

Studienskript



FORSCHUNGSMETHODIK UND DATENANALYSE

DLBWPFUD01

iu

INTERNATIONALE
HOCHSCHULE

FORSCHUNGSMETHODIK UND DATENANALYSE

IMPRESSUM

Herausgeber:
IU Internationale Hochschule GmbH
IU International University of Applied Sciences
Juri-Gagarin-Ring 152
D-99084 Erfurt

Postanschrift:
Albert-Proeller-Straße 15-19
D-86675 Buchdorf
media@iu.org
www.iu.de

DLBWPFD01
Versionsnr.: 001-2023-0901
N.N.

© 2023 IU Internationale Hochschule GmbH
Dieses Lernskript ist urheberrechtlich geschützt. Alle Rechte vorbehalten.
Dieses Lernskript darf in jeglicher Form ohne vorherige schriftliche Genehmigung der
IU Internationale Hochschule GmbH (im Folgenden „IU“) nicht reproduziert und/oder
unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet wer-
den.

Die Autor:innen/Herausgeber:innen haben sich nach bestem Wissen und Gewissen
bemüht, die Urheber:innen und Quellen der verwendeten Abbildungen zu bestimmen.
Sollte es dennoch zu irrtümlichen Angaben gekommen sein, bitten wir um eine dement-
sprechende Nachricht.

INHALTSVERZEICHNIS

FORSCHUNGSMETHODIK UND DATENANALYSE

Einleitung

Wegweiser durch das Studienskript	6
Literaturempfehlungen	7
Übergeordnete Lernziele	8

Lektion 1

Psychologie als empirische Wissenschaft	9
1.1 Was sind empirische Methoden?	11
1.2 Empirie in der psychologischen Forschung	12
1.3 Theorien und Modelle in der empirischen Psychologie	14
1.4 Forschung als Verbindung von Empirie und Theorie	18
1.5 Unterschiede zwischen qualitativen und quantitativen Methoden	20
1.6 Der Verlauf empirischer Forschungsprozesse und dessen Darstellung in Berichten	28
1.7 Gütekriterien empirischer Forschung	31
1.8 Moralische und ethische Aspekte empirischer Forschung	35
1.9 Themenübergreifende Erkenntnisziele der empirischen Psychologie	38

Lektion 2

Datenerhebung	41
2.1 Überblick zu den empirischen Methoden der Datenerhebung	42
2.2 Qualitative Interviews	45
2.3 Quantitative Umfragen mittels Fragebögen	51
2.4 Beobachtungen als Methode der Datensammlung	55
2.5 Experimente und Versuchspläne	62
2.6 Datenerhebungsverfahren in der kognitiven Psychologie	66
2.7 Psychophysiologische und neurowissenschaftliche Methoden	73
2.8 Der Einsatz standardisierter psychologischer Tests	75

Lektion 3

Datenaufbereitung: Qualitätskontrolle und Optimierung der gewonnenen Daten für die Auswertung	79
3.1 Bedeutung und Ziele der Datenaufbereitung	81
3.2 Schritte der Aufbereitung von quantitativen Daten	82
3.3 Schritte der Aufbereitung von qualitativen Daten	86

Lektion 4

Datenanalyse: Ausgewählte Auswertungsmethoden 91

4.1 Überblick zu den Verfahren der quantitativen Datenanalyse 93

4.2 Deskriptive Statistik und explorative Datenanalyse 95

4.3 Inferenzstatistik 110

4.4 Multivariate statistische Analysemethoden 123

4.5 Überblick zu den qualitativen Formen der Datenanalyse 134

4.6 Qualitative Text- bzw. Inhaltsanalyse von verbalen Daten 136

Verzeichnisse

Literaturverzeichnis 144

Abbildungsverzeichnis 151

EINLEITUNG

HERZLICH WILLKOMMEN

WEGWEISER DURCH DAS STUDIENSKRIPT

Dieses Studienskript bildet die Grundlage Ihres Kurses. Ergänzend zum Studienskript stehen Ihnen weitere Medien aus unserer Online-Bibliothek sowie Videos zur Verfügung, mit deren Hilfe Sie sich Ihren individuellen Lern-Mix zusammenstellen können. Auf diese Weise können Sie sich den Stoff in Ihrem eigenen Tempo aneignen und dabei auf lernspezifische Anforderungen Rücksicht nehmen.

Die Inhalte sind nach didaktischen Kriterien in Lektionen aufgeteilt, wobei jede Lektion aus mehreren Lernzyklen besteht. Jeder Lernzyklus enthält jeweils nur einen neuen inhaltlichen Schwerpunkt. So können Sie neuen Lernstoff schnell und effektiv zu Ihrem bereits vorhandenen Wissen hinzufügen.

In der IU Learn App befinden sich am Ende eines jeden Lernzyklus die Interactive Quizzes. Mithilfe dieser Fragen können Sie eigenständig und ohne jeden Druck überprüfen, ob Sie die neuen Inhalte schon verinnerlicht haben.

Sobald Sie eine Lektion komplett bearbeitet haben, können Sie Ihr Wissen auf der Lernplattform unter Beweis stellen. Über automatisch auswertbare Fragen erhalten Sie ein direktes Feedback zu Ihren Lernfortschritten. Die Wissenskontrolle gilt als bestanden, wenn Sie mindestens 80 % der Fragen richtig beantwortet haben. Sollte das einmal nicht auf Anhieb klappen, können Sie die Tests beliebig oft wiederholen.

Wenn Sie die Wissenskontrolle für sämtliche Lektionen gemeistert haben, führen Sie bitte die abschließende Evaluierung des Kurses durch.

Die IU Internationale Hochschule ist bestrebt, in ihren Skripten eine gendersensible und inklusive Sprache zu verwenden. Wir möchten jedoch hervorheben, dass auch in den Skripten, in denen das generische Maskulinum verwendet wird, immer Frauen und Männer, Inter- und Trans-Personen gemeint sind sowie auch jene, die sich keinem Geschlecht zuordnen wollen oder können.

LITERATUREMPFEHLUNGEN

ALLGEMEIN

Baur, N./Blasius, J. (Hrsg.) (2019): *Handbuch Methoden der empirischen Sozialforschung*. 2. Auflage, Springer, Wiesbaden.

Hussy, W./Schreier, M./Echterhoff, G. (2013): *Forschungsmethoden in Psychologie und Sozialwissenschaften*. 2. Auflage, Springer, Berlin/Heidelberg.

Schäfer, T. (2016): *Methodenlehre und Statistik. Einführung in Datenerhebung, deskriptive Statistik und Inferenzstatistik*. Springer, Wiesbaden.

LEKTION 1

Breuer, F. (2010): *Wissenschaftstheoretische Grundlagen qualitativer Methodik in der Psychologie*. In: G. Mey, K. Mruck (Hrsg.): *Handbuch Qualitative Forschung in der Psychologie*. Springer, Wiesbaden, S. 35–49.

LEKTION 2

Mey, G./Mruck, K. (2010) (Hrsg.): *Handbuch Qualitative Forschung in der Psychologie. Teil 3: Erhebung*. Springer, Wiesbaden, S. 423–552.

LEKTION 3

Kuckartz, U./Rädiker, S. (2019): *Datenaufbereitung und Datenbereinigung in der qualitativen Sozialforschung*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. 2. Auflage, Springer, Wiesbaden, S. 441–456.

Lück, D./Landrock, U. (2019): *Datenaufbereitung und Datenbereinigung in der quantitativen Sozialforschung*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. 2. Auflage, Springer, Wiesbaden, S. 457–471.

LEKTION 4

Backhaus, K. et al. (2018): *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. 15. Auflage, Springer Gabler, Berlin/Heidelberg.

Mey, G./Mruck, K. (2010) (Hrsg.): *Handbuch Qualitative Forschung in der Psychologie. Teil 4: Auswertung*. Springer, Wiesbaden, S. 553–752.

ÜBERGEORDNETE

LERNZIELE

Sie lernen im Kurs **Forschungsmethodik und Datenanalyse** alle wesentlichen Schritte innerhalb empirischer Forschungsprojekte im Rahmen der wissenschaftlichen Psychologie. Dieser Kurs bereitet Sie systematisch auf die Durchführung einer empirischen Studie (z. B. im Rahmen einer Bachelorarbeit) vor.

Es werden qualitative wie quantitative Forschungsmethoden, sowohl in der Datenerhebung als auch in der -auswertung behandelt. Dieses Wissen erlaubt Ihnen anschließend einen Vergleich zwischen beiden Ansätzen hinsichtlich Durchführung, Schwierigkeit und Erkenntnisgehalt.

Empirische Forschung kommt zudem nicht ohne theoretische Konzepte und Modelle aus. Den Zusammenhang zwischen Theorie (Modellbildung) und Empirie (Datenerhebung und Datenanalyse) in der Psychologie umfassend zu verstehen, ist daher ein Hauptlernziel dieses Kurses.

Dieser Studienbrief gibt eine breite Orientierung ohne Anspruch auf Vollständigkeit. Die Methoden der Erhebung und Auswertung können nicht detailliert in allen Schritten angesprochen werden. Dazu ist eine vertiefende Lektüre thematisch relevanter Beiträge in Lehr- und Handbüchern notwendig. In diesem Studienbrief werden Beispiele von Forschungsfragen, Datenerhebungen und Datenauswertungen besprochen. Das soll Lesern Hinweise für die eigene Forschungspraxis geben. Grundsätzliches Lernziel ist aber eine generelle Orientierung über die Methoden der empirischen Sozialforschung in der Psychologie. Für deren erfolgreiche Anwendung reicht die bloße Lektüre von Lehrtexten über Methoden nicht aus. Das Ausprobieren in der Praxis, verbunden mit einer kompetenten Beratung von erfahrenen Wissenschaftlern, ist der Königsweg, um empirische Forschungsmethoden zu erlernen und richtig anzuwenden.

LEKTION 1

PSYCHOLOGIE ALS EMPIRISCHE WISSENSCHAFT

LERNZIELE

Nach der Bearbeitung dieser Lektion werden Sie wissen, ...

- wie Theorie und Empirie in der Psychologie miteinander verknüpft sind.
- wie empirische Projekte vor der Datenerhebung konzipiert werden.
- welche erkenntnistheoretischen und ethischen Maßstäbe für Projekte gelten.
- worin die Unterschiede zwischen qualitativen und quantitativen Methoden bestehen.
- wie ein Forschungsbericht (z. B. eine Bachelorarbeit) aufgebaut ist.

1. PSYCHOLOGIE ALS EMPIRISCHE WISSENSCHAFT

Einführung

Am 9. März 2019 berichtete das Hamburger Abendblatt (Meyer-Wellmann 2019) über Verhandlungen zwischen dem Naturschutzbund Deutschland (NABU) und dem Hamburger Senat über einen Kompromiss zu der vom NABU initiierten Volksinitiative „Hamburgs Grün erhalten“. Der NABU hat mittlerweile genug Unterschriften gesammelt, um ein Volksbegehren (Volksabstimmung über ein Grünflächenschutzgesetz) zu starten, strebt jedoch mit der Hamburger Landesregierung eine gütliche Einigung an, die es ermöglichen würde, den Schutz der Grünflächen in die Regierungspolitik aufzunehmen, ohne neue Gesetze zu beschließen.

2017 hatte der NABU die Volksinitiative mit der spektakulären Aussage begründet, dass in Hamburg die Verkehrs- und Siedlungsflächen seit dem Jahr 2001 um 186 Hektar gewachsen seien. Entsprechend groß sei der Rückgang der Grünflächen. Der NABU berief sich dabei auf offizielle Statistiken der Umweltbehörde zur Flächennutzung. Allerdings stellte sich in der öffentlichen Diskussion bald heraus, dass der behördlichen Statistik nicht zu trauen ist (Meyer-Wellmann 2017). NABU und Hamburger Senat sehen deshalb als „entscheidende Voraussetzung für einen Kompromiss eine Einigung auf ein valides Verfahren zur Datenerhebung, mit dem man auch die Entwicklung des Grünanteils über die Jahre messen kann“ (Meyer-Wellmann 2019).

Was sind städtische Grünflächen, was sind Siedlungs- und Verkehrsflächen? Diese beiden gegensätzlichen Kategorien können unterschiedlich definiert werden. Ein öffentlicher Park, welcher der Erholung der Bevölkerung dient, ist ohne Zweifel eine Grünfläche. Ein großer städtischer Friedhof lässt sich jedoch weniger leicht einer der beiden Kategorien zuordnen – auf dem Gelände des Hamburger Friedhofs Ohlsdorf beispielsweise sind Auto- und Busverkehr erlaubt, dennoch werden die meisten Besucher diesen Ort subjektiv als eine Art Grünfläche empfinden, weil sie dort Ruhe, Besinnung und Erholung erleben. Aus behördlicher Sicht sind öffentlich genutzte Grünflächen zum Teil Verkehrsflächen, wie z. B. Sportplätze. Private Flächen, wie bepflanzte Balkone oder Dächer, sieht die Behörde in ihrer Statistik als Grünflächen, obwohl deren Erholungswert für die Bürger, die diese Flächen jeweils nicht selbst bewohnen, gering sein dürfte.

In der empirischen Umweltgeografie ist die Kategorie der Grünfläche ein Konstrukt. Dieses muss genau definiert werden, damit sein empirisches Vorkommen, z. B. die gemessene Grünfläche in der Stadt Hamburg, statistisch untersucht werden kann. Das Verfahren der Datenerhebung, so fordert der NABU, muss valide (gültig) sein. Alle Sichtweisen auf städtische Grünflächen (die von Bürgern, Umweltschutzverbänden und Behörden) sollten berücksichtigt werden.

In dieser Lektion geht es zwar nicht um Grünflächen, sondern um das Verhältnis von Theorie und Empirie in der Psychologie; dennoch verdeutlicht das Beispiel grundlegende Aspekte des Verhältnisses von Theorie und Empirie in den Wissenschaften. Die theoretischen Begriffe müssen, damit man in der Datenerhebung zu objektiven und validen Ergebnissen gelangen kann, zu Beginn einer empirischen Untersuchung eindeutig bestimmt sein.

Das wird in dieser Lektion an zwei Kategorien veranschaulicht, die in der wissenschaftlichen Psychologie im Fokus vieler empirischer Untersuchungen standen und stehen. Es sind dies die beiden Konstrukte von Motivation und Gedächtnis. Außerdem werden grundsätzliche Anforderungen an die Gestaltung der Erhebung psychologischer Daten besprochen, damit diese zu objektiven und gültigen Erkenntnissen führt.

1.1 Was sind empirische Methoden?

Der Begriff der Methode hat seinen Ursprung im altgriechischen Wort „metho-hodos“. Es kann übersetzt werden mit „einen Weg gehen“. Forschen ist das Einschlagen eines Forschungsweges, vom Anfang (der Forschungsfrage) bis zum Ziel (dem Forschungsergebnis). Während des Weges sind an Weggabelungen und Abzweigungen Richtungsentscheidungen zu treffen bezüglich der Auswahl und Anwendung der Erhebungs- und Auswertungsmethoden von empirischen Daten. Letztere sind in der Psychologie überwiegend Lebenserfahrungen, Lebensäußerungen und Lebensspuren von Menschen.

Die zu Beginn eines Forschungsprozesses formulierte Frage bestimmt den Kurs der Erhebung und Auswertung. In Methodenlehrbüchern werden die verschiedenen Formen der Datenerhebung und -auswertung systematisch und eingehend behandelt, die Formulierung einer untersuchungsleitenden Forschungsfrage als Forschungsschritt allerdings kommt in Lehrbüchern nur am Rande vor. Nun ist für die eigene Methodenanwendung aber insofern wichtig, die Kontur der leitenden Fragestellung zu schärfen, als dass sich aus ihr die Auswahl passender Methoden ergibt. Worum geht es in meiner Studie im Kern? Was ist mein Hauptkenntnisziel? Diese Fragen sollte sich jeder Student und Wissenschaftler in einer frühen Phase des Forschungsprozesses stellen und beantworten. Jedes Themenfeld lässt eine Vielzahl von Forschungsfragen zu.

Die Darstellung der empirischen Methoden in Lehrbüchern gleicht einem großen Instrumentenkoffer, der Anfänger leicht verunsichern kann. Manch Bachelor- oder Masterkandidat mag sich ob des riesigen Angebots verzweifelnd fragen, wie er bloß die für seine Studie richtige Auswahl treffen soll: Allein das Handbuch von Baur und Blasius (2014) zu den Methoden der empirischen Forschung ist über 1.100 Seiten stark und enthält 87 Beiträge. Ein Blick in ältere Methodenlehrbücher (z. B. Friedrichs 1990) zeigt, dass sich der Umfang solcher Lehrwerke vor gut dreißig Jahren noch auf ein Drittel davon beschränkte.

Grund für das Anwachsen der Methodenlehrbücher ist natürlich, dass Psychologen und Sozialwissenschaftler ständig neue Methoden erfinden. Dies sind zumeist keine Neuerfindungen im engeren Sinne, sondern Weiterentwicklungen und Verfeinerungen bereits vorhandener Methoden, die notwendig wurden, weil diese an neue Themen und Forschungsfragen angepasst werden mussten.

Methoden müssen zur Forschungsfrage und zum Forschungsziel passen. Das Kriterium der Passung ist zunächst recht vage, wird aber durch die in diesem Studienbrief behandelten Beispiele für den Leser Gestalt annehmen. Auf jeden Fall sollte deutlich werden, dass es keine per se besseren oder schlechteren empirischen Forschungsmethoden gibt, sondern eine solche Bewertung immer nur im Einzelfall vor dem Hintergrund der jeweiligen Forschungsfragen und Erkenntnisziele vorgenommen werden kann.

Anfängern fällt eine Orientierung im Methodendschungel zunächst schwer. Hilfreich ist hier das Wissen um den Umstand, dass komplexe Methoden auf einfache Methoden zurückgeführt werden können und aus diesen zusammengesetzt sind. Die drei Grundmethoden in der empirischen Psychologie sind Befragung bzw. Interview, Beobachtung und Experiment. Diese können wiederum in qualitativer oder quantitativer Form durchgeführt werden.

Die Auswahl der Auswertungsmethoden ist in einer empirischen Untersuchung abhängig von den Erhebungsmethoden. Wer sich z. B. für eine quantitative Form der Datenerhebung entschieden hat, dem stehen keine qualitativen Auswertungsmethoden zur Verfügung. Numerische Daten können nicht mit qualitativen Textanalysen ausgewertet werden, sondern allein mittels statistischer Verfahren, bei denen die Auswahl sehr groß ist. Die Datenaufbereitung wiederum hat zum Ziel, die gewonnenen Daten in eine fehlerfreie Form zu bringen, die eine optimale Grundlage für verschiedene Auswertungsverfahren darstellt.

Über die Methoden der empirischen Sozialforschung hinaus, die im Handbuch von Baur und Blasius dargestellt werden und im Schnittfeld von Psychologie, Soziologie und Wirtschaftswissenschaften zur Anwendung gelangen, existieren in der modernen Psychologie noch weitere spezielle Erhebungsmethoden, etwa solche der kognitiven Psychologie, neurophysiologische und neurowissenschaftliche Methoden sowie standardisierte psychologische Tests.

1.2 Empirie in der psychologischen Forschung

Die moderne Psychologie versteht sich als eine empirische Wissenschaft. Empirische Forschung ist eine Form wissenschaftlicher Praxis in modernen Gesellschaften. Wissenschaftliche empirische Forschung hat das Ziel der systematischen, intersubjektiv gültigen Erkenntnisgewinnung über die soziale und natürliche Umwelt sowie auch – besonders in der Psychologie – über das persönliche Innenleben von Menschen.

Die verschiedenen Formen und Techniken der empirischen Forschung werden als Methoden bezeichnet. Das Erlernen und Anwenden von empirischen Forschungsmethoden ist Teil des Studiums von empirisch ausgerichteten Wissenschaften. Die Gesamtheit der empirischen Forschungsmethoden in einer Wissenschaft bildet deren **Methodologie**, ihren Methodenkanon, der Studierenden und Doktoranden vermittelt wird.

Empirisch meint „auf Erfahrungen beruhend“ (vgl. zum Folgenden Hussy/Schreier/Echterhoff 2013, S. 3f.). Es geht dabei um die Gewinnung von Erkenntnissen über die Wirklichkeit, basierend auf Erfahrungen von Menschen. In das Forschungsergebnis fließen in der Psychologie sowohl die Erfahrungen der die Forschungen durchführenden Wissenschaftler ein als auch jene der untersuchten Personen.

Wissenschaftler sammeln in ihren Forschungsprozessen auf methodisch kontrollierte Weise Daten. In der Psychologie werden die Daten an Menschen, teilweise auch an Tieren gewonnen. Menschen werden zu ihren Erfahrungen befragt, z. B. in Interviews, oder sie werden als Versuchspersonen in Experimenten Erfahrungen ausgesetzt, die vom Experimentator gezielt herbeigeführt werden. Die verdichteten Erfahrungen der untersuchten Personen werden zu Daten, d. h. zu verbalen Aussagen in Interviews, zu Zahlen in Fragebögen oder Tests. Nach der Erhebung werden die Daten mit geeigneten Methoden ausgewertet.

Mit dem Anspruch, wissenschaftliche Erkenntnisse auf empirischer Basis zu gewinnen, orientiert sich die empirische Psychologie an den Naturwissenschaften. Zum Teil hat die Psychologie deren Forschungsmethoden auch direkt übernommen, wie etwa das Experiment und dazu passende statistische Auswertungsverfahren.

Die Instrumente der empirischen **Datenerhebung** (Fragebögen, Tests u. a.) haben das Ziel, die Empirie, also die Erfahrungen der untersuchten Fälle bzw. Menschen, so zu strukturieren, dass wissenschaftlich gültige Erkenntnisse entstehen. Die Instrumente wirken als ein Medium, das konkrete Erfahrungen von Menschen kanalisiert, damit sie zu den von den Wissenschaftlern anvisierten Untersuchungszielen passen. Daten werden von den Wissenschaftlern im Rahmen der Forschung in der methodisch kontrollierten Auseinandersetzung mit der Wirklichkeit hergestellt.

Datenerhebungen vollziehen sich in der Psychologie in der sozialen Interaktion mit untersuchten Menschen. Kommunikationsprozesse bilden den Rahmen der Datengewinnung. Das ist ein wesentlicher Unterschied zu klassischen naturwissenschaftlichen Untersuchungskonstellationen: Die Form der Daten wird durch soziale Austauschprozesse bestimmt. Die Weiterverarbeitung der Daten in der Dokumentation und Auswertung verfestigt deren sozialen Charakter:

[Daten] sind keine naturalistischen Abbilder aus der sozialen Welt, sondern basieren auf medial (technisch, textlich, mnemonisch etc.) konservierten Dokumentationen sozialer Wirklichkeit. Erst der Konservierungs-, Selektions-, Bewertungs-, Analyse-, Interpretations- und Darstellungsprozess während der Forschung macht diese im Feld gesammelten Dokumentationen dann schrittweise zu Daten, was ihnen [in der qualitativen wie in der quantitativen Forschung] einen grundlegend sozialen und rhetorischen Charakter verleiht (Meyer/Meier zu Verl 2014, S. 256).

Methodologie

Die Methodologie einer Wissenschaft ist zugleich ihre Methodenlehre.

Empirische Methoden

Sie werten Erfahrungen von Personen bzw. Fällen systematisch und methodisch kontrolliert aus, soweit diese Erfahrungen als Daten vorliegen.

