender) Messungen eine stabilere und repräsentative Schätzung eines Merkmals als eine einzelne Messung (höhere Messgenauigkeit) (Rushton, Brainerd & Pressley, 1983)
 - Leichtere Weiterverarbeitung & Kommunikation
- Clusteranalyse
Nutzung: zur Aufdeckung von Strukturen durch die Suche an Ähnlichkeiten in Variablen und Untersuchungseinheiten

Eingrenzung und Falsifikation von Kausalmodellen in der Korrelationsforschung

(Döring & Bortz, 2016, S. 518; Duller, 2013, 132f; Hussy et al., 2013, 176)

Das Vorliegen eines Zusammenhangs zwischen zwei Variablen bedeutet nicht zwingend, dass eine Kausalbeziehung vorliegt. (s. *Mögliche Kausalmodelle*)

→ **Korrelative Zusammenhänge dürfen nicht kausal interpretiert werden!**

Es besteht die Möglichkeit durch bestimmte korrelative Designs (Längsschnittstudien, insb. cross-lagged panel designs) und / oder inhaltliche Überlegungen die Anzahl kausaler Erklärungsalternativen einzuschränken bzw. zu falsifizieren (Bortz & Döring, 2006; Hussy et al., 2013). Die Frage nach der kausalen Interpretierbarkeit korrelativer Zusammenhänge hängt auch vom zugrundeliegenden Kausalitätskonzept ab (Hodapp 1984).

Schlüssige Kausalhypothesen lassen sich nur mit experimentellen Versuchsplänen unter Kontrolle von Störvariablen und Vermeidung von Konfundierungen prüfen. (Hussy et al., 2013, S. 176)

Wenn zwischen zwei Variablen keine positive oder negative Korrelation festgestellt werden kann und keine Messfehler vorliegen, dann ist die Hypothese eines Kausalzusammenhangs falsifiziert.

Liegt eine Korrelation zwischen zwei Variablen vor und kein Messfehler, können folgende, s. Abbildung 6, Kausalmodelle für die Interpretation der Zusammenhänge herangezogen werden.

Mögliche Kausalmodelle (vgl. Abbildung 12)

Am Beispiel von: x = übermäßiger Alkoholkonsum, y = Lebenserwartung

¹⁵ Faktoren = latente Variablen; inhaltlich sinnvolle Faktoren = hypothetisches Konstrukt

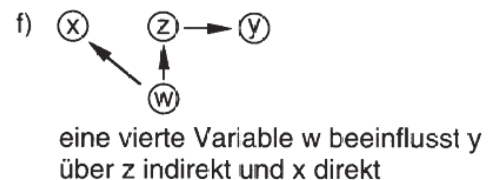
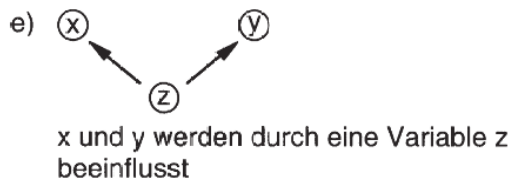
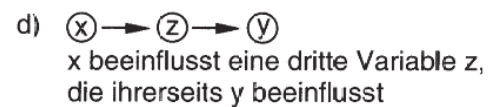
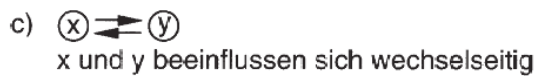
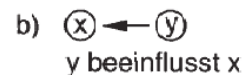
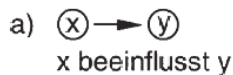


Abbildung 6: Kausalmodelle und ihre Stützung durch eine Korrelation (Quelle: (Bortz & Döring, 2006, S. 518))

Modell a): Übermäßiger Alkoholkonsum reduziert die Lebenserwartung. → einseitige Beeinflussung

Modell b): Eine geringe Lebenserwartung verursacht erhöhten Alkoholkonsum → einseitige Beeinflussung

Modell c): Übermäßiger Alkoholkonsum und eine geringe Lebenserwartung beeinflussen sich wechselseitig

Modell d): Intervention: Durch erhöhten Alkoholkonsum wird man arbeitsunfähig und damit arm. Armut (z) bedingt schlechte Ernährung, die das Leben verkürzt

Modell e): Scheinkausalität: Eine angeborene „Ich-Schwäche“ (z) erhöht die Anfälligkeit für lebensbedrohende Krankheiten und für Alkohol

Modell f): Stress (w) verursacht Trinken und Rauchen (z). Lebensverkürzend wirkt aber nur das Rauchen

Ausführliche Darstellung korrelativer Forschungsdesigns: (Bortz & Döring, 2006, 519f)

- **Querschnittsdesigns:** Die Erhebung von Merkmalsausprägungen erfolgt lediglich zu einem bestimmten Messzeitpunkt.
Nachteile: Zeitliche Generalisierbarkeit problematisch → externe Validität, die Interpretation der Gruppenunterschiede kann mit Kohorteneffekten¹⁶ konfundiert sein → geringe interne Validität (Bortz & Döring, 2006, S. 738)
- **Längsschnittsdesigns:** Die Untersuchungseinheiten werden wiederholt (verschiedene Messzeitpunkte) hinsichtlich derselben Variablen untersucht.
Unterscheidung: Trenduntersuchungen (unterschiedliche Stichproben, nacheinander aus derselben Population gezogen, werden untersucht) und Paneluntersuchungen (dieselbe Stichprobe wird über längere Zeit hinweg untersucht);
Für Variablen geeignet, die sich im Zeitverlauf verändern können; z.B. Einstellung eines Lehrers gegenüber seinen Schülern, Stabilität von Eigenschaften;
Nachteile: relativ großer untersuchungstechnischer Aufwand, lange Wartezeit bis zum Untersuchungsergebnis (Bortz & Döring, 2006, 519, 733)

¹⁶ Unter einer Kohorte wird eine Population verstanden, die im selben Zeitraum geboren wurden. Der Kohorten-, bzw. Generationeneffekt, beschreibt die erschwerte eindeutige Interpretation bezüglich einer Stichprobe mit unterschiedlichen Generationen. (Stangl, 2020))

- **Cross-lagged-panel designs** (Campbell 1963): „Verfeinerter“ Versuchsplan, um unterschiedliche Kausalmodelle in korrelativen Längsschnittstudien auszuschließen (s. Bsp. Bortz & Döring, 2006, S. 519f)

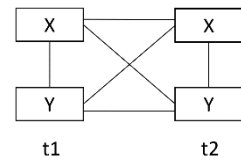


Abbildung 7: Schema Cross-lagged Panel-design

- **Pfadanalyse:** erweitert multiple Korrelations- und Regressionsrechnung, indem Zusammenhänge zwischen mehreren unabhängigen und abhängigen Variablen gleichzeitig berücksichtigt werden; diese werden in einem Pfadmodell grafisch dargestellt und es wird einer empirisch-statistischen Prüfung unterzogen (s. Bsp. Bortz & Döring, 2006, 520f)

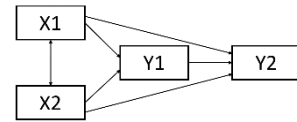


Abbildung 8: Schema Pfadanalyse

- Strukturgleichungsmodelle sind Pfadmodelle, die jeweils ein Messmodell und ein Strukturmodell umfassen; das Messmodell spezifiziert Beziehung zwischen den beobachteten Variablen und den hypothetischen Konstrukten; das Strukturmodell beschreibt die Zusammenhänge zwischen den theoretischen Konstrukten (Nachtigall, Kroehne, Funke & Steyer, 2003, S. 5)

2.1.6 Datenerhebung in den empirischen Wissenschaften im Überblick

Aufgabe des Kapitels: Was erwartet mich hier?

Häufig wird vernachlässigt, dass für die Datenerfassung verschiedenste Daten interessant sein können. Nach Auswahl der Untersuchungsart und des Forschungsansatzes werden jetzt passende Methoden mit gewünschten Daten ausgewählt. Kombinationen sind möglich. Eine weiterführende Recherche passend zur Methode und den Variablen wird empfohlen, um möglichst standardisierte, bereits nach Gütekriterien (s. 2.1.7) bewertete Erhebungsmethoden ausfindig zu machen.

Datenerhebungen werden in Primär- und Sekundärerhebungen eingeteilt. Sekundärerhebungen werden alle Datenerhebungen genannt, die nicht unmittelbar Daten erfassen, sondern nachträglich Daten aufbereiten und auswerten (z.B. veröffentlichte Studien, dessen Daten weiterverwendet werden). Primärerhebungen, also direkte Datenerhebungen, können eingeteilt werden nach Häufigkeit der Erhebung (einmalig, wiederholend), Art der Erhebung (Befragung, Beobachtung, Experiment / Test) und Auswahl der Zielgruppe (Zufalls- / bewusste Auswahl).

Die nachfolgende Klassifikation der Datenquellen von Kurt Pawlik (Pawlik, 2006) beschäftigt sich mit Primärerhebungen.

	Datenquelle	Erklärung	Beispiel	Erfassung
	Biographische und Aktuardaten	Soziodemographische Daten, die reaktionsobjektiv erfasst werden können	<ul style="list-style-type: none"> • Alter • Geschlecht • Beruf • Familienstand • Schul-/Berufsbildung • Krankengeschichte 	Biographischer Fragebogen, insb. anamnestisches Interview
2	Verhaltensspuren	Direkt beobachtbare Nachwirkungen (Produkte) menschlichen Verhaltens; Assoziation mit Persönlichkeitsmerkmalen möglich	<ul style="list-style-type: none"> • Äußeres Erscheinungsbild, z.B. Schuhe • Intentionale Verhaltensprodukte, z.B. Zustand des Arbeitsbereichs 	Analyse
3	Verhaltensbeobachtung	Methodisch kontrollierte und systematische Beobachtung aller visuell und akustisch wahrnehmbaren Aktivitäten / Veränderungen des Zustands einer Person	<ul style="list-style-type: none"> • Körperbewegung • Laut- & Sprachäußerungen • Physiologische Reaktionen, z.B. Schwitzen 	<ul style="list-style-type: none"> • Selbs- / Fremdbeobachtung • Reduktive Deskriptionen (Mees & Bolstad, 1977) • Weiterführende quantitative statistische Analyse der protokollierten Verhaltensdaten
4	Verhaltensbeurteilung	Subjektive und summarische Einschätzung und Bewertung der Häufigkeit, Intensität und Ausprägungsform des eigenen Verhaltens oder des Verhaltens einer anderen Person (meist in der Vergangenheit)	Bei einer öffentlichen Rede <ul style="list-style-type: none"> • Unsicherheit • Souveränität 	Subjektive Interpretation und Einschätzung komplexer Merkmale → numerisch und verbal verankerte Skalen
5	Ausdrucksverhalten	Ableiten von Gefühlen, Stimmungen und Affekten	<ul style="list-style-type: none"> • Variationen der Mimik • Stimme • Sprechweise • Ganzkörpermotorik • Handschrift 	Beobachtung & Analyse z.B. durch FACS (Facial Action Coding System): Registrierung mimischer Muskelbewegungen

6	Interview* *Fragebogen leichter	Zielgerichtete mündlicher Kommunikation zwischen 1/mehreren Befragern und 1/mehreren Befragten zur Informationssammlung über das Verhalten und Erleben der befragten Person(en) Zu berücksichtigen: (s. Qualitätskriterien der Umfrageforschung von Kaase, 1999) <ul style="list-style-type: none"> • Was geht hier vor? • Was will man von mir? • Wie soll ich mich am besten verhalten? • Warum will der Interviewer bestimmte Dinge wissen? 	Je nach Standardisierungsgrad: <ul style="list-style-type: none"> • Freie Exploration • Halbstrukturiertes Interview • Standardisierte Interviews 	Achtung: Vermeintlich einfach, jedoch komplex & schwierig valide Daten zu erhalten mit passender Konstruktion, Durchführung und Auswertung (s. Potenzielle Störgrößen: Gefährdung interner Validität)
7	Projektive Tests / Verfahren* *gut geeignet für Implizites	Mehrdeutiges Reizmaterial werden von Proband bearbeitet, um Aufschluss über wichtige psychologische Informationen zu geben	Thematische Apperzeptionstest: Implizite Machtmotivation	Test mit Proband, Beobachtung und Analyse Achtung. Einschränkung der Gütekriterien
8	Objektive Tests	An Gütekriterien überprüfte Stichproben sowie „undurchschaubare“ Messintention	<ul style="list-style-type: none"> • Leistungsaufgaben • „Flächengrößen vergleichen“ zur Bestimmung von Impulsivität vs. Reflexivität des Probanden (Kubinger & Ebenhö, 1996) 	Test mit Probanden Beobachtung und Analyse
9	Fragebogen*	Ein Erhebungsinstrument, bei dem festgelegte Antwortmöglichkeiten auf klar vorgegebene Fragen / Feststellungen angekreuzt werden *Besonderer Einsatz von Persönlichkeitsmerkmalen, Motivations- und Interessenvariablen		<ul style="list-style-type: none"> • Kurze Instruktion (Beispieltem) • Keine offene Beantwortung! • Ein Messwert: viele Antworten gesammelt • Selbst-/Fremdeinschätzung
10	Psycho-physiologische Diagnostik	Erfassung der Veränderung des Erlebens und Verhaltens durch das Kovariieren mit körperlich-organismischen Veränderungen	<ul style="list-style-type: none"> • Blutdruck • Herzfrequenz • Ausschüttung bestimmter Hormone • Aktivität in Hirnarealen 	Elektrophysiologische Registrierungsmethoden: <ul style="list-style-type: none"> • Elektrodermale Aktivität • Herz-Kreislauf-Aktivität • Muskelaktivität • Hirnaktivität Biochemische Messgrößen: <ul style="list-style-type: none"> • Hormonspiegel im Blut • Immunbiologische Analyse des Speichels • Messung der regionalen zerebralen Durchblutung /reg. zerebraler Stoffwechsel

Tabelle 4: Klassifikation der Datenquellen (in Anlehnung an Kurt Pawlik 2006)

Weiterführende Literatur:

- (Porst, 2014): Fragebogen. Ein Arbeitsbuch. Wiesbaden, VS Verlag für Sozialwissenschaften
- (Möhring & Schlütz, 2019): Die Befragung in der Medien- und Kommunikationswissenschaft
- (Heidmann, 1999): Aufgaben- und nutzerorientierte Unterstützung kartographischer Kommunikationsprozesse durch Arbeitsgraphik. Herdecke.
- (Bortz & Döring, 2006): ... Kapitel 4 (Quantitative Methoden der Datenerhebung), Kapitel 5.2 (Qualitative Methoden der Datenerhebung).
- (Hussy et al., 2013):... Kapitel 2 (Quantitative Erhebungsmethoden), Kapitel 6 (Qualitative Erhebungsmethoden)
- (Collins, 2003): Pretesting Survey Instruments: An Overview of Cognitive Methods. Quality of Life Research 12: 229 - 238.
- (Diekmann, 2018) : Sozialforschung: Grundlagen, Methoden, Anwendungen
- (Pawlik, 2006): Handbuch Psychologie
- (Schnell et al., 2018): Methoden der empirischen Sozialforschung
- (Zendas, 2013): Arbeitshilfe für die datenschutzgerechte Gestaltung wissenschaftlicher Umfragen. http://www.zendas.de/themen/umfragen/umfragen_arbeitshilfe.htm (30.10.2020)
- (FDZ Bildung): <https://www.forschungsdaten-bildung.de/zugang-forschungsinstrumente>
- (GESIS-Leibniz-Institut Für Sozialwissenschaften) – Guidelines: <https://www.gesis.org/gesis-survey-guidelines/home>
- (Bogner & Landrock, 2015): Antworttendenzen. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). DOI: 10.15465/gesis-sg_016
https://www.gesis.org/fileadmin/upload/SDMwiki/Antworttendenzen_Bogner_Landrock_08102015_1.1.pdf
- (Lenzner & Menold, 2015): Frageformulierung. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). DOI: 10.15465/gesis-sg_017
https://www.gesis.org/fileadmin/upload/SDMwiki/Frageformulierung_Lenzner_Menold_08102015_1.1.pdf
- (DGPs und BDP) Diagnostische Instrumente
https://www.dgps.de/fachgruppen/diff_psy_archiv/frames/frame-instrumente.html (30.10.2020)
- Studien als Referenz; (Psychdata); <https://www.psychdata.de/index.php?main=search&sub=browse#7> (30.10.2020)
- Etablierte diagnostische Instrumente:
 - Standardisierte psychologische Tests: Brähler, Holling, Leutner und Petermann (Brähler & Brickenkamp, 2002)
 - (GESIS-Leibniz-Institut Für Sozialwissenschaften): Studien planen und Daten erheben
<https://www.gesis.org/angebot/studien-planen-und-daten-erheben>
 - PSYNDEX (Literatur, Testverfahren, audiovisuelle Medien): <https://www.leibniz-psychology.org/>
 - Untersuchung grundlegender Persönlichkeitsmerkmale und Lebenszufriedenheit: sozioökonomisches Panel (SOEP, www.diw.de/soep)
- Fragebögen erstellen: (Wirtschaftspsychologische Gesellschaft):
 - <https://wpgs.de/fachtexte/frageboegen/> (30.10.2020)
 - <https://wpgs.de/fachtexte/frageboegen/umfrage-und-fragebogen-erstellen-ablauf-tipps-und-information/> (30.10.2020)
- Umfragen erstellen:
 - <https://www.evasys.de/startseite.html>
 - <https://www.limesurvey.org/de/>
- Komplexe Fragebogendesigns: <https://www.unipark.com/>
- Einfache Befragungen: <https://www.soscisurvey.de/>

2.1.7 Gütekriterien empirischer Untersuchungen berücksichtigen

Aufgabe des Kapitels: Was erwartet mich hier?

Über den gesamten Forschungsprozess, also von der Planung, der Durchführung, bis hin zur Auswertung sind wissenschaftliche Kriterien einzuhalten, um die Qualität der Forschung nachzuweisen. Diese Kriterien werden in der empirischen Forschung Gütekriterien genannt. Es ist essenziell, mindestens die Hauptkriterien quantitativer Forschung: Objektivität, Reliabilität und Validität zu kennen und in der Diskussion über die eigene Forschung anwenden zu können.

Gütekriterien: (Bortz & Döring, 2006, S. 729)

„Kriterien, um die Qualität von Untersuchungen, Datenerhebungsverfahren oder statistischen Methoden einzuschätzen.“

Die Bedeutung der Gütekriterien für die Qualität der Datenerhebung, also der Messung, wird hier insbesondere hervorgehoben (Hussy et al., 2013, S. 23; Lienert & Raatz, 1998 s. *Klassische Testtheorie (KTT) und probabilistische Testtheorie*). Je nach Methodeneinsatz können sich die Gütekriterien verändern bzw. etwas unterschiedlich definiert werden. Ein allgemeines Verständnis wird im Folgenden erläutert.

Gütekriterien quantitativer Datenerhebung (Hussy et al., 2013, S. 23; Lienert & Raatz, 1998)	Gütekriterien qualitativer Datenerhebung Bortz & Döring, 2006, S. 326; Hussy et al., 2013, S. 25)
<u>Hauptgütekriterien:</u> <ul style="list-style-type: none"> • Objektivität • Reliabilität • Validität <u>Nebengütekriterien:</u> <ul style="list-style-type: none"> • Normierung • Vergleichbarkeit • Ökonomie • Nützlichkeit von Tests 	<ul style="list-style-type: none"> • Objektivität, Reliabilität und Validität in jeweils modifizierter Formquantität • Verfahrensdokumentation • Argumentative Interpretationsabsicherung • Regelgeleitetheit • Nähe zum Gegenstand • kommunikative Validierung • Triangulation

Tabelle 5: Gegenüberstellung von Gütekriterien quanti- und qualitativer Datenerhebungen

Objektivität (Bortz & Döring, 2006, 195f)

Die Objektivität, das allgemeine Gütekriterium wissenschaftlicher Untersuchungen, beschreibt die Unabhängigkeit der Resultate von Versuchssituation und Versuchsleitern. „Verschiedene Forscher müssen unter den gleichen (Versuchs-) Bedingungen zu den gleichen Ergebnissen gelangen“ (Hussy et al., 2013, S. 23). (Döring & Bortz, 2016) nennen die Objektivität ein unproblematisches Testgütekriterium, da festgelegt werden kann, wie der Test durchzuführen, auszuwerten und das Ergebnis zu interpretieren ist und diese Informationen detailliert im Testhandbuch festgehalten werden kann (S. 195f). Unterschieden wird zwischen:

- Durchführungsobjektivität: die Ergebnisse einer Untersuchung sind unabhängig vom Versuchsleiter. Gegenmaßnahme: Standardisierung von Instruktionen und der Beantwortung von Rückfragen
- Auswertungsobjektivität: die Ergebnisse einer Untersuchung sind unabhängig von der auswertenden Person. Gegenmaßnahme: eindeutige Art der Item-Beantwortung verwenden
- Interpretationsobjektivität: die Ergebnisse einer Untersuchung sind unabhängig von der Person, die die Ergebnisse interpretiert. Gegenmaßnahme: Orientierung bei der Interpretation an vorgegebenen Vergleichswerten bzw. Normen

Bei qualitativen und projektiven Tests ist die Einhaltung dieses Gütekriteriums schwieriger. Es ist häufig erforderlich, die Objektivität empirisch zu prüfen.

Reliabilität (Bortz & Döring, 2006, 196f)

Reliabilität beschreibt die Zuverlässigkeit und Beständigkeit einer Untersuchung. „Reliabel ist ein Instrument dann, wenn es bei einem relativ gleichbleibenden Verhalten gleiche oder ähnliche Ergebnisse liefert“ (Hussy et al., 2013, S. 24).

Die Einschätzung der Reliabilität kann methodisch erfolgen. Beispiele hierfür sind:

- Retestreliabilität: zwei Ergebnisse eines Tests (unters. Zeitpunkte, selbe Stichprobe) werden korrelativ betrachtet.
- Paralleltestreliabilität: zwei Ergebnisse von zwei ähnlichen Tests (zeitnah, selbe Stichprobe) werden korrelativ betrachtet.

Validität (Bortz & Döring, 2006, 502ff, 743; Hussy et al., 2013, S. 24)

Die Validität beurteilt eine Untersuchung, bzw. ein Erhebungsinstrument danach, ob sie misst, was sie zu messen vorgibt. Die validen Ergebnisse einer Untersuchung hängen nicht von situativen oder personalen Merkmalen bei der Erhebung ab. Die Unterschiede zwischen den Messwerten verschiedener Personen sind hauptsächlich auf Unterschiede in der Ausprägung des zu messenden Konstrukts zurückzuführen. Validitätsaussagen sind immer relativ und vorläufig. Aus mehreren empirischen Untersuchungen, unterschiedlichen Methoden und u.U. verschiedenen theoretischen Perspektiven werden relativ valide Aussagen über empirische Zusammenhänge abgeleitet.