Datenerhebung

Datenerhebungen sind, im weiteren Sinne, Kommunikationsprozesse. Es werden Informationen von den untersuchten Fällen an die Wissenschaftler übertragen.

1.3 Theorien und Modelle in der empirischen Psychologie

Die Empirie als Untersuchungsfeld ist abzugrenzen von wissenschaftlichen Theorien, A-priori-Setzungen und normativen Modellen. In den Methodenlehrbüchern wird Empirie zuerst abgegrenzt von Theorien, die empirische Erfahrungen in verschiedenen Wirklichkeitsausschnitten zu abstrakten Konzepten verdichten, um die Welt der sinnlichen Wahrnehmung zu erklären. **Theorien** sind fundamentale Aussagen über Wirklichkeitsbereiche, die verschiedene empirische Phänomene in einem Forschungsfeld zusammenfassen und erklären. Albert Einstein schrieb zur Güte von wissenschaftlichen Theorien: „Eine Theorie ist desto eindrucksvoller, je größer die Einfachheit ihrer Prämissen ist, je verschiedenartigere Dinge sie verknüpft und je weiter ihr Anwendungsbereich ist“ (zitiert nach Calaprice 2007, S. 177).

Theorie

Eine Theorie ist ein System wissenschaftlicher Aussagen über Bereiche der Wirklichkeit. Sie wird aus der Empirie gewonnen und anhand der Empirie auf ihre Gültigkeit überprüft.

Theorien und Empirie stehen in einem hierarchischen Verhältnis zueinander. Theorien können im Gegensatz zur Empirie von Menschen nicht sinnlich wahrgenommen werden. Sie sind abstrakt und werden ihren Rezipienten nur durch bewusstes Durchdenken verständlich. In der Psychologie gibt es unzählige Theorien zur Erklärung psychologischer Phänomene. Theorien sorgen für eine Vorstrukturierung des empirisch zu untersuchenden Forschungsfeldes.

Beispiel: Die Attributionstheorie als Forschungsfeld

Ein Beispiel ist die ursprünglich aus der Sozialpsychologie stammende Theorie der **Attribution**, die inzwischen auch in andere Teildisziplinen der Psychologie Eingang gefunden hat. Die Attributionstheorie beschreibt die Formen der Ursachenzuschreibung von Effekten, die Menschen im Alltag an sich selbst oder in der Umwelt bei anderen Personen wahrnehmen (vgl. Bierhoff 1999). Attributionen sind zunächst Alltagstheorien von lebenden Menschen, die nach den Ursachen für eigene Erlebnisse und das Verhalten anderer fragen. So kann z. B. der eigene Erfolg oder Misserfolg beim Lernen von Studierenden internal der eigenen Person (den persönlichen Kenntnissen, den individuellen Lernanstrengungen) oder external äußeren Faktoren (didaktische Kompetenzen der Lehrenden, Organisation des Studiums) als Ursache zugeschrieben werden. Der Inhalt der einfachen Attributionstheorie ist eine Beschreibung grundsätzlich möglicher Richtungen der Ursachenzuschreibung.

Innerhalb einer empirischen Untersuchung in der pädagogischen Psychologie wäre hier erstens das reale Attributionsverhalten von Studierenden zu untersuchen. Dabei wirken die Grundkategorien der Attributionstheorie (internale oder externale Attribution) forschungsleitend. Sie beeinflussen die Fragestellung und den Forschungsprozess. Es könnte z. B. empirisch untersucht werden, ob sich männliche Studierende in ihrem Attributionsstil bei Versagens- oder Erfolgserlebnissen tendenziell von weiblichen unterscheiden. Die so gewonnenen empirischen Ergebnisse wirken in der Folge auf Theorien zurück. Diese werden umformuliert, ergänzt, eingegrenzt in Hinsicht auf ihren Geltungsbereich oder sogar verworfen.

Attributionstheorie

Die Attributionstheorie fragt, welche einfachen kausalen Annahmen Menschen im Alltag selbst als Erklärung des Verhaltens bei sich und anderen vornehmen.

Zweitens ist empirische Forschung abzugrenzen von A-priori-Setzungen. Dies sind die allgemeinen Voraussetzungen von empirischen Erfahrungen. Die Grundfrage ist: Was müssen Menschen mitbringen, um ihre Umwelt sinnlich adäquat wahrzunehmen und die wahrgenommene Umweltsituation richtig zu bewerten? Hier ist zunächst an anatomische und biologische Voraussetzungen zu denken, etwa das Gehirn. Von großer Bedeutung ist darüber hinaus die Sprache. Begriffe strukturieren die kognitive Verarbeitung von Informationen, die über die Sinne aufgenommen werden. Klassischerweise ist die Untersuchung der A-priori-Setzungen eine Aufgabe der Philosophie bzw. **Epistemologie**. In ihren Untersuchungen der Voraussetzungen menschlicher Erkenntnisprozesse verfährt diese historisch und begrifflich-diskursiv, also weitgehend ohne die hier behandelten empirischen Methoden.

Allerdings gibt es in der Psychologie auch empirische Studien zu den allgemeinen Voraussetzungen von Erkenntnisprozessen. So liefert die moderne empirische Säuglingsforschung wichtige Beiträge bei der Untersuchung der Apriori des menschlichen Erkennens. Sie operiert mit der Methode des Experiments, um herauszufinden, wie die vorsprachliche Verarbeitung von Sinnesreizen funktioniert und welche inneren Erkenntnispotenziale bei Säuglingen und Kleinkindern dem zugrunde liegen (vgl. Mausfeld 2005, S. 9f.).

Drittens schließlich ist die empirische Ausrichtung in den Wissenschaften von normativen Wissenschaftsansätzen und Modellen zu trennen. In der allgemeinen Wissenschaftslehre werden empirische Wissenschaften wie Psychologie und Naturwissenschaften von **normativ operierenden Wissenschaften** unterschieden (vgl. Schurz 2014, S. 32f.). Letztere stellen allgemeine Modelle und Grundsätze über die Wirklichkeit auf, denen die Praxis der wissenschaftlich ausgebildeten Experten in der jeweiligen Disziplin folgen soll. Beispiele in den Humanwissenschaften sind Jura und Didaktik. In deren Modellen wird beschrieben, wie Rechtsstreitigkeiten grundsätzlich zu entscheiden und welche Rechtsgüter dabei abzuwägen sind (Jura) bzw. wie bestimmte Lehrinhalte am besten zu definieren und zu vermitteln sind (Didaktik). Die Modelle in den Disziplinen Jura und Didaktik sind normative Setzungen, die durch interne wissenschaftliche Diskussionen der Fachexperten aufgestellt und weiterentwickelt werden. Beide Wissenschaften kennen keine empirischen Forschungsmethoden, und die Studierenden werden in Jura und Didaktik nicht in solchen ausgebildet.

Gleichwohl stehen diese beiden Disziplinen vor der Herausforderung, ihren jeweiligen Gegenstand in einer sich wandelnden Gesellschaft immer wieder neu zu bestimmen und damit neuen gesellschaftlichen und politischen Erwartungen Rechnung zu tragen. Die für sie relevanten gesellschaftlichen Umbrüche und die damit einhergehenden Veränderungen in den Erfahrungen der Menschen können beide Disziplinen nicht ignorieren. Deshalb greifen auch sie auf empirische Erkenntnisse aus ihren Nachbarwissenschaften zurück. Bei der Rechtswissenschaft, soweit es um Strafrecht geht, ist das vor allem die Kriminologie, die die Häufigkeiten verschiedener Delikte ermittelt. Im Fall der Didaktik ist das die pädagogische Psychologie, die Unterrichtsprozesse und Lernergebnisse empirisch erforscht.

Epistemologie

Die Epistemologie (auch Erkenntnistheorie genannt) ist ein Teilbereich der Philosophie und beschäftigt sich mit den Grundlagen von Erkenntnis und Wissen von Menschen. Die allgemeine Psychologie untersucht Fragen der Epistemologie.

Normativ operierende Wissenschaften

Als normativ operierende Wissenschaften bezeichnet man u. a. Jura und Didaktik. Ihre Theorien, Modelle und Aussagen dienen dazu, menschliches Handeln ethisch und rational zu regulieren.

Aufbau wissenschaftlicher Theorien in der Psychologie

Wissenschaftliche Theorien liegen in verschiedenen Formen vor. Es kann sich um Beschreibungen in Form verbaler Aussagen bzw. Sätze über die Wirklichkeit handeln. Beispielsweise ist die Attributionstheorie eine verbale Beschreibung möglicher Richtungen persönlicher Ursachenzuschreibungen. Ursachen für Verhaltenseffekte können in der Person (internale Attribution) oder in der Umwelt (externale Attribution) verortet werden.

In der wissenschaftlichen Forschung werden Theorien miteinander verbunden, um komplexe empirische Phänomene zu erklären. So wurde in der pädagogischen Psychologie die ursprünglich aus der Sozialpsychologie stammende Attributionstheorie mit der theoretischen Kategorie des Selbstbildes von Lernenden in Zusammenhang gebracht. Außerdem wurde der Umgang mit Erfolg und Misserfolg beim Lernen bzw. in Prüfungen mit in die theoretischen Konzepte aufgenommen. In der theoretischen Gesamtschau zeigt sich, dass Menschen mit einem negativen Selbstbild Misserfolge beim Lernen eher internal attribuieren („Ich kann das einfach nicht.“). Dagegen schreiben sie mögliche Lernerfolge eher den äußeren Umständen zu („Ich habe diesmal nur Glück gehabt und deshalb die Prüfung bestanden.“).

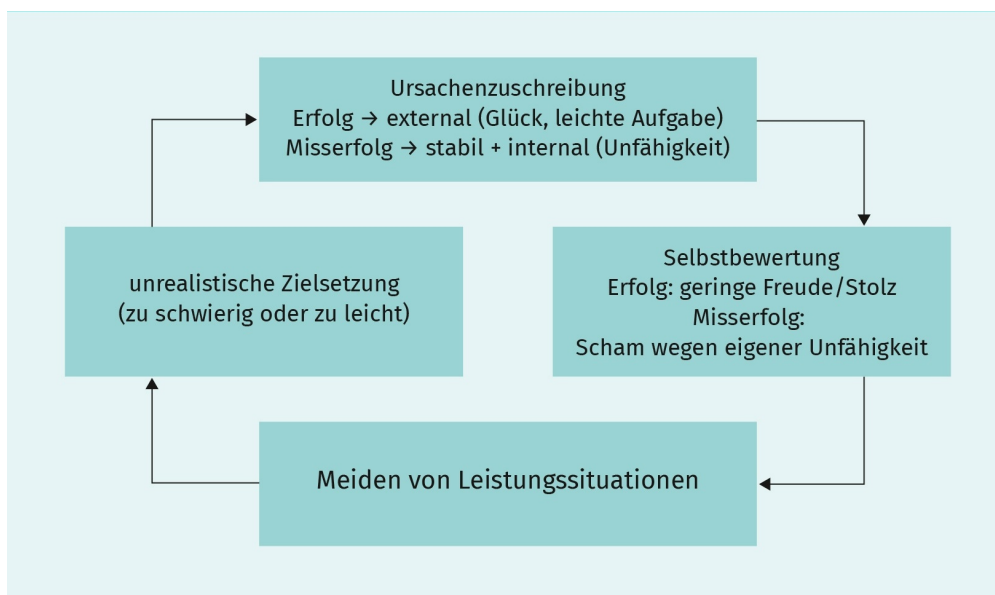
Komplexere psychologische Theorien bestehen aus mehreren theoretischen Kategorien, die empirische Phänomene, also das konkrete Erleben und Verhalten der Menschen im Alltag, erklären. Die motivationspsychologische Typologie des Umgangs mit Erfolg bzw. Misserfolg hat in der pädagogischen Psychologie so viel an Popularität gewonnen, weil sie verschiedene theoretische Konstrukte (Attributionsstil, Selbstbild, Umgang mit Erfolgen und Misserfolgen) plausibel miteinander verbindet (vgl. Hasselhorn/Gold 2017, S. 104f.). Zudem ist sie von einleuchtender Prägnanz und so auch für psychologische Laien verständlich.

Das Beispiel der auf Lernprozesse bezogenen, erweiterten Attributionstheorie belegt ein allgemeines Prinzip: Verschiedene psychologische Theorien werden miteinander verknüpft, damit sie die Empirie möglichst umfassend erklären. Psychologische Theorien bestehen zumeist aus mehreren theoretischen Konzepten bzw. Kategorien. Eine psychologische Theorie fristet kein Inseldasein, sondern ist Teil eines Netzwerks (vgl. Westermann 2016, S. 165f.). So ist z. B. die Attributionstheorie des Lernerfolgs außerdem eng verknüpft mit Anreiztheorien in der **Motivationspsychologie**. Anreiztheorien beziehen sich auf den Anreiz, etwas zu tun oder zu unterlassen, also auf Handlungsmotivationen. Lernende mit internaler Attribution von Misserfolgen vermeiden herausfordernde Lernsituationen aus Angst vor Misserfolg. Der Anreiz, neue Lernsituationen anzugehen, sinkt bei ihnen.

Motivation

Die Bereitschaft oder der Anreiz, etwas zu tun oder zu unterlassen, wird als Motivation bezeichnet. Motivation für ein positiv besetztes Handlungsziel führt zum Handeln. Wenn aus Sicht eines Akteurs negative Konsequenzen drohen, kommt es zum Vermeiden von Handlungen.

Abbildung 1: Verknüpfung von Attributionstheorie, Selbstkonzept und Anreiztheorien zur Erklärung der Auswahl von Lernaufgaben von misserfolgsorientierten Personen



Quelle: erstellt im Auftrag der IU, 2022 in Anlehnung an Hasselhorn/Gold 2017, S. 109f.

Modelle

Modelle haben die Funktion einer übergreifenden Konzeptbildung für einen bestimmten Forschungsbereich. Solche Modelle können **normativer** Natur sein, also Setzungen beinhalten, die das Handeln von Fachexperten in ihrer beruflichen Praxis als Psychologe steuern und legitimieren; sie können aber auch auf die **empirische** Forschung hin ausgerichtet sein oder auf empirischer Forschung beruhen. Modelle enthalten zumeist eine Liste von Begriffen und/oder Aspekten und beschreiben die Zusammenhänge zwischen diesen.

Ein Beispiel ist das Modell zur Lehrevaluation an Hochschulen des Psychologen Heiner Rindermann (2003). Dieses Modell beschreibt alle Aspekte der Lehr-Lern-Interaktion in der Hochschullehre. Es basiert auf umfangreichen empirischen Untersuchungen.

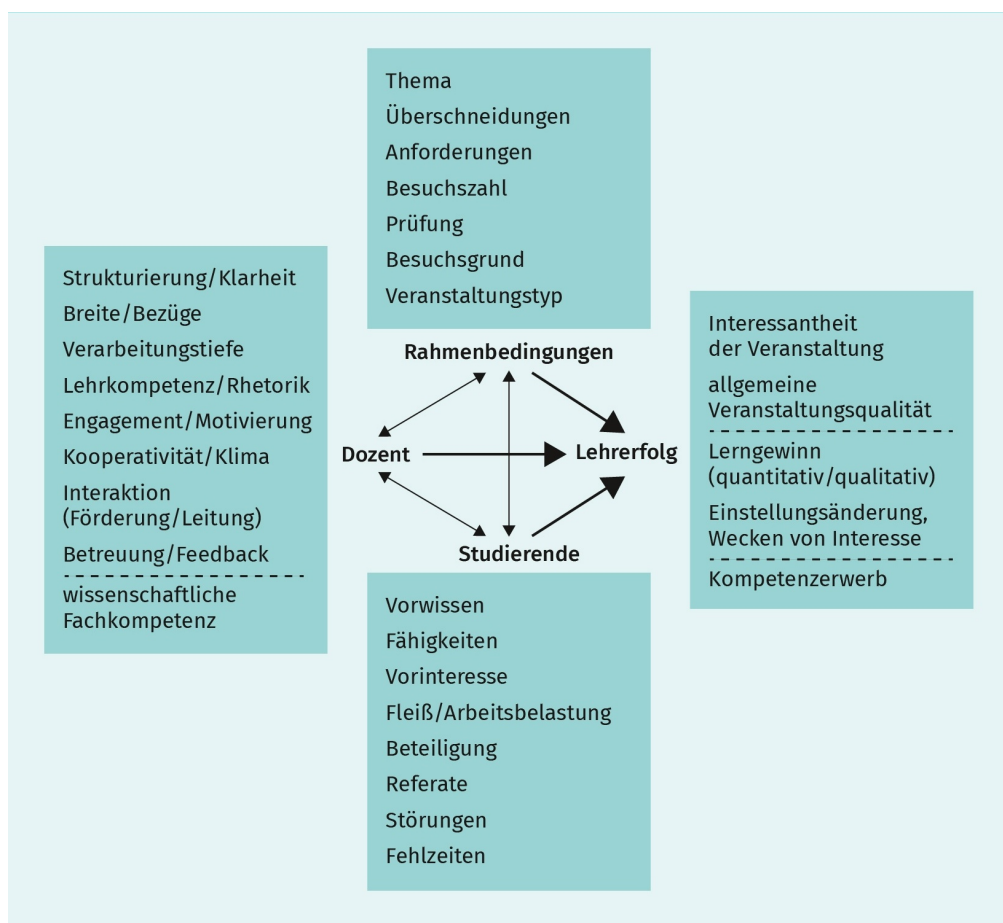
Normative Modelle

Sie regulieren wissenschaftliches Handeln und sind eine Vereinfachung komplexer Handlungsketten von Experten in der Berufspraxis.

Empirische Modelle

Sie sind als Konzepte eine Vereinfachung komplexer Zusammenhänge in der Empirie.

Abbildung 2: Multidimensionales Modell des Lehrerfolgs an Hochschulen



Quelle: Rindermann 2003, S. 236.

Das Modell gibt den konzeptionellen Rahmen vor für die Evaluation von Hochschullehre. Wer Evaluationen zur Lehre an Hochschulen durchführt, sollte sich an diesem Modell orientieren und zumindest einige Inhalte seiner Studie berücksichtigen. Die Entwicklung eines entsprechenden Fragebogens sollte sich daran orientieren, anderenfalls würden zentrale Standards der Fachdiskussion im Forschungsfeld der Evaluation von Hochschullehre verletzt.

1.4 Forschung als Verbindung von Empirie und Theorie

Damit die empirische Forschung zu Ergebnissen gelangt, die mit der Wirklichkeit übereinstimmen und einer Prüfung auf Wahrheit standhalten, kommen die in den Methodenlehrbüchern dargestellten Methoden zur Anwendung. Das sind systematisierte und kontrollierte Verfahren der Datenerhebung und -auswertung, die von Wissenschaftlern in

zurückliegenden Forschungen entwickelt wurden und in wissenschaftlichen Aufsätzen sowie zusammenfassend und kanonisch in Methodenlehrbüchern dokumentiert wurden. Diese Verfahren müssen bestimmten Gütekriterien genügen, damit sie als probate Mittel im wissenschaftlichen Erkenntnisprozess anerkannt werden. Gängige Lehrbücher zur empirischen Sozialforschung stellen die empirische Methodologie im Schnittpunkt von Psychologie und Soziologie dar (vgl. Baur/Blasius 2014, Kopp/Lois 2014). Diese beiden Wissenschaften haben die modernen empirischen Forschungsmethoden in den Humanwissenschaften hervorgebracht bzw. aus den Naturwissenschaften adaptiert.

Die praktische Anwendung der in Lehrbüchern dokumentierten Methoden geschieht nicht als gleichförmige technische Applikation, sondern vollzieht sich im Rahmen einer intensiven gedanklichen Auseinandersetzung der Forscher mit dem Forschungsthema während des Forschungsprozesses. Diese sind immer auch Denkprozesse. Der gesamte Forschungsprozess stellt eine komplexe geistige Auseinandersetzung der Forschenden mit ihren Gegenständen dar. Johann Wolfgang von Goethe, der den Beginn der modernen Naturwissenschaften miterlebte und sich kritisch gegenüber deren quantitativ-experimenteller Methodologie positionierte, definierte eine wissenschaftliche Methode als Form der Vermittlung zwischen einem Subjekt (Forschungsperson) und einem Objekt (Forschungsthema) (Goethe 1989). Methoden sind dem Forschungsgegenstand nicht äußerlich, sondern bestimmen diesen mit. Unterschiedliche Methoden führen zu unterschiedlichen Ergebnissen im gleichen Themenfeld. Damit verändern sich in der Folge auch Theorien und Modelle in dem Themenfeld.

Beispiel: Veränderung von Gedächtnistheorien durch empirische Forschung

Als Wissenschaft konstituierte sich die Psychologie mit dafür spezialisierten Professoren an deutschen und angelsächsischen Universitäten gegen Ende des 19. Jahrhunderts. Von Beginn an war für sie die empirische Erforschung des Gedächtnisses ein Thema. Als Begründer der experimentellen Gedächtnisforschung gilt der deutsche Psychologe Hermann Ebbinghaus (1850–1909) mit seiner Studie zum Silbengedächtnis (Ebbinghaus 1885). Er übernahm aus der damaligen Physiologie das quantifizierende Experiment als Untersuchungsmethode. Es hat bis heute in Psychologie und Medizin einen festen Platz in der empirischen Erforschung des Gedächtnisses. Der Ablauf ist immer ähnlich. Ein Versuchsleiter stellt einer Person Aufgaben zum Einprägen und Reproduzieren. Versuchsmaterial sind oft Listen (z. B. von Worten, Zahlen oder geometrischen Mustern). Wichtigster quantitativer Kennwert in der Auswertung ist die Anzahl der reproduzierten Elemente. Aus der Diskrepanz zwischen Einprägen und Reproduktion ergibt sich ein Leistungswert. Weitere quantitative Leistungsparameter können über Vergleiche verschiedener Zeitpunkte der Reproduktion (z. B. unmittelbar nach dem Einprägen, einen Tag später, eine Woche später etc.) errechnet werden.

Ab 1970 gewann in den USA ein anderer methodischer Ansatz bei der empirischen Erforschung des Gedächtnisses an Einfluss. Er wird heute als die Erforschung des autobiografischen Gedächtnisses bezeichnet (Pohl 2007). Der Ansatz beruht auf alltäglichen Erzählungen und Erinnerungen der Menschen. Die Untersuchungspersonen werden in Form qualitativer Interviews nach erinnerten Ereignissen und Erlebnissen befragt und sollen diese selbst aus der Erinnerung aufschreiben. Im Unterschied zum naturwissenschaftlich-experimentellen Ansatz wird den Untersuchungspersonen kein Material zum Einprägen

vorgegeben, das für alle gleich ist. Statt quantitativer Reproduktionszahlen interessiert vielmehr, welche qualitativen Merkmale Erlebnisse aufweisen, die besonders gut erinnert werden. Außerdem sollen die Versuchspersonen sich an länger Zurückliegendes erinnern, nicht an kurz zuvor Eingprägtes.

Dieser qualitative Forschungsansatz zeigt auf, welche Ereignisse und Erlebnisse des Alltags sich bei Menschen einprägen und welche nicht. In erster Regel sind es biografisch bedeutsame Ereignisse und emotional auffallende Erlebnisse, die im Gedächtnis haften bleiben. Bei bewussten Erinnerungen, also ob, wann und wie das vor längerer Zeit Erlebte als geistiges Bild im Bewusstsein eines Menschen auftaucht, spielen soziale Hinweisreize (engl. „clues“, im Deutschen auch als „Anker“ bezeichnet) eine entscheidende Rolle. Wenn z. B. in den Medien an bestimmte Ereignisse erinnert wird, deren wiederkehrender Jahrestag auf ein rundes Datum fällt, erinnern sich die Medienrezipienten, was sie an jenem Tag gemacht haben, als über das entsprechende Ereignis berichtet wurde.