Validität kann wie folgt untergliedert werden (Bortz & Döring, 2006, 53, 505, 502ff; Shadish et al., 2002):

	Erklärung	Beispiel	Maßnahmen
Interne Validität	Kausale Interpretation der Ergebnisse ist inhaltlich eindeutig (aufgetretener Effekt zwischen EG und KG kann eindeutig auf das Treatment zurückgeführt werden; Störfaktoren werden kontrolliert oder ausgeschaltet); Die interne Validität sinkt mit wachsender Anzahl plausibler Alternativerklärungen für die Ergebnisse.	Gruppendynamik; Umgebung: physikalische und räumlich-materiale Bedingungen, z.B. Größe eines Raumes, Temperatur, Lichtverhältnisse	Bei „Nullbefunden“ (kein signifikanter Unterschied zwischen EG und KG) wird die Hypothese verworfen, diese Schlussfolgerung ist nur nach Prüfung von Störgrößen valide!
Externe Validität	Generalisierbarkeit der Untersuchungsergebnisse ist möglich (bezüglich a) Personen, b) Treatments, c) Messinstrumente, d) Untersuchungsbedingungen) Die externe Validität sinkt mit wachsender Unnatürlichkeit bzw. abnehmender Repräsentativität der untersuchten Stichprobe.	Ergebnisse übertragbar auf: ○ Eine Grundgesamtheit aus denen VPs stammen ○ Ähnliche Treatments ○ Andere Operationalisierungen der erhobenen Variablen ○ Andere situative Bedingungen und kulturellen Kontexte	a) Per Zufall ausgewählte Probanden & hinreichend große Stichprobe; abhängig von wissenschafts-theoretischer Position der Forschenden & Fragestellung → Hohe Praxisorientierung: Empirische Überprüfung der Generalisierbarkeit durch Variation einzelner Faktoren (Cronbach & Shapiro, 1982; Shadish et al., 2002) d) Empirische Klärung durch Variation theoretisch relevanter Bedingungen

<p>Konstruktvalidität</p> <p>(zu externer Validität gehörig)</p>	<p>Validität der Messungen aller Variablen; Sind die theoretischen Konzepte der Hypothese in der Untersuchung angemessen erfasst worden?</p>	<p>Ungünstige Messinstrumente werden für die Messung gewählt: z.B. Messung der Fähigkeiten selbstregulierten Lernens durch einen Fragebogen, Ergebnisse werden leicht verfälscht durch bspw. sozial erwünschte Antworten; besser: Fremdbeobachtung</p>	<p>Ausreichende Explikation der verwendeten Konstrukte bzw. genaue Operationalisierungen kontrollieren; weitere Möglichkeiten von Datenerhebungen prüfen</p>
<p>Statistische Validität</p> <p>(zu interner Validität gehörig)</p>	<p>Korrekte bzw. passende Auswahl und Durchführung statistischer Analyseverfahren und Interpretation der resultierenden Ergebnisse (eng mit interner Validität verknüpft)</p>	<p>Falsches Analyseverfahren verwendet, z.B. Mittelwert (intervallskaliert) für nominalskalierte Variable bestimmen</p>	<ul style="list-style-type: none"> ○ Voraussetzungen beachten (z.B. Bestimmtes Skalenniveau) ○ Messinstrumente passend zur Erfassung der AV wählen (mit hoher Reliabilität) ○ Verteilung der AV-Werte betrachten (eingeschränkte Varianz) (Shadish et al., 2002, 42ff)

Tabelle 6: Validität - Erklärung, Beispiele und Maßnahmen

2.1.8 Auswahl & Anzahl der Untersuchungsteilnehmer

Aufgabe des Kapitels: Was erwartet mich hier?

Häufig wird die Auswahl der Untersuchungsteilnehmer unterschätzt. Die Auswahl und der Umfang einer Stichprobe sind für die Auswertung der Ergebnisse relevant, z.B. für die Generalisierbarkeit von Ergebnissen, und werden maßgeblich von der Untersuchungsart und die damit einhergehende gewünschte Aussagekraft einer Untersuchung geprägt.

Die Relevanz der Auswahl und Anzahl von Untersuchungsteilnehmern wird durch die Untersuchungsart geprägt. Hier gilt es zwischen explorativen (z.B. Vorstudien) und hypothesenprüfenden Studien (z.B. spezifische Hypothesen von Hauptstudien) zu unterscheiden. Durch Vorstudien werden Erkenntnisse gewonnen, die die Auswahl und Anzahl der Untersuchungsteilnehmer maßgeblich mitbestimmen. (Bortz & Döring, 2006, Abs. 7.1.1)

- **Explorative Untersuchung:** Die Auswahl der Untersuchungsteilnehmer aus der interessierenden Population ist weitgehend unerheblich. Interessante Hypothesen sind bereits aus Beobachtungen einzelner Teilnehmer ableitbar.
- **Deskriptive Untersuchung:** Hier bestehen keine Richtlinien zur Auswahl und Umfang einer Stichprobe, da lediglich Beobachtungen beschrieben werden.
- **Explanative Untersuchungen:** Um Hypothesen überprüfen und generalisieren zu können, ist die Repräsentativität von Stichproben dringend zu beachten! Unterschieden wird hierbei zwischen:
 - Unspezifische Hypothesen: Es bestehen keine genauen Richtlinien bzgl. der Stichprobengröße. Mit zunehmender Größe wächst jedoch die Wahrscheinlichkeit, eine Hypothese zu bestätigen. (Bortz & Döring, 2006, S. 71)
 - Spezifische Hypothesen: Stichprobengröße ist aus Tabelle ablesbar. (Bortz & Döring, 2006, S. 627)

Stichprobenumfang (Bortz & Döring, 2006, Kapitel 9, S.114f, 627, 736)

Ein angemessener Stichprobenumfang (n), bzw. optimaler Stichprobenumfang (n_{opt}), ist funktional verbunden mit und dadurch abhängig von: der Effektgröße eines Signifikanztests, der angestrebten Teststärke ($1 - \beta$) und dem angestrebten Signifikanzniveau (Alphafehlerniveau), s. Abbildung 9.

Um den optimalen Stichprobenumfang zu berechnen, wählt man zunächst die Parameter: Signifikanzniveau, Effektgröße und Teststärke aus. Zu beachten: Heterogene Zielpopulationen erfordern größere Stichproben als homogene Zielpopulationen.

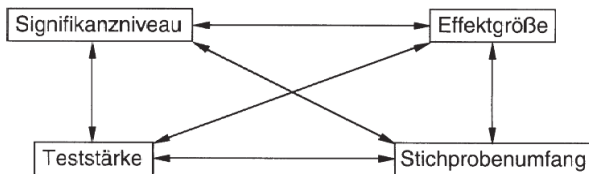


Abbildung 9: Wechselseitige Beziehungen im Signifikanztest (Bortz & Döring, 2006, S. 627)

Konventionen der empirischen Forschung: (Bortz & Döring, 2006, Kapitel 9, S.114f, 130, 627, 736)

- Signifikanzniveau: $\alpha = 0,01$ bzw. $\alpha = 0,05$ (α -Fehler-Niveau bei 1% bzw. 5%)
- Klassifikation von Effektgrößen: klein / mittel / groß
- Teststärke: $1 - \beta = 0,80$ (80%; angemessen für viele Fragestellungen)
- Optimaler Stichprobenumfang wird dementsprechend abgelesen / berechnet

Zentraler Grenzwertsatz: (Duller, 2013, S. 224)

Die Verteilung von Mittelwerten aus Stichproben des Umfangs n , die alle aus derselben Grundgesamtheit gezogen wurden, geht mit wachsendem Stichprobenumfang (zumindest approximativ) in eine Normalverteilung über. Bei genügend großen Stichprobenumfang (ab $n = 30$) kann deshalb mit Konfidenzintervallen gearbeitet werden. Somit ist ein minimaler Stichprobenumfang für die Normalverteilung nach dem Grenzwertsatz $n \geq 30$ möglich, im Experiment ab $n = 50$.

Stichprobenziehung: (Döring & Bortz, 2016, S. 714; Häder & Häder, 333ff)

Die Stichprobenziehung ist von der Forschungsfragestellung und den damit einhergehenden gewünschten Ergebnissen abhängig. Unterschieden wird zwischen der Stichprobe und Totalerhebung. Eine Stichprobe beschreibt eine Auswahl von Elementen aus der Grundgesamtheit, welche Rückschlüsse auf die zugrunde liegende Gesamtheit liefern soll. Die Auswahl findet idealerweise nach bestimmten mathematisch-statistisch begründeten Regeln statt. Bei einer Totalerhebung werden alle Elemente einer Grundgesamtheit in eine Untersuchung einbezogen. Eine Auswahl kann bewusst, willkürlich oder zufallsgesteuert sein. Zufallsstichproben / Wahrscheinlichkeitsauswahlen sind:

- **Einfache Zufallsstichprobe** (simple random samples, SRS)
Entspricht dem Urnenmodell: Alle Elemente der Grundgesamtheit haben die gleiche Wahrscheinlichkeit in die Stichprobe zu gelangen.
- **Mehrstufige Zufallsstichprobe**
Zufällige Auswahl auf einer ersten Ebene (Primary Sampling Units (PSUs), innerhalb dieser eine zweite Stichprobenziehung (Secondary Sampling Units (SSUs) stattfindet. (weitere Ebenen sind möglich: z.B. 1. Auswahl: Gemeinden, 2. Auswahl: Haushalte, 3. Auswahl: Personen in Haushalten)

- **Geschichtete¹⁷ Zufallsstichprobe**

(Voraussetzung hierfür: Informationen über Untersuchungseinheiten (min. ein Schichtungsmerkmal, das hoch mit untersuchten Merkmal korreliert) liegen vor und können in den Stichprobenplan aufgenommen werden. (Schichtung = Stratification)

Es werden Teilgruppen gebildet, in welchen eine separate Ziehung von Stichproben durchgeführt wird, dadurch können interessierende Subpopulationen entsprechend ihrem Anteil in der Grundgesamtheit vertreten werden (= **Proportionale Schichtung**). Es können auch geringer besetzte Schichten stärker berücksichtigt werden, um Schätzungen höherer Präzision angeben zu können (= **Disproportionale Schichtung**). Hierbei müssen bei Aussagen über die Grundgesamtheit Gewichte bei der Schätzung eingesetzt werden.

- **Klumpenstichprobe** (Cluster sample)

(Voraussetzung: Informationen über Untersuchungseinheiten liegen vor)

Klumpen (= natürliche Teilgruppen einer Population, z.B. Studierende je einer Universität) werden komplett erfasst.

Fehlerquellen von Zufallsstichproben:

Häufig ist ein Zufallsprozess bei der Ziehung nicht gegeben oder kann nicht sichergestellt werden.

- Zufallsfehler der Stichprobe (sampling variability), z.B. Streuung des Mittelwertes
- Systematische Fehler durch Auswahlverfahren, z.B. Elemente der Bevölkerung gehen mit untersch. Wahrscheinlichkeit in Stichprobe ein (falls bekannt: Gewichtung, falls nicht: Redressment)
- Verzerrungen (nicht durch das Auswahlverfahren verursacht) (nonsampling bias)
 - Messfehler
 - Fehlerquellen im Interview
 - Over-/Undercoverage (Auswahlpopulation enthält Elemente, die nicht zu der definierten Grundgesamtheit gehören/ Elemente, die zur definierten Grundgesamtheit gehören, werden nicht in der Auswahlpopulation erfasst)
 - Non-Response (Nichterreichbarkeit (Unit-Nonresponse), Verweigerung (Unit- und Item-Nonresponse))

Beurteilung der Güte einer Stichprobe:

[Repräsentanznachweis, also der Nachweis darüber, dass bestimmte Merkmale in einer Stichprobe mit derselben Häufigkeit vorkommen wie in der Grundgesamtheit, ist nicht ausreichend!]

1. Genaue Angaben zur Grundgesamtheit
2. Beschreibung der Auswahlgesamtheit
2. Ziehungsprozess beschreiben
3. Ausfälle benennen
4. Ausschöpfungsquote benennen (Bruttoausgangsstichprobe – stichprobenneutrale Ausfälle = bereinigte Stichprobe; Bereinigte Stichprobe – Unit-Nonresponse¹⁸ (Ausfallsursachen) = tatsächliche Stichprobe; Non-responsequote = 1-Ausschöpfungsquote, ca. bei 25-30%)
5. Verwendeten Instrumente beschreiben

¹⁷ Schichten sind nach Vorgabe bestimmter Merkmale/-kombinationen definierte Teilgruppen; z.B. Ärzte unter 40 Jahren

¹⁸ Unit-Nonresponse: Alle Angaben der Zielperson fehlen (durch z.B. nicht aufgefundene Adressen, Verweigerung, Krankheit, Datenerfassungsfehler, Datenaufbereitungsfehler)

Anwerbung von Untersuchungsteilnehmern (Bortz & Döring, 2006, S. 71)

Die Anwerbung der Untersuchungsteilnehmer ist sorgfältig zu planen, um die Verweigerungsrate bzw. Stichprobenverzerrung zu reduzieren. Förderliche Maßnahmen hierfür sind:

- Individuelle und persönliche Ansprache (schriftlich / mündlich)
- Untersuchungsvorhaben (wenn möglich) inhaltlich erläutern
- Angabe, wem die Untersuchung potenziell zugutekommt
- Bei interessanten Untersuchungsergebnissen: Angaben, wie und wann über individuelle Ergebnisse informiert wird

Interessant: Die Bereitschaft, an einer Untersuchung teilzunehmen, steigt nach Rosenthal und Rosnow (1975) mit möglichst hohem sozialen Status der potenziellen Probanden.

Merke:

Durch die Anwerbung von Untersuchungsteilnehmern werden „Erwartungshaltungen erzeugt, die die Reaktionen der Untersuchungsteilnehmer auf die spätere Untersuchungssituation nachhaltig beeinflussen“ (Bortz & Döring, 2006, S. 71).

Freiwillige Untersuchungsteilnehmer vs. Verweigerer (Bortz & Döring, 2006, 73f)

Es bestehen typische Merkmale freiwilliger Untersuchungsteilnehmer und situativer Determinanten der Freiwilligkeit, die in Tabelle 7 mit denen der Verweigerer verglichen werden. Je nach Forschungsfrage sind Verzerrungen dadurch u.U. möglich. Dann ist es sinnvoll, Maßnahmen zu ergreifen, wie z.B. gruppenspezifisch für eine Untersuchung zu werben. Sind Einflüsse zu erwarten, können diese als unabhängige Variablen miterfasst und hinterher als Kontrolle dienen. Zusätzlich sollten die Motive von Untersuchungsteilnehmern inhaltlich hinterfragt werden.

Merkmale	Ausprägung der freiwilligen Untersuchungsteilnahme im Vergleich zu Verweigerern
Schulische Leistungen (bei Schülern irrelevant)	Bessere Schulnoten
Einschätzung: eigener sozialer Status	Höherer sozialer Status
Intelligenz	Höhere Intelligenz (in meisten Untersuchungsergebnisse)
Bedarf an sozialer Anerkennung	Höherer Bedarf an sozialer Anerkennung
Eigenschaft: Geselligkeit	Geselliger
Einstellung bzgl. geschlechtsspezifisches Verhalten	Unkonventionellere Einstellung bzgl. geschlechtsspezifischem Verhalten
Geschlecht	Eher weibliche Personen
Eigenschaft: Autorität	Weniger autoritär
Tendenz zu konformem Verhalten	Schwächer ausgeprägte Tendenz zu konformem Verhalten

Tabelle 7: Untersuchungsteilnehmer - Unterscheidung von Freiwilligen zu Verweigerern

Hohe Bereitschaft einer freiwilligen Teilnahme durch:

(Bortz & Döring, 2006, S. 74)

- Interesse für den Untersuchungsgegenstand der Personen
- Einschätzung der Untersuchung als bedeutend (Je bedeutender, desto höher die Bereitschaft)
- Kleine persönliche Geschenke und Aufmerksamkeiten vor der Entscheidung der Teilnahme (Fördern mehr die Freiwilligkeit als Entlohnungen in Form von Geld!)
- Bekannte anwerbende Person
- Öffentliche Unterstützung der Untersuchung

Merke:

Eine entspannte Anwerbungssituation und Untersuchungsdurchführung, die die persönliche Freiheit und den Handlungsspielraum der Untersuchungsteilnehmer möglichst wenig einengen, tragen zur Reduktion von Verzerrungen (Stichwort: Psychologische Reaktanz) bei.

Werden ausschließlich **Studenten als Untersuchungsteilnehmer** ausgewählt, sind falsche Schlüsse für die Untersuchung möglich. Insbesondere bzgl. der Freiwilligkeit, ihrer Entwicklung, Sozialisation, Kognition und Intellekt ist die externe Validität zu problematisieren. (Bortz & Döring, 2006, S. 75)

🔍 (Häder & Häder)

2.1.9 Berücksichtigung von Störvariablen und Störfaktoren

(Bortz & Döring, 2006)

Aufgabe des Kapitels: Was erwartet mich hier?

Sehr wichtig, doch wenig beachtet, sind Einflüsse auf eine Erhebung, die nicht erfasst und somit nachträglich nicht nachvollziehbar sind. Diese Einflüsse stören die Untersuchung selbst bzw. deren Auswertung. Nachfolgend werden diese potenziellen störenden Einflussgrößen genauer erläutert, differenziert und es werden Maßnahmen beschrieben, wie man dagegen vorgeht. Dies ist besonders vor Beginn einer Untersuchung zu berücksichtigen!

Hinweis:

In der Literatur werden die Begrifflichkeiten Störgrößen, Störvariablen, Störfaktoren, etc. teilweise überschneidend, inkonsistent oder undefiniert genutzt.

Empfehlung: Klare Definition in eigener Forschungsarbeit; Orientierung an Verwendungsfeldern; Vermeidung ähnlicher Ausdrücke (Nutzung von „Störvariablen“, „Drittvariablen“ und „Störfaktoren“)

Verwendungsfelder:

Statistik: Störgrößen, Störparameter (im Sinne von Residuen)

Experimentelle Forschung: Störvariable, Störfaktoren („Faktor“= Synonym für unabhängige Variable)

Korrelative Forschung: Drittvariablen (*s. statistische Auswertung: Drittvariablenanalyse*)

Störvariablen:

Alle Variablen, die systematisch mit den Stufen der unabhängigen Variable variieren und Einfluss auf die abhängige Variable haben (Sachverhalt wird Konfundierung genannt).

Zu verhindern ist die systematische Variation einer potenziellen Einflussgröße mit den Stufen einer oder mehrerer unabhängigen Variablen.

Klassifikation von Störvariablen:

- Probandenmerkmale¹⁹ (z.B. bessere/schlechtere Konzentrationsfähigkeit aufgrund von Pausen)
- Situationsmerkmale²⁰ (z.B. laut/leise)
- Versuchsleitermerkmale (z.B. ermunterndes/entmutigendes Verhalten)

→ Störvariablen werden konstant gehalten bzw. kontrolliert, damit kein Einfluss auf das Testergebnis ausgeübt wird (Kontrolle des Effektes von Störvariablen durch experimentelle Techniken, wie Elimination, Konstanthalten, Parallelisierung, Randomisierung)

Ziel: Vergleichbarkeit der Untersuchungen (→ Vergleichbares Verhalten des Versuchsleiters gegenüber den Probanden, vergleichbare Situationsmerkmale für alle Probanden)

Drittvariablen: (Stangl, 2020)

Drittvariablen bezeichnen in einem Variablengefüge „jene Variablen, die weder als unabhängige noch als abhängige Variablen zu bezeichnen sind“. Beeinflusst eine Drittvariable den Zusammenhang zwischen der unabhängigen und abhängigen Variablen, so tritt die Drittvariable als Moderatorvariable in Erscheinung.

Werden in einem Experiment Drittvariablen nicht erfasst, bezeichnet man diese meist als Störvariablen und erwartet, dass deren Einflüsse sich gegenseitig aufheben.

Störfaktoren gehen mit der **Reduzierung der Validität** einher.

Gefährdung interner Validität: (Bortz & Döring, 2006, 502f, 523, 547)

- Sozialpsychologisch, z.B. Ersteindrucksbildung, Stereotype
- Motivationspsychologisch, z.B. Bedürfnisse nach Kontrolle und Komplexitätsreduktion, Selbstdarstellungstendenzen, Interesse
- Emotionale Gestimmtheit, z.B. Angst, Sympathie, Argwohn
- Habituelle Tendenzen: Persönlichkeitsmerkmale, z.B. Ängstlichkeit, Selbstwertschätzung, soziale Kompetenzen

Störfaktoren:

Unter Störfaktoren werden Einflussfaktoren verstanden, die neben der experimentellen Ursache ebenfalls Unterschiede in der abhängigen Variable verursacht haben können – und somit Alternativerklärungen zur Hypothese des Experiments darstellen. Auswirkend werden Unterschiede der Versuchs- und Kontrollgruppe nicht auf das Variieren der unabhängigen Variable zurückgeführt. *(Potenzielle Einflussfaktoren / Effekte (sozial)psychologischer Art)*

¹⁹ Probandenmerkmale (Versuchspersonenmerkmale): Alle Unterschiede der Probanden in den verschiedenen experimentellen Bedingungen (vers. Stufen der unabhängigen Variablen) sind potenzielle Störvariablen; z.B. Alter, Geschlecht, Intelligenz, Ausbildung, etc.; erst durch Konfundierung liegt Probandenmerkmal vor

²⁰ Potenzielle Situationsmerkmale: Tageszeit, Beleuchtung, Untersuchungsmaterial, Raumtemperatur

Zeitabhängige Störfaktoren

Folgende Einflussfaktoren („Confounder“) gefährden die interne Validität einer Untersuchung:

- **Externe zeitliche Einflüsse:** Bei Veränderungsprozessen können andere als die untersuchten Einflussgrößen, Ursache für die Wirkung auf die abhängige Variable haben; z.B.: In einer Untersuchung zum rückläufigen Fernsehkonsum von Kindern durch das weniger attraktive Fernsehangebot ist der rückläufige Fernsehkonsum auf neue Freizeitangebote zurückzuführen. Vor allem bei besonderen Anforderungen bzw. Belastungen in Experimenten
- **Reifungsprozesse** der Untersuchungsteilnehmer: Verhaltensänderungen der Untersuchungsteilnehmer durch ihre „Reifung“ (älter, hungriger, erfahrener, weniger aufmerksam etc.)
- **Testübung:** Das Untersuchungsinstrument beeinflusst das zu Messende; z.B.: Das Ausfüllen eines Einstellungsfragebogen verändert die zu messende Einstellung.
Kontrolle der Testeffekte und Wechselwirkungen zwischen Vortest und Treatment durch Nutzung: Vier-Gruppen-Designs von Solomon (Kontrollgruppe mit Vor- und Nachtest und eine nur mit Nachtest)
- **Mangelnde instrumentelle Reliabilität:** Das Untersuchungsinstrument erfasst das zu Messende nur ungenau oder fehlerhaft; z.B.: Testskala wurde nicht auf Eindimensionalität überprüft
- **Statistische Regressionseffekte:** Werden Veränderungshypothesen mit nicht zufällig ausgewählten Stichproben überprüft, kann es zu Veränderungen kommen, die statistisch bedingt sind.
- **Selektionseffekte:** v.a. bei quasiexperimentellen Untersuchungen; Beim Vergleich von nicht-randomisierten Gruppen, können durch Selbstselektion Gruppenunterschiede resultieren, die mit der geprüften Maßnahme nichts zu tun haben. Die Fragestellung ist hierbei relevant.
z.B. Zur Psychophysik sind keine bestimmte Zielgruppe und keine gruppenspezifischen Verhaltensweisen relevant (s. Westermann 2000), in späteren Forschungsstadien können gezielt andersartige Probandengruppen einbezogen werden, um Umstände und Probandengruppen der Theorie(n) zu überprüfen
- **Experimentelle Mortalität:** Ausfall von Untersuchungsteilnehmern (durch Interessenverlust, Umzug, Krankheit, Tod); Ist die Bereitschaft an einer Untersuchung (bis zum Ende) teilzunehmen, nicht unter allen Untersuchungsbedingungen gleich, kann es zu erheblichen Untersuchungsverfälschungen kommen; besonders bei Wiederholungsmessungen und Längsschnittstudien