Ein Beispiel ist der Jahrestag des Tsunamiunglücks in Asien aus dem Jahr 2004. Wenn am 26. Dezember die Medien an das Unglück erinnern, denken viele Menschen daran, was sie an jenem Tag gemacht haben, wie sie damals die Nachrichten verfolgt haben, an welchem Ort sie waren und mit welchen Menschen sie zu tun hatten. Die autobiografische Gedächtnisforschung führte zu einer Revision klassischer Modelle des Langzeitgedächtnisses, die den Fokus auf das verbale Einprägen legten. Neben dem verbalen Teil enthalten moderne Gedächtnismodelle einen episodischen Bereich. In diesem finden sich Bilder von Situationen, die erlebt wurden und für die betreffende Person von Bedeutung sind.

1.5 Unterschiede zwischen qualitativen und quantitativen Methoden

Seit Beginn der wissenschaftlichen Psychologie am Ende des 19. Jahrhunderts durchzieht sie der Streit um quantitative und qualitative Methoden. Welche Form der Datengewinnung ist besser geeignet, die menschliche Psyche empirisch einzufangen und wissenschaftlich zu analysieren? Lässt sich die Psyche in quantitativ messbare Teile und Variablen zerlegen oder sind ganzheitlich-verstehende Ansätze bei der psychologischen Untersuchung von Menschen angemessener? Ein stärker an den Naturwissenschaften orientierter Zugang setzt auf quantitative Methoden. Wer dagegen die Psychologie eher als Geisteswissenschaft bzw. als Sparte der Philosophie und Erkenntnistheorie sieht, favorisiert häufig qualitative Methoden.

Dokumentierte wissenschaftliche Dispute zwischen den führenden Fachvertretern in verschiedenen Epochen legen davon ein lebendiges Zeugnis ab. Zu nennen sind z. B. die Kontroversen zwischen Karl Bühler (1879–1963) und Wilhelm Wundt (1832–1920) vor dem Ersten Weltkrieg (Sekundärdarstellung in Vollmers 1992) oder zwischen Albert Wellek (1902–1974) und Peter Hofstätter (1913–1994) nach dem Zweiten Weltkrieg (Sekundärdarstellung in Métraux 1985). Die Pro- und Kontra-Argumente für qualitative oder quantitative Methoden zeigen bis in die heutige Zeit eine markante Kontinuität. Am Ende des 20. Jahrhunderts kam es sogar zu einer institutionellen Spaltung der wissenschaftlichen Psycholo-

gie: Qualitativ orientierte Forscher der Psychologie haben sich zur „Neuen Deutschen Gesellschaft für Psychologie“ zusammengeschlossen und bilden in der akademischen Community der Psychologen eine Opposition. Bei den jährlichen Tagungen der zahlenmäßig weitaus größeren „Deutschen Gesellschaft für Psychologie“ finden dagegen qualitativ arbeitende Psychologen kaum Gehör.

Worum ging und geht es bei diesem Streit im Wesentlichen, und welche Auswirkungen hat er auf die konkrete empirische Forschungspraxis in der Psychologie?

Qualität und Quantität sind zwei zentrale Merkmale aller materiellen Stoffe und Prozesse in der menschlichen Lebenswelt. Die quantitative Analyse von Themen und Objekten ermittelt deren Häufigkeiten und die numerischen Anteile und Verhältnisse der ihnen innewohnenden Teile. In der qualitativen Analyse von Objekten und Prozessen geht es demgegenüber um deren strukturelle Beschaffenheiten und Muster. Gesucht werden in der qualitativen empirischen Forschung die einen Gegenstand konstituierenden Merkmale; Ziel ist es, diese zu finden und präzise zu benennen. Anschließend können in einem Folgeschritt deren Anteile und Häufigkeiten statistisch aufgeschlüsselt werden. Quantitative und qualitative Forschungsansätze können sich also gegenseitig ergänzen. Typischerweise geht in einem solchen Mehrmethoden-Ansatz die qualitative Forschung der quantitativen voraus.

Eine Orientierung zu den Unterschieden zwischen qualitativen und quantitativen Methoden gibt die folgende Tabelle. Diese Unterschiede gelten allgemein in den Humanwissenschaften, unterscheiden also nicht nach Psychologie, Soziologie und Erziehungswissenschaft. Außerdem geht diese Tabelle nicht auf die konkrete Ausgestaltung der Datenerhebung und Auswertung ein.

Tabelle 1: Unterschiede zwischen qualitativer und quantitativer Forschung

qualitativ	quantitativ
konkret	abstrakt
Situationskontexte können berücksichtigt werden	Situationskontexte spielen eine untergeordnete Rolle
eine Phase in der empirischen Forschung	eine andere Phase in der empirischen Forschung
ganzheitlich	elementaristisch
induktiv	deduktiv
Stichprobe klein	Stichprobe groß
Unbekanntes entdecken in strukturellen Zusammenhängen	Bekanntes in quantitativen Abhängigkeiten erfassen

Quelle: erstellt im Auftrag der IU, 2022 in Anlehnung an Kleining 1982.

Konkret vs. abstrakt: Qualitative Forschung setzt an den vielfältigen Alltagserfahrungen von Individuen in verschiedenen Lebenskontexten an. Sie nimmt diese als Ausgangspunkt der Forschung. Bei der qualitativen autobiografischen Gedächtnisforschung werden Personen aufgefordert, die Erlebnisse aufzuschreiben, an die sie sich aus einem bestimmten Lebensjahr erinnern. Quantitative Forschung fokussiert demgegenüber mehr auf Strukturen und Zusammenhänge oder Trends und versucht, sie in Kategorien und Skalen darzustellen. Eine quantitative Umfrage, wie beispielsweise eine Meinungsumfrage in der Marktforschung, lässt Prognosen zum Kaufverhalten der Konsumenten zu.

Situationskontexte: Diese Unterscheidung knüpft an die Unterscheidung zwischen konkret und abstrakt an. Je abstrakter die Datenerhebung, desto weniger fließen die individuellen Lebenskontexte der Untersuchungspersonen in die Datenanalyse ein. Das quantitative Beispiel aus der Gedächtnisforschung belegt das eindrücklich. Dabei wird mit der Vorgabe von Silben oder Worten darauf abgezielt, die Einprägungssituation für alle untersuchten Personen identisch zu halten. Im autobiografischen Ansatz können die Personen dagegen bei der Darstellung der erinnerten Erlebnisse auf interindividuell unterschiedliche Situationen und Themen des Einprägens zugreifen.

Verschiedene Phasen in der Forschung: Qualitative Forschung geht quantitativer Forschung oft (nicht immer) voraus. Wer eine quantitative Forschungsarbeit verfolgt, muss sich vorher mit der Forschungsliteratur zu diesem Thema vertraut machen. Das ist ein qualitatives Erkunden des Themas. Ein Beispiel dafür ist Hermann Ebbinghaus, der Begründer der quantitativen Gedächtnisforschung. Er griff auf Überlegungen der Philosophen der Antike, des Mittelalters und der Neuzeit zurück, die sich vor ihm mit dem menschlichen Gedächtnis beschäftigt hatten. Besonders einflussreich war für ihn der griechische Philosoph Aristoteles, der das menschliche Gedächtnis als ein Bündel von Assoziationen zwischen Begriffen beschrieben hat. In der Methodenliteratur wird eine die Forschung vorbereitende Lektüre allerdings zumeist nicht explizit als Methode der empirischen Forschung beschrieben.

Ganzheitlich vs. elementaristisch: Qualitative Verfahren haben oft die menschliche Psyche als Ganzes zum Ausgangspunkt, ohne diese zu Beginn des Forschungsprozesses in Aspekte oder Kategorien zu zerlegen. Ein Beispiel aus der autobiografischen Gedächtnisforschung ist die Aufforderung an die Versuchspersonen, alles aufzuschreiben, was ihnen aus der Erinnerung an Ereignisse aus einem früheren Lebensabschnitt einfällt. Eine Zerlegung in Themen oder Gruppen von Situationen, an die sich die Person erinnern soll, findet vonseiten der Forscher, die die Aufgabe stellen, nicht statt. Die untersuchten Personen haben die freie Wahl, welchen Themen sie sich zuwenden möchten. Dagegen wird in der quantitativen Forschung zu Beginn eine Unterteilung des Forschungsthemas in verschiedene Aspekte vorgenommen, die in verbalen Aussagen und Skalen operationalisiert werden. So kann z. B. die Akzeptanz eines Produktes unter Käufern eingeteilt werden in die Akzeptanz des Designs, der technischen Funktion, des Preises etc.

Induktion
Geht als Denkprozess von
Einzelfällen aus und
gelangt zu Verallgemeine-
rungen.

Induktiv vs. deduktiv: „Induktion“ bedeutet Verallgemeinerung, **„Deduktion“** meint Ableitung. Beides sind Formen des Denkens, die im Alltag wie in der empirischen Forschung in enger Verzahnung von Menschen angewendet werden, um die Wirklichkeit adäquat zu verstehen. In der qualitativen Forschung herrscht als grobe Richtung des Denkens die Induktion vor, in der quantitativen dominiert tendenziell die Deduktion. In der qualita-

tiven Forschung werden aus den gewonnenen Daten induktiv verallgemeinernde Kategorien gewonnen, die die Empirie zusammenfassen, beschreiben und erklären; es entsteht eine empirisch begründete Theorie über einen Bereich der Wirklichkeit. Theorien stehen am Ende des Forschungsprozesses. Dagegen markieren in der quantitativen Forschung Theorien den Ausgangspunkt des Forschungsprozesses. Aus ihnen werden der Forschungsansatz und das Forschungsziel abgeleitet (deduziert).

Deduktion

Gelangt auf der Basis von generellen Konzepten zu konkreten Aussagen über Einzelfälle.

Stichprobe klein vs. groß: Qualitative Stichproben bestehen üblicherweise aus deutlich weniger Fällen als quantitative. Fälle können in der Psychologie Personen sein, aber auch andere Einheiten (z. B. Familien, Arbeitsgruppen oder Unternehmen). Quantitative Stichproben sind in der Regel grösser als qualitative Stichproben, weil sie strukturelle Beobachtungen anstellen.

Unbekanntes entdecken vs. Bekanntes in Abhängigkeiten erfassen: Aufgrund der geringen Vorgaben durch die Forscher bietet die qualitative Forschung die Chance, dass die Forschenden neue, für sie bisher unbekannte Aspekte eines Gegenstandes entdecken. Die Untersuchungspersonen sprechen oft Themen an, die zu Beginn der Studie nicht im Fokus der wissenschaftlichen Betrachtung lagen. In der quantitativen Forschung ist der Blick auf die Empirie dagegen durch das theoretische Vorverständnis und die Ableitung der Datenerhebung aus Theorien eingeschränkt. Ziel ist es, die Ausprägungen vorher definierter Merkmale in dem untersuchten Feld genau zu ermitteln.

**FAZIT:**

Bei der Erfassung und Analyse von Daten befasst sich die quantitative Forschung mit Zahlen und Statistiken, während sich die qualitative Forschung mit Wörtern und Bedeutungen befasst. Beides ist wichtig, um unterschiedliche Arten von Wissen zu erlangen.

Quantitative Forschung**Qualitative Forschung****Gemischte Ansätze**

Es gibt immer mehr Studien, die eine Mischung aus qualitativer und Quantitativer Forschung nutzen. Die zentrale Prämisse ist hierbei, dass die Verwendung qualitativer und quantitativer Ansätze in Kombination ein besseres Verständnis der Forschungsprobleme liefert als jeder Ansatz allein (Baur/Krapp/Bamberg 2017, Teile I und III).

Ein Forschungsbeispiel zum Unterschied qualitativ vs. quantitativ

Diese Unterscheidungen werden im Folgenden an einem kurzen imaginären Beispiel aus der empirischen Motivationsforschung verdeutlicht. Angenommen, uns interessiert die Motivation von Studierenden der Psychologie. Wir möchten herausfinden, warum sie Psychologie im Fernstudium studieren. Wie würde ein quantitativer Forschungsansatz aussehen, wie ein qualitativer? Dabei skizzieren wir die Vorüberlegungen, die zur Auswahl oder Konstruktion eines Fragebogens (quantitativ) bzw. eines Leitfadens (qualitativ) führen, der als Instrument der Datenerhebung genutzt werden könnte.

Die Forschungsfragestellung lautet: Welche Motive geben Studierende der Psychologie im Fernstudium als Gründe für ihre Studienwahl und Studiendurchführung an?

Ein quantitativer Ansatz zur Untersuchung der Studienmotivation

Am Ausgangspunkt der empirischen Untersuchung erfolgt eine theoretische Explikation und Eingrenzung. Theoretische Grundlage unserer Untersuchung soll die Selbstbestimmungstheorie (engl. „self determination theory“ = SDT) der amerikanischen Psychologen Richard Ryan und Edward Deci (2000) sein. Diese Theorie ist in der pädagogischen Psychologie außerordentlich populär und hat zahlreiche empirische Studien zum Lernverhalten und zur Lernmotivation hervorgebracht (vgl. Hasselhorn/Gold 2017, S. 108f.). Die SDT geht davon aus, dass die Motivation für jedwedes menschliches Handeln, nicht nur beim Lernen, auf drei universellen Grundbedürfnissen beruht:

- Bedürfnis nach Kompetenz,
- Bedürfnis nach sozialer Eingebundenheit,
- Bedürfnis nach Autonomie.

Für ein positives Lernerlebnis ist bei allen Menschen in allen Lernsituationen die Erfahrung der Befriedigung dieser drei Grundbedürfnisse entscheidend. Werden diese drei Grundbedürfnisse in den erlebten Lernsituationen erfüllt, entwickelt sich die lernende Person positiv. Es kommt zu einer Zunahme an Kreativität und Problemlösekompetenz. Das stärkt das Durchhaltevermögen beim Lernen. Die SDT steht in der Tradition des Menschenbildes der humanistischen Psychologie, die für psychisches Wachstum das Umsetzen der **intrinsischen Motivation** als entscheidend sieht. Ein empirisches Forschungsprojekt, das sich an der SDT orientiert, klammert deshalb zwangsläufig Aspekte der **extrinsischen Motivation** (Geld, Karriere, gesellschaftlicher Status) aus.

Die drei theoretischen Kategorien bilden die Basis für die Konstruktion eines Fragebogens zur quantitativen Untersuchung der Studienmotivation. Dazu ist eine Operationalisierung der drei Kategorien erforderlich. Für jede Kategorie werden mehrere Umschreibungen bzw. Aussagen formuliert. Dies wird auch als Bildung empirischer Indikatoren bezeichnet. Aus diesen ergeben sich in einem weiteren Schritt die Fragen des Fragebogens. Die folgende Tabelle enthält plausible Indikatoren der drei Kategorien der SDT.

Intrinsisch vs. extrinsisch

Intrinsisch und extrinsisch sind zwei gegensätzliche Motivationspole. Diese Unterscheidung ist grob gleichzusetzen mit der von inneren (persönlichen) und äußeren (aus der Umwelt stammenden) Motivationsquellen.

Tabelle 2: Aus der SDT abgeleitete Indikatoren für die empirische Forschung

Kategorie der SDT	empirische Indikatoren bzw. Umschreibungen
Kompetenz	<ul style="list-style-type: none">• effektive Einwirkung auf Umgebung• Selbststeuerung des Lernens• erlebter Zuwachs an Kreativität
soziale Eingebundenheit	<ul style="list-style-type: none">• Zugehörigkeitsgefühl zu einer Gruppe• Wertschätzung durch andere Personen• gemeinsame Aufgabenbewältigung in einer Gruppe
Autonomie	<ul style="list-style-type: none">• Freiwilligkeit des Lernens• keine Unterwerfung unter Autoritäten• Eigenständigkeit im Denken

Quelle: erstellt im Auftrag der IU, 2022 in Anlehnung an Ryan/Deci 2000 und Hasselhorn/Gold 2017.

Die Indikatoren führen direkt zu den Fragen eines Fragebogens. Der folgende Fragebogen ist ein erster Entwurf zur Untersuchung der Forschungsfragestellung. Zu jedem der neun Indikatoren wird eine Frage formuliert. Die Reihenfolge der Fragen orientiert sich an der Abfolge der Indikatoren in obiger Tabelle.

Abbildung 3: Quantitativer Fragebogen zur Studienmotivation basierend auf der SDT

Alter _____ Geschlecht: M <input type="checkbox"/> W <input type="checkbox"/> Semesterzahl _____				
Ich studiere das Fach Psychologie an dieser Fernuniversität aus folgenden Gründen:			4 = Trifft vollkommen zu. 1 = Trifft gar nicht zu.	
1. Meine Kompetenzen bei der Bewältigung privater und beruflicher Aufgaben nehmen dabei zu.	(1)	(2)	(3)	(4)
2. Ich kann das Studium überwiegend nach meinen Vorstellungen gestalten.	(1)	(2)	(3)	(4)
3. Ich gewinne vielfältige Ideen und Anregungen für mein weiteres Leben.	(1)	(2)	(3)	(4)
4. Ich erlebe einen solidarischen Austausch mit anderen Studierenden.	(1)	(2)	(3)	(4)
5. Ich erfahre positive Rückmeldungen durch andere Menschen.	(1)	(2)	(3)	(4)
6. Ich habe beim Lernen viele Gelegenheiten zur Kooperation mit anderen Studierenden.	(1)	(2)	(3)	(4)
7. Ich empfinde mein Lernen während des Studiums überwiegend als frei und ungezwungen.	(1)	(2)	(3)	(4)
8. Die Lehrenden betrachten mich als gleichberechtigte/n Partner/in.	(1)	(2)	(3)	(4)
9. Ich kann im Studium meinen eigenen Ideen und Interessen nachgehen.	(1)	(2)	(3)	(4)

Quelle: erstellt im Auftrag der IU, 2022.

Dieser kurze Entwurf müsste für eine komplette empirische Studie, wie z. B. in einer Bachelorarbeit, mit weiteren Fragen zu den drei Kategorien der SDT ergänzt werden. Das Beispiel verdeutlicht den konzeptionellen Weg hin zum quantitativen Fragebogen: Eine allgemeine Forschungsfrage und theoretische Überlegungen sind der Ausgangspunkt. In quantitativen Fragebögen wird zur Beantwortung der Fragen in Psychologie und Sozialwissenschaften gerne eine sogenannte LIKERT-Skala verwendet. Sie hat im vorliegenden Beispiel vier Abstufungen (Werte 1 bis 4).

Bei der Wahl bzw. Konstruktion des empirischen Forschungsinstruments sollte bereits an Auswertungsziele gedacht werden. Im Beispiel könnten die im Fragebogen enthaltenen **soziodemografischen Merkmale** (Alter, Geschlecht, Semesterzahl) als Gruppierungsvariablen für die Auswertung fungieren. Quantitative Analysen würden sich auf die Unterschiede bei den neun Aussagen bzw. Variablen zur Motivation bezüglich verschiedener Altersgruppen, Geschlechter und der Studiendauer beziehen. Im ersten Schritt der quantitativen Auswertung wäre mit einer Datei aller Befragten in einem Statistikprogramm zu arbeiten und Häufigkeitsanalysen für alle zwölf Variablen durchzuführen.

Soziodemografische Daten

Diese enthalten Informationen zur Verteilung typischer sozialer Merkmale von der Stichprobe. Bei Personen sind das z. B. Alter, Geschlecht oder Einkommen, bei Unternehmen Jahresumsatz oder Zahl der Mitarbeiter.

Ein qualitativer Ansatz zur Untersuchung der Studienmotivation

Qualitative Forschung verzichtet zu Beginn der Forschung auf theoretische Setzungen. Offenheit der Forschungsperson gilt als leitendes Forschungsprinzip (Kleining 1982, S. 8). Die Forschenden sollen sich durch die Erlebnisse und Erfahrungen der untersuchten Personen leiten lassen, nicht durch wissenschaftliche Theorien. Ausgehend von einer offenen Fragestellung überlegen sich die Wissenschaftler mittels freier Assoziationen, ähnlich wie in der Kreativtechnik des Brainstormings, welche Aspekte des Themas aufgrund ihrer eigenen Lebens- und Forschungserfahrungen relevant sein könnten. Aus dieser Liste entstehen die Fragen eines Interviewleitfadens.

Bei der Frage, wie Studierende ihre eigene Motivation im Studium der Psychologie sehen, könnten Überlegungen der Forschenden vor der Datenerhebung zu folgenden Aspekten führen:

- persönliche Interessen und Vorlieben,
- einschneidende biografische Erlebnisse,
- persönliche Werte und Lebensziele.

Die Fragen eines qualitativen Interviewleitfadens sind offen formuliert und fordern die Untersuchungspersonen zum Nachdenken auf. Sie sind so angelegt, dass sie längere Statements provozieren. Aus der Sammlung von relevanten Aspekten lassen sich die folgenden drei Fragen eines Interviewleitfadens formulieren:

1. Was sind Ihre Motive für die Aufnahme und weitere Durchführung Ihres Psychologiestudiums an der Fernuniversität? Bitte nennen Sie alle aus Ihrer persönlichen Sicht relevanten Gründe, auch wenn Ihnen diese unter Umständen trivial oder unwissenschaftlich erscheinen.
2. Gab es irgendwelche Situationen oder Ereignisse in Ihrem Leben, die Sie in Ihrer Entscheidung für dieses Studium nachhaltig beeinflusst haben?
3. Was versprechen Sie sich persönlich von diesem Studium für Ihr weiteres berufliches und privates Leben?

Dieser Interviewleitfaden wäre zu kurz für die Anwendung in einer empirischen Studie. Der Umfang von Interviewleitfäden hängt immer vom Thema ab. Erfahrungswerte des Autors dieses Studienbriefs sind: Bei qualitativen Bachelorarbeiten beläuft sich der Umfang eines Interviewleitfadens auf etwa acht bis zwölf Fragen, bei Masterarbeiten auf zwölf bis fünfzehn. Hinzu kommen Fragen zur Erhebung soziodemografischer Informationen, ähnlich wie in quantitativen Fragebögen.

Das Interview wird in der Regel aufgenommen und liegt als Audiodatei vor. Vor der Auswertung wird es teilweise oder vollständig transkribiert. Die schriftlichen Protokolle sind die Basis der Auswertung. Diese wird als qualitative Text- und Inhaltsanalyse durchgeführt.

Bewertung der beiden Ansätze zur Untersuchung der Studienmotivation

Der quantitative Ansatz lässt sich fachwissenschaftlich gut begründen. Die SDT hat als Motivationstheorie viele Anhänger in Psychologie und Wirtschaft. Der qualitative Ansatz ist umfassender und lässt Raum für intrinsische und extrinsische Motive. Außerdem lässt er die Möglichkeit zu, dass die Befragten Motive äußern, die auf die Besonderheiten des Faches Psychologie abzielen. Beispielsweise wurde die Begründung, dass die Studierenden sich selbst und andere Menschen besser verstehen möchten, mit dem quantitativen Fragebogen nicht angesprochen. Experten der Psychologie sehen das jedoch als ein sehr wichtiges fachliches und persönliches Studienziel (vgl. Ernst 2014). Eine Ergänzung des quantitativen Fragebogens um diese Einschätzung von Experten wäre leicht durchführbar. In den Fragebogen könnte dies als zusätzliche Aussage aufgenommen werden. Eine andere Möglichkeit ist, den Befragten die Möglichkeit einzuräumen, selbst ein weiteres Motiv schriftlich zu formulieren. Die Aufnahme einer oder weniger offener Fragen in einen quantitativen Fragebogen ist möglich. Qualitative Fragen sind aber nicht quantitativ auswertbar und müssen separat inhaltlich analysiert werden.

1.6 Der Verlauf empirischer Forschungsprozesse und dessen Darstellung in Berichten

Die Reihenfolge der wesentlichen Forschungsschritte ist in allen empirischen Forschungsprozessen ähnlich:

1. Lektüre von Lehrbüchern und Fachartikeln zum Forschungsthema
2. Eingrenzen des Themas bzw. Bestimmen der Hauptforschungsfrage
3. Konstruktion oder Auswahl der Datenerhebungsinstrumente
4. Definition der zu untersuchenden Stichprobe
5. Durchführung der Datenerhebung in der Stichprobe
6. Datenaufbereitung für die Auswertung
7. Auswertung der Daten
8. Interpretation der Ergebnisse
9. Schreiben des Forschungsberichts

Die Anfangsphase (Punkte 1 und 2) führt zur Hauptforschungsfrage. Das Finden der Fragestellung ist ein sukzessives Schärfen der Fragestellung. Für eine Person ist oft ein von ihr beobachtetes Problem im Alltag oder Beruf der erste Anstoß. Durch die intensivere Beschäftigung damit sowie durch begleitende Lektüre konturiert sich für sie allmählich die genaue Fragestellung. Die verschiedenen Forschungsphasen vom Beginn (dem Finden der Fragestellung) bis zum Ende (der Datenauswertung) werden dokumentiert und begründet im schriftlichen Forschungsbericht, der nach Abschluss der Datenauswertung in seiner endgültigen Form verfasst und ggf. publiziert wird. Ein Forschungsbericht, wie z. B. eine Bachelor- oder Masterarbeit, ist ein Protokoll des Forschungsprozesses. Er stellt dessen Verlauf chronologisch in seinen wesentlichen Etappen dar.

Die wesentlichen Gliederungspunkte eines Forschungsberichts sind:

1. Stand der Forschung im jeweiligen Fachgebiet
2. Hauptforschungsfragen
3. Konstruktion oder Auswahl der Datenerhebungsinstrumente
4. Charakteristika der Stichprobe
5. Durchführung der Datenerhebung
6. Datenaufbereitung
7. Auswertung der Daten
8. Interpretation der Ergebnisse

Die schriftliche Darstellung der acht genannten Forschungsetappen folgt wissenschaftlichen Konventionen, auf die im Folgenden kurz eingegangen wird.

Stand der Forschung: Hier werden Theorien und empirische Forschungsergebnisse aus anderen Studien zusammenfassend dargestellt, die für die eigene Forschung relevant sind. Diese Aspekte werden in Berichten zur quantitativen Forschung zumeist an den Anfang gestellt, weil sich daraus der eigene Forschungsansatz ableitet. In der eher induktiv ausgerichteten qualitativen Forschung stehen diese Aspekte oft am Ende des Berichts. Die eigenen Forschungsergebnisse werden auf relevante Theorien bezogen, um diese zu ergänzen, zu relativieren oder sogar zu verwerfen.

Hauptforschungsfragen: Wie sind die Forschungsfragen entstanden und wie sind sie formuliert? Die Entstehung der Forschungsfragen kann auf ganz persönliche Motive und Eindrücke zurückgehen. Diese können im Forschungsbericht benannt werden. Wichtig ist aber eine Begründung der Forschungsfrage mit wissenschaftlichen Bezügen (auf andere Studien, auf behandelte Themen in Lehrveranstaltungen, auf psychologische Fachdiskurse). Die Forschungsfragen müssen explizit formuliert werden, und eine zusätzliche Erläuterung ihrer Inhalte oder Hintergründe verbessert das Verständnis der Leser des Forschungsberichts.

Datenerhebungsinstrument: Der im Forschungsprozess verwendete Fragebogen, Test oder Interviewleitfaden muss vollständig im Forschungsbericht enthalten sein, entweder im Inhalt oder im Anhang. Es entspricht den Konventionen, dessen Auswahl und Konstruktion zu begründen, etwa durch Verweise auf wissenschaftliche Theorien.

Stichprobe: Die ausgewählten Fälle werden in ihren wichtigsten demografischen Merkmalen kurz beschrieben, ebenso die Form der **Stichprobenbildung**. Stichproben sollen möglichst repräsentativ sein. Diese Forderung ist in quantitativer und qualitativer Logik unterschiedlich zu erfüllen. Quantitative Stichproben müssen repräsentativ sein für eine Grundgesamtheit, für die statistische Aussagen getroffen werden. Eine Grundgesamtheit wären z. B. alle Wahlberechtigten eines Landes. Qualitative Stichproben müssen repräsentativ für ein Thema sein, zu dem eine Theorie oder ein Modell entwickelt wird. Dazu ist in der Stichprobe das Einholen unterschiedlicher Sichtweisen auf das Thema erforderlich.

Stichprobenbildung
Eine Stichprobe von Fällen ist eine Teilmenge aus einer größeren Gesamtheit von Fällen. Die Stichprobe wird durch ein spezielles Auswahlverfahren gebildet.

Datenerhebung: Wann und wo die Datenerhebungen stattfanden, diese Informationen gehören in jeden Forschungsbericht. Außerdem interessieren Mitteilungen über die besondere Form der Datenerhebung, etwa dass es sich um eine Online- oder eine Telefon-

befragung gehandelt hat. Eine kurze Begründung zur Eignung der gewählten Erhebungsmethoden für die Fragestellung, mit Bezug auf Methodenliteratur (warum z. B. ein qualitativer Ansatz gewählt wurde), rundet die Darstellung ab.

Datenaufbereitung: Hier wird in der quantitativen Sozialforschung beschrieben, mit welchem Statistikprogramm gearbeitet wurde und welche Maßnahmen ergriffen wurden, um die Vollständigkeit und Fehlerfreiheit der Daten zu sichern. In der qualitativen Sozialforschung gehen die Forscher zudem auf die von ihnen angewendeten Transkriptionsregeln ein.

Datenauswertung: Zunächst wird die gewählte Auswertungsmethode mit ihren Zielen erläutert. Verzichten sollten die Autoren eines Berichts auf eine umfassende Erläuterung der einzelnen Schritte einer Auswertungsmethode. Es ist nicht üblich, die Schritte einer Inhaltsanalyse oder die SPSS-Befehle wie in Lehrbüchern zu erläutern. Kern des Berichts sind vielmehr die Ergebnisse. Bei einer qualitativen Inhaltsanalyse sind das die ermittelten Kategorien, bei statistischen Analysen die statistischen Kennwerte.

Interpretation der Ergebnisse: Die gefundenen Ergebnisse sprechen in der Regel nicht automatisch für sich, sondern bedürfen einer zusätzlichen Interpretation, Einschätzung, Begründung oder Relativierung. Die Interpretation der Ergebnisse kann vor dem Hintergrund der Besonderheiten des Forschungsfeldes erfolgen. Sie geht dann von den Erfahrungen der Forschenden aus, soweit diese bisher noch nicht im Forschungsbericht thematisiert wurden. An dieser Stelle ist der Bezug auf Theorien und Ergebnisse aus anderen Studien sinnvoll.

Forschungsbericht
Ein Forschungsbericht ist ein schriftlicher Bericht, z. B. eine wissenschaftliche Hausarbeit, eine Bachelorarbeit oder ein Aufsatz in einer Fachzeitschrift.

Die quantitativen Anteile der verschiedenen Kapitel eines **Forschungsberichts** sind unterschiedlich. Der Zweck empirischer Forschung ist die Gewinnung von Erkenntnissen. Demzufolge sollte die Darstellung und Interpretation der Ergebnisse quantitativ den größten Anteil des Berichts einnehmen. Die Themen Stichprobenumfang und Stichprobengewinnung können oft auf einer Seite abgehandelt werden, ebenso die Datenaufbereitung. Bedeutung und Umfang der anderen Aspekte hängen vom gewählten Thema und den Methoden ab. Eher unkonventionelle Methoden, wie z. B. qualitative Beobachtungen, benötigen unter Umständen einen größeren Platz in ihrer Beschreibung und Begründung.

Der damit beschriebene Aufbau eines Berichts ist dessen typische Form. Meyer und Meier zu Verl (2014, S. 251f.) unterscheiden bei der Ergebnispräsentation von empirischer Forschung zwischen einem konventionellen und einem experimentellen Stil. Der konventionelle Stil orientiert sich an den Normen eines Fachs, der experimentelle überschreitet sie oder will sie bewusst zurückweisen. Konventionelle Darstellungen stellen das Finden der Fragestellung, die Datenerhebung und die Datenauswertung in ihren Grundzügen dar. Experimentelle Formen orientieren sich nicht daran. Sie berichten nicht systematisch über die Forschungsphasen, sondern verdichten oder verfremden Forschungsprozess und -ergebnis. Beispiele sind vor allem künstlerische Darstellungen (Rollenspiel, Bild, Collage) oder die Weiterverarbeitung der Forschungsergebnisse in journalistischer oder schriftstellerischer Form (Essay, Roman). Allerdings zeigt die Erfahrung, dass Betreuer von Bachelorarbeiten für derartige Abenteuer oft wenig aufgeschlossen sind; die Gliederung der Bachelorarbeit sollte immer mit den Betreuern abgesprachen werden.

Die Qualität der Darstellung im Forschungsbericht

Die Frage einer optimalen Darstellung der Ergebnisse in Forschungsberichten wird in der Literatur zur empirischen Sozialforschung nur am Rande behandelt. Das Verfassen und der Aufbau des Berichts gehören nicht zu den empirischen Methoden. Die Forschungsergebnisse und die Qualität in deren Darstellung verbinden die Leser der Berichte mit den Verfassern. Der Autorenname eines Berichts „haftet“ für dessen guten oder schlechten Inhalt. Wer seine Forschungsergebnisse an andere weitergibt, sollte sich im Rahmen einer internen Qualitätsprüfung anhand des Manuskripts die folgenden Fragen stellen:

- Sind meine Darlegungen für mögliche Leser überzeugend?
- Habe ich meine Forschungsergebnisse gut zusammengefasst?
- Werde ich mit meinem Bericht die Leser erreichen?
- Werde ich bei den Lesern etwas bewirken, z. B. das Interesse für meine Ergebnisse wecken?

Anfänger sollten sich bei der Beantwortung dieser Fragen von pragmatischen Kriterien leiten lassen. Um anderen Personen eine Überprüfung der Qualität zu ermöglichen, die ausschließlich über Transparenz der Forschung möglich ist, muss diese in sich stimmig sein. Im Forschungsbericht muss der Bezug zwischen den erhobenen Daten und den Ergebnissen für Leser erkennbar sein. In der quantitativen Forschung bestehen die Ergebnisse aus statistischen Kennwerten, in der qualitativen aus Kategorien und Zitaten als Erläuterung oder Beleg. Bei beiden Varianten sollten die Interpretation und die Diskussion der Befunde von deren Darstellung getrennt sein. Im Übrigen hat eine lebendige Darstellung von Forschungsablauf und -ergebnissen, die den inneren Zusammenhang beider Aspekte bewahrt und stilistisch gut geschrieben ist, auf jeden Fall einen positiven Effekt auf die Bewertung von Abschlussarbeiten.

1.7 Gütekriterien empirischer Forschung

Gütekriterien der empirischen Sozialforschung sind Objektivität, Reliabilität (Zuverlässigkeit) und Validität (Gültigkeit). Diese stammen ursprünglich aus der Testpsychologie (vgl. Lienert/Raatz 1998, S. 7f.) und wurden dann von der empirischen Methodenlehre übernommen, zunächst von der quantitativen Richtung, später auch von der qualitativen (systematisch zuerst von Kleining 1982). Unter qualitativ orientierten Psychologen wird zum Teil die Auffassung vertreten, dass andere Gütekriterien als die klassischen drei zur Bewertung qualitativer Methoden von Relevanz sind (Flick 2014, S. 416f.).

Auf jeden Fall haben die drei klassischen Gütekriterien in der quantitativen und qualitativen Forschung jeweils andere Bedeutungen. In der quantitativen Forschung werden die Gütekriterien in erster Linie mit Bezug auf die Datenerhebung und die eingesetzten Messinstrumente bestimmt. In der qualitativen Forschung ist der Referenzpunkt dagegen das Forschungsergebnis, das in Form von verbalen oder grafischen Protokollen bzw. Kategorien von Inhaltsanalysen vorliegt.

Objektivität in der quantitativen Forschung

Bei der Anwendung quantitativer psychologischer Fragebögen und Tests in der empirischen Sozialforschung wird nach Objektivität gestrebt. Wissenschaftler, Testanwender und Testauswerter dürfen bei der Datenerhebung und Datenauswertung nicht verfälschend auf die Ergebnisse Einfluss nehmen. In der Methodenliteratur wird zwischen Durchführungs-, Auswertungs- und Interpretationsobjektivität unterschieden (Lienert/Raatz 1998, S. 7). Die Durchführungsobjektivität wird in erster Linie durch standardisierte Testinstruktionen gesichert, an die sich alle halten müssen, die an dem Testeinsatz auf Forschungsseite beteiligt sind. Die Auswertungsobjektivität ist die Forderung, dass alle Auswertenden den gleichen Schritten folgen müssen, damit sie bei gleichen Daten zu identischen Ergebnissen gelangen. Bei standardisierten Tests ist die Auswertungsobjektivität durch die standardisierte Testauswertung zumeist mittels Schablonen gesichert. Interpretationsobjektivität bezieht sich auf die Interpretation der gemessenen Testwerte. Die Interpretationen sollten bei gleichen Ergebnissen ähnlich sein.

Objektivität

Als objektiv gelten empirische Forschungsergebnisse, wenn sie frei von verzerrenden Einflüssen sind, die potenziell von den Personen ausgehen können, die eine Studie durchführen.

Objektivität ist in den Wissenschaften und der Philosophie ein viel diskutiertes Erkenntnisproblem. Viele Philosophen und Erkenntnistheoretiker haben sich darüber bereits den Kopf zerbrochen. Ein objektiver Erkenntnisprozess hat den Anspruch, dass die Ergebnisse für sich genommen wahr und intersubjektiv gültig sind. Unabhängig von den subjektiven Meinungen, Ansichten und Interessen der Forschungspersonen, die eine Studie durchführen, sollen die Ergebnisse einer Untersuchung allgemeine Geltung beanspruchen können. An psychologische Experimente wird deshalb immer wieder die Forderung gestellt, dass die Ergebnisse die gleichen sein müssen, wenn eine andere Forschungsperson bei einer anderen Untersuchungsstichprobe die gleiche Versuchsanordnung anwendet. Experimentelle Ergebnisse sollen replizierbar sein, dann gelten sie als objektiv wahr. Replikationsstudien von psychologischen Experimenten haben allerdings oft zu abweichenden Ergebnissen gegenüber den Originalstudien geführt (vgl. Paulus 2017).

Reliabilität in der quantitativen Forschung

Reliabilität

Reliabel sind quantitative Forschungsergebnisse, wenn sie auf fehlerfreien Messinstrumenten beruhen.

Reliabilität bedeutet in der quantitativen Sozialforschung, dass Messfehler möglichst auszuschalten bzw. gering zu halten sind. Reliabilität bezieht sich auf die Zuverlässigkeit des eingesetzten Datenerhebungsinstruments. Diese wird in der Regel vor dessen Einsatz im Rahmen von Vorstudien überprüft. Über Korrelationsstudien wird für das Messinstrument ein Reliabilitätskoeffizient berechnet.

Die vier Varianten der Bestimmung der Reliabilität sind:

- **Split-Half-Reliabilität:** Der Fragebogen bzw. Test wird in zwei Hälften mit der gleichen Anzahl an Fragen bzw. Aufgaben zerlegt. Für beide Teile wird an einer Stichprobe das Gesamtergebnis ermittelt und beide Ergebnisse werden miteinander korreliert.
- **Retest-Reliabilität:** Der gleiche Fragebogen bzw. Test wird an der gleichen Stichprobe zweimal durchgeführt. Der ermittelte **Korrelationskoeffizient** spiegelt das Maß der Übereinstimmung beider Messungen wider.
- **Paralleltest-Reliabilität:** Den Untersuchungspersonen werden nacheinander zwei sehr ähnliche Tests vorgelegt. Der ermittelte Koeffizient ist ein Maß für die Ähnlichkeit der beiden Verfahren.

- **Interne Konsistenz:** Jede Aufgabe eines Tests bzw. jede Frage eines Fragebogens wird mit jeder anderen des gleichen Tests bzw. Fragebogens korreliert. Basis sind die Daten der Messung in einer Stichprobe. Für die Bestimmung der Reliabilität gilt gewissermaßen jede Aufgabe als eigenständiger Test. Eine hohe mittlere Korrelation zwischen allen Aufgaben belegt die Homogenität eines Erhebungsinstruments.

Die vier Arten der Reliabilitätsbestimmung erfolgen mit unterschiedlichen statistischen Formeln, die sich in entsprechenden Lehrbüchern finden (z. B. in Westermann 2016, S. 287f.). In allen Varianten wird mittels Korrelationen ein Maß für die innere Stabilität und Homogenität eines Forschungsinstruments gewonnen. Im Allgemeinen gelten Korrelationen über 0,8 als akzeptabel für den Einsatz in der empirischen Forschung, Korrelationen über 0,9 als gut bis sehr gut. Ein Reliabilitätskoeffizient von 1 gehört zu einem Test, der frei von Messfehlern ist.

Validität in der quantitativen Forschung

Das dritte Gütekriterium, die **Validität**, bezieht sich auf die inhaltliche Gültigkeit der gemessenen Ergebnisse. Es geht um die Frage, ob die in empirischen Studien untersuchten Theorien und Kategorien durch die Messinstrumente hinreichend erfasst werden. Im Allgemeinen wird zwischen einer internen und einer externen Validität unterschieden (Westermann 2016, S. 290f.). Letztere meint, dass die Prüfung der Validität anhand zusätzlicher externer Kriterien erfolgt. Die Prüfung der internen Validität ihrerseits erfolgt anhand von Kriterien, die im Rahmen der durchgeführten Forschungsstudien diskutiert bzw. erhoben werden.

Die Berechnung der externen Validität erfolgt über Korrelationskoeffizienten. Beispielsweise werden bei einer Stichprobe die Ergebnisse eines psychologischen Tests auf die Ergebnisse anderer Tests oder Kriterien bezogen, zu denen Werte in der Stichprobe vorliegen (z. B. Schulnoten oder Mitarbeiterbeurteilungen). Die Kriterien für einen ausreichend hohen Korrelationskoeffizienten werden in der Forschungsliteratur niedriger angesetzt als bei der Reliabilität. Im klassischen Lehrbuch zur Testkonstruktion des Psychologen und Mediziners Gustav Lienert (1920–2001) gelten bereits Korrelationen ab 0,5 als ausreichend für die Validität empirischer Untersuchungen (Lienert/Raatz 1998, S. 270).

Die interne Validität bezieht sich auf den Zusammenhang zwischen den empirischen Datenanalysen zugrundeliegenden Theorien und den eingesetzten Messinstrumenten. Diese sollten inhaltlich zu den Konstrukten bzw. Kategorien der Referenztheorien passen. Bezeichnet wird diese Form als **Inhaltsvalidität** oder Konstruktvalidität. Die Inhaltsvalidität wird in der Forschungspraxis durch Einschätzungen von Experten bestimmt.

Die Untersuchung der Konstruktvalidität ist erheblich aufwendiger. Im Zentrum steht die Ermittlung der statistischen Differenz zwischen konvergenter und diskriminanter Validität. Daten aus Fragebögen bzw. Tests, die sich auf dasselbe Konstrukt beziehen, müssen hoch miteinander korrelieren (konvergente Validität). Daten aus Instrumenten, die sich auf verschiedene Konstrukte beziehen, sollten dagegen nicht miteinander korrelieren (diskriminante Validität).

Korrelationskoeffizient

Ein Maß des Zusammenhangs zweier Variablen X und Y, das unabhängig vom Maßstab der Variablen X und Y immer einen Wert zwischen -1 und +1 annimmt. Ist der Korrelationskoeffizient hoch, so treten hohe X-Werte häufig mit hohen Y-Werten auf. Ein negativer KK hingegen zeigt an, dass hohe X-Werte häufig mit niedrigen Y-Werten auftreten.

Validität

Valide sind die Ergebnisse einer Untersuchung, wenn die durch Forschungsfragen anvisierten Forschungsziele empirisch exakt erfasst worden sind.

Inhaltsvalidität

Der Begriff „Inhaltsvalidität“ bezeichnet die offensichtliche Validität. Mit Inaugenscheinnahme des Messinstruments wird dessen Validität von Experten beurteilt.

Die drei Gütekriterien in der qualitativen Forschung

In der qualitativen Sozialforschung finden standardisierte Messinstrumente keine Verwendung. Objektivität, Reliabilität und Validität können nicht über Korrelationskoeffizienten oder andere statistische Verfahren berechnet werden. Sie werden vielmehr konsensorientiert zwischen Forschenden und den untersuchten Personen besprochen und bestimmt. Statt von Objektivität spricht man in der qualitativen Sozialforschung daher bevorzugt von Intersubjektivität: Andere Forscher, die eine parallele Studie durchführen, sollten zu ähnlichen Ergebnissen gelangen.

Reliabilität bezieht sich auf die innere Struktur der Daten, wenn die Ergebnisse der durchgeführten Inhaltsanalyse vorliegen. Alle gesammelten Befunde müssen sich plausibel in das Gesamtbild einfügen. Es dürfen am Ende nicht Aussagen aus den verbalen Protokollen zurückbleiben, die außerhalb des Kategoriensystems stehen, das die verbalen Daten aller Untersuchungspersonen (bei Interviews) oder Untersuchungssituationen (bei Beobachtungen) verdichtet. Für die Reliabilität gilt damit die 100-Prozent-Forderung der internen Passung bzw. Stimmigkeit aller erhobenen Daten (Kleining 1982, S. 21).

Die Reliabilität der Ergebnisse einer qualitativen Inhaltsanalyse ist quantitativ mit Korrelationskoeffizienten bestimmbar, wenn es mehrere Auswerter für die gleichen Daten gibt. Zwischen ihren Ergebnissen wird das Maß der quantitativen Übereinstimmung bzw. Abweichung errechnet. In der Fachsprache heißt das Interrater- oder Intercoder-Reliabilität.

Validität ist in der qualitativen Sozialforschung, wie die Reliabilität, ein Gütekriterium in Bezug auf das Forschungsergebnis. Die Validität ist hoch, wenn viele unterschiedliche Sichtweisen von den untersuchten Fällen im Laufe der Forschung thematisiert wurden und im Ergebnis Widerhall finden. Handlungsanweisung für die Forschung ist in diesem Zusammenhang die Variation der Perspektiven in der Stichprobenbildung, d. h., es werden möglichst unterschiedliche Fälle untersucht (Kleining 1982, S. 10). Außerdem gibt es im Idealfall eine Triangulation der Methoden (Flick 2014, S. 418). Das bedeutet, dass beim gleichen Thema verschiedene Methoden (Experimente, Beobachtungen und Befragungen) zum Einsatz kommen. Die Validität ist in der qualitativen Forschung ein ausschließlich internes Güte Merkmal. Die Ergebnisse sind allein gültig für die untersuchte Stichprobe.