Bei Felduntersuchungen zusätzlich (Cook & Campbell 1979):

(Verfälschung der Kontrollgruppe durch das Untersuchungsgeschehen)

- **Empörte Demoralisierung** (Resentful Demoralization): Kontrollgruppe (KG) erfährt, dass Experimentalgruppe (EG) günstigere bzw. vorteilhaftere Behandlung erfährt → Reaktion: Neid, Ablehnung, Empörung beeinträchtigte Reaktionen
- **Kompensatorischer Wettstreit** (Compensatory Rivalry; Synonym: John-Henry-Effekt): Wahrgenommene Ungleichheit in der Behandlung der EG führt zu Verhaltensänderungen durch z.B. Motivation, Ehrgeiz, Anstrengungsbereitschaft der Untersuchungsteilnehmern
- **Kompensatorischer Ausgleich** (Compensatory Equalization): Versuchsleiter bemerkt Ungerechtigkeiten im Vergleich von KG und EG und versucht diese mit gezielten Maßnahmen auszugleichen
- **Treatmentdiffusion** (Treatment Diffusion): KG erhält Kenntnisse über Geschehen in der EG und versucht die Reaktionen der EG zu antizipieren und imitieren

Gefährdung externer Validität:

- **Mangelnde instrumentelle Validität:** Das Untersuchungsinstrument erfasst nicht das, was es eigentlich erfassen sollte; z.B.: Pazifismusskala von vor 40 Jahren aus den USA wird heute in Deutschland eingesetzt
- **Stichprobenfehler:** Untersuchungsergebnisse einer Stichprobe dürfen nicht auf Grundgesamtheiten verallgemeinert werden, für die die Stichprobe nicht repräsentativ ist
- **Experimentelle Reaktivität:** v.a. bei Laboruntersuchungen; Ergebnisse sind zunächst nur unter den Bedingungen valide, unter denen sie ermittelt wurden; z.B.: Angstreaktionen im Labor haben eine andere Qualität wie im Alltag
- **Pretesteffekte:** Einschränkung der Generalisierbarkeit der Untersuchungsbefunde, wenn Pretests die Sensitivität oder das Problembewusstsein der Untersuchungsteilnehmer verändern
Andere Reaktionen vorgetesteter Untersuchungsteilnehmer als nicht zuvor getestete
- **Hawthorne-Effekte** (Bortz & Döring, 2006, S. 504): Verhaltensänderung von Untersuchungsteilnehmern durch das Bewusstsein, Teilnehmer einer Untersuchung zu sein

Systematik von Störfaktoren (Campbell & Stanley, 1967; Sarris & Reiß, 2005)

Die in Tabelle 8 aufgezeigte Systematik kann bei jeder Art kausaler Schlussfolgerungen, die aus empirischen Befunden gezogen werden könnten, als Art Kategoriensystem für die Diskussion der Validität der Studie bzw. der möglichen Störfaktoren verwendet werden. Hierbei werden einige der zuvor aufgelisteten Störfaktoren nochmals aufgegriffen. [in zeitlicher Abfolge eines Experiments dargestellt]

	Bezeichnung der Störgröße	Erläuterung	Beispiel
1	Auswahlverfahren der Untersuchung	Auswahl der Probanden, die an der Untersuchung teilnehmen → externe Validität Aufteilung auf Versuchs- und Kontrollgruppe(n) → interne Validität	Freiwilligkeit garantiert nicht Repräsentativität der Probanden; falls Konfundierung mit Treatment: Randomisierung oder Parallelisierung, ggf. Vortest durchführen
2	Experimenteller Dropout	Ausfall von Probanden während der Untersuchung (→ allg. die externe und gruppenspezifisch die interne Validität (Vergleichbarkeit der Gruppen geht verloren))	Insbesondere bei mehrtägigen bzw. Langzeituntersuchungen (vgl. „Panel-Dropout“)
3	Testeffekte	Einflüsse des Vortests auf das Verhalten während des Experiments oder bei dem Nachtest → interne Validität	Vortest als „advance organizer“ bei Lehr-Lern-Untersuchungen; Trainingseffekte wiederholter Tests
4	Veränderung der verwendeten Hilfsmittel	Geringe zeitliche Stabilität der Geräte, Beobachter, Auswerter und/oder Testverfahren, die eingesetzt werden, oder durch Wechsel der Messverfahren (zwischen Vor- und Nachtest)	Verschleiß von Messgeräten (z.B. Recorder); Veränderung der Handhabung von Kategorien bei Beobachtungen bzw. Testauswertungen; Unterschiede zwischen verschiedenen Methoden
5	Reaktive Effekte in der Untersuchung	Bewusstsein der Teilnahme an einer Untersuchung → gruppenspezifisch interne Validität;	Placebo- und Hawthorne-Effekt

		Generalisierbarkeit auf andere Situationen	
6	Zwischenzeitliches Geschehen	Einflüsse von außen während des gesamten Untersuchungszeitraumes Δt auf die abhängige Variable (z.B. Einstellung Y) → interne Validität	Ereignisse aus politischer oder institutioneller Ebene (z.B. Streik, Entlassungen, Gewaltverbrechen); äußere Faktoren z.B. Spritpreiserückgang während Studie über ÖPNV-Nutzung
7	„Natürliche“ Änderungsprozesse („Reifung“)	Biologisch oder physiologisch bedingte Veränderungen, die im gesamten Untersuchungszeitraum Δt aller Untersuchungsgruppen auftreten → interne Validität	Konzentrationsverlust bei monotoner Tätigkeit, Hunger, Anstrengung, Langleweiligkeit, Müdigkeit; individuelle Entwicklung insbesondere bei Kindern und älteren Menschen; Ermüdung
8	Statistische Regression	Statistisch zu erwartende Änderung des individuellen Messwertes durch die sog. „Regression zum Mittelwert“	Insbesondere bei der Verwendung von Extremgruppen

Tabelle 8: Systematik von Störfaktoren nach (Campbell & Stanley, 1967)

Kontrolltechniken potenzieller Störvariablen:

- **Elimination** (Einflüsse einer Variable auf Null setzen; z.B. Lärm ausschalten)
- **Konstanthalten** (alle Maßnahmen zur Standardisierung der Untersuchungssituation; insb. bei Probanden- und Situationsmerkmale; z.B. sinnvoll bei bestehenden Unterschieden in abhängiger Variable vor Treatment)
Vorexperimentelle Befragung empfohlen
 - **Randomisierung**
 - **Parallelisierung**
- **Systematische Variation** (Störvariable als weitere unabhängige Variable (Kontrollfaktor) untersuchen; falls möglich)
- **Zufällige Variation** (Möglichst viele Ausprägungsgrade potenzieller Störvariablen werden realisiert, Probanden werden diesen zufällig zugeordnet)
- **Blindversuche** (Versuchsleiter kennt die Hypothesen, die der Untersuchung zugrunde liegen, nicht)

Möglichkeiten des Konstanthaltens von Störvariablen sind in *Tabelle 9* aufgeführt. (s. Abbildung 5)

Maßnahme	Erklärung	Kontrolle Störgrößen	Vorteile	Nachteile
Randomisierung	Zufällige Einteilung der Probanden in EG und KG	Werden konstant gehalten (personenbezogene Störgrößen); Prinzip des statistischen Fehlerausgleichs, Neutralisierung personenbezogener Störvariablen	Einfach, hohe interne Validität	Große Stichprobe (≥ 60 Probanden) notwendig
Parallelisierung (Matching)	Einteilung der Probanden, sodass ähnlicher Mittelwert und ähnliche Verteilung der Störgrößen in Gruppen besteht	Werden konstant gehalten Bsp.: Bei Störgröße „Lernerfolg“: 2 Probanden mit ähnlichem Notendurchschnitt in Mathe werfen Münze für Gruppen-einteilung	Bei kleinen Stichproben	Messung von potenziellen Störvariablen vor Untersuchung

Laborunter-suchung* *Ethische Gren-zen berücksich-tigen		Bessere Kontrolle als im Feld	<ul style="list-style-type: none"> • Veränderung des Verhaltens durch Studienteilnahme (Angepasstes Verhalten) → Problem externer Validität (Inwiefern sind gefundene Effekte auf die Realität übertragbar?)
---	--	-------------------------------	---

Tabelle 9: Umgang mit Störvariablen

2.1.10 Potenzielle Einflussfaktoren / Effekte (sozial)psychologischer Art

Aufgabe des Kapitels: Was erwartet mich hier?

Weitere Einflussfaktoren, alltagspsychologischer Art beispielsweise, können eine Untersuchung ebenfalls verzerren. Besonders als Ingenieur sind sozialpsychologische Effekte wenig bekannt und sollen hier näher beschrieben werden.

Je nach Forschungsfrage und Untersuchungsart sind diese potenziellen Einflussfaktoren individuell zu berücksichtigen.

Es bestehen viele, teils mehr oder weniger gut erforschte, Effekte in der Psychologie, die Einfluss auf die Wahrnehmung, Einstellung und subjektive Empfindungen des Menschen nehmen und somit Untersuchungen mit Probanden verfälschen können. Im Folgenden wird eine Auswahl dieser dargestellt (s. Tabelle 10).

Name	Beschreibung
Primacyeffekt	Urteilsverzerrung: Erster Eindruck bleibt deutlicher im Kopf eines Menschen als die nachkommenden. (Asch, 1952)
Recencyeffekt	Urteilsverzerrung: Letzter Eindruck bleibt deutlicher im Kopf eines Menschen. Bei Aufzählungen wird zuletzt genanntes stärker wahrgenommen. Gegenmaßnahme: Probanden bitten, sich alle Informationen zu merken
Fundamentaler Attributionsfehler / Akteur-Beobachter-Fehler	Erklärung von Verhalten von Personen: Personen schreiben Verhalten anderer Personen eher personalen, das eigene Verhalten situationalen Faktoren zu (Ross, Jones & Nisbett, 1972)
Negativitätsbias	Tatsächliche Eigenschaften werden von innerer Einstellung „verfälscht“ → Fehler und Urteilsverzerrungen (Renner, 2002)
Fehler beim alltagspsychologischen Umgang mit Wahrscheinlichkeiten	Menschen besitzen keinen intuitiven Umgang mit Wahrscheinlichkeiten Empfehlung: Youtube-Videos von Gigerenzer zu Statistikfallen, anschauliche Beispiele verwenden (Gigerenzer, 2015)
Unzureichende Prüfung alltagspsychologischer Vermutungen	Kein systematisches Prüfen von Hypothesen; Menschen orientieren sich an Erfahrungen oder Autoritäten
Pygmalion- / Rosenthal-Effekt	Tendenz soziale und psychische Wirklichkeit aufgrund bloßer Vermutungen zu erzeugen
Barnum Effekt	Neigung von Menschen, vage und allgemeingültige Aussagen über die eigene Person so zu interpretieren, dass sie als zutreffende Beschreibung empfunden werden
Arousal-Theorie	Stimuli mit mittlerem Erregungspotenzial („arousal“) werden positiver bewertet als Stimuli mit schwacher oder starker Erregung

	(Berlyne, 1974) (Bortz & Döring, 2006, S. 17)
Erwartungseffekte bei Versuchsleitern	Kenntnis der Untersuchungshypothese beim Versuchsleiter wirkt sich auf das Verhalten und die Leistung von Probanden und damit auf die Versuchsergebnisse aus (s. Selbsterfüllende Prophezeiung) Gegenmaßnahme: Doppelblindversuch (Bortz & Döring, 2006, 82ff)
Ermüdungseffekt	Gefahr, dass bei derartig aufwendigen Versuchsreihen die Ergebnisse durch Ermüdungs- oder Übungseffekte verfälscht werden. (Bortz & Döring, 2006, S. 168)
Reaktanzeffekt	Empirische Untersuchungen versetzen Personen in soziale Situationen, die zu Einschränkung ihrer persönlichen Handlungsfreiheit erleben. Es kann zu Abwehrmechanismen kommen. (Bortz & Döring, 2006, S. 74) Theorie der psychologischen Reaktanz (Brehm, 1966)
Unconscious Bias	Stereotype und unbewusste Vorurteile schränken die objektive Wahrnehmung des Menschen ein
Hawthorne-Effekt	Das Bewusstsein, Teilnehmer einer wissenschaftlichen Untersuchung zu sein, verändert das Verhalten. (Beispiel: Die Näherin einer Großschneiderei, die erfahren hat, dass ihre Leistungen in einer arbeitsanalytischen Untersuchung ausgewertet werden, verhält sich anders als unter normalen Umständen.) (Bortz & Döring, 2006, S. 504) (Roethlisberger & Dickson, 1964).
Spezifitätsproblematik / Individual-spezifische Reaktion	Messproblem; Unabhängig von verschiedenen Stresssituationen reagieren Probanden mit einem für sie typischen Reaktionsmuster Zu unterscheiden: stimulusspezifische (von bestimmten Umweltbedingungen bei allen Individuen in gl./ähnl. Weise hervorgerufen) und motivationsspezifische (durch spezifischen Motivationszustand beim Individuum hervorgerufen) Reaktionen Gegenmaßnahme: Autonome Reaktionsspezifität (Bortz & Döring, 2006, S. 279)
Ausgangswertproblematik	Messproblem: Bezüglich Veränderungsmessungen (Messungen vor und nach bestimmten Treatments und die hierbei auftretende Abhängigkeit zwischen Ausgangswert und der Differenz aus Verlaufs- und Ausgangswert (Veränderungswert) Ausgangswertgesetz (Wilder 1931): „Je stärker vegetative Organe aktiviert sind, desto stärker ist ihre Ansprechbarkeit auf hemmende Reize und desto niedriger ihre Ansprechbarkeit auf aktivierende Reize.“ Gegenmaßnahme: Genaue Planung bei Messungen biophysiologicaler Werte (Bortz & Döring, 2006, ab S. 279)
Anker	Einschätzungen orientieren sich an zuvor festgesetzten Ankern Gegenmaßnahme: Möglichst: Vermeidung von Ankersetzungen; Abfrage von möglichen Ankern als unabhängige Variablen
Confirmation Bias	Tendenz, nur diejenigen Untersuchungen zu veröffentlichen die die Wunschhypothese unterstützen; generell: Menschen suchen nach Bestätigung ihrer bisherigen Annahmen (Bortz & Döring, 2006, S. 695)
Selbsterfüllende Prophezeiung	Erwartungshaltung des Versuchsleiters gegenüber den Probanden beeinflusst das Versuchsergebnis (Bortz & Döring, 2006, S. 686)

Tabelle 10: Potenzielle Einflussfaktoren in Form von (sozial)psychologischen Effekten

2.1.11 Untersuchungsplan erstellen


Aufgabe des Kapitels: Was erwartet mich hier?

Mit diesem Kapitel sollen alle bisher getroffenen Entscheidungen zusammengetragen und in einem Untersuchungsplan zusammengetragen werden. Wie genau dieser aussieht ist zweitrangig, der Ablauf der Untersuchung soll hier festgehalten werden und dient der Vorbeugung von Fehlern, wie beispielsweise dem Vergessen einer Pause während eines Experiments für den Versuchsleiter oder der rechtzeitigen Vorbereitung von Untersuchungsobjekten.

Um alle bisher geplanten Schritte festzuhalten und für die anstehende Durchführung der Untersuchung vorbereitet zu sein, wird ein standardisierter Untersuchungsplan empfohlen, der sowohl für Vorstudien als auch für Hauptuntersuchungen, unabhängig von der Untersuchungsart genutzt werden kann. Eine Vorlage, die nach den Bedürfnissen des/der Forschenden angepasst werden kann, wird hier angeboten.

Empfehlung:

Untersuchungsplan mit anderen, möglichst erfahrenen Forschenden durchsprechen: Die eigene Perspektive ist sehr subjektiv und durch Beschäftigung mit dem Forschungsthema voreingenommen. So kann eine objektivere Einschätzung erfolgen und Fehler vermieden werden.

 Vorlage Untersuchungsplan: *Anhang B: Untersuchungsplan*

2.2 Durchführung

2.2.1 Messen und messtheoretische Probleme

Aufgabe des Kapitels: Was erwartet mich hier?

Während der Durchführung der Datenerhebung kommt es häufig zu messtheoretischen Problemen. Diese werden hier benannt und erklärt, sowie deren Grundlage, das Messen selbst. Nicht erschrecken: Fachvokabular findet hier seinen Platz, um bewusst das Messen aus der nicht-ingenieurwissenschaftlichen Sicht einzuführen.

Messen: (Bortz & Döring, 2006, S. 65; Hussy et al., 2013, S. 65)

Übertragung der Realität in Zahlen, eine sog. strukturerhaltende Abbildung empirischer Relationen zwischen Objekten in numerische Relationen (Zahlen).

Messen als zentrales Element in Untersuchungen:

Ein Hypothetisches Konstrukt soll messbar (Merkmalsausprägungen von Objekten Zahlen zuordnen), Daten erhoben (entspricht Messwerten; mit geeigneten Methoden) und analysiert werden (mit statistischen Verfahren)

Die Qualität der Messung von Konstrukten wird gemessen an den **Gütekriterien**: Objektivität, Reliabilität und Validität. (s. klassische Testtheorie (KTT) und probabilistische Testtheorie)

Die Quantifizierung bzw. Messung der Merkmale (Datenerhebung) ist bereits in der Planungsphase zu klären. Dabei sind die **Skalenniveaus** der Variablen und die mit der Operationalisierung eng verknüpften messtheoretischen Probleme zu berücksichtigen.

[Mit Zunahme des Skalenniveaus nehmen zulässige Transformationen ab und nehmen mathematische Operationen zu]

Messtheoretische Probleme: (Bortz & Döring, 2006, S. 65; Gigerenzer, 1981; Orth, 1974; Steyer & Eid, 2001)

- **Repräsentationsproblem** (*Lässt sich ein gegebenes empirisches Relativ durch ein numerisches Relativ repräsentieren?*)

Eigenschaften des empirischen Relativs prüfen:

- Verhältnis: sog. Transitivität vs. Intransitivität: wenn $a > b$ und $b > c$, dann auch $a > c$ (bei physikalischen Messungen gegeben; bei z.B. Spielerstärke von Fußballern nicht) (s. Sedlmeier & Renkewitz, 2018, S. 58)

Nach Lösung des Problems liegen vor:

- Empirisches Relativ E (z.B.: Geschlecht)
- Numerisches Relativ N (z.B.: 1, 2)
- Zuordnungsregel bzw. Abbildungsfunktion f (Nominalskala); E, N und f konstituieren eine Skala

- **Eindeutigkeitsproblem**

(Skalenniveaus, mit höherem Skalenniveau: Abnahme zulässiger Transformation & Zunahme mathematischer Operationen); Achtung vor Übertragungsfehlern; nicht trivial!
Eine Messung ist umso eindeutiger, je weniger Transformationen möglich sind.

- **Bedeutsamkeitsproblem** (Gigerenzer, 1981)

Welche mathematischen Operationen führen zu empirisch sinnvollen Aussagen?

2.2.2 Skalenniveaus berücksichtigen

Aufgabe des Kapitels: Was erwartet mich hier?

Die Skalenniveaus sind ab dem Zeitpunkt der Operationalisierung von größter Bedeutung und begleiten den Forschungsprozess. Insbesondere bei der Wahl der Methode der Datenerhebung und der -auswertung stellt das Skalenniveau einer Variable wichtige Randbedingungen. Dieser Stellenwert soll hier betont und die einzelnen Skalen genau erläutert werden.

In der Methodenlehre wird das Konzept der Skalenniveaus (s. Tabelle 11) verwendet, um systematisch eine sinnvolle und angemessene Form der Quantifizierung von Merkmalen auszuwählen, also über die Art der Abbildung von Merkmalen durch Zahlen zu entscheiden. Die zugrundeliegende Überlegung lautet hierbei: *Welcher Art ist das interessierende Merkmal oder Ereignis und welche mathematische Kenngröße soll erfasst werden?* (Hussy et al., 2013, 65, 70)

Niveau	Beobachtbare Relationen / bedeutungsvolle Aussagen	Zulässige Transformationen	Mögliche statistische Operationen (Beispiele)	Beispiele
Nominal-Skala	Äquivalenzrelation: Gleichheit / Ungleichheit	Eindeutige Transformation	Modus (Häufigkeitsverteilungen)	Kategorien: Geschlecht, Studienfächer, Nationalitäten
Ordinalskala	Ordnungsrelation: größer / kleiner; besser / schlechter	Streng monotone Transformation (rangfolgebewahrende)	Median (Prüfung von Zusammenhangs- und Unterscheidungshypothesen)	Rangreihen: Schulnoten, akademische Abschlüsse
Intervallskala	Äquidistanz: Gleichheit von Differenzen (kein Nullpunkt)	Positiv lineare Transformation $y = a \cdot x + b$ mit $a > 0$	Mittelwert	IQ-Skala, Ratingskalen
Ratioskala	Verhältnisrelation: doppelt, dreimal so schnell wie... (natürlicher Nullpunkt)	Ähnlichkeits-transformationen $y = a \cdot x$ mit $a > 0$	Geometrisches Mittel	Zeitmaße, Längenmaße, Gewichtsmaße, Einkommen, Reaktionszeiten
Absolutskala	Natürliche Maßeinheit	keine	alle	Häufigkeiten: z.B. Anzahl der Unterbrechungen an einem Arbeitsplatz

Tabelle 11: Übersicht Skalenniveaus

Hinweis:

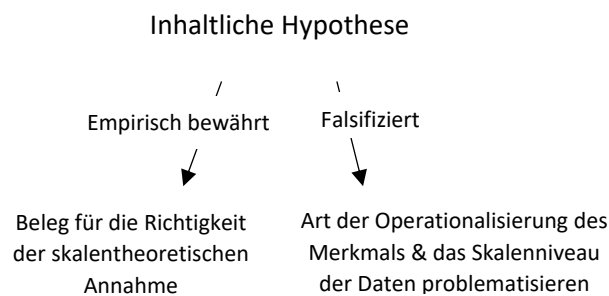
→ Es ist sinnvoll, ein möglichst hohes Skalenniveau, mindestens Intervallskalenniveau (ab hier metrisch), zu erzielen, damit der statistischen Auswertung von Daten keine bzw. weniger Grenzen gesetzt sind (z.B. beim arithmetischen Mittel oder Regressionsanalysen).