Kommunikative Validierung

Bei der kommunikativen Validierung werden die Forschungsergebnisse und Schlussfolgerungen daraus den Untersuchungspersonen selbst zur Prüfung vorgelegt und mit ihnen diskutiert.

Gerne wird in der qualitativen Sozialforschung auch die **kommunikative Validierung** angewendet (Flick 2014, S. 413f.). Die Gültigkeit der Ergebnisse wird im Konsens allgemeinverbindlich festgestellt. Dazu ist eine Gesprächsrunde zwischen allen Beteiligten, vor allem mit den untersuchten Personen, erforderlich. Diese findet nach Abschluss der Datenauswertung statt. In ihr werden die vorläufigen Ergebnisse von den Wissenschaftlern präsentiert und Rückmeldungen dazu eingeholt. Unter Umständen werden die Ergebnisse revidiert und es findet eine weitere Gesprächsrunde zur Absicherung ihrer Gültigkeit statt.

Objektivität, Reliabilität und Validität sind normative epistemologische Anforderungen an die Methoden und Ergebnisse empirischer Sozialforschung. Sie sollen, unabhängig von Versuchspersonen und Versuchsleiter, innere Stimmigkeit und Wahrheit der ermittelten Ergebnisse sichern. Wer als Forscher die drei Gütekriterien vor, während und nach dem empirischen Forschungsprozess selbst genau bestimmt, der führt eine interne Qualitäts-

kontrolle durch. Kritisch ist anzumerken, dass der damit verbundene Aufwand groß ist und die detaillierte Beachtung und Umsetzung der Gütekriterien die Ressourcen einer einzelnen Person leicht überfordert. Für studentische Forschungsprojekte und Abschlussarbeiten gilt daher, dass grobe Überlegungen und Abschätzungen zu den Gütekriterien wertvoll sind. Eine eigenständige Berechnung (bei quantitativen Projekten) oder diskursive Bestimmung (bei qualitativen Projekten) ist aus zeitlichen Gründen oft nicht möglich.

1.8 Moralische und ethische Aspekte empirischer Forschung

Heiligt der höhere moralische Zweck die angewendeten, jedoch als unmoralisch empfundenen Mittel? Ist es moralisch vertretbar, Untersuchungspersonen, beteiligte Institutionen und die Öffentlichkeit zu täuschen, wenn dadurch wissenschaftliche Erkenntnisse gewonnen werden, die zu einer Aufklärung der allgemeinen Öffentlichkeit führen oder der Behebung von gesellschaftlichen Missständen den Weg bereiten? Das ist eine ethische Grundfrage in der empirischen Sozialforschung. In den 60er- und 70er-Jahren gab es besonders in den USA eine Reihe prominenter Soziologen und Psychologen, die die von ihnen untersuchten Personen über die Absichten und Methoden ihrer Untersuchungen massiv getäuscht haben.

So setzte sich in der US-amerikanischen Soziologie und Psychologie die Methode der verdeckten **teilnehmenden Beobachtung** mehr und mehr durch. Die Wissenschaftler ließen sich unter Vorgabe einer falschen Identität und mittels Simulation psychiatrischer Symptome oder krimineller Taten in psychiatrische Kliniken und Gefängnisse einweisen. Ziel war es, das dortige Leben unverfälscht und unvoreingenommen zu untersuchen und die Rolle des Patienten bzw. Häftlings selbst zu spielen. Eine vorherige Absprache mit der Klinik- bzw. Gefängnisleitung hätte vermutlich authentische Erfahrungen der Forschungspersonen verhindert. Zumindest wäre das Erleben der entsprechenden Rollen und der damit einhergehenden psychischen Auswirkungen so nicht passiert, da die Verantwortlichen in den jeweiligen Institutionen Einfluss auf die Untersuchungen genommen hätten – sie hätten ihre Mitarbeiter vor der Untersuchung instruiert und den Wissenschaftlern nur ausgesuchte Fälle präsentiert. Dieses wäre ihr gutes Recht und vermutlich sogar ihre Pflicht als Leiter einer Institution gewesen, die sozialwissenschaftlich untersucht wird. Zudem wäre es auch in ihrem Interesse als Leiter gewesen, wissenschaftliche Ergebnisse zu produzieren, die einem möglichen Imageschaden für ihre Institution in der Öffentlichkeit vorbeugen.

Teilnehmende Beobachtung
In der teilnehmenden Beobachtung wirkt die Forschungsperson selbst in den von ihr untersuchten Situationen mit.

Im Sinne des diese Methode propagierenden kanadischen Soziologen Erving Goffman (1922–1982) handelt es sich bei psychiatrischen Kliniken und Gefängnissen um totale Institutionen (Goffman 1973), deren Regeln und Rituale von den Forschenden von innen erfahren werden müssen, um sie wissenschaftlich zu analysieren. Verbunden war damit der Anspruch, Missstände öffentlichkeitswirksam aufzudecken und für grundlegende Reformen dieser Institutionen zu sorgen. In Deutschland wurde die Methode der verdeckten teilnehmenden Beobachtung durch den Journalisten Günter Wallraff und seinen Bestseller „Ganz unten“ (Wallraff 1985) einer breiten Öffentlichkeit bekannt.

Der amerikanische Psychologe Stanley Milgram (1933–1984) täuschte bei seinen Experimenten zur Unterordnung unter Autoritäten seine Versuchspersonen (vgl. Friedrichs 2014, S. 83f.). Diese glaubten, sie würden an Lernexperimenten teilnehmen. Auf Anordnung des autoritär auftretenden Versuchsleiters quälte eine beträchtliche Anzahl der Teilnehmer andere Menschen mit vermeintlichen Elektroschocks, wenn diese nicht exakt die zu lernenden Worte reproduzierten. Stanley Milgram rechtfertigte sein Vorgehen damit, dass nur so die Unterwerfung unter Autoritäten psychologisch zu studieren gewesen sei. Hätten er und seine Mitarbeiter die über Zeitungsannoncen angeworbenen Untersuchungspersonen über die wahren Ziele der Experimente in Kenntnis gesetzt, hätten die Personen vermutlich erst gar nicht an dem Experiment teilgenommen. Der amerikanische Psychologenverband American Psychological Association (APA) war anderer Ansicht und schloss Stanley Milgram mehrere Jahre aus dem Fachverband aus. Allerdings wurde er bald wieder in Ehren aufgenommen. Ein Grund dafür war vermutlich, dass seine Experimente weltweit einen enormen Widerhall fanden.

Moralische Bewertungen fallen im Rückblick anders aus als solche, die ein Handelnder in der betreffenden Situation vornimmt. Welche konkreten moralischen Abwägungen Goffman und Milgram vor ihren Untersuchungen anstellten, lässt sich heute nicht mehr rekonstruieren. Im Rückblick jedoch scheint ihr Täuschungsverhalten ethisch vertretbar, da sie mit ihren Forschungen wichtige Erkenntnisse erzielt haben, die für demokratische Gesellschaften, welche auch ihre Institutionen demokratisch gestalten müssen, von Nutzen sind. Getäuscht haben die beiden Forscher allein Menschen in ihrem unmittelbaren Untersuchungsumfeld, nicht jedoch die wissenschaftliche oder allgemeine Öffentlichkeit.

Bewusste Täuschungen der wissenschaftlichen und der allgemeinen Öffentlichkeit dahingegen sind moralisch verwerflich und werden bei Bekanntwerden von den zuständigen Fachgesellschaften sanktioniert. Eine ganze Reihe bekannter Psychologen wurde in den zurückliegenden Jahren des Datenbetrugs überführt. Ein bekanntes Beispiel ist der niederländische Sozialpsychologe Diederik Stapel (geb. 1966), der im Anschluss seinen Lehrstuhl an der Universität Tilburg räumen musste. Er hat Diskriminierungen und Aggressionen gegenüber Fremden untersucht, wozu er u. a. verdeckte teilnehmende Beobachtungen an niederländischen Bahnhöfen durchführte. Im Ergebnis empfahl er in Gesprächen mit niederländischen Regionalregierungen, die für die Bahnhöfe zuständig waren, bauliche Veränderungen. Wie er in seinem Buch selbst einräumte, standen hinter seinem Verhalten egoistische Motive: Er wollte durch seinen Betrug mehr öffentlichen Einfluss gewinnen und seinen Bekanntheitsgrad erhöhen.

Ethisch korrekte Durchführung empirischer Forschung

Fachgesellschaften definieren ethische Standards in der Wissenschaft, die für ihre Mitglieder verpflichtend sind. Studierende sollten sich in der empirischen Forschung an ethische Grundregeln halten, die in Übereinstimmung mit den entsprechenden Standards der psychologischen Fachgesellschaften stehen.

Drei zentrale ethische Prinzipien in der empirischen Sozialforschung:

- **Transparenz gegenüber den untersuchten Personen:** Diese müssen über die Ziele der empirischen Untersuchung vorher aufgeklärt werden. Vor, während und nach der Datenerhebung dürfen die untersuchten Personen nicht getäuscht und ihre Aussagen nicht verfälscht werden.
- **Freiwilligkeit:** Die untersuchten Personen dürfen nicht zur Teilnahme an einer Untersuchung gezwungen werden. Es muss ihnen mitgeteilt werden, dass ihre Teilnahme freiwillig ist und sie sich dagegen entscheiden können.
- **Anonymisierung:** Durch die Untersuchung und mögliche spätere Veröffentlichung der Ergebnisse dürfen den untersuchten Personen keine Nachteile in ihrem beruflichen oder privaten Umfeld entstehen. Deshalb wird ihnen zu Beginn der Befragung oder Beobachtung zugesichert, dass ihre Mitteilungen vertraulich behandelt und in einer späteren Veröffentlichung ihre Namen nicht genannt werden.

Anonymisierung
In der Fachliteratur wird zwischen Anonymisierung und Pseudonymisierung unterschieden. Letztere meint, dass die Namen der untersuchten Personen durch Codes (Buchstaben, Zahlen) ersetzt werden.

Die Bedeutung gesetzlicher Bestimmungen zum Datenschutz

Bundesregierung und Landesregierungen sowie die Europäische Kommission formulieren die Gesetze, die in Deutschland gelten. Gesetze regeln auch moralische und ethische Fragen und regulieren das Verhalten der Bürgerinnen und Bürger eines Landes. Es gibt bisher keine Gesetze, die von der Politik speziell mit Blick auf die empirische Sozialforschung formuliert wurden. Von Bedeutung für den Kontakt mit den Untersuchungspersonen und den sonstigen an der Forschung beteiligten Personen und Institutionen ist in Deutschland natürlich das Grundgesetz. Dessen Artikel müssen im empirischen Forschungsprozess beachtet werden. Wichtige Grundgesetzartikel sind in diesem Zusammenhang selbstverständlich Artikel 1 („Die Würde des Menschen ist unantastbar.“) und Artikel 2 (allgemeine Persönlichkeitsrechte). Aus diesen beiden Artikeln leitet sich das Recht jedes Bürgers auf informationelle Selbstbestimmung ab. Dies ist das Recht, grundsätzlich selbst über die Preisgabe und Weiterverwendung seiner persönlichen Daten zu bestimmen (vgl. Mühlchen 2014, S. 93).

Gesetze zum Datenschutz fußen auf diesem Rechtsprinzip. Ein Beispiel ist die 2018 in Kraft getretene und öffentlich vehement diskutierte Datenschutz-Grundverordnung der Europäischen Kommission. Für die empirische Sozialforschung ist sie kaum von Relevanz, da sie sich auf Geschäftsbeziehungen zwischen Unternehmen bzw. von Kunden und Unternehmen bezieht. Sie hat z. B. die Kunden sozialer Netzwerke im Internet im Blick und stärkt deren Rechte gegenüber den digitalen Unternehmen. Zwischen den Untersuchungspersonen und den Forschenden existiert in empirischen Forschungsprojekten in der Regel aber keine geschäftliche Beziehung.

Für Studierende ist es in der Regel nicht zwingend erforderlich, sich vor der Durchführung ihrer empirischen Forschungsprojekte detailliert mit rechtlichen Fragen zu beschäftigen und die entsprechenden Gesetze oder Verordnungen zu lesen. Die strikte Befolgung der oben aufgeführten drei Grundsätze (Transparenz, Freiwilligkeit und Anonymisierung) reicht zumeist aus.

1.9 Themenübergreifende Erkenntnisziele der empirischen Psychologie

Psychologische Forschung ist eingebunden in einen größeren gesellschaftlichen Rahmen. Diesen füllen Gruppen, Verbände, Organisationen und Institutionen aus. Interessensgruppen und Stakeholder richten Erwartungen an die psychologische Forschung, vergeben Aufträge und setzen dabei selbst Forschungsziele. So vergeben etwa Professoren Themen für Bachelor- und Masterarbeiten. Drittmittelprojekte bestimmen das wissenschaftliche Leben vieler forschenden Psychologen. Industrie, Wirtschaft, Verwaltungen und Regierungen finanzieren Forschungsaufträge. Ausschreibungen definieren Anforderungen an die Forschenden und den Forschungsprozess. Sie bestimmen über die Themen und damit indirekt auch über die anzuwendenden Methoden.

In verschiedenen Forschungskontexten kommen der empirischen Psychologie verschiedene Aufgaben zu. Oft sind diese in den Hauptforschungsfragen eines empirischen Projekts enthalten. Die Forschenden sollten deshalb selbst klären, welche übergeordneten Erkenntnisziele in ihren Forschungsaufträgen und Forschungsfragen stecken. In den Lehrtexten der empirischen Sozialforschung werden im Allgemeinen der psychologischen Forschung die folgenden Aufgaben zugeschrieben: Beschreiben, Entdecken, Erklären, Vorhersagen oder Bewerten (vgl. zum Folgenden Hussy/Schreier/Echterhoff 2013, S. 11f.).

Beschreiben: Psychologische und soziale Verhältnisse in ihren quantitativen und/oder qualitativen Relationen zu beschreiben, das gilt als eine wesentliche Aufgabe empirischer Forschung. Beschreiben heißt, sich ein Bild über einen Wirklichkeitsbereich zu verschaffen, idealerweise möglichst vollständig und objektiv.

Entdecken: Die Entdeckung geht über die Beschreibung hinaus. Neue, unbekannte Sachverhalte geraten in den Blick, die bisher nicht in der Theorie oder in anderen empirischen Studien Berücksichtigung fanden. Die Entdeckung geht von der Empirie aus. Die Forschenden machen neue Erfahrungen, die sich zunächst nicht durch die gängigen fachspezifischen Schablonen beschreiben lassen.

Erklären: Hier spielen Modelle und Theorien eine wichtige Rolle. Diese fungieren als Erklärungsbrille, zu der die Empirie die passenden Daten liefert. Sie kann aber auch dazu führen, dass die bisher verwendeten theoretischen Erklärungen für einen Wirklichkeitsbereich als falsch, unpassend oder zu kurz greifend bewertet werden.

Vorhersagen: Forschung kann auch auf die Zukunft gerichtet sein. Dann geht es oft um Prognosen, wie sich Menschen, Organisationen, Umsätze oder politische Präferenzen in einem Wirklichkeitsausschnitt in einem definierten Zeitraum entwickeln werden. Dann genügt die Datenerhebung zu einem Zeitpunkt in einer Stichprobe nicht. Zusätzlich sind weitere Datenanalysen erforderlich. In diesem Zusammenhang kommen statistische oder mathematische Modelle der Prognostik zum Einsatz – wie etwa die logistische Regression –, oder es werden Daten zu mehreren Zeitpunkten an der gleichen Stichprobe erhoben und daraus Trends für die Zukunft extrapoliert.

Bewerten: Bewertungen sind der Kernauftrag bei Evaluationen (vgl. ausführlich Döring 2014). Zu evaluieren sind für Psychologen und Sozialwissenschaftler die Wirkungen von psychologischen oder sozialen Interventionen, etwa von Psychotherapien, medizinischen Präventionsprogrammen oder Rehabilitationsmaßnahmen für Kranke, Benachteiligte oder Strafgefangene. Hier gelangen diverse Methoden der empirischen Sozialforschung zum Einsatz. Ziel ist immer eine umfassende Bewertung des Evaluationsgegenstandes. Erforderlich ist ein abschließendes, empirisch begründetes Urteil, das die verschiedenen empirischen Bewertungskriterien, deren qualitative und/oder quantitative Ausprägungen bestimmt worden sind, in einen intersubjektiv gültigen Zusammenhang stellt.



RÜCKGRIFF AUF DAS BEISPIEL STUDIENMOTIVATION ALS ERLÄUTERUNG

Bei dem dargestellten Beispiel zur Studienmotivation ist das übergeordnete formale Erkenntnisinteresse eine Beschreibung. Es sollen die verschiedenen Motivationslagen in einer Stichprobe von Studierenden beschrieben werden. In der quantitativen Variante stehen die Beschreibungskategorien theoretisch fest; unbekannt und zu untersuchen sind deren Ausprägungen. In der qualitativen Skizze wird zusätzlich die Möglichkeit des Entdeckens von etwas Neuem mitgedacht. Entdeckt werden können unerwartete Motive von Studierenden.



ZUSAMMENFASSUNG

Empirische Forschung in der Psychologie ist eng verflochten mit theoretischen Modellen und Theorien. Aus ihnen leiten sich Fragestellungen und Erkenntnisziele ab. Die Ergebnisse eines Forschungsprozesses können zu Veränderungen in den bestehenden psychologischen Theorien führen.

Damit die Ergebnisse korrekt ausfallen, ist während des Forschungsprozesses eine Orientierung an den drei Gütekriterien Objektivität, Reliabilität und Validität erforderlich. Im Umgang mit dem Forschungsfeld und den untersuchten Personen spielen außerdem ethische, moralische und juristische Aspekte eine Rolle, wie z. B. der Datenschutz.

Es werden qualitative und quantitative Methoden unterschieden. Sie unterscheiden sich u. a. in den Datenformen (verbale Protokolle vs. Zahlen), in der Erkenntnisrichtung (Gemeinsamkeiten in den Daten finden vs. Unterschiede zwischen Fällen bzw. Gruppen messen) und dem Ausmaß des Einbezugs von Situationskontexten (deutlich vs. gering) bei der

Untersuchung von Fällen bzw. Personen. Dementsprechend unterschiedlich sind die Formen der Datenerhebung und der Datenauswertung.

LEKTION 2

DATENERHEBUNG

LERNZIELE

Nach der Bearbeitung dieser Lektion werden Sie wissen, ...

- wie unterschiedliche Erhebungsmethoden funktionieren und wie sie kombiniert werden können.
- wie sich qualitative und quantitative Erhebungsmethoden in ihren Stärken und Schwächen voneinander unterscheiden.
- welche Erhebungsmethoden bei bestimmten Fragestellungen und/oder Erkenntniszielen zu bevorzugen sind.
- wie typische Fehler in der Datenerhebung vermieden werden können.

2. DATENERHEBUNG

Einführung

„Leben achtsame Menschen nachhaltiger?“ Diese interessante Frage ist zugleich die Überschrift eines kurzen Artikels der Zeitschrift „Psychologie heute“ in der Februarausgabe von 2019 (Wetzel 2019). Die Annahme erscheint auf den ersten Blick plausibel. Wer bewusster lebt, müsste auch ökologisch bewusster konsumieren. Achtsamkeit und Konsumverhalten sind jedoch zwei unterschiedliche Forschungsbereiche.

Wie können sie also in der psychologischen Datenerhebung miteinander verbunden werden? Die Autorin berichtet über eine komplexe empirische Studie von drei deutschen bzw. Schweizer Umweltpsychologen (Geiger/Grossman/Schrader 2019), in der verschiedene Erhebungsmethoden miteinander kombiniert wurden. Mit den Teilnehmern der Studie wurde zunächst ein Meditations- und Achtsamkeitstraining durchgeführt, das die innere Haltung zum persönlichen Konsum verändern sollte. Vor und nach Ende des Trainings wurden die Teilnehmer mittels eines standardisierten Fragebogens zu ihrer Achtsamkeit, ihren ethischen Werten und ihrem konkreten Konsumverhalten befragt. Nach dem Training führten die Teilnehmer ein Tagebuch zur Achtsamkeit und zum Konsumverhalten im Alltag. Sieben Monate nach Trainingsende gab es eine dritte Befragungsrunde zu den Themen Achtsamkeit, Werte und Konsum.

Diese anspruchsvolle Studie hatte zum Ziel, achtsames Verhalten zu fördern und auf das alltägliche Konsumverhalten zu übertragen. Dabei wurden als Erhebungsmethoden Experimente, Befragungen und Selbstbeobachtungen kombiniert. In dieser Lektion werden diese drei und weitere psychologische Erhebungsmethoden beschrieben, zunächst in ihren einfachen Formen, danach in komplexeren Varianten und Kombinationen. Übrigens sind die Ergebnisse der Studie der drei Umweltpsychologen ernüchternd. Trotz des Achtsamkeitstrainings und der Anleitung zur Selbstbeobachtung mittels Tagebüchern änderte sich das Konsumverhalten in Bezug auf Kleidung und Ernährung nicht (Wetzel 2019, S. 6).

2.1 Überblick zu den empirischen Methoden der Datenerhebung

Es gibt unendlich viele Fragestellungen für empirische Forschungen in der Psychologie. In diesem Studienbrief werden einige allgemeine Hinweise anhand idealtypischer Beispiele gegeben, in welche Methoden der Datensammlung bestimmte Forschungsfragen im Allgemeinen münden. Die Methoden der **Datenerhebung** dienen dazu, die Daten an den für die empirische Untersuchung ausgewählten Fällen in einer Stichprobe zu sammeln. Die Erhebungsmethoden legen den Rahmen für den Forschungszugriff auf die untersuchten Fälle fest. Die zu Beginn eines Forschungsprozesses formulierte Frage bestimmt die Richtung von Erhebung und Auswertung.

Datenerhebungen
Sie finden an Fällen (Personen, Gruppen, Unternehmen u. a.) statt. An diesen wird die quantitative oder qualitative Ausprägung von Merkmalen (Variablen) ermittelt.

Fragestellungen bei psychologischen Interventionen als Beispiel

In der Psychologie wird häufig die Wirkung von Interventionen überprüft. Interventionen können z. B. Fortbildungen für Mitarbeiter in Unternehmen, verschiedene Präventionsprogramme im Gesundheitsbereich oder unterschiedliche Formen von Psychotherapie für ähnliche psychische Beeinträchtigungen sein.

Die Hauptforschungsfrage ist bei Interventionen oft ein Wirkungsvergleich. Eine vergleichende Wirkungsanalyse zielt auf das Ausmaß der durch die Intervention erreichten positiven Effekte in verschiedenen Fallgruppen, die an unterschiedlichen Varianten der Interventionen partizipieren. Die Datenerhebung teilt die untersuchten Fälle in Gruppen ein, z. B. bei verschiedenen Fortbildungen in die Teilnehmer von Seminar A, Seminar B, Seminar C usw. Daraus entsteht ein **quantitativ-experimenteller Versuchsplan** mit unabhängiger Variable (Seminarform) und abhängiger Variable (positive Effekte bei den Seminarteilnehmern, z. B. die Selbsteinschätzung von Lerneffekten). Quantitativ-experimentelle Versuchspläne bestehen aus mehreren Versuchs- und Kontrollgruppen. In der Auswertung wird eine vergleichende quantitative Analyse zwischen den Gruppen durchgeführt. So kristallisieren sich wirksame und weniger wirksame Interventionsformen heraus.