Bei einigen Variablen ist das Skalenniveau umstritten: *Sind beispielsweise Testwerte, Einstellungsmessungen, Schätz-(Rating-)Skalen ordinal- oder intervallskaliert?*

In der üblichen Forschungspraxis in der Psychologie wird auf eine empirische Prüfung der jeweiligen Skalenaxiomatik verzichtet. „Die meisten Messungen sind **Per-fiat-Messungen** (Messungen `durch Vertrauen´), die auf Erhebungsinstrumenten [...] basieren, von denen man annimmt, sie würden das jeweilige Merkmal auf einer Intervallskala messen. Das ist umso unproblematischer, je mehr Ausprägungen eine ordinale Variable besitzt; Achtung für Überinterpretationen (Bortz & Döring, 2006, S. 70; Hussy et al., 2013, 66ff)

Annahme: (Bortz & Döring, 2006, S. 70)

Es wird angenommen, dass die Variablen einer inhaltlichen Hypothese skalentheoretisch richtig eingeordnet wurden, wenn eine Hypothese sich bewährt. Wird diese jedoch falsifiziert, muss das angenommene Skalenniveau der Variablen nochmals überprüft und angepasst werden.



2.3 Auswertung

2.3.1 Datenanalyse & Hypothesenprüfung

Aufgabe des Kapitels: Was erwartet mich hier?

Je nach den ausgewählten Datenerhebungsmethoden und den Eigenschaften der Variablen, sind unterschiedliche Auswertemethoden zu wählen.

Generell kann deskriptive Statistik immer verwendet werden, um die Stichprobe zu beschreiben, darauf aufbauend sind dann weitere Möglichkeiten denkbar. In den Exkursen werden die deskriptive und induktive Statistik detailliert erläutert. Diese sollen den Einstieg in die Materie erleichtern. Da viele Grundlagenbücher sehr umfangreich sind, ist hier eine Auswahl der wichtigsten empirischen Datenanalysemöglichkeiten aufgeführt. Weitere Recherchen zur individuellen Datenerhebung ist sinnvoll.

Die Analyse der erhobenen Daten mit statistischen Methoden führt zu zwei wichtigen Aspekten: dem der Erkenntnis und dem der Kommunikation. Dies sollte in der Forschung nicht unterschätzt und die methodische Auswahl dementsprechend offengelegt werden. Das erleichtert zum einen das Nachvollziehen des Prozesses und der Ergebnisse und zum anderen dient es weiterer Forschungen als Grundlage (auch bei Verwerfung einer Hypothese).

Nachstehend wird die grundlegende Vorgehensweise der Datenanalyse, insb. bei einer Hypothesenprüfung, mit einer Auswahl an anwendungsorientierten Hinweisen beschrieben. Zu beachten sind die unterschiedlichen Möglichkeiten, v.a. innerhalb der Inferenzstatistik, die es sorgfältig abzuwägen gilt. Insgesamt ist je nach gewähltem Forschungsansatz und Erhebungsmethode (s. *Tabelle 1*) die Datenauswertung unterschiedlich und sollte bereits in der Planung der Datenerhebungsmethoden mitgedacht werden.

Die Datenauswertemethode ist abhängig von: (Bortz & Döring, 2006, S. 741)

- der ausgewählten Datenerhebungsmethode / Datenquelle
- dem Skalenniveau der zu untersuchenden Variablen (s. *Tabelle 11*)

Teilbereiche der Statistik: (Bortz & Döring, 2006, S. 741; Duller, 2013, S. 9)

- **Deskriptive Statistik** (beschreibende Statistik): Ausgangspunkt jeder Datenanalyse; Darstellung von Stichprobenergebnissen, Aussagen über unmittelbar vorliegenden Daten (Aufbereitung von Daten in Form von Tabellen und Grafiken, Berechnung statistischer Kennzahlen univariater, bivariater und multivariater Zusammenhänge)
- **Explorative Statistik**²¹: Untersuchung von Daten, um Auffälligkeiten und daraus ableitend Datenstrukturen, möglichen Fragestellungen und Hypothesen zu entdecken bzw. zu generieren; Überprüfungen der Hypothesen finden unter Anwendung der Methoden der induktiven Statistik statt
- **Induktive Statistik** (Inferenzstatistik, schließende Statistik; hypothesenprüfend): Rückschlüsse von einer repräsentativen Stichprobe auf die Grundgesamtheit; Schätzung von Populationsparametern auf Basis von Stichprobenkennwerten und Testen von Populationshypothesen mit Hilfe von Signifikanztest (Wahrscheinlichkeitsaussagen)

²¹ Häufig als EDA, explorative Datenanalyse, bezeichnet

Vorgehensweise der Datenanalyse:

Ausgangslage: Verschiedene Arten von Daten durch die eingesetzten Erhebungsinstrumente

1. **Datenaufbereitung** (je nach Datenquelle unterschiedlich)

z.B. bei Interviewdaten: Transkription²²

2. **Codierung:**

Daten werden in Zahlen transformiert → Eindeutige Zuordnung von Symbolen zu bestimmten Merkmalsausprägungen / qualitativen Informationen → genaue Dokumentation bei repräsentativen Stichproben in Skalenhandbüchern, damit eine Reanalyse mit neuen Methoden gewährleistet ist; Berücksichtigung der Skalenniveaus und Typen von Variablen (Programme: z.B. SPSS / R; Nachteil: Verlust an Detailinformationen)

3. **Datenbereinigung**

z.B. Paper-pencil-Daten

- Fehler- / Plausibilitätsprüfung (z.B. „22“ statt „2“)
- Umgang mit fehlenden Werten (missing values) [Entstehung durch unvollständige Informationsangaben]:

Auffüllen nach MCAR-Test (bei zufällig fehlenden Werten wird in SPSS Algorithmus eingesetzt) (Little & Rubin, 2012; Lüdtke, Robitzsch, Trautwein & Köller, 2007) oder Selektion (falls ganze Seiten o. Abschnitte fehlen) (in SPSS: „MISSING VALUES“ Kommando)

3 Gründe für fehlende Werte: „verweigert“, „weiß nicht“, „keine Ahnung“; höchste Werte werden in Codierung verwendet

- Einspaltige Variablen: 7 („verweigert“), 8 („weiß nicht“), 9 („keine Ahnung“)
- Zweispaltige Variablen: 97 („verweigert“), 98 („weiß nicht“), 99 („keine Ahnung“)
- Dreispaltige Variablen: 997 („verweigert“), 998 („weiß nicht“), 999 („keine Ahnung“)

4. **Datenmodifikation**

= Rekodierungen²³ von Variablen (in SPSS: „RECODE“ Kommando)

- Variablenneubildungen, Ziel: leichteres Datenmanagement, z.B. Umrechnung Geburtsjahr in Alter (in SPSS: „COMPUTE“ Kommando)
- **Bildung von Skalenwerten / Summenscores**
Mehrere Items werden nach Kriterien aggregiert; Vorteil: höhere Messgenauigkeit (in SPSS: neue Variable mit „COMPUTE“ Kommando erzeugen und zum Datensatz hinzufügen)
Theoretisch relevante Variablenneubildungen können auch Typenvariablen, die aus mehreren Merkmalskombinationen mehrerer Variablen gebildet werden (in SPSS: „IF“ Kommando)

²² Transkription: Verschriftlichung auf der Basis von Audioaufzeichnungen

²³ Rekodierung: Kategorien von Variablen werden umdefiniert bzw. zusammengefasst

5. Deskriptive und explorative Statistiken

- Häufigkeitsverteilungen der Variablen – Betrachtung einzelner Variablen als Vorstufe zur Analyse bivariater und multivariater Zusammenhänge (relativ, prozentual, kumuliert²⁴; gruppierte Daten²⁵)

Grafische Darstellung: relationale²⁶ / absolute²⁷ Interpretierbarkeit

- Nichtmetrische Variablen: Säulendiagramm
 - Metrische Variablen: Histogramm (Achtung: Prinzip der Flächentreue!) (Mögliche Verteilungen: symmetrisch /schief, unimodal/bimodal/polymodal, steil/flach)
- Ausreißer und Extremwerte²⁸
 - Identifizieren (z.B. durch Häufigkeitstabellen, Boxplots, Minimum und Maximum, Daten auf Normalverteilung testen bzw. Schiefe der Daten prüfen) und
 - Umgang mit diesen festlegen (abhängig von geplanten weiteren Analysen) z.B. durch robuste Statistiken = spezielle Analysemethoden: problematische Extremwerte werden berücksichtigt und in verzerrender Wirkung neutralisiert) (Erceg-Hurn & Mirosevich 2008) oder z.B. durch Orientierung an Daumenregeln: Residuen²⁹ betrachten, bspw. bei linearer Regression: Keine standardisierten Residuen > 3 und > -3 zulassen; höchstens 1% der standardisierte Residuen $> 2,5$ / $< -2,5$ und nur 5% > 2 / < -2 (Field 2013)

6. Inferenzstatistik (bei Hypothesenprüfung)

Prüfung (statistischer) Hypothesen (Achtung: Skalenniveau)

Nullhypothese: Annahme, Zusammenhang besteht nicht: $H_0: \mu_1 = \mu_2$ (populationsbezogen)

Alternativhypothese: Annahme, Zusammenhang besteht $H_1: \mu_1 \neq \mu_2$

Möglichst repräsentative Stichprobe aus der Grundgesamtheit mit jeweiligen Ausprägungen; ermittelte Mittelwerte als Schätzung der jeweiligen Populationsmittelwerte;

Signifikanztest, z.B. t-Test³⁰ für unabhängige Stichproben → Prüfung, ob Mittelwerte sich bedeutsam voneinander unterscheiden → Signifikanz / Bedeutung wird statistisch durch sog. bedingte Wahrscheinlichkeit $p(\text{Ergebnis}|H_0)$ abgeschätzt

²⁴ Kumulierte Häufigkeitsverteilungen sind erst ab Ordinalskalenniveau sinnvoll zu interpretieren!

²⁵ Gruppierte (klassierte) Daten: Merkmale mit sehr vielen Ausprägungen werden in Gruppen zusammengefasst → Vorteil: übersichtlichere Häufigkeitstabellen, Nachteil: Informationsverlust; Grundregeln bei Klassenbildung: überschneidungsfrei, Vermeidung leerer Klassen, gleiche Klassenbreite (nicht immer sinnvoll); Klassenbreite = Differenz der oberen und unteren Klassengrenze; Klassenmitte = arithmetisches Mittel aus der Klassengrenze

²⁶ Relational interpretierbar: betreffende Daten gewinnen nicht für sich, sondern im Vergleich ihre Bedeutung → interessant: Zeitvergleich und Merkmalszusammenhänge

²⁷ Absolut interpretierbar: betreffende Daten für sich bedeutend; prinzipiell unabhängig vom Erhebungsinstrument; alle Ratioskalen absolut interpretierbar; Variablen, deren Ausprägungen sich eindeutig empirischen Sachverhalten zuordnen lassen, z.B. Religionszugehörigkeit

²⁸ Ausreißer und Extremwerte liegen 2-3 Standardabweichungen über/unter dem Mittelwert der Verteilung einer Variable

²⁹ Residuen: Abweichung vom Modell

³⁰ T-Test: gebräuchlicher Signifikanztest zur Prüfung von Hypothesen über Populationsparameter

Bedeutung/Signifikanz → statistisch durch bedingte Wahrscheinlichkeit $p(\text{Ergebnis} / H_0)$ abgeschätzt; vorgegebene Irrtumswahrscheinlichkeit α meist bei 5% / 1% → „Es kann sich um keinen Irrtum handeln“ → Nullhypothese wird verworfen, Alternativhypothese wird vorläufig angenommen; mehr Gewissheit kann durch mehrere Studien mit verschiedenen Stichproben und verschiedenen Methoden erlangt werden

Exkurs: Ausgewählte quantitative Analysemethoden

Deskriptive Statistik (Bortz & Schuster, 2010; Bühner & Ziegler, 2017; Diaz-Bone, 2019; Duller, 2013; Hartung, Elpelt & Klösener, 2009; Henze, 2018; Holling, 2010; Sedlmeier & Renkewitz, 2018; Stelzl, 2005)

Deskriptive Statistik beschäftigt sich mit der Beschreibung von Merkmalen (univariate Analysemethoden) und Merkmalszusammenhängen (bivariate und multivariate Analysemethoden). Der Exkurs bietet eine Übersicht über gängige Methoden und soll einen ersten Zugang zur deskriptiven Statistik darstellen. Während der Datenanalyse sind weitere Literaturrecherchen sinnvoll.

Univariate Analysemethoden

- Lagemaße: (Erfassung des zentralen Werts von **univariaten Verteilungen**)

- **Modus M**

Definition: Merkmalsausprägung einer Variablen, die am häufigsten auftritt

Voraussetzung: Nominalskalenniveau

[größte Dichte einer Verteilung; bei gruppierten metrischen Werten entspricht Modus dem Mittelwert der häufigsten Klasse]

- **Median \tilde{x}**

Definition: Merkmalsausprägung, die eine geordnete Reihe aus n Messwerten (x_1, x_2, \dots, x_n) in zwei gleiche Teile aufteilt; Vorteil: unempfindlich gegenüber Ausreißern

Voraussetzung: min. Ordinalskalenniveau

Berechnung: Für ungerade n : $\tilde{x} = x_{\frac{n+1}{2}}$

$$\text{Für gerade } n: \tilde{x} = \frac{1}{2} \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right)$$

[Bei Häufigkeitstabellen für ungruppierte Daten entspricht der Median der Ausprägung, bei der die kumulierte relative Häufigkeit erstmals den Wert von 50% überschreitet; bei gruppierten Daten auf Ordinalskalenniveau genauso; bei metrisch gruppierten Daten wird der feinberechnete Median verwendet]

Allgemeine Interpretationsformel zur Ermittlung des feinberechneten Medians (für rel. Häufigkeiten):

$$\tilde{x} = O_{k-1} + \frac{b_k}{p_k} \cdot (0,5 - cp_{k-1})$$

n = Stichprobenumfang (intervallskaliert) [Grundgesamtheit = N]

k = Einfallsklasse; Klasse, in der die kumulierte Häufigkeit 50% erreicht

O = Obergrenze

b = Klassenbereich

p = relative Häufigkeit

cp = kumulierte relative Häufigkeit

[Häufigkeiten sind beliebig oft stückelbar z.B. in vier (Quartile), zehn (Dezite) oder hundert (Perzentile) Teile; allg.: Quantile; Quantilsrechnung analog zur Medianberechnung, nur 0,5 wird entsprechend ersetzt.]

- **Arithmetisches Mittel \bar{x}** (umgangssprachlich: „Mittelwert“)

Definition: Summe der n Messwerte (x_1, x_2, \dots, x_n) , geteilt durch ihre Anzahl; Summe aller Abweichungen der Messwerte von ihrem arithmetisches Mittel ist null ($\sum_{i=1}^n (x_i - \bar{x}) = 0$); Summe aller Abweichungsquadrate vom arithmetisches Mittel ist der minimal mögliche Wert ($\sum_{i=1}^n (x_i - \bar{x})^2 = \text{minimal}$)

Berechnung: $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$

Voraussetzung: min. Intervallskalenniveau

Hinweis: dichotome Merkmale können wie metrische behandelt werden

▪ **Geometrisches Mittel \bar{x}_{geo}**

Definition: n-te Wurzel des Produkts aller n Beobachtungswerte (x_1, x_2, \dots, x_n) einer Variable →
Nutzung: Berechnungen bzgl. Zuwächsen, Wachstumsfaktoren o.ä. Veränderungen im Zeitablauf

Berechnung: $\bar{x}_{geo} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$

Voraussetzung: min. Ratioskalenniveau

• Streuungsmaße: (Erfassung der Variabilität / Heterogenität von **univariaten Verteilungen**)

▪ **Spannweite R (Range)**

Definition: Differenz zwischen dem größten und dem kleinsten Merkmalswert; Nachteil: relativ informationsarm, anfällig für Ausreißer

Voraussetzung: min. Intervallskalenniveau

▪ **Quartilsabstand Q**

Definition: Abstand zwischen 1. und 3. Quartil: $Q = Q_3 - Q_1$; umfasst die mittleren 50% einer geordneten Messwertreihe

→ Veranschaulichung am leichtesten durch grafische Darstellung kumulierter relativer Häufigkeiten

▪ **Box-Plot**

Definition: Visualisierung der Fünf-Punkte-Zusammenfassung einer Verteilung, bestehend aus: $x_{min}, Q_1, \tilde{x}, Q_3, x_{max}$ (Minimum, 1. Quartil, Median, 3. Quartil, Maximum)

Voraussetzung: min. Intervallskalenniveau (Quartilsabstand auch bei ordinalen Daten berechenbar)

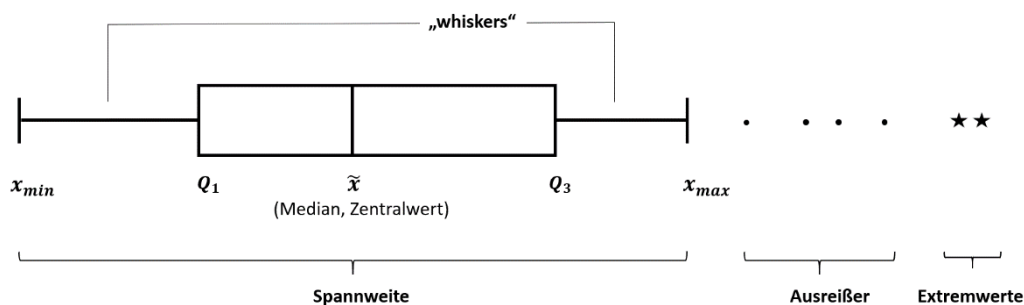


Abbildung 10: Schematische Darstellung eines Box-Plots

[Minimal- und Maximalwerte werden bis max. 1,5 fache der Schachtellänge beim modifizierten Box-Plot eingezeichnet.]

▪ **Varianz s^2**

Definition: Summe der Abstandsquadrate geteilt durch die Anzahl der Messwerte

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

Voraussetzung: min. Intervallskalenniveau

▪ **Standardabweichung s**

Definition: Quadratwurzel aus der Varianz

Berechnung: $s = \sqrt{s^2}$

▪ Berechnung der Varianz und Standardabweichung bei (un)gruppierten Häufigkeitstabellen:

- Ungruppierte Häufigkeitstabellen: $s^2 = \frac{1}{n} \sum_{k=1}^k (x_k - \bar{x})^2$
- Gruppierte Häufigkeitstabellen: $s^2 = \frac{1}{n} \sum_{k=1}^k n_k (m_k - \bar{x})^2$

n = Häufigkeit der Ausprägungen

k = einzelne Ausprägungen

m_k = Klassenmitte der Ausprägungen

▪ **Variationskoeffizient V**

Definition: Quotient aus der Standardabweichung einer Verteilung und ihrem Mittelwert; einheitslose Kennziffer, nur für positive Messwerte definiert

Berechnung: $V = \frac{s}{\bar{x}}$

Hinweis: Variablen mit hohen Mittelwerten weisen oft auch höhere Standardabweichungen auf. Die Streuung wächst mit der Größe der Messwerte. → V hilft davon unabhängig vergleichen zu können.

▪ **Simpsons D**

Definition: Einheitslose Kennziffer zur (vergleichenden) Einschätzung der Heterogenität nominalskalierten Variablen; schwierig zu interpretieren

Berechnung: $D = \frac{K}{K-1} \cdot (1 - \sum_{k=1}^k p_k^2)$

K = Anzahl der Variablenausprägungen

p_k = relative Häufigkeit der Ausprägung K

D = 0, wenn alle Beobachtungswerte in Modalkategorie sind, max. Streuung

D = 1, wenn alle Kategorien die gleiche Häufigkeit aufweisen, vollständige Gleichverteilung, keine Streuung [D=0,5 bei dichotomem Merkmal, D=0,3 bei trichotomem Merkmal, D=0,25 bei Merkmal mit 4 Ausprägungen]

Bivariate Analysemethoden

Um bestmöglich die Informationen der erhobenen Daten auszuschöpfen, bestehen für die Analyse des Zusammenhangs zweier Merkmale für nominale, ordinale und metrisch skalierte Daten jeweils eigens konzipierte Koeffizienten.

Zwei Variablen können hierbei in den Beziehungen zueinanderstehen: Kontingenz, Assoziation und Korrelation.

Statistische Zusammenhänge³¹ werden nach folgenden drei Kriterien unterschieden:

1) **Stärke** (für alle Skalenniveaus geeignet): Je stärker ein Zusammenhang zwischen zwei Variablen, desto weniger Fehler treten bei der Vorhersage der einen durch die andere Variable auf.
Ausprägungen: Zusammenhangslosigkeit / schwacher/mittlerer/starker Zusammenhang / vollständiger Zusammenhang

2) **Richtung** (ab Ordinalskalenniveau geeignet)

Ausprägungen: positiver Zusammenhang.: je größer x, desto größer y /
negativer Zusammenhang.: je größer x, desto kleiner y

3) **Form**

Ausprägungen: monotone und nichtmonotone Zusammenhang.: je, desto – Aussagen (Voraussetzung: min. Ordinalskalenniveau) /

lineare und nichtlineare Zusammenhänge (Voraussetzung: min. Intervallskalenniveau)

Ausgangspunkt der bivariaten Datenanalyse sind bivariate Häufigkeitstabellen. Sie lassen die Betrachtung von Zeilen- und Spaltensummen (=Randverteilungen / marginale Verteilungen) und damit auch von Zeilen- und Spaltenprozenten (=bedingte / konditionale Verteilungen) zu; also die Betrachtung der Verteilung einer Variablen unter verschiedenen Bedingungen, die sich aus Ausprägungen einer anderen Variable ergeben.

Liegen völlig identische bedingte Verteilungen zweier Merkmale vor, sind diese voneinander unabhängig und man spricht von der statistischen Unabhängigkeit.³² Eine Kreuztabelle mit dieser Eigenschaft heißt Indifferenzmatrix.

In Tabelle 12 sind die wichtigsten Zusammenhangsmaße im Überblick aufgeführt. Die Zusammenhangsmaße haben gemeinsam, dass sie normiert (meist zw. 0 und 1), inhaltlich klar interpretierbar und möglichst unabhängig von der Anzahl gebildeter Variablenkategorien sind.

Bezeichnung	Skalenniveau
Prozentsatzdifferenz (d%); Chi ² -basierte Maße (Phi, Cramérs V)	Beide Variablen min. auf Nominalskalenniveau
Konkordanzmaße (Kendalls tau-a und tau-b, Gamma)	Beide Variablen min. auf Ordinalskalenniveau
Pearsons r	Beide Variablen min. auf Intervallskalenniveau

Tabelle 12: Ausgewählte Zusammenhangsmaße im Überblick

³¹ Def. statistischer Zusammenhang = Jede Abweichung von der Indifferenzmatrix

³² Def. Statistische Unabhängigkeit: In einer bivariaten Häufigkeitstabelle liegt statistische Unabhängigkeit vor, wenn alle erwarteten Häufigkeiten mit den beobachteten Häufigkeiten übereinstimmen.

- Zusammenhangsmaße für nominalskalierte Merkmale

→ Aussage nur über Stärke des Zusammenhangs

- **Prozentsatzdifferenz**

Definition: Für die Bestimmung des Zusammenhangs zweier Variablen in Vier-Felder-Tabellen geeignet (=bivariate Häufigkeitstabelle mit 2 dichotomen Merkmalen).