Quantitativ-experimenteller Versuchsplan

Besteht aus mindestens einer unabhängigen und einer abhängigen Variablen. Die Werte der abhängigen Variablen werden als Daten erhoben und ausgewertet.

Eine offene Fragestellung, ohne die Definition unabhängiger und abhängiger Variablen, ist bei der Untersuchung von Interventionen ebenfalls möglich. Die Forschungsfrage wäre dann, nach welchen Kriterien eine bestimmte Intervention von den Teilnehmenden bewertet wird. Alle Teilnehmenden der Intervention werden am Ende des Interventionsprogramms gebeten, eine umfassende Bewertung der Intervention schriftlich (in einem Fragebogen) oder mündlich (in Interviews) abzugeben. Das Erkenntnisinteresse besteht darin, herauszufinden, welche Bewertungsdimensionen den Teilnehmern des besuchten Seminars überhaupt wichtig oder geläufig sind. Ziel ist es, die wichtigsten Dimensionen der Bewertung zu bestimmen. Das wäre eine qualitative Studie mit Interviews als Erhebungsmethode.

Stehen die Dimensionen der Bewertung durch vorherige Forschungen in einem Themenfeld vor der Intervention bereits fest, empfiehlt sich der Einsatz quantitativer Fragebögen. Für gängige psychologische Interventionen wie Trainings zum Stressabbau, Anti-Angst-Programme, Selbstkontrolltrainings u. ä. liegen Tests in standardisierter Form vor.

Individuelle Erkenntnisinteressen und Werthaltungen der Wissenschaftler in Bezug auf das Untersuchungsfeld haben ebenfalls Einfluss auf die Entscheidung der geeigneten Methode der Datenerhebung. Gelten die zu untersuchenden Fälle qua beruflicher Erfahrung oder durch besondere Betroffenheit als besondere Experten für das Forschungsthema, kommen offene qualitative Interviews als Methode infrage. Die Forschenden sehen sich gegenüber den Befragten in der Rolle der Lernenden. Qualitative Experteninterviews haben sich so als Methode in der Wirtschaftspsychologie und Karriereforschung etabliert. Eine Alternative dazu sind Verfahren der Wissensdiagnostik in der kognitiven Psychologie.

Sind Wissenschaftler der Auffassung, dass die untersuchten Fälle am meisten über sich und das Untersuchungsthema verraten, wenn diese sich in ihrer lebensweltlichen Umgebung, also etwa am Arbeitsplatz oder in ihrem privaten Umfeld befinden, so sind oft Formen der Beobachtung die Methode der Wahl. Geht es dagegen um die Überprüfung genau

definierter Wirkungszusammenhänge in sehr elaborierten, in der entsprechenden Forschungsliteratur dargestellten theoretischen Zusammenhängen, bieten sich Experimente an – dies umso mehr, wenn die entsprechende Forschungstradition in einem Bereich experimentell ausgerichtet ist. Ein Beispiel sind die Experimente zur Analyse von Reproduktionsleitungen des Langzeitgedächtnisses.

Grundformen der Datenerhebung: Befragung, Beobachtung und Experiment

Es gibt in der Psychologie drei Grundformen der Datensammlung: Befragungen, Beobachtungen und Experimente. Dies gilt für die qualitative wie für die quantitative Forschung, wenngleich die entsprechenden Datenformen unterschiedlich sind. Die Grundformen beinhalten drei verschiedene Modi der methodisch kontrollierten Gestaltung des **Untersuchungsfeldes** durch die Forschenden. Die drei Grundformen legen die generelle Umgangsweise der Wissenschaftler mit den von ihnen untersuchten Fällen bzw. Personen bei der Datengewinnung fest. Die Forschenden führen mit den untersuchten Personen entweder ein Gespräch zu einem Thema oder sie beobachten, wie Menschen sich in bestimmten Situationen (z. B. am Arbeitsplatz, bei Freizeitbeschäftigungen u. a.) verhalten. Beim Experiment, der dritten Methode, setzen die Untersucher die Fälle gezielt bestimmten Situationen bzw. Reizen aus, die sie als Forscher selbst herstellen. Die untersuchten Personen sind gezwungen, sich mit diesen Bedingungen auseinanderzusetzen. Sie müssen darauf reagieren und zeigen Verhaltensweisen der Bewältigung. Ursache-Wirkungs-Zusammenhänge werden in der Psychologie mit Experimenten erforscht.

Untersuchungsfeld

Als Untersuchungsfeld wird der Wirklichkeitsbereich bezeichnet, für den die untersuchten Fälle stehen bzw. aus dem die Stichprobe gezogen wurde. Werden z. B. die Mitarbeiter einer Abteilung eines Unternehmens befragt, so ist das Untersuchungsfeld die entsprechende Abteilung.

Befragungen: Befragungen zielen darauf ab, das Wissen, die Meinungen oder die Einstellungen von Personen zu bestimmten Themen einzufangen. Qualitative Befragungen sind häufig Interviews, also direkte Gespräche von Angesicht zu Angesicht. Online- oder telefonische Befragungen sind jedoch ebenso möglich. Grundlage für den Ablauf des Interviews ist ein thematischer Leitfaden mit offenen Fragen. Quantitative Befragungen ohne direkten Kontakt zu den Befragten sind Umfragen, bei denen die untersuchten Personen vorgegebene Antwortmöglichkeiten markieren oder Ziffern bzw. kurze Aussagen eintragen. Vor dem Internetzeitalter wurden solche Befragungen überwiegend schriftlich oder telefonisch durchgeführt. Heute dominiert als Durchführungsform die Onlinevariante.

Beobachtungen: Studien, in denen die Wissenschaftler die von ihnen untersuchten Personen begleiten und deren Verhalten registrieren, ohne selbst auf diese gezielt einzuwirken, sind **Beobachtungsstudien**. Feldbeobachtungen werden direkt in dem Lebens- bzw. Arbeitsumfeld der untersuchten Personen durchgeführt. Verschiedene Formen von Feldbeobachtungen werden nach der Rolle unterschieden, die Wissenschaftler dabei einnehmen. Sie können sich weitgehend passiv verhalten oder kooperieren vor Ort mit den untersuchten Fällen, z. B. als Kollegen am Arbeitsplatz.

Beobachtungsstudien

Bei Beobachtungsstudien im Umfeld der untersuchten Personen füllen die Wissenschaftler entweder selbst eine soziale Rolle aus oder bleiben passive Beobachter.

Experimente: Bei Experimenten werden die Situationen, die die Wissenschaftler untersuchen möchten, von ihnen gezielt hergestellt. Dies hat den Vorteil, dass die Wissenschaftler steuern können, was sie untersuchen möchten und was nicht. Oft wird nach dem Ort der Durchführung psychologischer Experimente zwischen Labor- und Feldexperimenten unterschieden. Zu den experimentellen Untersuchungen zählen zudem alle Untersuchungsformen, in denen unabhängige und abhängige Variablen definiert werden, auch

wenn die Wissenschaftler selbst nicht die unabhängigen Variablen sukzessive im Experiment variieren. Ein Beispiel sind die genannten vergleichenden Wirkungsstudien bei Interventionen.

Zusammengesetzte Erhebungsmethoden

Viele der in Methodenlehrbüchern präsentierten komplexeren Erhebungsmethoden sind Weiterentwicklungen der drei Grundformen oder Kombinationen zweier oder sogar aller drei Varianten. So ist z. B. der in der qualitativen Sozialforschung häufig verwendete Ansatz der **Grounded Theory** eine Kombination aus qualitativer Beobachtung und Befragung. Zugleich ist es ein Forschungskonzept, das Erhebung und Auswertung aufs Engste miteinander verknüpft.

Grounded Theory
Eine Theorie, deren Grundlage (engl. „ground“) die empirisch gewonnenen Daten der Stichprobe sind, wird als Grounded Theory bezeichnet. Die gefundene Theorie steht am Ende der qualitativen Datenanalyse.

Experimente enthalten bei genauerer Betrachtung mindestens eine weitere Methode. Beispielsweise hat Milgram in seinen klassischen Experimenten zur Unterordnung unter Autoritäten seine Versuchspersonen heimlich beobachtet und ihre verbalen Reaktionen während der Experimente auf Tonband aufgezeichnet. Bei experimentellen Studien ist es bisweilen üblich, die Versuchspersonen hinterher zu ihren Erlebnissen zu befragen, um ihr Verhalten während der Experimente besser einschätzen zu können. Solche postexperimentellen Befragungen hat Milgram ebenfalls durchgeführt.

2.2 Qualitative Interviews

Qualitative Interviews haben zum Ziel, subjektive Einstellungen, Meinungen oder Erlebnisse der Befragten zu erheben. Das detaillierte Wissen von Menschen zu einem Thema kann mit qualitativen Interviews ebenfalls gut exploriert werden. Qualitative Interviews werden im direkten Kontakt („face to face“), per Telefon oder im Internet (E-Mail, Chat) geführt. Grundlage ist ein Leitfaden mit offenen Fragen.

Offene Fragen sind sogenannte W-Fragen („Wieso, wann, weshalb, wozu, wo haben Sie ...?“). Sie geben den Befragten die Möglichkeit zu längeren Ausführungen und Erläuterungen – auch zu solchen, mit denen die Forscher bzw. Interviewer vorher nicht gerechnet haben.

In der qualitativen Sozialforschung gibt es verschiedene Einteilungen qualitativer Interviews. Unterschieden werden können die Interviewformen z. B. anhand des Grades der thematischen Vorstrukturierung durch die Forscher bzw. Interviewer zu Erhebungsbeginn. Das führt zu drei Varianten: **Leitfadeninterview**, **narratives Interview** und **episodisches Interview**. Das größte Ausmaß an thematischer Vorstrukturierung liegt beim Leitfadeninterview vor, das geringste beim narrativen Interview. Das episodische Interview liegt zwischen den beiden Polen starker und geringer Vorstrukturierung.

Leitfadeninterview
Ein Leitfadeninterview basiert auf einer schriftlichen Liste zuvor formulierter offener Fragen, die im Interview abgearbeitet werden und so eine Grundstruktur vorgeben.

Beim Leitfadeninterview, oft auch Experteninterview genannt, wird eine Liste zuvor formulierter Fragen abgearbeitet, die das Thema strukturieren. Davon haben alle annähernd die gleiche Relevanz für das Forschungsthema. Beim narrativen Interview hat die aller-

Narratives Interview

Ein narratives Interview beginnt mit einer Erzählungen generierenden Eingangsfrage. Hinterher werden dazu noch Verständnisfragen gestellt.

Episodisches Interview

Das episodische Interview ist eine Mischung aus Leitfaden- und narrativem Interview. Die Befragten sollen markante Erlebnisse aus ihrem Leben erzählen. Ergänzend dazu werden Fragen gestellt.

erste Frage dagegen das größte thematische Gewicht und legt die weitere Strukturierung des Themas in die Hände der Befragten. Diese Eingangsfrage hat das Ziel, dass der Befragte eine längere Geschichte erzählt (Narration).

Beispielsweise kann eine narrative Eingangsfrage lauten: „Bitte erzählen Sie mir ausführlich, was Sie damals bewogen hat, diese zentrale Entscheidung in Ihrem Leben (z. B. Auswanderung, Umschulung, Adoption eines Kindes u. a.) zu treffen.“ Während des anschließenden ausführlichen Statements notiert sich der Interviewer mögliche Nachfragen, die der Vertiefung des Verständnisses der rezipierten Erzählung dienen. Die Nachfragen beziehen sich in erster Linie auf die Erzählung und sprechen keine neuen Themen an. Das episodische Interview zielt ebenfalls auf längere Darstellungen der Befragten. Hier wird mit zuvor formulierten Fragen nach verschiedenen markanten Erlebnissen (Episoden) in Bezug auf das Forschungsthema geforscht.

Merkmale von Leitfaden- und Experteninterviews

Im Alltag hat jeder schon einmal Experten befragt und sich zuvor oft auch eine kurze Fragenliste, also einen Leitfaden, zurechtgelegt, dies zumeist im Kopf. Beispiele sind die Befragung des Hausarztes zu Symptomen und Krankheitsbildern oder das Befragen von Bankmitarbeitern zu Anlagemöglichkeiten oder Krediten. Die Befragten sind die Experten. Ihnen wird die Kompetenz zugesprochen, richtige und kenntnisreiche Antworten zu geben, die den Horizont der Fragenden erweitern. Die Befragung dient dem Einholen umfassender Informationen zu einem Thema.

Diese Merkmale der Befragung von Experten zu Themen des Alltagslebens gelten auch für wissenschaftliche Experteninterviews. Die Forscher sind wie Fragensteller im Alltag in der Rolle der Lernenden – diejenigen, die Auskunft geben, besitzen die Expertise auf dem Fachgebiet, über das gesprochen wird.

Expertise wird in der Regel mit einer besonders hohen beruflichen Qualifikation oder langjähriger Erfahrung in einem Metier gleichgesetzt. Doch auch Laien können durch spezielle Erfahrungen zu Experten werden. Beispiele aus der Psychologie sind Patienten in der klinischen Psychologie, die zu ihren Störungen (Ängste, Depressionen) sehr kompetent Auskunft geben können. In der Wirtschaft geben die von neuen Management- oder Arbeitsmethoden betroffenen Mitarbeiter den Interviewern oft profundere Informationen über die betrieblichen Auswirkungen der neuen Arbeitsregularien als Führungskräfte. Auf jeden Fall ist ihre Perspektive eine andere als die des Managements oder Vorstands. Sie sind Betroffene und erleben hautnah die Vor- und Nachteile betrieblicher Umstrukturierungen.

Der Weg zum thematischen Leitfaden

Helfferrich (2014) definiert einen Leitfaden für Interviews wie folgt:

Der Leitfaden ist eine vorab vereinbarte und systematisch angewandte Vorgabe zur Gestaltung des Interviewablaufs. Er kann sehr unterschiedlich angelegt sein, enthält aber immer als optionale Elemente (Erzähl-)Aufforderungen, explizit vorformulierte Fragen, Stichworte für frei formulierbare Fragen und/oder Vereinbarungen für die Handhabung von dialogischer Interaktion für bestimmte Phasen des Interviews (Helfferrich 2014, S. 560).

Steht die Hauptforschungsfrage fest, sollte vor der Datenerhebung der Leitfaden konstruiert werden. Dazu sind Kreativtechniken, wie z. B. das Brainstorming, hilfreich. Die Forscher überlegen, welche Aspekte zu dem Thema gehören, und listen diese auf. Der nächste Schritt ist die Überführung dieser Aspekte in offene Fragen, die im Interview an die zu befragenden Personen gerichtet werden (vgl. zur Erstellung eines Leitfadens Helfferich 2014, S. 566f.).

Neben dem Inhalt der Fragen des Leitfadens werden vor der Datenerhebung auch Überlegungen zur sprachlichen Passung des Leitfadens auf die Zielgruppe angestellt. Das betrifft vor allem die Formulierung und Verständlichkeit der Fragen, aber auch Überlegungen zur Länge des Leitfadens. Formulierungen für Kinder sind andere als für Erwachsene, Formulierungen für Führungskräfte andere als für Mitarbeiter.

Was sollte bei der Formulierung von Interviewfragen vermieden werden?

In einem Leitfaden sind fast alle Fragen zum Forschungsthema erlaubt, soweit sie nicht die Persönlichkeitsrechte der Befragten verletzen – und auch dann können die Befragten die Beantwortung einer Frage immer noch ablehnen. Richtlinie für die Formulierung der Fragen eines Leitfadens ist das Prinzip der Offenheit in der qualitativen Sozialforschung (Kleining 1982, S. 8). Daraus ergeben sich drei Regeln, welche Frageformen in qualitativen Interviews unbedingt zu vermeiden sind:

- **keine geschlossenen Fragen:** Das sind Fragen, die mit nur einem Wort bzw. nur mit Ja oder Nein zu beantworten sind („Haben Sie schon einmal X gesehen?“).
- **keine Suggestivfragen:** Das sind Fragen, die den Befragten bestimmte Meinungen oder Ansichten nahelegen („Sind Sie nicht auch der Meinung, dass ...?“).
- **keine Fragen mit quantitativen Vergleichen:** Das sind Fragen, die Vergleiche zwischen Personen oder Gegenständen fordern („Finden Sie X besser als Y?“).

Suggestivfragen

Diese Fragen wollen auf die Meinungen, Ansichten und sogar auf das Handeln ihres Gegenübers einwirken, indem sie bereits bestimmte Annahmen, Meinungen o. ä. vorgeben (suggerieren).

Suggestivfragen sind in der qualitativen wie der quantitativen Sozialforschung verboten. Geschlossene Fragen und Vergleiche sind indes ein wesentlicher und sinnvoller Teil quantitativer Fragebögen. In der qualitativen Sozialforschung wird offenen Fragen mit Interrogativpronomen (W-Fragen) der Vorzug gegeben. Die Befragten erhalten so die Gelegenheit zu längeren Ausführungen. Sie schildern ihre persönliche Sicht auf das Thema.

Um sicherzugehen, dass die Sichtweisen der Befragten im Interview ihren Niederschlag finden, empfiehlt es sich, am Ende jedes Interviews eine Frage nach fehlenden, zuvor noch nicht angesprochenen Themen zu stellen. Eine mögliche Formulierung der Abschlussfrage ist: „Gibt es noch etwas, was bisher nicht behandelt wurde, das aus Ihrer Sicht aber wichtig ist?“ Antworten die Befragten mit Ja, werden sie zum Erzählen aufgefordert und es kann sich eine längere Interviewphase zu einem neuen Aspekt des Forschungsthemas anschließen.

Strikte Rollenaufteilung im Interview

Allen drei Formen qualitativer Interviews (Leitfadeninterview, narratives Interview und episodisches Interview) ist eine streng definierte Rollenaufteilung zwischen Interviewer und Befragten gemeinsam. Damit unterscheiden sich qualitative Interviews von Alltagsge-

Interviewer

Als Interviewer muss nicht unbedingt der Wissenschaftler fungieren, der das Projekt durchführt. Oft sind es auch Personen, die speziell für die Interviews angeworben, geschult und bezahlt werden.

sprächen. Mit eigenen Äußerungen (Meinungen, Ansichten, Bewertungen des Gesagten) muss sich der **Interviewer** zurückhalten. Er befindet sich in der Rolle eines nicht wertenden Zuhörers. Dass das Gesagte inhaltlich verstanden wurde, sollte von ihm signalisiert werden (etwa durch „Mhm“ oder durch explizite Verbalisierungen wie: „Das habe ich jetzt verstanden“). Nachfragen bei Nichtverstehen sind notwendig, damit das Interviewprotokoll hinterher vollständig ausgewertet werden kann (Helfferich 2014, S. 564 und 574).

Es ist ein besonderes Merkmal der qualitativen Sozialforschung, dass Phasen der Erhebung und Auswertung stärker ineinander greifen als in der quantitativen Forschung. In Letzterer wird erst nach Abschluss der Erhebung mit der Auswertung begonnen. In qualitativen Interviewstudien ist es dagegen üblich, nach einzelnen Interviews informell Auswertungen vorzunehmen. Interviewer sind oft angehalten, nach Ende eines Interviews schriftlich zu notieren, welche Besonderheiten ihnen während des Interviews aufgefallen sind. Beispielsweise wird festgehalten, welche Themen durch die Befragten neu aufgeworfen wurden und welche Fragen des Leitfadens sich als wenig ergiebig erwiesen haben. Für die ausstehenden Interviews wird der Leitfaden unter Umständen entsprechend geändert.

Qualitative Befragungen in schriftlicher Form

Qualitative schriftliche Befragungen ohne direkte Anwesenheit von Forschungspersonen bzw. Interviewern werden in der akademischen Forschung weitaus seltener eingesetzt als mündliche Interviews. Grund dafür ist vor allem, dass im direkten Kontakt Befragte besser zu motivieren sind, Antworten zu geben. Wenn Interviewer Interesse signalisieren, fühlen sich die Befragten wertgeschätzt. Außerdem können durch unmittelbares Nachfragen Unklarheiten und Missverständnisse schnell ausgeräumt werden. Durch die Weiterentwicklung der Computertechnik werden heute jedoch über das Internet mittels Onlinebefragungen sehr viele Menschen erreicht. Dies führte in den letzten Jahren zu einem Aufschwung schriftlicher Interviews, oft auch in der Mischform zwischen quantitativer und qualitativer Befragung. Das Versenden von Fragebögen auf postalischem Weg entfällt so.

Ein kurzer Vergleich von mündlichen und schriftlichen Befragungen

Qualitative Leitfadeninterviews sind für qualitative studentische Forschungsarbeiten oft die Methode der Wahl. Sie sind relativ leicht zu erlernen und durchzuführen. Werden Experten zu einem Thema befragt, ist es sicher, dass die Interviewprotokolle gut auswertbare Daten enthalten. Interviews sind zudem eine ethisch gut vertretbare Methode, wenn der Datenschutz eingehalten wird. Die Befragten wissen, dass sie untersucht werden, und willigen explizit in eine Kooperation ein. Sie teilen zudem nur das mit, was sie äußern möchten, und können zu bestimmten Fragen die Auskunft verweigern. Darüber hinaus profitieren sie bisweilen selbst von der Erhebung, da sie bestimmte Themen systematisch neu durchdenken, was zu einer Selbstklärung und Neuorientierung im eigenen Verhalten führen kann. Der Leitfaden sollte nicht zu viele Details enthalten, sonst besteht die Gefahr der Überfrachtung. Die Befragten entscheiden selbst, ob und wie sie Fragen vertiefen.

Schriftliche Befragungen ohne direkten persönlichen Kontakt zu den Befragten sind gut geeignet für die psychologische Markt- und Meinungsforschung in Form von Onlinebefragungen. Seitens der Befragten ist dafür kein Expertenwissen erforderlich. Die Methode ist mit dem Einsatz des Internets geeignet, rasch eine große Zahl von Meinungen zu einem

eng umgrenzten Thema, etwa die Bewertung eines neuen Produkts, zu erhalten. Qualitative Kunden- oder Nutzerbefragungen werden heute häufig auf diese Weise durchgeführt. Zu bedenken ist dabei allerdings, dass der schriftliche Onlinefragebogen kurz gehalten sein muss, da viele Kunden bzw. Nutzer nicht bereit sind, für Umfragen längere Textpassagen zu verfassen.

Gruppenbezogene Verfahren qualitativer Befragung

Gruppenbezogene Befragungen sind eine besondere Variante qualitativer Interviews. Grundlage ist auch hier ein zuvor erstellter thematischer Leitfaden. Die Gruppe kommt zu gemeinsamen Treffen zusammen, die von einem Leiter moderiert werden. Sie ist homogen bezüglich eines für die Fragestellung zentralen Merkmals. Beispielsweise bilden alle Mitglieder eines Arbeitsteams in einer Firma, alle Kunden des gleichen Unternehmens oder alle von der gleichen psychischen Beeinträchtigung Betroffenen in diesem Punkt eine homogene Gruppe. Die Gruppengröße wird so gewählt, dass alle Gruppenmitglieder während einer Gruppensitzung ausführlich zu Wort kommen können. Das ist eher bei kleinen Gruppen gewährleistet. Größere Gruppen sind hingegen oft schwer zu moderieren, die einzelnen Teilnehmer geraten für die Moderatoren aus dem Blick. Vogl (2014, S. 584) nennt als Obergrenze eine Anzahl von zwölf Teilnehmern für eine Gruppendiskussion.