Tabelle 13: Beschriftung Vier-Felder-Tabelle

Berechnung: $d\% = |100 \cdot \left(\frac{a}{a+c} - \frac{b}{b+d}\right)|$

Normierung: 0 (=Zusammenhangslosigkeit) $\leq d\% \leq 100$ (=perfekter Zusammenhang); Einheit in Prozentpunkten

- **Odds-Ratio**

Definition: Kreuzproduktverhältnis in Vier-Felder-Tabellen (s. Tabelle 13) zur Quantifizierung der Stärke der Zusammenhänge; aufbauend auf Proportionen bedingter (relativer) Häufigkeiten; wird auch Odds-bedingte Chancen genannt und wird durch das Verhältnis der beiden Häufigkeiten einer Spalte definiert

Berechnung: $OR = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a \cdot d}{b \cdot c}$

Normierung: Chancenverhältnis

=1, wenn bedingte Chancen gleich sind, also statistische Unabhängigkeit herrscht (Spaltenprozentwerte identisch);

>1, wenn Produkt der Hauptdiagonalen > Produkt Nebendiagonalen

<1, wenn Produkt der Hauptdiagonalen < Produkt Nebendiagonalen

χ^2 -basierte Zusammenhangsmaße

- **Chi-Quadrat-Koeffizient (χ^2)**

Definition: Messung der Stärke des Zusammenhangs zweier Variablen beruhend auf dem Vergleich der empirischen Kreuztabelle (beobachtete Werte n_{ij}) mit der Indifferenzmatrix (Erwartungswerte e_{ij}); abhängig von Fallzahl und Größe der Kontingenztabelle; Nachteil: schwierig interpretierbar; Werte der einzelnen Maße nicht unmittelbar miteinander vergleichbar

Berechnung: $\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$

k = Zeilen

l = Spalten

Normierung: $0 \leq \chi^2 \leq \infty$

0 bei unabhängigen Variablen, > 0 bei abhängigen Variablen

- **Phi-Koeffizient (ϕ)**

Definition: „Punkt-Korrelationskoeffizient“ genannt, weil sein Wert dem linearen Korrelationskoeffizienten für zwei dichotome Merkmale (Vier-Felder-tabelle, s. Tabelle 13) entspricht.

Berechnung: $\phi^2 = \frac{\chi^2}{n} \rightarrow \phi = \sqrt{\frac{\chi^2}{n}} = \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$

Normierung: $0 \leq \phi \leq 1$

ab Ordinalskalenniveau ist das Vorzeichen interpretierbar:

positiv = Hauptdiagonale > Nebendiagonale

negativ = Hauptdiagonale < Nebendiagonale;

0 entspricht statistischer Unabhängigkeit, beobachtete Werte entsprechen den erwarteten:

1 entspricht perfekter Abhängigkeit (dann $\chi^2 = n$), 2 Diagonalzellen der Kreuztabelle sind leer;

(s. Abbildung 11)

▪ Cramérs V

Definition: Zur Messung der Stärke des Zusammenhangs zweier Variablen, bei Mehrfeldertabellen verwendbar

Berechnung: $V = \sqrt{\frac{\chi^2}{\chi_{max}^2}} = \sqrt{\frac{\chi^2}{n \cdot \min(k-1, l-1)}}$

Normierung: $0 \leq V \leq 1$

0 entspricht statistischer Unabhängigkeit, Zusammenhangslosigkeit

1 entspricht perfekter Abhängigkeit, perfektem Zusammenhang

Stärkegrad von Zusammenhängen (für ϕ und V)

0,00 – 0,05	kein bzw. zu vernachlässigender Zusammenhang
0,06 – 0,15	schwacher Zusammenhang
0,16 – 0,35	mittlerer Zusammenhang
> 0,36	starker Zusammenhang

Abbildung 11: Stärkegrad von Zusammenhängen (Phi und Carmérs V)

▪ Lambda λ - Maß der prädikativen Assoziation

Für das Prinzip der prädikativen Assoziation zur Bestimmung der Stärke eines Zusammenhangs wird das PRE-Modell verwendet $\rightarrow \lambda = \text{PRE-Maß}$;

Das Modell der proportionalen Fehlerreduktion (PRE = Proportional Reduction of Error) gibt an, wie stark sich die Vorhersage einer Variable (Y) durch Kenntnis einer anderen Variable (X) verbessern lässt (prozentuale Verkleinerung des Vorhersagefehlers).

Berechnung: $PRE = \frac{E_1 - E_2}{E_1} = 1 - \frac{E_2}{E_1}$

E_1 = Vorhersagefehler ohne Kenntnis von X

E_2 = Vorhersagefehler mit Kenntnis von X

$\lambda_{yx} = 1 - \frac{E_2}{E_1} = 1 - \frac{\sum_{j=1}^l (n_{.j} - \max(n_{ij}))}{n - \max(n_{i.})}$ in Spalte j

λ_{yx} = Zusammenhang zwischen Y (abhängiger Variable) und X (unabhängiger Variable); asymmetrisches Maß (unabhängige Variable = Spaltenvariable) \rightarrow zeilenweise Berechnung

Normierung: $0 \leq PRE \leq 1; 0 \leq \lambda \leq 1$

$\lambda = 0$, wenn Vorhersagefehler nicht reduziert wird (unabhängige Variable besitzt keine Prognosekraft)

$\lambda = 1$, wenn kein Vorhersagefehler auftritt

- Zusammenhangsmaße für ordinalskalierte Merkmale

→ Aussage nur über Stärke und Richtung des Zusammenhangs

Grundlagen:

Zusammenhangsmaße zur Erfassung von Stärke und Richtung heißen Konkordanzmaße. Sie beruhen auf dem Vergleich von Wertepaaren, also Merkmalskombinationen von min. zwei Fällen, die miteinander verglichen werden.

Gesamtzahl der möglichen Paare: $\frac{n(n-1)}{2}$, n = Stichprobenumfang

Positiver Zusammenhang: Je größer der Wert von x, desto größer der Wert von y

Negativer Zusammenhang: Je größer der Wert von x, desto kleiner der Wert von y

Paartypen:

- **Konkordante Paare:** alle Paare stehen in positivem Verhältnis; Häufigkeit des Auftretens = C
- **Diskordante Paare:** alle Paare stehen in negativem Verhältnis; Häufigkeit des Auftretens = D
- **Paare mit Bindungen in X:** Anzahl der Paare mit gleichen x-, untersch. Y-Werten = T_x
- **Paare mit Bindungen in Y:** Anzahl der Paare mit gleichen y-, untersch. x-Werten = T_y
- **Paare mit Bindungen in X und Y:** Anzahl der Paare mit gleichen x- und y-Werten = T_{xy}

- **Kendalls Tau-a (τ_a)**

Berechnung:
$$\tau_a = \frac{C-D}{\frac{n(n-1)}{2}} = \frac{C-D}{C+D+T_x+T_y+T_{xy}}$$

Normierung: $-1 \leq \tau_a \leq 1$, Maximalwerte nur dann, wenn keinerlei Bindungen vorhanden sind

- **Kendalls Tau-b (τ_b)**

Berechnung:
$$\tau_b = \frac{C-D}{\sqrt{(C+D+T_x) \cdot (C+D+T_y)}}$$

Normierung: $-1 \leq \tau_b \leq 1$, Extremwerte nur in symmetrischen Kontingenztafeln erreichbar (gl. Spalten- und Zeilenanzahl)

$\tau_b = 0$, wenn kein Zusammenhang besteht

$-1 \leq \tau_b < 0$, wenn negativer (strikt monotoner) Zusammenhang

$0 < \tau_b \leq 1$, wenn positiver (strikt monotoner) Zusammenhang

- **Kruskals Gamma (γ)**

Verwendung bei größeren als Vier-Felder-Tabellen

Berechnung:
$$\gamma = \frac{C-D}{C+D}$$

Normierung: $-1 \leq \gamma \leq 1$

$\gamma = 0$, wenn kein Zusammenhang besteht

$-1 \leq \gamma < 0$, wenn negativer (schwach monotoner) Zusammenhang

$0 < \gamma \leq 1$, wenn positiver (schwach monotoner) Zusammenhang

PRE-Interpretation von γ

Verwendung zur / zum

- Vorhersage der Rangordnung der Untersuchungseinheiten beim 1. Merkmal ohne Kenntnis der Rangordnung beim 2. Merkmal; $E = 0,5(C + D)$
- Vorhersagefehler mit Kenntnis der Rangordnung beim zweiten Merkmal (überwiegen diskordante Paare, werden alle konkordanten Paare falsch vorhergesagt und

umgekehrt, man richtet sich nach der Rangordnung, die am häufigsten vorkommt;
 $PRE = \frac{E_1 - E_2}{E_1}$ entspricht dem Prozentwert der Paare, die man über reines Raten hinaus richtig vorhersagen kann, wenn man die Rangordnung im zweiten Merkmal berücksichtigt
 z.B.: $PRE = 0,35 \rightarrow$ Die Vorhersage der Variable x verbessert sich um 35%, wenn man die eine weitere Variable heranzieht

▪ Somers d

Asymmetrisches Maß

Berechnung:

$$d_x = \frac{C-D}{C+D+T_x} \quad \text{zur Vorhersage der unabhängigen Variable x (bei Kenntnis y)}$$

$$d_y = \frac{C-D}{C+D+T_y} \quad \text{zur Vorhersage der abhängigen Variable y (bei Kenntnis x)}$$

• Bivariater Zusammenhang metrischer Variablen

\rightarrow Aussage über Stärke und Richtung eines linearen Zusammenhangs

Darstellung: Kreuztabelle (gruppierte Daten), Streudiagramm (ungruppiert)

▪ Kovarianz s_{xy} (durchschnittliche Kreuzproduktsumme)

Berechnung: $s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$

▪ Pearson r

Berechnung: $r = \frac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}}$

Normierung: $-1 \leq r \leq 1$

$r = -1$, perfekte negative Korrelation

$r = 0$, es besteht kein linearer Zusammenhang

$r = 1$, perfekte positive Korrelation

$-1 < r < 0$, positive Korrelation

$0 < r < 1$, negative Korrelation

Konventionen von Cohen (1988) zur Stärke von Korrelationen

um +/- 0,10 schwacher Zusammenhang

um +/- 0,30 mittlerer Zusammenhang

um +/- 0,50 starker Zusammenhang

▪ Darstellung: Lineares Regressionsmodell

Regressionsfunktion: $\hat{y}_i = a + bx_i$, mit $b = \frac{s_{xy}}{s_x^2}$ und $a = \bar{y} - b\bar{x}$

\hat{y}_i = Vorhersagewert für y

Festlegung der eindeutigen Lage der Regressionsgeraden: $e_i = y_i - \hat{y}_i$

\rightarrow Methode der kleinsten Quadrate: $\sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \text{Minimum}$

▪ Regression

Standardisierungen der Variablen \rightarrow mittelwertbereinigen ($\bar{x} = 0; s = 1$), Regressionsgerade verläuft durch den Ursprung des Koordinatensystems

$$z_x = \frac{x - \bar{x}}{s_x}, z_y = \frac{y - \bar{y}}{s_y}, r = s_{z_x z_y} = b_s \quad [-1 \leq b_s \leq 1]$$

▪ **Determinationskoeffizient r^2**

Aussagekraft, wie gut x in der Lage ist, y zu erklären → dies ist besser, je kleiner die Residuen³³

Berechnung:
$$r^2 = \frac{\sum(\hat{y}_i - \bar{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} = PRE = 1 - \frac{E_2}{E_1}$$

$\sum(\hat{y}_i - \bar{y}_i)^2$ = erklärte Varianz

$\sum(y_i - \hat{y}_i)^2 = E_2$ = nicht erklärte Varianz

$\sum(y_i - \bar{y}_i)^2 = E_1$ = Gesamtvarianz

Normierung: $0 \leq r^2 \leq 1$

$r^2 = 1$, perfekte (positive/negative) Korrelation

$r = 0$, fehlende Korrelation (Zusammenhangslosigkeit)

ab $r^2 = 0,25 \rightarrow 25\%$ aufgeklärte Varianz: sehr hoch

PRE-Interpretation des Determinationskoeffizienten

• Zusammenhangsmaß zwischen einer nominalen und einer metrischen Variable

▪ Eta η

Ausgangspunkt: Mittelwert der metrischen Variablen in verschiedenen Gruppen (bedingter Mittelwert); Unterscheiden sich diese vom Gesamtmittelwert, besteht ein statistischer Zusammenhang. Ist kein Unterschied vorhanden, besteht kein statistischer Zusammenhang. Die Gesamtstreuung setzt sich aus der erklärten Streuung (zwischen den Gruppen) und der nicht erklärten Streuung (innerhalb der Gruppen) zusammen.

Erklärte Streuung (zwischen den Gruppen): $\sum_{i=1}^k (\bar{y}_i - \bar{y})^2$ = Abweichungen der Gruppenmittelwerte \bar{y}_i vom Gesamtmittelwert \bar{y}

Nicht erklärte Streuung (innerhalb der Gruppen): $\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2$ = Abweichungen der einzelnen Messwerte vom jeweiligen Gruppenmittelwert \bar{y}_i ;

Unterschiedlichkeit, die nicht auf Gruppenzugehörigkeit zurückgeht: Gesamtstreuung: $\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y})^2$

Anmerkung: Bei dichotomen Variablen ($k=2$) ist $\eta = r$

Berechnung:
$$\eta^2 = \frac{\sum_{i=1}^k n_i \cdot (\bar{y}_i - \bar{y})^2}{\sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y})^2} = \frac{\text{erklärte Streuung}}{\text{Gesamtstreuung}}$$

▪ Eta² als PRE-Maß η^2

Berechnung:
$$\eta^2 = 1 - \frac{E_2}{E_1} = 1 - \frac{\text{nicht erklärte Streuung}}{\text{Gesamtstreuung}} = \frac{\text{erklärte Streuung}}{\text{Gesamtstreuung}}$$

Das PRE-Maß Eta kann zur Vorhersage der abhängigen Variable genutzt werden:

- Vorhersage von y mit Kenntnis von x; Fehler E_2 = Streuung (Varianz) innerhalb der Gruppen
- Vorhersage von y ohne Kenntnis von x; Fehler E_1 = Varianz (Gesamtstreuung von y)

Anmerkung: Bei dichotomen Variablen ($k=2$) ist $\eta = r$

• Zusammenhangsmaße zwischen zwei metrischen Merkmalen: r^2 und η^2 im Vergleich:

R und Eta liefern Hinweise auf die Linearität einer Beziehung zwischen Merkmalen auf der Grundlage bedingter Mittelwerte der „abhängigen“ Variable

- r^2 und η^2 stimmen überein: Beziehung ist linear
- $r^2 < \eta^2$: Beziehung ist nicht linear

³³ Ein Residuum entspricht der Abweichung zwischen vorhergesagtem und tatsächlichem Wert (sog. Vorhersagefehler)

Drittvariablenanalyse

Vorgehen: **Teilgruppenanalyse** zur Drittvariablenkontrolle

1. Erstellung Kausalhypothese
2. Bivariaten Zusammenhang bestimmen
3. Bildung von min. 2 Teilgruppen nach z (Kontrollvariable)
4. Berechnung der teilgruppenspezifischen Zusammenhänge
5. Vergleich der Zusammenhänge in Teilgruppen mit Gesamtgruppe
6. Kausalanalytische Interpretation

Kausalanalytische Interpretationsmuster:

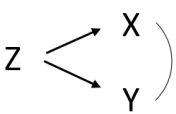
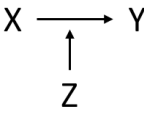
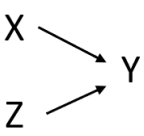
Interpretationsmuster	Erklärung	Statistischer Hinweis	Beispiel (Hypothese + Variablen + kausale Interpretation)
Scheinkausalität 	X und Y haben eine gemeinsame Ursache	Zusammenhang besteht nur in Gesamtheit, tendenziell kein Zusammenhang in Teilgruppen	<i>„Nur wer genügend Einkommen besitzt, kann sich hochkulturelle Freizeitorientierung leisten.“</i> X = Einkommen Y = Hochkulturelle Freizeitorientierung Z = Bildung → „Einkommen und Hochkultur sind nur scheinbar kausal miteinander verbunden. Bildung ist eine starke Hintergrundvariable.“
Intervention / Mediation $X \rightarrow Z \rightarrow Y$	Drittvariable geht Y zeitlich direkt voraus, nicht aber X	Zusammenhang besteht nur in Gesamtheit, tendenziell kein Zusammenhang in Teilgruppen Achtung: Kausalinterpretation kann nur inhaltlich unterschieden werden.	<i>„Frauen sind bei der Verkehrswahl umweltbewusster als Männer.“</i> X = Geschlecht Y = Verkehrsmittelwahl Z = Autobesitz → „Das Geschlecht wirkt sich nicht direkt, sondern über den Umweg des Autobesitzes auf die Verkehrsmittelauswahl aus.“ [Inhaltlich ist die kausale Interpretation des Autobesitzes als gemeinsame Ursache nicht sinnvoll.]
Interaktion / Moderation ³⁴ 	Stärke des Effekts von X und Y abhängig von Drittvariable Z	Deutlicher Unterschied des Zusammenhangs zwischen X und Y in Teilgruppen	<i>„Gewaltbereitschaft gegen Fremde hängt von der Fremdenfeindlichkeit ab.“</i> X = Fremdenfeindlichkeit Y = Gewaltbereitschaft Z = Soziales Umfeld → „Das soziale Umfeld trägt zur (De)Eskalation von Gewalt bei.“
Additive Multi-kausalität 	Additiver Effekt: X und Z haben eigenen Effekt auf Y	Zusammenhänge in Teilgruppen unterscheiden sich nicht wesentlich von dem in Gesamtheit (Kontrollvariablen: Z, X)	<i>„Die Einschätzung der persönlichen wirtschaftlichen Lage hängt von der Furcht vor dem Arbeitsplatzverlusts ab.“</i> X = Furcht vor dem Arbeitsplatzverlust Y = Persönliche wirtschaftliche Lage Z = Allgemeine wirtschaftliche Lage → „Die Einschätzung der persönlichen wirtschaftlichen Lage hängt von der Furcht vor dem Arbeitsplatzverlust und unabhängig davon von der allgemeinen wirtschaftlichen Lage ab.“

Abbildung 12: Kausalanalytische Interpretationsmuster bei der Drittvariablenkontrolle: Teilgruppenanalyse

³⁴ Moderator-Hypothesen können experimentell / korrelativ überprüft werden; korrelativ durch eine hierarchische, moderierte Regression

Partielle Korrelation

Berechnung Partialkorrelation:

$$r_{xy.z} = \frac{r_{xy} - r_{yz} \cdot r_{xz}}{\sqrt{(1 - r_{yz}^2) \cdot (1 - r_{xz}^2)}}$$

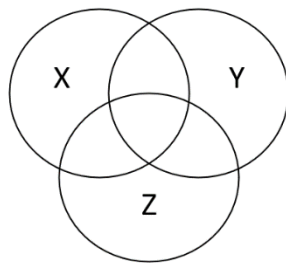
$r_{xy.z}$ = Zusammenhang zwischen x und y unter Einfluss von z

Vorgehensweise:

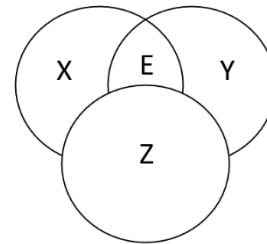
- Datenerhebung (zu allen Variablen, Voraussetzung intervallskaliert)
- Datenanalyse: Partialkorrelation → Statistisches Ausschalten der Kontrollvariable Z³⁵
(= Herausparsialisieren von Z; Anteile von Z auf X und Y werden rausgerechnet)

Mögliche Ergebnisse:

- Bestehender Zusammenhang: $r_{xy.z} \neq 0$ (deutliche Unterscheidung von Null)
→ Kausale Beziehung zwischen X und Y oder
- Verschwinden des Zusammenhangs: $r_{xy.z} = 0$ (gleich Null)
→ Gemeinsame Ursache (Scheinkausalität) / Intervention



Bivariate Korrelation: X, Y und Z korrelieren miteinander; sie haben eine „gemeinsame“ Varianz (alle Überschneidungsbereiche im Venn-Diagramm)



Partielle Korrelation: X und Y korrelieren unter Kontrolle von Z → Formale Schreibweise: $r_{xy.z}$
E = Gemeinsame Varianz von X und Y ohne Einfluss von Z

Auspartialisieren: Regressionsanalytische „Bereinigung“ der Variablen von Y und Z

Regressionsgerade: $\hat{Y} = a + b \cdot Z$

\hat{Y} = vorhergesagter, prognostizierter Wert

a = Konstante

b = Regressionskoeffizient

Residuen (nicht erklärter Anteil): $e_i = Y_i - \hat{Y}$

Y_i = reale Y-Werte

[Residualvariable (=Fehler-/Partialvariable) mathematisch zwingend von Z linear unabhängig]

→ Eine Partialkorrelation ist die bivariate Korrelation zwischen den betreffenden Residualvariablen

³⁵ Soll eine nominalskalierte Variable Z kontrolliert werden, muss diese zuvor „dummy-codiert“ werden (Bortz und Döring, 2006, S. 510f, Box 8.2)

Multiple Regressionsfunktion (Vorhersagegleichung)

$$\hat{Y} = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + b_3 \cdot X_3 + b_m \cdot X_m$$

b_0 =Regressionskonstante,

b_m =Regressionsgewichte/-koeffizienten (Interpretation: Effekt der betreffenden unabhängigen Variablen auf die abhängigen Variablen bei Kontrolle (=Konstanthaltung) aller anderen unabhängigen Variablen; bei additiver Multikausalität verändern sich Regressionkoeffizienten nicht/nur gering; bei Scheinkausalität / Intervention gehen Regressionkoeffizienten gegen Null)

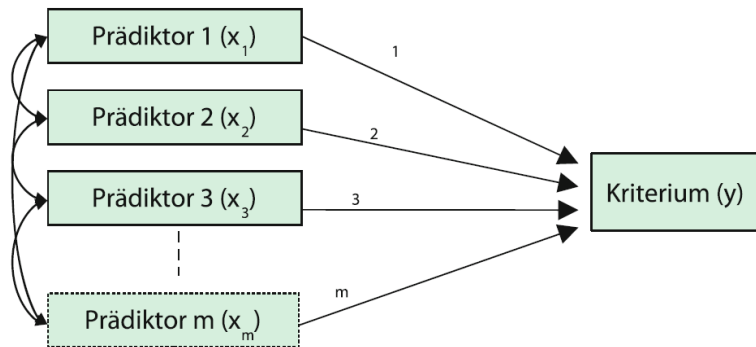


Abbildung 13: Schematische Darstellung einer multiplen Regression (Hussy, 2013, S. 177)

Optimierungskriterium der multiplen Regressionsfunktion:

$$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \text{Minimum}$$

Vorteile der multiplen Regressionsanalyse³⁶

(gegenüber bivariater Korrelationen, diese sind in diesem Fall nicht sinnvoll)

- **Identifikation redundanter Prädiktoren:**

Prädiktoren können untereinander korrelieren → Redundanzen; durch multiple Regressionsanalyse können Prädiktoren identifiziert werden, die substantiellen Beitrag zur Vorhersage des Kriteriums leisten (Bsp. bei Wolfrad & Doll, 2005);

Achtung: Falls Prädiktoren zu hoch interkorrelieren, kann Validität der Regressionsanalyse gefährdet sein (*Problem der Multikollinearität*)

- **Identifikation von Interaktionseffekten:**

Interaktionen zwischen den unterschiedlichen Prädiktoren (auch Interkorrelationen genannt) können in multiplen Regressionsanalysen im Rahmen von Moderator-Hypothesen (→ moderierte Regression geprüft werden (eine bivariate Regression kann dies nicht leisten)

- **Identifikation von Suppressionseffekten:**

Auftretende Suppressoren können in multiplen Regressionsanalysen aufgedeckt werden (im Gegensatz zu bivariaten Korrelationen). Das sind Prädiktorvariablen, die den Vorhersagebeitrag einer / mehrerer anderer Variablen erhöhen, indem sie für die Vorhersage irrelevante Varianzanteile unterdrücken (Bsp. bei Paulhus, Robins, Trzesniewski & Tracy, 2004)

³⁶ Die multiple Regression ist eine lineare Regression mit mehreren Prädiktoren

Regressionsanalyse mit standardisierten Werten

Standardisierung der Variablen X (z-Werte): $z_1 = \frac{x_1 - \bar{x}}{s_{x_1}}$

Für z-Werte gilt: Mittelwert = 0, Standardabweichung = 1

Vorteil: Standardisierte Werte sind unabhängig von Skalierungen und Kodierungen (unterschiedliche Messbereiche)

$$\hat{z}_y = b_1^* * z_1 + b_2^* * z_2 + b_3^* * z_3 + b_m^* * z_m$$

\hat{z}_y = geschätzter standardisierter y-Wert

b_m^* = standardisierte Regressionsgewichte, Vorteil: Vergleichbarkeit untereinander

z_m = standardisierte X-Werte

[Wenn z_1 um 1 Standardabweichung steigt, steigt \hat{z}_y auch um 1 Standardabweichung, wenn alle anderen Werte konstant bleiben.]

Multiple Korrelation & Bestimmtheitsmaß

Multiple Korrelation R: Produkt-Moment-Korrelation der Vorhersagevariable mit der abhängigen Variable

Bestimmtheitsmaß (PRE-interpretierbar): $R^2 = \frac{\text{erklärte Varianz}}{\text{Gesamtvarianz}}$

Weiterführende Literatur:

- Andy Field (2018): Discovering Statistics Using SPSS. (Field, 2018)
- Bös, Hänsel, Schott (2004): Empirische Untersuchungen in der Sportwissenschaft. Planung – Auswertung – Statistik. (Bös et al., 2004)

Exkurs: Induktive Statistik / Inferenzstatistik: (Bortz & Schuster, 2010; Bühner & Ziegler, 2017; Diaz-Bone, 2019; Duller, 2013; Hartung et al., 2009; Henze, 2018; Holling, 2010; Sedlmeier & Renkewitz, 2018; Stelzl, 2005)

Induktive Statistik ermöglicht den Rückschluss von Informationen einer Stichprobe auf die Grundgesamtheit (Abbildung 14), also die Verallgemeinerung von Stichprobenergebnissen mit einem berechenbaren Irrtumsrisiko. Die Methoden basieren auf wahrscheinlichkeitstheoretischen Grundlagen, weshalb eine Auseinandersetzung mit den Grundbegriffen und Denkweisen der Wahrscheinlichkeitsrechnung empfohlen wird (Duller, 2013, S. 159). Der Exkurs soll einen ersten Überblick hierfür liefern.

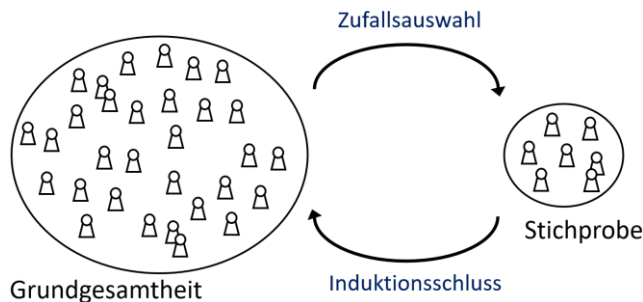


Abbildung 14: Schema Induktive Statistik

Eine tatsächlich realisierte Zufallsauswahl (*Zufallsstichprobe*) der Stichprobe gilt als Voraussetzung für den Induktionsschluss.

Zentrale Begriffe:

- Zufallsexperiment: Eine Handlung, die
 - die (theoretisch) beliebig oft wiederholbar,
 - deren Resultat eines von möglichen Ereignissen (Ereignisraum) und
 - deren Resultat unbekannt ist.
- Wahrscheinlichkeit: Jedem Ereignis (A) eines Zufallsexperiments kann eine Wahrscheinlichkeit $P(A)$ zugeordnet werden. Summe der Wahrscheinlichkeiten aller Ausprägungen = 1.
- Wahrscheinlichkeit eines Ereignisses entspricht der relativen Häufigkeit (=Grenzwert der Wahrscheinlichkeit), mit der das Ereignis bei unendlicher Wiederholung des Zufallsexperiments eintreten würde.
- Zufallsvariable: Variable, deren Werte die Ereignisse eines Zufallsexperiments symbolisieren (diskret / stetig)
- Wahrscheinlichkeitsverteilung: entspricht Häufigkeitsverteilung von Zufallsvariablen
- Kennwerteverteilung/Stichprobenverteilung: Wahrscheinlichkeitsverteilung von Stichprobenkennwerten; wichtiger Verteilungstyp → Wahrscheinlichkeiten berechenbar & Induktionsschlüsse möglich (Schätzen und Testen)

Stichprobenziehung als Zufallsexperiment:

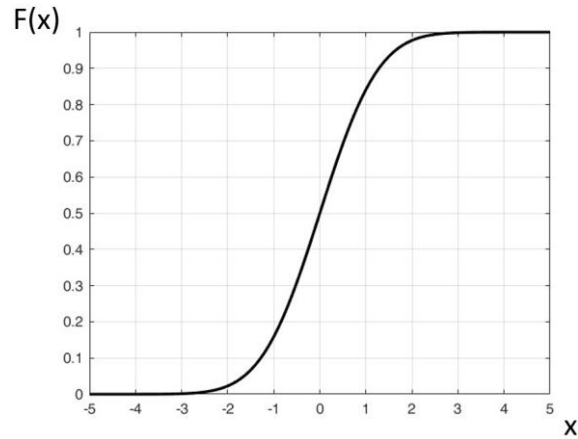
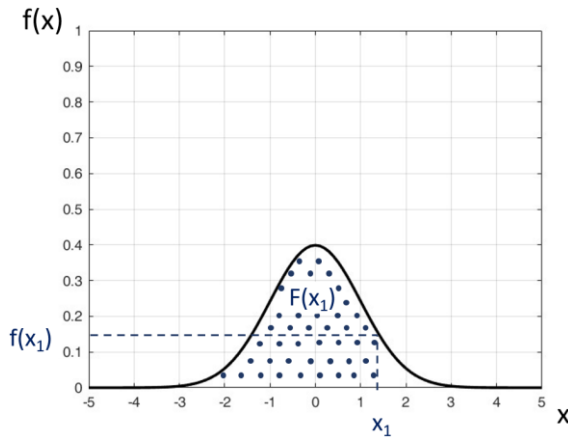
Urnenmodell als Prototyp → Ziehung mit / ohne Zurücklegen (je größer Grundgesamtheit, desto unwichtiger)

Maßzahlen metrischer Zufallsvariablen: [P_k = relative Häufigkeit in der Stichprobe]

- Erwartungswert: Mittelwert einer Zufallsvariablen $\bar{x} = \sum_{k=1}^k P_k \cdot x_k$
- Zufallsvariablen: Mittelwert einer Zufallsvariablen $E(X) = \mu_x = \sum_{k=1}^k P_k \cdot x_k$
- Varianz einer Zufallsvariablen: $VAR(X) = \sigma_x^2 = \sum_{k=1}^k P_k \cdot (x_k - \mu_k)^2$
- Standardabweichung einer Zufallsvariablen: $\sigma_x = \sqrt{\sigma_x^2}$

- Erwartungswert der Stichprobenmittelwerte entsprechen Mittelwert der Grundgesamtheit: $\mu_{\bar{x}} = \mu$
- Standardabweichung des Stichprobenmittelwertes = Funktion der Standardabweichung der Grundgesamtheit und des Stichprobenumfangseiner Zufallsvariablen: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
- Standardabweichung des Mittelwertes = Standardfehler des Mittelwertes (nimmt proportional zur Wurzel von n ab → Wurzel-n-Gesetz)

Dichtefunktion $f(x)$ und Verteilungsfunktion $F(x)$ einer stetigen Zufallsvariable (Normalverteilung)



Keine genauen Wahrscheinlichkeiten;
Gesamtfläche entspricht Wahrscheinlichkeit von 1;
Wahrscheinlichkeiten entsprechen Integral der Flächenanteile der Dichtefunktion

Kumulierte Wahrscheinlichkeitsverteilung:
Wahrscheinlichkeit, dass Zufallsvariable x einen Wert kleiner/gleich x_1 annimmt

Gaußsche Normalverteilung:

Eigenschaften: glockenförmig, symmetrisch um Modalwert/Hochpunkt, asymptotische Annäherung der Verteilung an x-Achse

- Dichtefunktion der Normalverteilung: $f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
 x = Wert der Zufallsvariable X
 μ = Mittelwert (Erwartungswert) der Zufallsvariable (beeinflusst Lage der Verteilung)
 σ = Standardabweichung der Zufallsvariablen (beeinflusst Breite der Verteilung)
 $N(0,1)$; $N(\mu|\sigma)$ Standardnormalverteilung
 → Alle Standardverteilungen lassen sich durch Verschieben/Strecken/Stauchen in eine Standardnormalverteilung überführen → z-Transformation
- Z-Transformation (Umrechnungsformel) $z = \frac{x-\mu}{\sigma}$; für z-Werte gilt: $\mu = 0$; $\sigma = |1|$
 Mittelwertermittlung (Umformung): $z = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \rightarrow \bar{x} = \mu + z \cdot \sigma_{\bar{x}}$
- Kennwerte in Population und Stichprobe

Kennwert	Population	Stichprobe
Arithmetisches Mittel	μ	\bar{x}
Standardabweichung	σ	s
Varianz	σ^2	s^2
Fallzahl	N	n

Abbildung 15: Kennzeichnung von Kennwerten in Stichproben und Population

Verteilungen

• Chi-Quadrat-Verteilung

= Verteilung einer Summe von standardnormalverteilten Zufallsvariablen Z ; einfacher statistischer Test zur Untersuchung von Häufigkeiten

Anwendung: Kreuztabellen

$$\chi^2_{df=n} = Z_1^2 + Z_2^2 + \dots + Z_n^2$$

df = Zahl der Freiheitsgrade (degrees of freedom) bzw. Zahl der voneinander unabhängigen Variablen
Mit zunehmender Anzahl von Freiheitsgraden geht die Chi-Quadrat-Verteilung in die Normalverteilung über (ab df=30)

• T-Verteilung

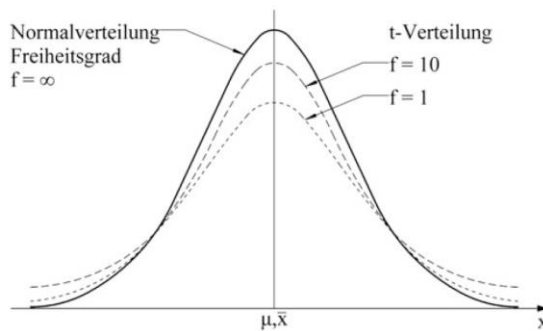


Abbildung 16: Normalverteilung und T-Verteilung (Schiefer & Schiefer, 2018, S. 33)

$$T = \frac{Z}{\sqrt{\frac{\chi^2}{df}}}$$

(aus standardnormalverteilten und chi-quadrat-verteilten Zufallsvariable konstruiert)
Anwendung: Vergleich von Mittelwerten, Signifikanzprüfung von Korrelationskoeffizienten und Regressionskoeffizienten

Mit zunehmender Anzahl von Freiheitsgraden geht die T-Verteilung in die Normalverteilung über (ab df=30); für kleine Stichproben, die eine unbekannte Streuung aufweisen, ungeeignet

• F-Verteilung (Fischer-Verteilung $F(m,n)$)

$$Z = \frac{x}{\frac{m}{y}}, F = \frac{\sigma_1^2}{\sigma_2^2}$$

Anwendung: Vergleich von Streuungen, Varianzen, Regressionsmodelle

Voraussetzung: normalverteilte Grundgesamtheit N

(Schiefer & Schiefer, 2018, ab S. 76)

Weiterführende Literatur:

- Kühnel & Krebs (2014): Statistik für die Sozialwissenschaften. (Kühnel & Krebs, 2014)
- Jann (2005): Einführung in die Statistik. (Jann, 2005)

Schätzung als Induktionsschluss (Verallgemeinerung von Stichprobenergebnissen)

Punktschätzung = Schätzung von Populationswerten mit Hilfe von Schätzern

Die Stichprobenvarianz unterschätzt die Populationsvarianz (bias); bei großen Stichproben kann die Verzerrung vernachlässigt werden

$$\hat{\sigma}_x^2 = \frac{n}{n-1} \cdot s_x^2 \approx s_x^2$$

Intervallschätzung = Angabe eines Wertebereichs, der den Populationswert vermutlich enthält.

Standardfehler des Mittelwertes: $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$

Beispiel:

Aus GG wird einfache Zufallsstichprobe (n=36) gezogen und durchschnittlicher IQ von 112 ermittelt. Gesucht wird durchschnittlicher IQ in der Grundgesamtheit mit gewisser Wahrscheinlichkeit; 1. Standardfehler des Mittelwertes berechnen mit $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$; 2. Festlegung der Irrtumswahrscheinlichkeit z.B. $\alpha = 0,025$ und zugehörigem z-Wert, hier 1,96; 3. Z-Transformation: $z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$; 4. Zusammenführung der Ergebnisse: 116,9 und 107,1

Konfidenzintervall: Symmetrischer Wertebereich um einen Stichprobenkennwert, der Populationswert mit bestimmter Wahrscheinlichkeit enthält (Konfidenzniveau = $1 - \alpha$)

- Konfidenzintervall des Mittelwertes bei bekannter Varianz der Populationswerte

$$\bar{X} \pm z_{(1-\frac{\alpha}{2})} \cdot \sigma_{\bar{x}} = \bar{X} \pm z_{(1-\frac{\alpha}{2})} \cdot \frac{\sigma_x}{\sqrt{n}}$$

Standardfehler des Mittelwertes: $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$

- Konfidenzintervall des Mittelwertes bei unbekannter Varianz der Populationswerte (Normalfall)

$$\bar{X} \pm z_{(1-\frac{\alpha}{2})} \cdot \hat{\sigma}_{\bar{x}} = \bar{X} \pm z_{(1-\frac{\alpha}{2})} \cdot \frac{\hat{\sigma}_x}{\sqrt{n}} = \bar{X} \pm z_{(1-\frac{\alpha}{2})} \cdot \frac{s_x}{\sqrt{n-1}}$$

Standardfehler des Mittelwertes muss aus Stichprobe geschätzt werden: $\hat{\sigma}_{\bar{x}} = \frac{s_x}{\sqrt{n-1}} \approx \frac{s_x}{\sqrt{n}}$

Irrtumswahrscheinlichkeit α : Wahrscheinlichkeit, dass Populationswert nicht im Konfidenzintervall liegt

Übliche Werte für Konfidenzniveau / Irrtumswahrscheinlichkeit:

Konfidenzniveau ($1-\alpha$)	Irrtumswahrscheinlichkeit (α)	$\alpha/2$	$z_{(1-\frac{\alpha}{2})}$
0,90	0,10	0,05	+1,64
0,95	0,05	0,025	+1,96
0,99	0,01	0,005	+2,58

Schätzung von Anteilswerten

- Konfidenzintervall für den Anteilswert

$$p \pm z_{(1-\frac{\alpha}{2})} \cdot \sigma_p = p \pm z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$$

Standardfehler der relativen Häufigkeit (Schätzung, wenn Anteilswert in Grundgesamtheit unbekannt ist):

$$\hat{\sigma}_p = \sqrt{\frac{p \cdot (1-p)}{n}}$$

Voraussetzung: Kennwertverteilung des Anteilswertes ist normalverteilt

- Bestimmung des benötigten Stichprobenumfangs

- Mittelwerte:
$$n = \frac{4 \cdot z^2 \left(1 - \frac{\alpha}{2}\right) \cdot \sigma_{\bar{x}}^2}{d^2}$$
- Anteilswerte:
$$n = \frac{4 \cdot z^2 \left(1 - \frac{\alpha}{2}\right) \cdot \pi(1-\pi)}{d^2}$$

d = Breite des Konfidenzintervalls

Mit wachsendem n nimmt die Genauigkeit einer Schätzung zu (Konfidenzintervall werden – bei konstantem α – kleiner) und die Irrtumswahrscheinlichkeit ab (bei konstanter Intervallbreite). Die Genauigkeit nimmt nicht proportional zu n zu.

Allgemeine Struktur von Konfidenzintervallen:

$$\text{Schätzer} \pm \left(1 - \frac{\alpha}{2}\right) - \text{Quantilwert der Kennwerteverteilung} \cdot \text{Standardfehler}$$

- Anwendung der Normalverteilung (Bortz, 1999, S. 78; Kühnel und Krebs, 2001, Kap. 8)

Schätzung von Mittelwerten: X ist normalverteilt oder die Stichprobe ist größer als n = 30

Schätzung von Anteilswerten: $n \cdot p \cdot (1 - p) > 9$

p = Anteilswert der Stichprobe

Sollzustand bei Schätzung des Anteilswertes: $n > 60$

Methodik von Signifikanztests

Grundfrage: *Sind Unterschiede beim Induktionsschluss von Stichprobenergebnissen auf die Grundgesamtheit zufällig entstanden oder nicht?*

Ablauf von Signifikanztests

- Hypothesenformulierung (Nullhypothese & Alternativhypothese³⁷ schließen sich aus)
- Bestimmung der Teststatistik und -verteilung
 - Prüfgröße = Zu testender Parameter der Grundgesamtheit (steht in H_0/H_1); z.B. Mittelwert
 - Teststatistik = Größe, die zur Durchführung eines Signifikanztests berechnet wird (enthält relevante Prüfgröße); z.B.)
$$Z = \frac{\bar{x} - \mu}{\frac{\sigma_{\bar{x}}}{\sqrt{n}}} = \frac{\bar{x} - \mu}{\frac{\sigma_{\bar{x}}}{\sqrt{n-1}}} \approx \frac{\bar{x} - \mu}{\frac{s_{\bar{x}}}{\sqrt{n}}}$$
 - Testverteilung = Kennwertverteilung der Teststatistik; z.B. Annähernde Standardnormalverteilung der Teststatistik Z für $n \geq 30$
- Festlegung der Irrtumswahrscheinlichkeiten³⁸ (Signifikanzniveau)

Übliche Signifikanzniveaus:

$\alpha > 0,05$	(nicht signifikant)
$\alpha \leq 0,05$	(signifikant)
$\alpha \leq 0,01$	(sehr signifikant)
$\alpha \leq 0,001$	(hoch signifikant)

Ablehnungsbereich (kritischer Bereich) beginnt mit dem Wert der Teststatistik, der gerade noch mit der Nullhypothese zu vereinbaren ist.

³⁷ Alternativhypothese entspricht der Forschungshypothese; die Nullhypothese wird in dessen Abhängigkeit erstellt

³⁸ Irrtumswahrscheinlichkeit / Signifikanzniveau entspricht der Wahrscheinlichkeit, dass die Nullhypothese abgelehnt wird, obwohl sie in Wirklichkeit zutrifft

4. Berechnung der Teststatistik
5. Entscheidung über Annahme / Ablehnung der Nullhypothese

Alternative zu Schritten 3-5: Berechnung des empirischen Signifikanzniveaus

Empirisches Signifikanzniveau p

= Wahrscheinlichkeit, dass Teststatistik – bei Gültigkeit von H_0 – den berechneten/noch „extremere“ Wert aufweist; das entspricht dem Signifikanzniveau, auf dem die Nullhypothese abgelehnt werden kann; häufig in Statistikprogramm berechnet

Je kleiner p ausfällt, desto signifikanter ein Testergebnis

Hypothesenarten: Einseitig gerichtete und zweiseitig ungerichtete
Hypothesenpaare zum Populationsmittelwert

- | | | |
|---------------------------|-----------------------|---------------------------------------|
| (1) $H_0: \mu_x = \mu$ | $H_1: \mu_x \neq \mu$ | → zweiseitige, ungerichtete Hypothese |
| (2) $H_0: \mu_x \leq \mu$ | $H_1: \mu_x > \mu$ | → einseitige, gerichtete Hypothese |
| (3) $H_0: \mu_x \geq \mu$ | $H_1: \mu_x < \mu$ | → einseitige, gerichtete Hypothese |

μ_x = tatsächlicher Populationswert eines Merkmals X (Schluss aus Stichprobe)
 μ = zu testender Mittelwert (hypothetisch festgelegt)

Fehlerarten bei statistischen Entscheidungen:

- α – Fehler (Fehler 1. Art): zufällige Ablehnung der Nullhypothese
- β – Fehler (Fehler 2. Art): zufällige Bestätigung der Nullhypothese

		In der Population gilt die:	
		H_0	H_1
Entscheidungen aufgrund der Stichprobe zugunsten der:	H_0	Richtige Entscheidung	β -Fehler
	H_1	α -Fehler	Richtige Entscheidung

Abbildung 17: α - und β -Fehler bei statistischen Entscheidungen (Bortz & Döring, 2006, S. 498)

Weiterführende Literatur:

(Diekmann, 2018), Kühnel und Krebs, 2001, Kap. 9

Ausgewählte Signifikanztests (Hussy et al., 2013, 179ff)

- Tests für Mittelwerte
- Test eines Populationsmittelwertes (eine Gruppe)
Unterscheidet sich ein Stichprobenmittelwert überzufällig von einem vermuteten Populationsmittelwert?
- Mittelwertvergleich zwischen zwei Gruppen (Mittelwertdifferenztest, t-Test)
Unterscheiden sich die Stichprobenmittelwerte zweier Gruppen überzufällig?
- Mittelwertvergleich: mehr als zwei Gruppen (Varianzanalyse)
Unterscheiden sich Gruppenmittelwerte insgesamt überzufällig?
- Tests zur Prüfung auf Unabhängigkeit
- Tests zur Prüfung Verteilung
- Tests zur Prüfung auf Streuung
- Tests zur Prüfung von Zusammenhängen

Hinweis:

Die Tests können eingeteilt werden nach parametrischen und nichtparametrischen Verfahren, also der Unterscheidung, ob die Verteilung normalverteilt ist (parametrisch) oder nicht. (*StatistikGuru*)

Anmerkungen zu Signifikanztests

- Signifikanz stark abhängig von der Stichprobengröße
- Nicht-Signifikanz eines Testergebnisses ist kein sicherer Nachweis, dass die Nullhypothese zutrifft
- Statistische Signifikanz und Relevanz sind zu unterscheiden!
[Relevante Bereiche (aus Fachwissen) werden zunächst festgelegt; trotz signifikantem Testergebnis kann das Ergebnis außerhalb dieses Bereichs liegen und somit nicht relevant]
- Testspezifikationen sind vor der Durchführung des Tests festzulegen (zwei-/einseitig)

t-Test

Beachten: Unabhängige / abhängige Stichproben (Ab hier für unabhängige Stichproben³⁹)

Hypothesenpaare zum Populationsmittelwert

- | | | |
|-----------------------------|-------------------------|---------------------------------------|
| (1) $H_0: \mu_1 = \mu_2$ | $H_1: \mu_1 \neq \mu_2$ | → zweiseitige, ungerichtete Hypothese |
| (2) $H_0: \mu_1 \leq \mu_2$ | $H_1: \mu_1 > \mu_2$ | → einseitige, gerichtete Hypothese |
| (3) $H_0: \mu_1 \geq \mu_2$ | $H_1: \mu_1 < \mu_2$ | → einseitige, gerichtete Hypothese |

Teststatistik für den t-Test

- $n < 30$ in beiden Stichproben → T-Verteilung wird genutzt

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}} \text{ mit [Standardfehler der Mittelwertdifferenz: } \hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} \text{]}$$

- $n \geq 30$ in beiden Stichproben → Teststatistik annähernd normalverteilt

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

Verwendung bei unbekannter Varianz: geschätzter Standardfehler

³⁹ Unabhängige Stichproben: Auswahl der Elemente einer Stichprobe (/Teilgruppe) erfolgt unabhängig von Auswahl der Elemente einer anderen Stichprobe.