Gruppendiskussionen als Methode werden in fast allen Lehrbüchern zur empirischen Sozialforschung beschrieben. Ihren wissenschaftlichen Ursprung hat diese Erhebungsmethode in der US-amerikanischen Markt- und Meinungsforschung der Nachkriegszeit. Die damaligen Sozialpsychologen und Soziologen bezeichneten ihren Ansatz als Fokusgruppe oder fokussiertes Interview. Der Fokus der Gruppe wird durch einen von den Untersuchern gesetzten Diskussionsanreiz bestimmt. Beispielsweise wird zu Beginn der Gruppensitzung ein Werbefilm gezeigt, danach diskutieren die Gruppenmitglieder über diesen Werbefilm anhand der Fragen eines Leitfadens.

Der Leitfaden ist bei **Gruppendiskussionen** ähnlich wie bei qualitativen Interviews. Jedoch befindet sich der Gruppenleiter hier in einer anderen Rolle: Gruppendiskussionen müssen moderiert werden. Der Moderator muss Diskussionen initiieren, weiterlaufen lassen und ggf. Konflikte schlichten. Immer geht es darum, die Gruppe als Ganzes im Blick zu haben. Der Moderator darf sich nicht in Einzelgespräche verwickeln lassen. Jede Gruppendiskussion entfaltet überdies eine spezielle Gruppendynamik. Für den Moderator ist oft nicht vorhersehbar, in welche Richtungen Diskussionen verlaufen und ob es zu persönlichen Auseinandersetzungen kommt, die ein Eingreifen durch die Moderation erfordern. Diskussionen können sich als besonders lebendig oder extrem mühsam erweisen. In beiden Fällen ist ein anderes Verhalten des Moderators gefordert.

Gruppendiskussion

In einer Gruppendiskussion werden die gemeinsamen Erfahrungen aller Gruppenmitglieder thematisiert, z. B. ähnliche Erfahrungen am Arbeitsplatz oder ähnliche Erfahrungen aufgrund einer gesellschaftlichen Stellung, z. B. arbeitslos zu sein.

Bezüglich des Aufwandes in der Datenerhebung haben Gruppendiskussionen den Vorteil, dass oft nur wenige Sitzungen erforderlich sind, unter Umständen sogar nur eine einzige. Befragungen von Arbeitsteams sind in Form von Gruppendiskussionen daher effizienter als Einzelinterviews. Gruppendiskussionen bieten den Vorteil der komprimierten Datenerhebung. Ein Nachteil kann sein, dass sich die Gruppenmitglieder in einer Diskussion mit ihnen bekannten Personen, z. B. den Arbeitskollegen und Vorgesetzten, weniger offen

äußern als in Einzelinterviews. Wenn die Betroffenen Nachteile befürchten, falls ihre Kollegen oder Vorgesetzten von ihren Aussagen erfahren, halten sie sich im Allgemeinen stärker zurück als in schriftlichen Befragungen oder persönlichen Interviews.

Dialogische Gruppenintrospektion

Introspektion
Die nach innen gerichtete Beobachtung einer Person bezeichnet man als Introspektion. In der Psychologie wird diese Selbstbeobachtung in der Regel an markante Außenreize (Filme, Bilder, komplexe Denkaufgaben) gekoppelt. Die Person soll beobachten, welche Gedanken und Gefühle durch den Außenreiz bei ihr innerlich ausgelöst wird.

Eine neuere Form der gruppenbezogenen Datenerhebung ist die dialogische **Introspektion** in Gruppen (Burkart/Kleining/Witt 2010). Sie verbindet die Methode der retrospektiven Selbstbeobachtung aus der kognitiven Psychologie mit der Fokusgruppe: Die Teilnehmer werden zu Beginn der Sitzung einem ihnen zuvor unbekanntem medialen Reiz ausgesetzt, z. B. einem Film oder einem Musikstück. Anschließend notieren sie die an sich selbst wahrgenommenen Gedanken und Empfindungen. Alle Selbstbeobachtungsprotokolle werden vorgelesen und in ein gemeinsames Gruppenprotokoll überführt. Dieses wird in der Auswertung mittels qualitativer Text- bzw. Inhaltsanalyse zu Kategorien verdichtet. Eine Diskussion über die Wahrheit oder Angemessenheit der Selbstbeobachtungen oder gar ein Streit mit dem Austausch von Pro- und Kontraargumenten finden bei der dialogischen Introspektion nicht statt. Es geht um ein vollständiges Zusammentragen verschiedener Eindrücke in der Datensammlung ohne Bewertungen und diskursives Abwägen.

Befragung von Texten (qualitative Sekundäranalysen)

Sekundäranalyse
In der Sekundäranalyse liegen die Datenprotokolle bereits vor. Sie wurden in anderen Zusammenhängen erstellt. Es kann sich um wissenschaftlich erhobene Daten handeln, aber auch um nicht wissenschaftliche.

Texte sind Träger von Informationen und geben Menschen Auskünfte zu ihren Fragen zu Themen, die in den Texten behandelt werden. Qualitative Datenanalysen können auch an Texten und Protokollen ansetzen, die nicht in systematischen Datenerhebungen erstellt wurden. Derartige Texte sind oft in forschungsunabhängigen Kontexten ohne Forschungsabsicht entstanden (z. B. als Autobiografie oder als Zeitungstext). In der empirischen Sozialforschung wird diese Herangehensweise im Allgemeinen als **Sekundäranalyse** bezeichnet. Der Begriff der Sekundäranalyse wurde ursprünglich in der quantitativen Sozialforschung geprägt. Er meint dort, dass bereits erhobene und/oder veröffentlichte Statistiken unter neuen Forschungsfragestellungen analysiert werden. In der qualitativen Sozialforschung ist der Begriff der Sekundäranalyse vorhandener Daten jedoch erst seit kurzer Zeit gebräuchlich. Aufgrund der Kontextgebundenheit qualitativer Daten ist umstritten, ob eine Loskoppelung von der Erhebung in der qualitativen Forschung über eine Sekundäranalyse, ähnlich wie bei quantitativen Daten, möglich ist (vgl. Medjedović 2014, S. 226f.).

Mögliche Texte für eine qualitative Sekundäranalyse sind u. a. Protokolle von beruflichen Gesprächen, Arbeitsplatzbeschreibungen, autobiografische Aufzeichnungen, journalistische Artikel in Zeitungen und Zeitschriften, wissenschaftliche Gutachten sowie Protokolle von Chats im Internet. Das Erkenntnisziel ist immer ein tieferes Verständnis der fremden Texte vor dem Hintergrund einer leitenden Forschungsfrage. Wer z. B. erforschen möchte, welches Image eine politische Partei oder ein Unternehmen in der medialen bzw. öffentlichen Wahrnehmung genießt, muss entsprechende Darstellungen in Medien als Datenquelle nutzen. Zunächst ist zu entscheiden, welche Medien untersucht werden (Presse, Fernsehen oder Internet), danach ist aus der Vielzahl an Texten zu der betreffenden Partei bzw. über das Unternehmen in einem Medienbereich eine eingrenzende Auswahl zu treffen. Diese Auswahl ist nichts anderes als die Bildung einer Stichprobe von Fällen. Die

Datenerhebung bei der Analyse fremder Texte besteht also aus der Vornahme einer medialen Eingrenzung und der Auswahl relevanter Texte aus einem größeren Korpus. Sobald das geschehen ist, beginnt die qualitative inhaltsanalytische Auswertung.

2.3 Quantitative Umfragen mittels Fragebögen

Quantitative Befragungen mittels Fragebögen gehören zu den meistverwendeten Methoden in der Sozial- und Evaluationsforschung. Befragt werden Menschen zu Wissen, Meinungen, Einstellungen und Verhaltensweisen. Jeder quantitative Fragebogen enthält außerdem zu Beginn oder am Ende einige Fragen zu soziodemografischen Variablen (Alter, Geschlecht, Schulabschluss, Staatsangehörigkeit, Familienstand, Beruf u. a.).

Auf dem Fragebogen nehmen die Befragten Eintragungen vor. Sie markieren etwas oder setzen Zahlen oder Worte ein. Die Fragen sind von den Wissenschaftlern vorformuliert und für alle Befragten identisch (standardisiert). Häufig werden standardisierte Fragebögen über einen längeren Zeitraum eingesetzt, um Dienstleistungen oder Produkte kontinuierlich zu bewerten. Das sind z. B. Umfragen zur Kundenzufriedenheit, zur Mitarbeiterzufriedenheit oder kontinuierliche Evaluationen von Fortbildungen und Seminaren.

Quantitative Befragungen werden heute oft online durchgeführt. Das hat den Vorteil, dass die entsprechende Software aus den Antworten der Befragten eine Datentabelle generiert, die als Grundlage für die statistische Auswertung fungiert. Die Daten müssen also nicht von den Wissenschaftlern selbst in das Statistikprogramm eingegeben werden. Zusätzlich können in etlichen Statistikprogrammen die Auswertungen vorprogrammiert werden, sodass die Ergebnisse nach der Eingabe sehr schnell vorliegen oder fortlaufend aktualisiert werden können.

Quantitative Befragungen sind, verglichen mit den meisten anderen hier besprochenen Methoden, leichter zu lernen, routinemäßiger einsetzbar und ökonomischer auswertbar. Das verführt Anwender allerdings leicht zu Quick-and-Dirty-Lösungen. Sie entwickeln Fragebögen minderer Qualität, die nicht den Anforderungen der **statistischen Gütekriterien** genügen.

Fragenformulierung

Worauf ist für Anfänger bei der Formulierung der Fragen zu achten? Es gibt einige Grundregeln, denen jeder folgen sollte, der einen quantitativen Fragebogen entwirft:

1. Jede Aussage sollte kurz sein und nur einen Aspekt ansprechen.
2. Es gilt, möglichst ohne Verneinungen in den Formulierungen auszukommen.
3. Die vorgegebenen Antworten zum Ankreuzen auf einer Skala müssen trennscharf sein.

Statistische Gütekriterien

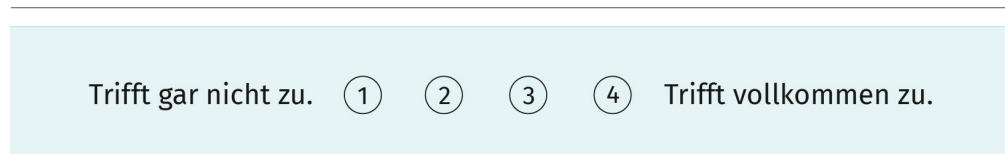
Statistische Gütekriterien sind Objektivität, Reliabilität und Validität. Wenn Fragebögen nach der Konstruktion einem Pretest unterzogen werden, werden die Gütekriterien zum Teil quantitativ (statistisch) ermittelt, vor allem über Korrelationen. Fragebögen, die im Pretest schlechte Werte bei den Gütekriterien erzielen, werden nicht verwendet oder umgearbeitet.

Antwortformate

Das Antwortformat ist die Form, in welcher die Antwort von den Befragten eingefordert wird. Es gibt verschiedene Antwortformate mit den jeweils damit verbundenen Skalen in standardisierten Befragungen (vgl. Reinecke 2014, S. 605f.). Im Rahmen dieses Studienbriefs spreche ich drei Formen an: LIKERT-Skala, dichotome Antwortformate und Häufigkeitsangaben von Verhalten.

Wenn es darum geht, Themen, Gegenstände (Produkte) oder eigene Erlebnisse zu bewerten, ist die LIKERT-Skala in der empirischen Sozialforschung das häufigste Antwortformat. Sie besteht aus mehreren linear ansteigenden Zahlen als Stufen.

Abbildung 4: Vierstufige LIKERT-Skala

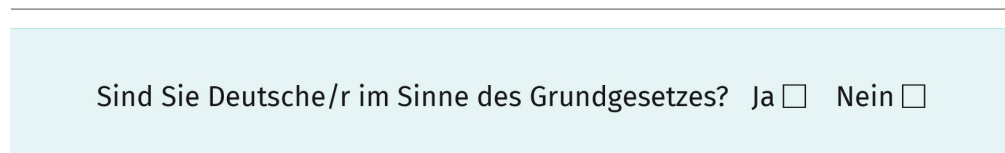


Quelle: erstellt im Auftrag der IU, 2022.

Die Bezeichnung dieser Skalenform geht auf deren Erfinder, den amerikanischen Sozialwissenschaftler Rensis Likert (1903–1981), zurück. LIKERT-Skalen werden auch als Ratingskalen bezeichnet, weil es bei ihnen um Einschätzungen (engl. „ratings“) geht. Die Anzahl der Stufen variiert in den empirischen Anwendungen; es sollten jedoch mindestens drei sein, denn lediglich zwei Stufen ergäben eine dichotome Skala. Nach oben gibt es keine absolute Grenze. Das Sozio-oekonomische Panel (SOEP), eine regelmäßig vom Deutschen Institut für Wirtschaftsforschung (DIW) durchgeführte Befragung deutscher Haushalte zum Lebensstandard, an der zahlreiche Wirtschaftswissenschaftler und Soziologen mitarbeiten, benutzt bei der Befragung der Personen eine 11-stufige LIKERT-Skala von 0 bis 10 (DIW 2018, S. 4). Das ist die höchste LIKERT-Stufenanzahl in einer empirischen Studie, die dem Autor dieses Studienbriefs bekannt ist.

Dichotome Antwortformate hingegen erwarten von den Befragten ein genaues Ja oder Nein als Antwort. Oft geht es um Klassifikationen bzw. eindeutige Zuordnungen.

Abbildung 5: Beispiel einer Ja/Nein-Zuordnung



Quelle: erstellt im Auftrag der IU, 2022.

Wenn es um Einstellungen und Bewertungen geht, kann eine dichotome Skala um eine neutrale Kategorie ergänzt werden, die den Unentschlossenen und Unsicheren die Möglichkeit der Positionierung gibt.

Dichotomes Antwortformat
Ein dichotomes Antwortformat fordert von den Befragten die Entscheidung zwischen zwei Alternativen (Ja oder Nein, trifft zu oder trifft nicht zu).

Abbildung 6: Beispiel einer Ja/Nein-Zuordnung mit neutraler Option

Sind Sie für die Einführung einer Geschwindigkeitsbegrenzung von 130 km/h auf deutschen Autobahnen? Ja Nein Weiß nicht

Quelle: erstellt im Auftrag der IU, 2022.

Eine andere Gruppe typischer Fragen in einem Fragebogen bezieht sich auf die Häufigkeit bestimmter Verhaltensweisen. Je nach Forschungsfrage kann es erforderlich sein, möglichst ganz genaue Zahlen zur Nutzung, zum Konsum oder zu anderen Verhaltensweisen (z. B. Dosis bei Suchtverhalten) zu erhalten. Ein Beispiel wäre die Untersuchung des Alkoholkonsums in der medizinischen bzw. psychologischen Suchtforschung.

Abbildung 7: Beispiel: Wie oft in der Woche trinken Sie Alkohol?

niemals an 1 bis 2 Tagen an 3 bis 5 Tagen jeden Tag

Quelle: erstellt im Auftrag der IU, 2022.

Diese vier Aussagen zur Auswahl sind trennscharf, da sich die Anzahl der Tage bei den Antwortvorgaben nicht überlappt. Allerdings wird damit weder etwas über die Menge des getrunkenen Alkohols ausgesagt noch über das Getränk (Bier, Wein, Schnaps). Die Frage nach dem Alkoholkonsum pro Woche müsste in mehrere Fragen aufgeteilt werden, um zu genauen Zahlen zu führen.

Skalenniveau

Durch das Antwortformat einer Variablen wird auch deren **Skalenniveau** mitbestimmt. „Skalenniveau“ ist eine Bezeichnung für die numerische Komplexität einer Variablen. Jede Variable liegt auf einem bestimmten Skalenniveau vor. In komplexer werdender Reihenfolge sind das: Nominal-, Ordinal-, Intervall- und Verhältnisskala. Je höher ihr Skalenniveau, desto mehr statistische Auswertungsformen sind bei einer Variablen möglich.

- **Nominalskala:** Bei diesen Variablen ist nur eine Zuordnung der Fälle zu Ausprägungen möglich, ohne dass zwischen den Ausprägungen ein quantitativer Unterschied vorliegt. Beispiele sind die Nationalitäten von Personen oder ihre Telefonnummern. Die Ausprägungen der Werte stehen gleichberechtigt nebeneinander. Die Telefonnummer 0162-767676 ist nicht kleiner als 0162-767699. Für die Ausprägungen solcher Variablen können statistisch nur Häufigkeiten ermittelt werden, z. B. die Anzahl der Personen aus einem Land, aber keine Kennwerte, wie z. B. ein Mittelwert.
- **Ordinalskala:** Hier werden Zuordnungen für Werte einer Variablen getroffen, die eine Größer-Kleiner-Relation enthalten. Ein markantes Beispiel ist die Variable höchster Bildungsabschluss. Ein Abitur ist höher zu bewerten als ein Realschulabschluss. Gleiches gilt für das Ergebnis eines Wettlaufs in Bezug auf die Plätze. Der Sieger steht über dem Zweitplatzierten, dieser über dem Drittplatzierten. Allerdings sind die zeitlichen

Skalenniveau

Das Skalenniveau einer Variablen liegt in einer von vier Varianten vor: Nominal-, Ordinal-, Intervall- oder Verhältnisskala. Intervall- und Verhältnisskalen werden als metrische Skalen bezeichnet.

Abstände nicht einheitlich. Ebenso sind die Abstände zwischen Bildungsabschlüssen nicht numerisch definiert. Deshalb können die Ausprägungen einer Variablen zwar in eine Rangreihe gebracht werden, aber es kann nicht mit den Abständen zwischen den Werten gerechnet werden. Etliche statistische Berechnungen sind für eine Ordinalskala nicht möglich, z. B. keine Mittelwerte und keine Standardabweichungen.

- **Intervallskala:** Im Unterschied zur Ordinalskala sind die Abstände zwischen den Skalenpunkten, den möglichen Ausprägungen aller Werte, einheitlich numerisch definiert. Außerdem gibt es einen Nullpunkt. Ein Beispiel ist der Intelligenzquotient (IQ). Bei Intervallskalen sind die meisten statistischen Auswertungsverfahren anwendbar, wie die gängigen Maße der zentralen Tendenz (Mittelwert, Standardabweichung). Der Nullpunkt der Skala ist aber eine willkürliche Setzung. Deshalb sind Aussagen über Relationen (z. B. ein IQ von 120 sei 20 % höher als einer von 100) nicht möglich.
- **Verhältnisskala:** Der Nullpunkt ist natürlich gegeben. Beispiele sind: Zahl der Kinder in einem Haushalt, Haushaltsbruttoeinkommen oder das Gewicht einer Person. Die anwendbaren statistischen Verfahren decken sich mit denen bei Intervallskalen, da die Abstände zwischen den Ausprägungen der Werte definiert sind. Zusätzlich sind aber auch Aussagen über numerische Relationen und Anteile möglich. Beispiele sind: Eine Person wiegt doppelt so viel wie eine andere oder sie verdient 20 % mehr als eine andere.

Der Weg zum quantitativen Fragebogen

Die Hauptthemen eines selbst erstellten Fragebogens sollten sich an den Theorien und Modellen in einem Forschungsgebiet orientieren. Nur dann ist die durchgeführte empirische Studie anschlussfähig an andere Untersuchungen. Eine Orientierung an einem psychologischen Modell wäre z. B. die Entwicklung eines Fragebogens zur Evaluation der Lehre im akademischen Bereich, der das Modell von Heiner Rindermann (2003, S. 236) als Ausgangspunkt hat. Die Themen eines selbst entwickelten Fragebogens zur Lehrevaluation enthalten verschiedene Aspekte von Rindermanns Modell (Fachkompetenz des Dozenten, eingesetzte Lernmedien, Vorwissen der Lerner zum Thema, subjektiver Lernerfolg u. a.).

Für die Entwicklung eines quantitativen Fragebogens ist fast immer eine vorherige Auseinandersetzung mit der Forschungsliteratur in dem entsprechenden Themengebiet erforderlich. Eine sinnvolle Alternative ist oft der bewusste Verzicht auf eine Eigenentwicklung und der Rückgriff auf bewährte Fragebögen oder Tests. Möglicherweise sind aber das Untersuchungsfeld und die Fragestellungen so speziell, dass eine Neukonstruktion eines Fragebogens notwendig ist. Vor dessen Einsatz sollte ein **Pretest** durchgeführt werden. Mehrere Personen füllen den neuen Fragebogen aus und geben dem Konstrukteur eine Rückmeldung, ob alles verstanden wurde und der Fragebogen fehlerfrei ist.

Pretest

Ein Vortest, der an einer kleinen Stichprobe vor der eigentlichen Studie durchgeführt wird, wird als „Pretest“ bezeichnet. Er dient der Qualitätsverbesserung des Erhebungsinstrumentes.

2.4 Beobachtungen als Methode der Datensammlung

Die qualitative Beobachtung im sozialen Feld hat eine lange Tradition in verschiedenen Humanwissenschaften, besonders in Soziologie, Ethnologie und Kulturanthropologie (Knoblauch 2014, S. 521). Sie ist die dominierende Methode in der Feldforschung bei der Untersuchung von (für den Forschenden) fremden Institutionen, Organisationen, Kulturen und Gesellschaften. Verbunden ist damit in der Regel ein längerer Aufenthalt der Forscher im Untersuchungsfeld.

Damit Verhaltensweisen von Personen in einem fremden Umfeld korrekt beobachtet, kategorisiert und analysiert werden können, muss sich der Wissenschaftler in diesem Feld, ähnlich wie ein Reisender in einem Land mit einer anderen Kultur, erst einmal orientieren und sich mit den Sitten und Gebräuchen der Akteure vertraut machen. Danach erst kann er entscheiden, was er wie beobachten möchte. Dazu entwirft er ein Beobachtungskonzept. Dieses beinhaltet zu Beginn zwei methodische Grundentscheidungen (vgl. Thierbach/Petschick 2014, S. 856):

- Wird die Beobachtung verdeckt (heimlich) oder offen (transparent für die Beobachteten) durchgeführt?
- Verhält sich der beobachtende Wissenschaftler passiv (nicht-teilnehmend) oder übernimmt er Aufgaben der täglichen Praxis im Feld?

Bei der verdeckten Beobachtung wissen die Beobachteten nicht, dass sie Untersuchungsobjekte sind. Auch wenn dies auf den ersten Blick ethisch dubios erscheint, kann es dafür gute Gründe geben. Bei der offenen Durchführung sind die Rollen der Beobachter, die Formen der Datenerhebung und mögliche Folgen aus den Ergebnissen den untersuchten Personen bekannt. Die Untersuchung ist mit den Beobachteten bzw. den für sie verantwortlichen Personen und Vorgesetzten abgesprochen. Die Unterscheidung zwischen teilnehmender und nicht-teilnehmender Beobachtung bezieht sich auf die Frage, ob die Beobachter selbst eine Funktion im Alltag des beobachteten Feldes übernehmen. Eine teilnehmende Beobachtung am Arbeitsplatz bedeutet für einen Beobachter die Übernahme von Arbeitsaufgaben. Er schlüpft in die Rolle des Kollegen der Beobachteten.