- Ablehnungsbereiche: Nullhypothese wird abgelehnt, wenn z
 - bei einseitigen Tests: $z_{1-\alpha}$ kritischen Wert überschreitet ($H_1: \mu_1 \leq \mu_2$)
 - bei einseitigen Tests: $z_{1-\alpha}$ kritischen Wert unterschreitet ($H_1: \mu_1 \geq \mu_2$)
 - bei zweiseitigen Tests: $|z_{1-\alpha/2}|$ kritischen Wert überschreitet ($H_1: \mu_1 \neq \mu_2$)

Bei einseitigen Tests wird die Nullhypothese viel leichter verworfen als bei zweiseitigen Tests.
 [z.B.: Kritischer Wert liegt bei Irrtumswahrscheinlichkeit von $\alpha=0,05 \rightarrow z\text{-Wert}=1,1 \rightarrow 86,5\%$ der Fläche der Standardnormalverteilung \rightarrow bei einseitigem Test: $\alpha=13,5\%$, bei zweiseitigem Test: $\alpha/2=13,5\%$]

Test eines Anteilswert in der Grundgesamtheit

Hypothesen

- | | | |
|---------------------------|-----------------------|---|
| (1) $H_0: \pi_x = \pi$ | $H_1: \pi_x \neq \pi$ | \rightarrow zweiseitige, ungerichtete Hypothese |
| (2) $H_0: \pi_x \leq \pi$ | $H_1: \pi_x > \pi$ | \rightarrow einseitige, gerichtete Hypothese |
| (3) $H_0: \pi_x \geq \pi$ | $H_1: \pi_x < \pi$ | \rightarrow einseitige, gerichtete Hypothese |

π = Anteilswert (zu testender/hypothetischer Anteilswert; Ausgangsgröße)

π_x = tatsächlicher Populationsanteil (Schluss aus Stichprobe)

Teststatistik

$$Z = \frac{p_x - \pi}{\sqrt{\frac{\pi \cdot (1-\pi)}{n}}}$$

p_x = Stichprobenanteil (bisherige Annahme) standardnormalverteilt; Voraussetzung: $n \cdot p \cdot (1-p) > 9$

Kritische Werte und Ablehnungsbereiche

- | | | | |
|------------------------------|------|------------------------------|---|
| (1) $Z < \frac{z_\alpha}{2}$ | oder | $Z > z_{1-\frac{\alpha}{2}}$ | \rightarrow zweiseitige, ungerichtete Hypothese |
| (2) $Z > z_{1-\alpha}$ | | | \rightarrow einseitige, gerichtete Hypothese |
| (3) $Z < z_{1-\alpha}$ | | | \rightarrow einseitige, gerichtete Hypothese |

Test des Korrelationskoeffizienten

= Prüfung der Hypothese, dass in der Grundgesamtheit kein linearer Zusammenhang zwischen zwei metrischen Variablen besteht (Korrelationskoeffizient „Rho“ = 0)

Hypothesen

$H_0: \rho = 0$ $H_1: \rho \neq 0$

Teststatistik

$$Z = r \cdot \sqrt{\frac{n-2}{1-r^2}}$$

r = bivariater Korrelationskoeffizient (Stichprobe)

Stichprobenverteilung:

Bei kleinen Stichproben folgt die Teststatistik einer t-Verteilung; bei großen Stichproben (> 30 Fälle) ist Z annähernd standardnormalverteilt

Empirisches Signifikanzniveau: $\alpha < 0,000$

Chi-Quadrat-Unabhängigkeitstest

= breit anwendbarer Test auf Unabhängigkeit zweier (nominalskaliertes) Merkmale
 Datengrundlage: Bivariate Kreuztabelle mit $k \times l$ Zellen (Stichprobenergebnisse)

Hypothesen (nur ungerichtet)⁴⁰:

H_0 : Es besteht kein Zusammenhang

H_1 : Es besteht ein Zusammenhang

Teststatistik

Entspricht dem Chi-Quadrat-Koeffizienten (*Chi-Quadrat-Koeffizient* (χ^2))

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

Die Teststatistik folgt annähernd einer Chi-Quadrat-Verteilung mit $(k-1)(l-1)$ Freiheitsgrade (df).

Die Nullhypothese wird abgelehnt, falls $\chi^2 > \chi_{df, 1-\alpha}^2$

Signifikanz ist nicht gleich Relevanz (Jeder noch so irrelevanter Zusammenhang wird signifikant, wenn der Stichprobenumfang nur groß genug ist).

Empirische Signifikanz: $p < 0,005$

Regression: Overall-F-Test

= Test des multiplen Korrelationskoeffizienten: Geprüft wird die Nullhypothese; dass das Gesamtmodell in der Grundgesamtheit keine Erklärungskraft hat (der multiple Korrelationskoeffizient weicht nicht signifikant von Null ab)

Hypothesen:

$H_0: \beta_j = 0$ für alle j

$H_1: \beta_j \neq 0$ für mindestens ein j

Teststatistik und ihre Verteilung

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-J-1}{J}$$

J = Anzahl der unabhängigen Variablen

[Unter H_0 folgt F einer $F(m, n)$ -Verteilung mit $m=J$ und $n=n-J-1$ Freiheitsgraden]

Test der Regressionsgewichte - t-Test des Regressionskoeffizienten

= Test der einzelnen Regressionskoeffizienten einer Regressionsgleichung, ob diese jeweils in der Grundgesamtheit eine Wirkung auf die abhängige Variable ausüben oder nicht (signifikant von Null abweichen oder nicht)

$$\hat{Y} = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + b_3 \cdot X_3 + b_m \cdot X_m$$

Hypothesen:

$H_0: \beta_j = 0$

$H_1: \beta_j \neq 0$

→ Der zur Variable X_j zugehörige Regressionskoeffizient β_j ist in der Grundgesamtheit gleich 0 bzw. unterscheidet sich in der Grundgesamtheit von 0

Teststatistik und ihre Verteilung

$$T = \frac{\beta_j}{\sigma_{\beta_j}}$$

σ_{β_j} = Standardfehler des Regressionskoeffizienten j

Unter H_0 folgt T einer t -Verteilung mit $df=n-j-1$ Freiheitsgraden; bei großen Stichproben ($df > 30$) folgt T einer Standardnormalverteilung

⁴⁰ Gerichtete Hypothesen sind beim Chi-Quadrat-Unabhängigkeitstest nicht möglich.

Weiterführende Literatur:

- Andy Field, Discovering Statistics Using SPSS, 5. Auflage, Sage 2018. (Field, 2018)
- Literaturempfehlungen zur Datenanalyse: Übersicht zu verschiedenen Tests zur Überprüfung von statistischen Tests: Entscheidungsbaum von Vorberg und Blankenberger (s. http://www.psych.uni-halle.de/abteilungen/allgemeine_psychologie/mitarbeiter/blankenberg/entscheidungsbaum/)

Hilfsmittel für die statistische Auswertung

Bei der Auswahl einer geeigneten Methode, um die erfassten Daten statistisch auszuwerten, hilft der Statistikbaum.

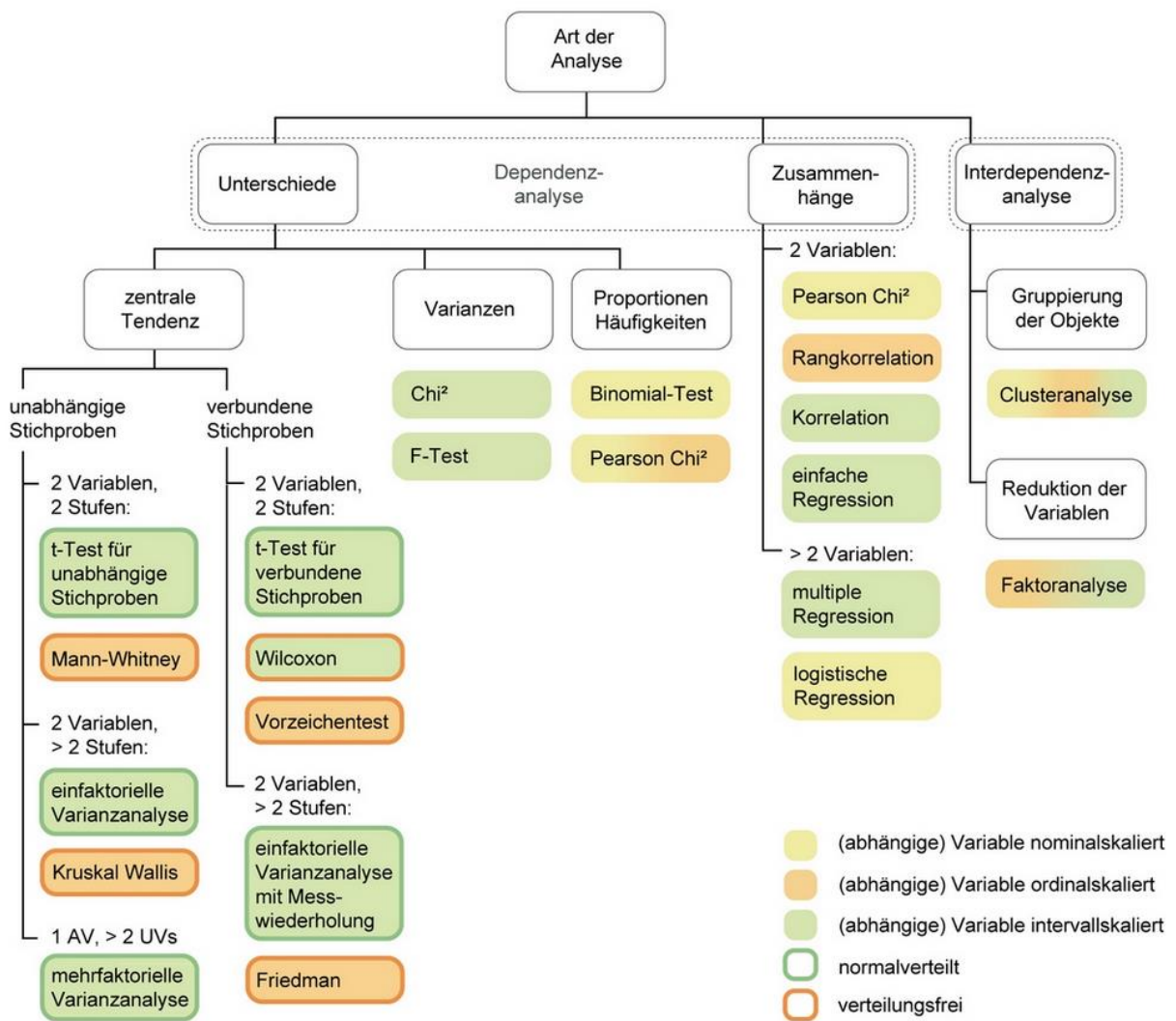


Abbildung 18: Statistikbaum (Methodenberatung Uni Zürich; <https://www.methodenberatung.uzh.ch/de.html>)

Statistik-Software:

- SPSS (= Statistical Package for Social Scientists)
- Mplus (für z.B. Strukturgleichungsmodelle geeignet; s. www.statmodel.com)
- LISREL (= Linear Structural RELations; s. www.ssicentral.com)
- AMOS (integriert in SPSS)
- R (OpenSource; s. www.r-project.org)

Weiterführende Literatur:

Statistikseiten, um alles Wichtige nochmals genau nachzulesen; besonders empfehlenswert: Beispielsätze, wie Ergebnisse in Veröffentlichungen integriert werden.

- StatistikGuru
- Methodenberatung Uni Zürich

Oberflächlicher, aber ein guter Überblick: <https://etools.fernuni.ch/entscheidungsbaum/>

2.3.2 Interpretation & Diskussion der Ergebnisse***Aufgabe des Kapitels: Was erwartet mich hier?***

Die Ergebnisse der Auswertung werden zuletzt diskutiert und bewertet. Wie man hierbei vorgeht und was zu beachten ist, wird näher betrachtet. Empfehlung: Abstand nach Beendigung der Untersuchung nehmen und dann nochmals mit „offenen“ Augen einen Blick auf alles werfen.

Übersicht über zu behandelte Themen in der Ergebnisdiskussion:

1. Verbale Übersetzung der statistischen Tests
2. Interpretation der Befunde im Kontext des theoretischen Bezugsrahmen
3. Vergleich eigener Befunde mit Ergebnissen aus vorangegangenen Untersuchungen
Interpretation von Abweichungen zu einerseits eigenen Hypothesen, andererseits anderen Studien; notwendige Klärung: einzelne Schritte der Untersuchung sowie mögliche Einflüsse verschiedene Bedingungs- und Störfaktoren
4. Vorschläge zur Weiterentwicklung des methodischen Vorgehens + Modifikation des verwendeten theoretischen Bezugsrahmen + Hinweise zu weitführender Forschung

Inhaltlich zu diskutieren:

- Besonderheiten freiwilliger Untersuchungsteilnahme (Bortz & Döring, 2006, S. 74)
→ *Resultate durch Verweigerungen möglicherweise verfälscht?*

Bewertungskriterien wissenschaftlicher Untersuchungen (Bortz & Döring, 2006, S. 40)

Anhand der Gütekriterien (s.

Gütekriterien empirischer Untersuchungen, Tabelle 6)

- Signifikanzniveau der Auswertungen: *Auf welcher Grundlage werden Hypothesen bestätigt/verworfen?*
- Kontrollierte Störfaktoren: *Werden viele potenzielle Stör- bzw. Einflussfaktoren berücksichtigt?*
- Externe Bewertungskriterien / -tabellen zur Orientierung

2.4 Stichwortartige Zusammenfassung

Idealtypischer Ablauf einer empirischen Untersuchung

Aufgabe des Kapitels: Was erwartet mich hier?

Was chronologisch verlaufend in der empirischen Untersuchung beachtet werden muss, wird hier kurz und grob zusammengefasst. Dies kann als Art Checkliste zur Kontrolle vor Beginn einer Untersuchung sowie danach als Gedankenstütze genutzt werden.

Empfehlung: Zu Beginn individuell mit themenspezifischen Anforderungen ergänzen.

1. Forschungsfragestellung

- Bedingungen der Wahl einer Fragestellung:
 - Motivation
 - Verfügbarkeit der Forschungsressourcen
 - Möglichkeit der Publikation
- Recherche: bereits theoretische Ansätze, empirische Studien (lesen + nachdenken)

2. Theorie & Hypothesen

- Anlehnung an bestehende Theorien → Forschungslücke vorhanden, Mehrwert klären
- Literaturrecherche (Empfehlung: aktuelle Übersichtsarbeiten zum Thema, danach empirische Artikel im ausgewählten Thema)
- Arten von Hypothesen (Bortz & Schuster, 2010; Bunge, 2012b; Hussy et al., 2013)
 - Unterscheidungshypothese
 - Zusammenhangshypothese
 - Veränderungshypothese
 - Hypothesen in Einzelfalluntersuchungen
 - Aggregathypothesen (Aussagen über Eigenschaften einer Population; Kontrast zu Einzelfallhypothesen)

3. Operationalisierung und Untersuchungsplanung

= Festlegung von Beobachtungs- und Messvorschriften zur Erfassung von Indikatoren für hypothetische Konstrukte

- Achtung: Immer verschiedene Möglichkeiten:
 - verschiedene Datenquellen
 - verschiedene Messinstrumente → unbestimmte und offene Operationalisierung eines Konstrukts
 - verschiedene Möglichkeiten der Hypothesenuntersuchung
- Standardisierungen bevorzugen (Brähler & Brickenkamp, 2002)
(Vorteile: Vergleiche möglich, Gütekriterien bereits nachgewiesen, Aufwand einer Konstruktion neuer diagnostischer Instrumente ersparen), falls nicht möglich: Adaptierung des Vorhandenen
- unumgänglich: Bestimmung psychometrische Gütekriterien bei Entwicklung von Forschungsinstrumenten, idealerweise in Vorstudien (DGPs und BDP, 2004, S. 59)

4. Durchführung und Datenerhebung

- Vortests / Pilotstudien zur Erprobung und Modifizierung der geplanten Versuchsordnung und sämtliche Messinstrumente

- Achtung: Genaue Planung und sorgfältige Umsetzung der empirischen Untersuchung (in Versuchs- / Untersuchungsplan festhalten)
 - **Ethische** und **rechtliche** Aspekte der geplanten Studie
 - **Probanden:** Hypothesenabhängige Auswahl? Welche Population soll abgebildet werden?
Bedingungen:
 - Zufallsauswahl
 - Repräsentativität der Stichprobe → aufwändig und kostenpflichtig

[Gelegenheitsstichproben möglich: Zusammenstellung der Probanden nach Verfügbarkeit; Teils Generalisierungen dabei möglich, z.B. bei alltagspsychologischen Themen: Wahrnehmung, Gedächtnis, Lernen oder allgemeinen biophysiologicalen Themen]
 - **Treatment:** Untersuchung der Wirkung von Treatments (z.B. Trainingsmaßnahmen, Reizkonstellationen oder situative Bedingungen)
Systematische Variation von Treatments; mögliche Wirkung von Störgrößen muss kontrolliert werden oder die willkürliche Manipulation (der unabhängigen Variablen) ausgeschlossen werden; falls dies nicht möglich ist: korrelative Designs (Kausalinterpretation nicht / nur teils möglich);
Beachtung und Dokumentation von Störvariablen und Störfaktoren (insb. bei experimentellen Untersuchungen) → möglichst identischer Ablauf der Studiendurchführung pro Proband (für die Vergleichbarkeit & zusammenfassende Betrachtung und Analyse);
Ein standardisierter Ablauf nach „Drehbuch“ ist notwendig; *Wer sagt, macht und führt wann wie was durch und wann werden Daten wie erhoben?*
 - **Setting:** Ort / Kontext der Untersuchung mit physikalischen, räumlichen, ökologischen und sonstigen Bedingungen; Potenzieller Einfluss auf die Durchführung und Datenerhebung
 - Unterscheidung: Labor / Feld
 - Einfluss auf die Datenerhebung: Studien mit Messäquivalenz (Kubinger 2009)
 - Vorbereitung der Versuchsräume, -material, technische Hilfsmittel (Videokamera)
 - **Schulung** (+ggf. Nachschulung) der Versuchsleiter
 - Sicherstellung und genaue **Dokumentation** der Datenerhebung
 - **Zeitlicher Ablauf:**
 - Erhebung von Daten
 - Durchführung und Treatments
 - 1 oder mehrere Zeitpunkte der Datenerhebung → abhängig von Fragestellung
 - Auswahl **statistischer Verfahren** zur Analyse der Daten

5. Datenanalyse und Hypothesenprüfung

6. Interpretation und Diskussion

1. Verbale Übersetzung der statistischen Tests
2. Interpretation der Befunde im Kontext des theoretischen Bezugsrahmen
3. Vergleich eigener Befunde mit Ergebnissen aus vorangegangenen Untersuchungen; Interpretation von Abweichungen zu einerseits eigenen Hypothesen, andererseits anderen Studien; notwendige Klärung: einzelne Schritte der Untersuchung sowie mögliche Einflüsse verschiedener Bedingungs- und Störfaktoren
4. Vorschläge zur Weiterentwicklung des methodischen Vorgehens, Modifikation des verwendeten theoretischen Bezugsrahmen sowie Hinweise zu weiteführender Forschung

ANHANG.

Anhang A: Checkliste

Checkliste zur Einordnung und Klärung wichtiger Randbedingungen

Klärungsbedarf	Erklärung	Eintragen	Geklärt
Verantwortliche	Wer trägt die Verantwortung?		<input type="checkbox"/>
Ansprechpartner	Sparingpartner o.Ä. zur Unterstützung		<input type="checkbox"/>
Stakeholder	Industriebeteiligung? Einflussnehmer auf Inhalt, Form, Methode und Ziele der Untersuchung → Unterschiedliche / einheitliche Zielvorstellungen?		<input type="checkbox"/>
Beteiligung weiterer Forschungseinrichtungen	Verteilung von Aufgaben, Interessen und Rollen?		<input type="checkbox"/>
Forschungsphase	Vorstudie / Hauptstudie /... → Resultierende Unterschiede?		<input type="checkbox"/>
Erfahrungshintergrund	Wie viele Vorkenntnisse im Themenfeld und in der Forschungsmethodik sind bereits vorhanden; Wie viel Vorarbeit für die Untersuchung ist notwendig?		<input type="checkbox"/>
Ressourcenbedarf	Wie viel Zeit, Mitarbeiter und weitere Kapazitäten stehen für die Untersuchung zur Verfügung?		<input type="checkbox"/>
Angestrebtes Ergebnis	Klare Zielformulierung bereits vorhanden? Formulieren!		<input type="checkbox"/>
Publikationsabsicht	Werden Ergebnisse in Journals oder Konferenz veröffentlicht?		<input type="checkbox"/>

Tabelle 14: Checkliste Randbedingungen

„Hallo zusammen, ich bin ... und freue mich sehr, dass du / ihr euch entschieden habt, mich in meiner Forschung als Studienteilnehmer zu unterstützen.“

Vorab: Ist es für alle in Ordnung, wenn wir uns duzen? (*Bestätigen lassen*) Sehr gut.

Ich werde gleich kurz meine Forschung vorstellen, welchen Beitrag diese Studie dabei hat einordnen und euch den Ablauf der Untersuchung erklären.

Zunächst setze ich dich/euch hiermit in Kenntnis, dass du/ihr auf freiwilliger Basis hier anwesend bist/seid und den Versuch jederzeit ohne negativen Konsequenzen abbrechen kannst/könnt.

Um das zu bestätigen, bitte ich dich/euch diesen Vertrag nochmals genau durchzulesen und zu unterschreiben. Damit gibst/gibt du/ihr dein/euer Einverständnis und es kann losgehen.“

Verträge verteilen, ausfüllen lassen und bedanken:

„Das hat ja super geklappt. Vielen Dank dafür. Dann kann es nun ja weitergehen.“

Aufklärung über Forschungs- und Untersuchungsziel:

„In meiner Forschung geht es um Hierbei hilft mir die Studie zu untersuchen, wie *
Das Besondere an der heutigen Untersuchung ist“

*Bei Täuschungen Aufklärung über Zweck der Studie weglassen und darüber informieren, dass nach Erhebungszeitraum die Möglichkeit besteht, sich informieren zu lassen. Falls dies erwünscht ist, soll bestmöglich die E-Mail-Adresse und der Wunsch benachrichtigt zu werden auf dem Vertrag angegeben werden. Nach dem Erhebungszeitraum sollte dementsprechend eine E-Mail mit den wichtigsten Informationen zum Zweck und dem Ergebnis der Studie verschickt werden.

Ablauf der Untersuchung:

„Kommen wir nun zum Ablauf der Untersuchung. Die Untersuchung wird insgesamt circa ... min gehen.“

Klären:

- Aufgabenstellung für den/die ProbandIn
- Was wird gemessen während der Untersuchung?
- Wird Ergebnis oder Leistung der/des ProbandinEn bewertet?
(Achtung: Kann Ergebnisse je nach Fragestellung verfälschen)
- Aufnahme von Ton oder Bild? → Für welche Zwecke werden Aufnahmen verwendet?
- Welche Funktion hat VersuchsleiterIn während Untersuchung?
(z.B. nur als Ansprechpartner bei Fragen; Achtung: Potenzielle Störvariable)
- Wann am besten Fragen stellen?
- Notfallszenario: Was soll Proband machen, wenn etwas schief geht oder sie/er sich nicht wohl fühlt? (z.B. „Stopp“ als Signalwort einführen)
- Vergütung, falls vorhanden (Achtung: Potenzielle Störvariable)

**Fragen zulassen:**

„Gibt es bis dahin Fragen?“

Antwortmöglichkeiten vorbereiten – mögliche Fragen denkbar zu:

- Dauer der Untersuchung
- Anonymität & Datenschutz
- Veröffentlichung der Ergebnisse
- Vergütung für Probanden
- Kooperationspartner

Verabschiedung:

„Wir sind nun auch am Ende der Untersuchung angekommen.
Vielen Dank für deine/eure Teilnahme!“

Falls relevant:

„Wenn du/ihr angegeben hast/habt, dass du über den Zweck und die Ergebnisse der Untersuchung informiert werden willst/wollt, wirst/werdet du/ihr ab dem ... von mir entsprechend per E-Mail hören.“

Falls relevant (z.B. bei subjektiven Einschätzungen):

„Zur Einschätzung der Qualität der Daten und der Erhebungsmethode(n) selbst, würde ich dich /euch bitten, diesen kurzen Fragebogen auszufüllen, um mir **Feedback zur Untersuchung** zu geben. Wenn du/ihr magst/mögt, kannst/könnt du/ihr mir anschließend kurz erzählen, was unverständlich war oder wobei du/ihr Probleme hattet.“

Ausfüllen lassen (Vorab Fragebogen zur Untersuchung vorbereiten)

Notizen:

Anhang C: Informationspflicht Versuchsteilnehmer

Vertrag zwischen Versuchsleiter/in und Versuchsteilnehmer/in

(→ in Anlehnung an (Hussy et al., 2013, S. 46))

Versuchsteilnehmer ID: _____

Liebe/r Untersuchungsteilnehmer/in,

es folgt eine Information über deine Rechte und Pflichten als Versuchsperson sowie über unsere Verpflichtungen als Versuchsleiter/in:

1. Freiwilligkeit der Teilnahme

Du hast dich freiwillig dafür entschieden, an dieser Untersuchung teilzunehmen. Wir danken dir sehr für diese Bereitschaft, mit deiner Mitarbeit zum Zustandekommen unserer Untersuchung beizutragen.

2. Schutz vor Schädigung und Belastung

Außerdem möchten wir versichern, dass wir uns bemühen, dir in diesem Experiment keine körperlichen oder seelischen Belastungen zuzumuten und dass das Wohlergehen der Versuchsteilnehmer/innen Vorrang vor reinem Interesse an signifikanten Ergebnissen hat.

3. Abbruchrecht

Sollten nun entgegen unseren Bemühungen während des Versuchs Belastungen auftreten, die du für zu schwerwiegend erachtest, so hast du die Möglichkeit, die Untersuchung abzubrechen. Aufgrund des Abbruchs musst du aber keinerlei negative Konsequenzen befürchten.

4. Potenzielle Täuschung bzw. unvollständige Information [*Abschnitt nur bei Täuschung nutzen*]

In diesem Experiment ist es aus methodischen Gründen leider nicht möglich, die TeilnehmerInnen vollständig vor der Durchführung des Experiments über die Ziele und Vorgehensweise aufzuklären.

5. Recht auf postexperimentelle Aufklärung

Wir versichern, dass nach Abschluss der Erhebungsphase der einzelnen Experimente auf Nachfrage sämtliche gewünschten Informationen über Ablauf, Zweck und Ergebnis des Versuchs gegeben werden können (Hierfür bitte Telefonnummer u.o. E-Mail-Adresse angeben: _____)

6. Pflichten als Versuchsperson

Diesen geschilderten Rechten jeder Versuchsperson stehen auch nachvollziehbare Pflichten gegenüber. Die Planung und Durchführung einer Untersuchung erfordert viel Zeit und Mühe. Es ist deshalb wichtig, dass du versuchst, die Aufgabenstellungen der Untersuchung so gut wie möglich zu erfüllen. Dazu gehört, dass du offen und ehrlich auf die Fragen antwortest und den Versuch ernst nimmst. Eine uninteressierte und oberflächliche Mitarbeit gefährdet die Erreichung der Untersuchungsziele erheblich, und es wäre sehr schade, wenn dadurch die ganze Arbeit, die in die Vorbereitung des Experiments gesteckt worden ist, umsonst gewesen wäre.

7. Gewährleistung der Anonymität

Abschließend wollen wir darauf hinweisen, dass deine Antworten streng vertraulich behandelt werden.

Wir danken dir noch einmal sehr für deine Teilnahme!

Dieser Vertrag wurde gelesen und zur Kenntnis genommen:

Unterschrift der Versuchsleiter/innen

Unterschrift der Versuchsperson

Anhang D:

Exkurs: Möglichkeiten zur Generierung von Theorien und Hypothesen

Dieser Exkurs ist als Inspiration zu empfehlen, wenn Schwierigkeiten bei der Findung von Theorien und Hypothesen auftreten. Falls diese bereits formuliert sind, kann der Exkurs mit gutem Gewissen beim Lesen übersprungen werden.

Im Stand der Forschung wird die Generierung von Theorien oder Hypothesen selten berücksichtigt. Gigerenzer, Pierce und Walach bieten hierbei methodische Anhaltspunkte. McGuire (1997) wird eine besondere Rolle zugesprochen, insbesondere in den psychologischen Forschungsmethoden, mit seinen 49 Heuristiken. Im Folgenden sind diese und weitere Möglichkeiten angeführt, die zur Anregung für die Generierung von Hypothesen und Theorien dienen sollen.

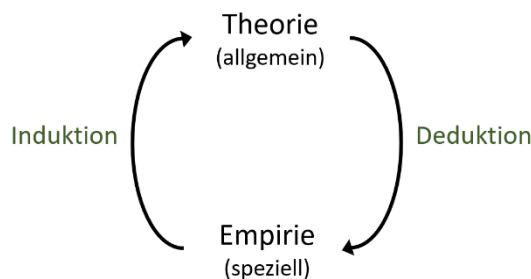


Abbildung 19: Prinzipien der Induktion und Deduktion

- **Induktion:** Schlussfolgerung aus Einzelfällen (Empirie) auf das Allgemeine (Theorie)

Beispiel

Fall: Diese Bohnen sind aus dem Sack. Resultat: Diese Bohnen sind weiß.
→ Regel: Alle Bohnen aus diesem Sack sind weiß.

- **Deduktion:** Ableitung des Besonderen aus dem Allgemeinen (Regeln, Gesetzmäßigkeiten, Modelle, Theorien)

Regel: Alle Bohnen aus dem Sack sind weiß. Fall: Diese Bohnen sind aus dem Sack.
→ Resultat: Diese Bohnen sind weiß.

- **Abduktion:** Vom Allgemeinen und Beschreibenden wird ein einzelner Fall in bestehende Kategorien eingeteilt.

Regel: Alle Bohnen aus diesem Sack sind weiß. Resultat: Diese Bohnen sind weiß.
→ Fall: Diese Bohnen sind aus diesem Sack.

- **Erklärungsmodell: Hempel & Oppenheim (1948):**

Struktur deduktiv-nomologischer Erklärungen (DN-/HO-Schema):

Explanans: allg. Gesetze (/ Hypothesen / theoretische Annahmen) und Antezedenzbedingungen

Explanandum: Beschreibung, logische Ableitung; Adäquatheitsbedingungen müssen korrekt sein: logisch, min. 1 allg. Gesetz, empirischer Gehalt im Explanans, gut bewährt (Hempel, 1977)

- **Humphreys Modell probabilistischer Kausalerklärungen (1989):**

Erklärungsmodell multikausaler Bedingtheit; Vorhersagen lediglich nach Wahrscheinlichkeit möglich

- (McGuire, 1997): **49 Heuristiken**⁴¹ zur Generierung von Theorien und Hypothesen, welche in folgende Kategorien zur Klassifikation eingeteilt sind:
 - **Sensitivität gegenüber ungewöhnlichen oder provozierenden Phänomenen** (Heuristics simply calling for sensitivity to provoke natural occurrences)
z.B. Ausnahmefälle in bisherigen Theorien, Grounded-Theory-Ansatz: Intensive Beobachtung u./o. Interviews von Einzelfällen ohne Vorannahmen und Erwartungen
 - **Einfache konzeptionelle Analysen** (Heuristics involving simple conceptual analysis)
z.B. Kausalitätsrichtung bestehender Hypothese umdrehen, Gedankenexperimente
 - **Komplexe konzeptionelle Analysen** (Heuristics calling for complex conceptual analysis)
z.B. Analogiebildung - „Tools-to-Theories-Heuristik (Gigerenzer 1988): wissenschaftliche Werkzeuge werden auf wissenschaftliche Fragestellungen übertragen, bspw. menschlicher Geist wird in Analogie zum Computer als informationsverarbeitendes System konzipiert; Varianzanalyse als Analogie (z.B. Attributionstheorie - Harold Kelley)
 - **Reinterpretation bestehender Forschungsergebnisse** (Heuristics demanding reinterpretations of past research)
z.B. Widersprüche in Ergebnissen verschiedener empirischer Studien, bereits gegebenen psychologischen Kenntnissen
 - **Datensammlung und Reanalyse bestehender Datensätze** (Heuristics necessitating collecting new or reanalyzing old data)
z.B. Übersicht bestehender Studien schaffen
- **Kreativitätsmethoden**
[Hinweis: Wenn die Hypothesengenerierung als Problemlösungsprozess betrachtet wird, kann der SPALTEN-Prozess beim Sortieren der einzelnen Schritte helfen]
- Geeignete **Randbedingungen** schaffen: z.B. intensive Beschäftigung mit einem Thema, dann Entspannung
- **Einflussfaktoren** „manipulieren / stimulieren / anpassen“: Person, Prozess, Umfeld: Persönlichkeit, Situation, Motivation, Gruppe

Weiterführende Literatur:

(Gigerenzer, 1988, S. 91–100; Walach & Stillfried, 2013)

⁴¹ Heuristiken: Findestrategien

Literaturverzeichnis

- Akremiti, L., Baur, N. & Fromm, S. (2011). *Datenanalyse mit SPSS für Fortgeschrittene 1. Datenaufbereitung und uni- und bivariate Statistik* (3., überarbeitete und erweiterte Auflage). Wiesbaden: VS Verlag für Sozialwissenschaften / Springer Fachmedien Wiesbaden GmbH Wiesbaden. <https://doi.org/10.1007/978-3-531-93041-1>
- Asch, S. E. (1952). *Social psychology*. New York: Prentice Hall.
- Asendorpf, J. B. (2000). Idiographische und nomothetische Ansätze in der Psychologie. *Zeitschrift für Psychologie / Journal of Psychology*, 208(1-2), 72–90. <https://doi.org/10.1026//0044-3409.208.12.72>
- Berlyne, D. E. (1974). *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation* (Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation). Oxford, England: Hemisphere.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-33306-7>
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (Springer-Lehrbuch, 7., vollständig überarbeitete und erweiterte Auflage). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-12770-0>
- Brähler, E. & Brickenkamp, R. (Hrsg.). (2002). *Brickenkamp-Handbuch psychologischer und pädagogischer Tests* (3., vollst. überarb. und erw. Aufl.). Göttingen: Hogrefe Verl. für Psychologie.
- Bühner, M. & Ziegler, M. (2017). *Statistik für Psychologen und Sozialwissenschaftler* (2., aktualisierte und erweiterte Auflage). Hallbergmoos: Pearson.
- Bunge, M. (2012a). *Scientific research i*. [Place of publication not identified]: Springer-Verlag Berlin An.
- Bunge, M. (2012b). *Scientific research i*. [Place of publication not identified]: Springer-Verlag Berlin An.
- Campbell, D. T. & Stanley, J. C. (1967). *Experimental and quasi-experimental designs for research* (2. print; Reprinted from "Handbook of research on teaching"). Boston: Houghton Mifflin Comp.
- Conrad, E. & Maul, T. (1981). *Introduction to experimental psychology*. New York: Wiley.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation. Design & analysis issues for field settings*. Boston: Houghton Mifflin.
- Cronbach, L. J. & Shapiro, K. (1982). *Designing evaluations of educational and social programs* (A Joint publication in the Jossey-Bass series in social and behavioral science & in higher education, 1st ed.). San Francisco: Jossey-Bass.
- Dennis, S. & Kintsch, W. (2007). Evaluating Theories. In *Critical thinking in psychology* (S. 143–159). New York, NY, US: Cambridge University Press.
- Deutsche Vereinigung für Sportwissenschaft e.V. (2003). *Berufsethische Grundsätze für Sportwissenschaftl/innen*.
- DGPs und BDP. (2004). *Revision der auf die Forschung bezogenen ethischen Richtlinien*.
- Diaz-Bone, R. (2019). *Statistik für Soziologen* (UTB Sozialwissenschaften, 5., überarbeitete Auflage).
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-41089-5>

- Duller, C. (2013). *Einführung in die Statistik mit EXCEL und SPSS. Ein anwendungsorientiertes Lehr- und Arbeitsbuch* (Springer-Lehrbuch, 3., überarb. Aufl. 2013). Berlin, Heidelberg, s.l.: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-37859-1>
- Eckard Kämper. (2016). *Risiken sozialwissenschaftlicher Forschung ? Forschungsethik, Datenschutz und Schutz von Persönlichkeitsrechten in den Sozial- und Verhaltenswissenschaften*. RatSWD Working Paper, No. 255.
- Flick, U. (2016). *Sozialforschung. Methoden und Anwendungen : ein Überblick für die BA-Studiengänge* (Rororo Rowohlt's Enzyklopädie, Bd. 55702, 3. Auflage). Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.
- Garnefeld, I. (2009). *Kundenbindung durch Weiterempfehlung. Eine experimentelle Untersuchung der Wirkung positiver Kundenempfehlungen auf die Bindung des Empfehlenden* (Gabler Edition Wissenschaft Angewandte Marketingforschung, 1. Aufl.). Zugl.: Paderborn, Univ., Diss., 2008, u.d. T.: Garnefeld, Ina: Kundenempfehlungen als Quelle der Kundenbindung. Wiesbaden: Gabler Verlag / GWV Fachverlage GmbH Wiesbaden. <https://doi.org/10.1007/978-3-8349-9945-0>
- Gigerenzer, G. (1981). *Messung und Modellbildung in der Psychologie* (Uni-Taschenbücher Psychiatrie, Pädagogik, Soziologie, Psychiatrie, Bd. 1047). München: Reinhardt.
- Gigerenzer, G. (2015). *Das Einmaleins der Skepsis. Über den richtigen Umgang mit Zahlen und Risiken* (M. Zillgitt, Übers.). München: Piper.
- Häder, M. (2009). Der Datenschutz in Den Sozialwissenschaften. *RatSWD Working Paper 90/2009*.
- Häder, M. (2019). *Empirische Sozialforschung. Eine Einführung* (Lehrbuch, 4. Auflage).
- Häder, M. & Häder, S. Stichprobenziehung in der quantitativen Sozialforschung. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 333–348).
- Hartung, J., Elpelt, B. & Klöser, K.-H. (2009). *Statistik. Lehr- und Handbuch der angewandten Statistik ; [mit zahlreichen durchgerechneten Beispielen]* (15., überarb. und wesentlich erw. Aufl.). München: Oldenbourg. <https://doi.org/10.1524/9783486710540>
- Henze, N. (2018). *Stochastik für Einsteiger*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-22044-0>
- Holling, H. (Hrsg.). (2010). *Handbuch Statistik, Methoden und Evaluation* (Handbuch der Psychologie, Bd. 13). Göttingen: Hogrefe. Verfügbar unter: <http://elibrary.hogrefe.de/9783840918483/U1>
- Huber, O. (2019). *Das psychologische Experiment. Eine Einführung : mit fünfundfünzig Cartoons aus der Feder des Autors* (7., überarbeitete Auflage).
- Hussy, W. & Jain, A. (2002). *Experimentelle Hypothesenprüfung in der Psychologie*. Göttingen: Hogrefe Verlag für Psychologie.
- Hussy, W., Schreier, M. & Echterhoff, G. (2013). *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (Springer-Lehrbuch, 2., überarbeitete Auflage). Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-34362-9>
- Kubinger, K. D. & Ebenhöf, J. (1996). *Arbeitshaltungen : kurze Testbatterie ; Anspruchsniveau, Frustrationstoleranz, Leistungsmotivation, Impulsivität, Reflexivität ; Manual*. Swets Test Services. Verfügbar unter: <https://books.google.de/books?id=SiYASwAACAAJ>
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (Grundlagen Psychologie, 6. Auflage). Weinheim: Beltz. Verfügbar unter: http://www.content-select.com/index.php?id=bib_view&ean=9783621278454
- Little, R. J. A. & Rubin, D. B. (2012). *Statistical analysis with missing data* (3. rev. ed.). Chichester, West Sussex: Wiley Blackwell.

- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen. In *Psychologische Rundschau* (58 (2007)).
- McGuire.. Creative Hypothesis Generating in Psychology: Some Useful Heuristics. *Annual Review of Psychology*, 1997(48), 1–30.
- Mees, U. & Bolstad, O. D. (Hrsg.). (1977). *Verhaltensbeobachtung und Verhaltensmodifikation. Anwendungsmöglichkeiten im pädagogischen Bereich* (1. Aufl.). Stuttgart: Klett.
- Nachtigall, C., Kroehne, U., Funke, F. & Steyer, R. (2003). Pros and cons of structural equation modeling.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17(11), 776–783. <https://doi.org/10.1037/h0043424>
- Orth, B. (1974). *Einführung in die Theorie des Messens* (Studientext). Stuttgart: Kohlhammer.
- Paulhus, D. L., Robins, R. W., Trzesniewski, K. H. & Tracy, J. L. (2004). Two Replicable Suppressor Situations in Personality Research. *Multivariate Behavioral Research*, 39(2), 303–328. https://doi.org/10.1207/s15327906mbr3902_7
- Pawlik, K. (Hrsg.). (2006). *Handbuch Psychologie. Wissenschaft, Anwendung, Berufsfelder ; mit 46 Tabellen*. Heidelberg: Springer. Verfügbar unter: <http://www.loc.gov/catdir/enhancements/fy0826/2006436319-d.html>
- Rat für Sozial- und Wirtschaftsdaten. (2017). *Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften*.
- Renner, K.-H. (2002). *Selbstinterpretation und Self-Modeling bei Redeängstlichkeit* (Schriftenreihe Lehr- und Forschungstexte Psychologie, N.F.,11). Teilw. zugl.: Bamberg, Univ., Diss., 2000. Göttingen: Hogrefe Verl. für Psychologie.
- Ross, Jones & Nisbett. (1972). *Ross; Jones; Nisbett*.
- Rushton, J. P., Brainerd, C. J. & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin*, 94(1), 18–38. <https://doi.org/10.1037/0033-2909.94.1.18>
- Sarris, V. & Reiß, S. (2005). *Kurzer Leitfaden der Experimentalpsychologie* (Pearson Studium ps psychologie). München: Pearson-Studium.
- Sedlmeier, P. & Renkewitz, F. (2018). *Forschungsmethoden und Statistik. Für Psychologen und Sozialwissenschaftler* (3., aktualisierte und erweiterte Auflage). Hallbergmoos: Pearson. Verfügbar unter: <https://ebookcentral.proquest.com/lib/gbv/detail.action?docID=5583851>
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, Calif.: Wadsworth Cengage Learning.
- Sixtl, F. (1993). *Der Mythos des Mittelwertes. Neue Methodenlehre der Statistik*. München: Oldenbourg.
- Stangl. (2020). *Kohorteneffekt*. Verfügbar unter: <https://lexikon.stangl.eu/3930/kohorte-kohorteneffekt/>
- Steinhoff, L. (2014). *Loyalitätswirkung des geschenkten bevorzugten Kundenstatus. Eine theoretische und empirisch-experimentelle Analyse* (Applied Marketing Science / Angewandte Marketingforschung). Zugl.: Paderborn, Univ., Diss., 2013. Wiesbaden: Springer Gabler. <https://doi.org/10.1007/978-3-658-06357-3>
- Stelzl, I. (2005). *Fehler und Fallen der Statistik. Für Psychologen, Pädagogen und Sozialwissenschaftler* (Standardwerke aus Psychologie und Pädagogik, Reprints, Band 1). Münster, New York, München, Berlin: Waxmann.

- Steyer, R. & Eid, M. (2001). *Messen und Testen. Mit Übungen und Lösungen ; mit 20 Tab* (Springer-Lehrbuch, 2., korrigierte Aufl.). Berlin: Springer.
- VDI. (2002). *Ethische Grundsätze des Ingenieurberufs*. Zugriff am 22.11.2020. Verfügbar unter: https://www.vdi.de/fileadmin/pages/mein_vdi/redakteure/publikationen/VDI_Ethische_Grundsätze.pdf
- WMA Generalversammlung. (2013). *WMA Deklaration von Helsinki - Ethische Grundsätze für die medizinische Forschung am Menschen*.
- Wolfrad, U. & Doll, J. (2005). Geschlecht und Persönlichkeit als Prädiktoren der Internetnutzung. In K.-H. Renner (Hrsg.), *Internet und Persönlichkeit. Differentiell-psychologische und diagnostische Aspekte der Internetnutzung* (Internet und Psychologie, Bd. 8). Göttingen: Hogrefe.