Damit gibt es vier verschiedene Möglichkeiten der Ausgestaltung der Beobachterrolle (verdeckt und teilnehmend, verdeckt und nicht-teilnehmend, transparent und teilnehmend, transparent und nicht-teilnehmend). In der Regel wird von den Beobachtern die entsprechende Rolle während der gesamten Untersuchung beibehalten. Die Art der Kombination der beiden Dimensionen Offen-/Verdecktheit und Teilnahme oder nicht ist nicht beliebig. Wer verdeckt beobachtet, entscheidet sich aus Gründen der Unauffälligkeit zumeist für eine teilnehmende Rolle, da sonst die Gefahr der Entdeckung besteht. So hat z. B. der deutsche Journalist Günter Wallraff bei den Recherchen für seinen Bestseller „Ganz unten“ (Wallraff 1985) selbst als Leiharbeiter mit falscher Identität bei der Firma gearbeitet, deren menschenverachtende Praktiken er aufgedeckt hat.

Für Beobachter im sozialen Feld ist es unmöglich, alles zu jeder Tageszeit und überall zu beobachten. Dafür sind die untersuchten Institutionen, Organisationen und Kulturen in der Regel viel zu groß und zu komplex. Deshalb sollten Bereiche in einem Beobachtungsplan definiert werden, auf die sich der besondere Fokus der Beobachtung richtet. Eine mögliche Untergliederung eines sozialen Feldes umfasst die folgenden vier Bereiche (nach Döring/Bortz 2016, S. 326):

1. **Beobachtungsorte (observation sites):** Wo soll beobachtet werden? Wo genau im Feld, an wie vielen unterschiedlichen oder ähnlichen Orten?
2. **Beobachtungszeiten (observation times):** Wann und wie lange soll beobachtet werden?
3. **Beobachtungsobjekte (observation objects):** Welche Personen (z. B. man selbst oder andere Personen), Gruppen, Ereignisse oder Gegenstände sollen beobachtet werden?
4. **Beobachtungseinheiten (observation units):** Welche konkreten Aspekte der Beobachtungsobjekte sollen hinsichtlich ihrer Qualität, Häufigkeit oder Intensität beobachtet werden?

Zur Beobachtung gehört ein Beobachtungsbogen. Darin wird eingetragen, was beobachtet wird. So entsteht ein Beobachtungsprotokoll. Bei einer unstrukturierten qualitativen Beobachtung gleicht dieser Bogen zu Beginn einem weißen Blatt Papier. Die Unterteilung nach Themen und Kategorien, etwa im Sinne obiger vier Bereiche, erleichtert die Beobachtung im Feld, weil sie Hinweise gibt, worauf zu achten ist und was protokollierungswürdig sein kann.

Möchten Forscher aussagefähige Daten zu einem Forschungsthema gewinnen, ist die reine Beobachtung im sozialen Feld in der Regel deutlich zeit- und kostenaufwendiger als die Durchführung qualitativer Interviews mit einzelnen Personen. Zudem sind viele subjektive Erlebensphänomene (Gefühle, Gedanken, Einstellungen, Meinungen, Motive) einer Fremdbeobachtung nicht unmittelbar zugänglich, sondern müssen erfragt werden.

Die wissenschaftliche Beobachterrolle in einem komplexen Feld stellt an die Forschenden besondere Herausforderungen (vgl. zum Folgenden Thierbach/Petschick 2014, S. 861f.). Anders als bei Interviews gelangen die Daten nicht ohne Weiteres über Verbalisierungen von den untersuchten Fällen zu den Wissenschaftlern. Allein der Beobachter ist für das Gelingen der Datensammlung verantwortlich. Bei den vielfältigen Prozessen und Interaktionen, die im Feld zu beobachten sind, ist ein hohes Maß an Aufmerksamkeit, Sensibilität und Konzentration auf das Wesentliche erforderlich.

Zudem besteht bei der Variante der teilnehmenden Beobachtung die Gefahr einer Absorption durch die täglichen Aufgaben. Der Beobachter verlässt die Wissenschaftlerrolle und verschmilzt immer stärker mit der Praxis im sozialen Feld. Tägliche Aufgaben nehmen ihn unter Umständen so in Anspruch, dass das Forschungsziel aus dem Horizont seiner Aufmerksamkeit gerät oder kaum mehr Zeit bleibt für genaue Beobachtungen und deren gewissenhafte Protokollierung.

Beobachterfehler

Menschen nehmen andere Menschen unterschiedlich wahr. Diese Grundtatsache des Alltagslebens gilt auch für die wissenschaftliche Beobachtung. Wenn z. B. ein Teammeeting am Arbeitsplatz beobachtet wird, mit dem Fokus auf das Verhalten des Teamleiters, fallen die Beobachtungsprotokolle für die identische Sitzung bei verschiedenen Beobachtern unterschiedlich aus. Die vom Leiter bei dieser Gelegenheit geäußerte Kritik an den Teammitgliedern könnte z. B. in verschiedenen Beobachtungsprotokollen als laut, überraschend, unberechtigt oder herabsetzend protokolliert werden.

Beim Beobachten vermischen sich Wahrnehmung und Interpretation. Schon Adjektive oder Adverbien wie die obigen sind interpretativ aufgeladen. Aufgrund ihrer Vorerfahrungen neigen Menschen dazu, das Verhalten anderer zu interpretieren, dies zum Teil falsch und ohne ausreichende Datengrundlage. Ursachen sind naive Verhaltenstheorien, die sich durch den Kontakt mit anderen im Laufe des Lebens bei Personen gebildet haben. Wenn etwa andere Menschen korrekt und ordentlich angezogen sind, neigen einige Beobachter zu der Annahme, dass diese ebenso korrekt und ordentlich ihre Aufgaben am Arbeitsplatz erledigen und genauso im Sozialverhalten sind. Als Beobachterfehler ist das der sogenannte Überstrahlungseffekt, auch HALO-Effekt genannt.

Verschiedene typische Beobachterfehler werden in einigen Methodenlehrbüchern ausführlich beschrieben, einschließlich möglicher Gegenmaßnahmen (z. B. Döring/Bortz 2016, S. 331f.). Letztlich sind Beobachterfehler jedoch nicht vollständig zu vermeiden. Um sie zu reduzieren, helfen Beobachterschulungen, der Einsatz eines strukturierten (statt eines unstrukturierten) Beobachtungsbogens sowie der Einsatz mehrerer Beobachter in derselben Situation. Dadurch gewinnen die Ergebnisse von Beobachtungen einen höheren Grad an Objektivität.

Non-reaktive Verfahren: Beobachtung ohne Sichtkontakt

Dass das Verhalten von Akteuren in einem sozialen Feld beobachtet werden kann, ohne dass die Beobachter die untersuchten Personen sehen, mutet paradox an, ist aber in der empirischen Forschung möglich. Bei **non-reaktiven Verfahren** besteht kein visueller oder mündlicher Kontakt zwischen den Beobachtenden und den Beobachteten, der Prozess der Beobachtung selbst ist ein indirekter. Ihr Gegenstand sind dementsprechend auch nicht die Menschen selbst, sondern Dinge, die von ihrem Dasein und Handeln Zeugnis geben, etwa Zeitungsanzeigen, Telefonbucheinträge oder Abnutzungserscheinungen an technischen Geräten.

Stanley Milgram wurde durch seine Gehorsamkeitsexperimente bekannt, beschäftigte sich danach jedoch mit non-reaktiven Verfahren, um die Alltagsferne psychologischer Experimente zu überwinden. Non-reaktive Verfahren haben den Vorteil, dass das natürliche Alltagsverhalten der Menschen nicht durch die Fragen eines Interviewers oder die Interventionen eines Experimentators beeinflusst wird.

Milgram erfand als non-reaktives Verfahren die Technik der verlorenen Briefe (engl. „lost-letter-technique“). Seine Mitarbeiter platzierten frankierte Briefe in verschiedenen Stadtteilen New Yorks, die mit verschiedenen Adressen bekannter Organisationen versehen

Non-reaktive Verfahren
Gemeinsames Merkmal aller non-reaktiven Verfahren ist, dass die untersuchten Personen nicht wissen, dass sie untersucht werden. Bei ihrem Verhalten werden sie nicht direkt von Wissenschaftlern beobachtet oder befragt.

waren. Es musste für die New Yorker Bevölkerung so aussehen, als hätten Postautos oder Briefzusteller die Briefe unterwegs verloren. Die Zahl der von der Bevölkerung bei der Post abgegebenen oder in Briefkästen eingeworfenen Briefe pro Adressat werteten die Forscher als Hinweis auf die positive oder negative Einstellung der Bevölkerung gegenüber den Adressaten.

Grounded Theory als kombiniertes Verfahren nicht-teilnehmender Beobachtung und informeller Befragung

Dieser qualitative Ansatz ist eng verbunden mit zwei bedeutenden Sozialwissenschaftlern des letzten Jahrhunderts, den beiden US-amerikanischen Soziologen Anselm Strauss (1916–1996) und Barney Glaser (geb. 1930). Sie haben die Grounded Theory als empirisches Forschungskonzept aus der Taufe gehoben. Als Studierende wurden beide in zwei Richtungen der amerikanischen Soziologie ausgebildet, Chicagoer Schule und symbolischer Interaktionismus. Diese soziologischen Richtungen betonen die Bedeutung menschlichen Handelns bei der Gewinnung soziologischen Wissens. Die Chicagoer Schule der Stadtforschung untersuchte in der Stadt Chicago in der ersten Hälfte des 20. Jahrhunderts die wechselseitige Anpassung der Menschen an ihre soziale und natürliche Umwelt. Untersuchungseinheiten waren größere Regionen und Gruppen.

Der symbolische Interaktionismus studierte wenige Jahre später die Prozesse der zwischen Menschen stattfindenden Aushandlung und Verständigung durch Sprache und nichtsprachliche Symbole (symbolische Interaktion), wodurch sich in Gruppen und Organisationen Regeln und Routinen etablieren. Beide Richtungen interessierten sich besonders für die Interaktionen und das Sozialverhalten in Gemeinschaften (Gruppen, Organisationen, Institutionen). Diese suchten die Sozialforscher selbst auf. Methodisch handelt es sich um eine Form nicht-teilnehmender Beobachtung. Die Beobachter übernehmen keine Aufgaben im Feld.

Glaser und Strauss entwickelten diese beiden Ansätze der Soziologie weiter und schufen eine komplexe Methodologie, in der Erhebungsphasen und Auswertungen abwechseln. Die Forscher sind dabei im sozialen Feld und sammeln die Daten über einen längeren Zeitraum hinweg. Sie beobachten die Akteure, registrieren deren Interaktionen und führen Gespräche zu bestimmten Forschungsfragen. Dies geschieht überwiegend informell bei sich ad hoc bietenden Gelegenheiten ohne systematischen Leitfaden.

Das erste große Forschungsfeld von Glaser und Strauss waren Krankenhäuser in Kalifornien. Es war ihr zentrales Forschungsanliegen, einen besseren Umgang der Ärzte und Pfleger mit schwer kranken und sterbenden Patienten zu unterstützen. Ihr bekanntestes Buch trägt den Titel „Interaktion mit Sterbenden“ (dt. 1974). Das empirische Herangehen der beiden Soziologen an das Untersuchungsfeld der Krankenhäuser war offen und pragmatisch. Keine ausformulierte wissenschaftliche Theorie zum Sterben, kein starres methodisches Konzept sollte als Filter wirken bei ihrer Auseinandersetzung mit existenziellen Fragen des Lebens und Sterbens und des professionellen Umgangs damit durch das Personal in Krankenhäusern.

Die methodischen Instrumente der Grounded Theory sind seit den ersten Veröffentlichungen von Glaser und Strauss vor fast 50 Jahren im Wesentlichen die gleichen geblieben (detailliert beschrieben in Strübing 2014).

Die drei methodischen Prinzipien der Grounded Theory

1. **Ausführliche Fallbeschreibungen:** Protokolle von umfangreichen Beobachtungen (z. B. alle Arbeitsabläufe während eines Tages in einem Arbeitsteam), Gespräche zwischen Akteuren (z. B. Protokolle von Teammeetings oder persönliche, auch informelle Gespräche zwischen Teammitgliedern), Interviews mit Akteuren.
2. **Stichprobenbildung durch Vergleich unterschiedlicher Fälle:** Die untersuchten Fälle müssen unterschiedlich sein und werden zielgerichtet anhand bestimmter Kriterien ausgewählt. Bei ihrer klassischen Untersuchung zum Sterben in Krankenhäusern gingen die beiden Wissenschaftler davon aus, dass auf verschiedenen medizinischen Stationen, in denen Patienten dem Tode nahe sind, je nach medizinischer Spezialisierung unterschiedlich mit dem Sterben umgegangen wird. Untersucht wurden eine Station mit zu früh geborenen Babys (damals noch mit hoher Todesrate), eine Krebsstation, eine geriatrische Station und eine Unfallstation.
3. **Memos der Forschenden:** Die Wissenschaftler schreiben Gedächtnisprotokolle über ihre Gedanken bei wichtigen Beobachtungen im Feld und formulieren Vorsätze und Pläne für die weitere Datenerhebung.

Die Auswertung der verschiedenen Datenprotokolle (Einzelfallbeschreibungen, Beobachtungsprotokolle, informelle Interviews und Memos) beginnt schon in der Erhebung. Die Auswertung ist das Codieren. Es ist eine pragmatische Form der Text- und Inhaltsanalyse, die auf detaillierte Ablaufregeln verzichtet. Sie beruht auf einem trennscharfen Kategoriensystem, dessen Erstellung die Hauptaufgabe der qualitativen Datenauswertung ist (vgl. Mayring 2010; Mayring/Gahleitner 2010).

Quantitative Beobachtungen

Bisher wurden ausschließlich qualitative Formen der Beobachtung dargestellt. In der Soziologie und Ethnografie haben qualitative Beobachtungen eine lange Tradition. Quantitative Beobachtungen finden dagegen Anwendungen in Psychologie und Wirtschaft. Oft geht es darum, Einzelpersonen in verschiedenen Verhaltensparametern zu bewerten. Dabei werden die Beobachtungssituationen genau definiert. Beispielsweise wird das Verhalten von Lehrern in einer Schulstunde oder von Therapeuten in einer Therapiesitzung von nicht beteiligten Beobachtern, verdeckt oder offen, in verschiedenen Variablen (Aussagen auf einem standardisierten Beobachtungsbogen) eingeschätzt. Das Erkenntnisziel ist die quantitative Bewertung beruflicher Leistungen.

Im Bereich von Personalauswahl und Personalentwicklung haben **Assessment-Center** mit Beobachtern ihren festen Platz. In ihnen geht es darum, die untersuchten Personen bestimmten Prüfungen – häufig sind das Simulationen von Arbeitsaufgaben – auszusetzen. Die Beobachter bewerten das gezeigte Verhalten anhand verschiedener Kriterien auf einem Beobachtungsbogen. Nach Durchlaufen aller Aufgaben eines Assessment-Centers erhält der Prüfling eine Gesamtbewertung als Rückmeldung über seinen Leistungsstand. Verbunden ist damit oft eine Entscheidung, etwa ob ein Bewerber eingestellt wird oder ob

Assessment-Center
AC sind Verfahren der Einschätzung (Assessment) von Personen in Bezug auf ihre beruflich relevanten Kompetenzen. Elemente eines AC sind Rollenspiele zur sozialen Kompetenz und Aufgaben zur Problemlösung.

ein Beschäftigter bestimmte Personalentwicklungsmaßnahmen durchlaufen sollte. Beobachtungen bei Assessment-Centern verdeutlichen Grundprinzipien quantitativer Beobachtungen im beruflichen Umfeld:

- Beobachtet und eingeschätzt werden Arbeitsleistungen von Einzelpersonen.
- Die Beobachtungssituationen sind eng umrissen.
- Die Inhalte der quantitativen Beobachtungsbögen sind die typischen Anforderungen und Tätigkeiten an einem bestimmten Arbeitsplatz.

Quantitative Beobachtungen bei eng umrissenen Szenarien und Simulationen

Das Verfahren des Assessment-Centers wurde auch von Vollmers und Kindervater eingesetzt, die in einer mehrjährigen Evaluationsstudie die Entwicklung beruflicher Handlungskompetenzen von benachteiligten Auszubildenden im Einzelhandel untersucht haben (Vollmers/Kindervater 2009). Im Folgenden wird eine Aufgabe aus dem eingesetzten Assessment-Center als Beispiel erläutert, unter Einbezug des für die Aufgabe verwendeten Beobachtungsbogens.

Die Aufgabe ist ein Rollenspiel, in dem ein Verkaufsgespräch im Einzelhandel simuliert wird (Vollmers 2009). Es dient der Diagnostik berufsbezogener sozialer Kompetenzen. In dem Rollenspiel wirken ein Verkäufer und ein Kunde mit. Die Beobachter bewerten nur den Verkäufer. Dieser wird gespielt von dem zu testenden Auszubildenden. Die Rolle des Kunden übernimmt ein geschulter Mitarbeiter des Unternehmens bzw. der pädagogischen Einrichtung für Behinderte, das bzw. die den Jugendlichen ausbildet.

Simuliert wird ein Konflikt zwischen Kunde und Verkäufer. Der Kunde möchte ein Sonderangebot kaufen, von dem er durch eine Anzeige erfahren hat. Dieses ist aber bereits ausverkauft. Der Verkäufer soll den Kunden zum Kauf eines teureren Produkts überreden. Der Auszubildende, der den Verkäufer spielt, hat die Rollenspielinstruktion erst unmittelbar vor der Aufgabe gelesen. Das Rollenspiel dauert maximal zehn Minuten und ist nur eine Aufgabe innerhalb eines komplexen Assessment-Centers, das der Auszubildende an einem Tag absolviert. Bei dieser Aufgabe, dem Rollenspiel Verkauf, verwendeten die Beobachter den folgenden quantitativen Beobachtungsbogen. Er enthält 14 verschiedene Aussagen.

Abbildung 8: Beobachtungsbogen im Rollenspiel Verkauf

Kompetenzbereiche	5 = kann Azubi sehr gut 1 = kann Azubi gar nicht				
1. beachtet Höflichkeitsregeln und Umgangsformen	5	4	3	2	1
2. geht auf realistische Kundenwünsche ein	5	4	3	2	1
3. bleibt bei Verhandlungen über Sonderangebot freundlich	5	4	3	2	1
4. vertritt eigenen Standpunkt und kann den Kunden überzeugen	5	4	3	2	1
5. tritt in Gesprächen/Verhandlungen sicher auf	5	4	3	2	1
6. bleibt in hektischen Situationen ruhig und gelassen	5	4	3	2	1
7. lässt sich von schwierigen Kunden nicht aus der Ruhe bringen	5	4	3	2	1
8. kann sich verständlich ausdrücken	5	4	3	2	1
9. berät Kunden sachgerecht	5	4	3	2	1
10. bleibt so lange im Gespräch, bis Ware verkauft ist	5	4	3	2	1
11. versteht schriftliche Anweisungen/Texte	5	4	3	2	1
12. zeigt Interesse an Beratung von Kunden	5	4	3	2	1
13. präsentiert die Waren ansprechend	5	4	3	2	1
14. Das äußere Erscheinungsbild ist sauber und gepflegt.	5	4	3	2	1

Quelle: Vollmers 2009.

Äußerlich ähnelt dieser Beobachtungsbogen einem quantitativen Fragebogen mit einer LIKERT-Skala. Im Unterschied dazu enthält der Beobachtungsbogen Fremdeinschätzungen von Beobachtern über die untersuchten Personen. Bei diagnostischen Untersuchungen in Psychologie und Wirtschaft zu Leistungen im Beruf wird ein Fremdbeobachtungsbogen gerne um eine parallele Selbsteinschätzung zu den gleichen Punkten ergänzt. Die untersuchten Personen füllen nach der Aufgabe den Selbstbeobachtungsbogen aus. Dem Verhältnis von Fremd- und Selbsteinschätzungen gilt dann das besondere Untersuchungsinteresse der Wissenschaftler. So war es auch beim in diesem Abschnitt erwähnten Projekt (vgl. zu Unterschieden der Selbst- und Fremdeinschätzung sozialer Kompetenzen bei Auszubildenden Vollmers/Kindervater 2010, S. 521f.).

Qualitativ oder quantitativ? Ein Fazit zu empirischen Beobachtungen

Quantitative Beobachtungen fokussieren sich auf das Verhalten und die Leistungen von Einzelpersonen. In der Psychologie sind die untersuchten Fälle Personen. Es interessiert ihr Verhalten in einem von den Forschern genau umrissenen Verhaltensbereich, z. B. das Verhalten bei bestimmten Aufgaben am Arbeitsplatz. Dagegen ist bei qualitativen Beobachtungen der Untersuchungsfokus umfassender. Es interessieren die Abläufe in einem komplexen sozialen System, z. B. einem Krankenhaus. Qualitative Beobachtungen fokussieren sich stärker auf die Interaktionen in einem größeren sozialen und institutionellen Kontext, dessen Abläufe, Regeln und Rituale untersucht werden.

2.5 Experimente und Versuchspläne

Experimentator
Ein Experimentator definiert und verändert gezielt die unabhängige Variable und misst die resultierenden Effekte der abhängigen Variablen bei den untersuchten Fällen.

Mit ihrer Entstehung als wissenschaftliche Disziplin im 19. Jahrhundert hat die Psychologie das Experiment als Methode aus der Medizin bzw. Physiologie adaptiert. In einem Experiment werden von einem **Experimentator** Versuchsbedingungen (unabhängige Variable) systematisch variiert, um bei den untersuchten Personen bestimmte Verhaltensweisen (abhängige Variable) hervorzurufen.

Es gibt mehrere Klassifikationsraster, um psychologische Experimente einzuteilen. Die Unterscheidung zwischen Labor- und Feldexperiment rekurriert auf den Ort der Durchführung, im psychologischen Labor oder im sozialen Feld bzw. in der Lebensumgebung der untersuchten Personen.

Laborexperimente markierten die Geburtsstunde der akademischen Psychologie. Im Allgemeinen gilt die Gründung des ersten psychologischen Labors an der Universität Leipzig im Jahr 1879 durch den Mediziner Wilhelm Wundt (1832–1920) als ihr Gründungsakt. Wundt und seine Mitarbeiter untersuchten sinnesphysiologische Prozesse, d. h. die Wahrnehmung und kognitive Verarbeitung der im Experiment vorgegebenen Farben, Töne oder Worte. Sie registrierten dazu quantitative Parameter bei den Versuchspersonen, vor allem die Reaktionszeit. Das ist die Zeit, die eine Person benötigt, um eine Taste zu drücken, nachdem sie einen bestimmten Reiz erkannt hat. Bis heute ist die Reaktionszeitmessung eine Methode der kognitiven Psychologie.

Wundt war vermutlich der erste Psychologe, der das Experiment als psychologische Forschungsmethode ausführlich beschrieb und dafür methodische Standards definierte (Eifler 2014, S. 196). Er war ein Pionier der Psychophysiologie. Äußerlich ähnelten seine Experimente im Leipziger Labor durchaus denen der modernen Psychophysiologie, wenngleich die von ihm verwendeten Messapparaturen verglichen mit den heutigen primitiv anmuten.

Feldexperimente finden außerhalb des Labors unter Bedingungen statt, die das normale Alltagsleben der Versuchspersonen charakterisieren. Die Versuchsbedingungen werden von den Experimentatoren bestimmt. Allerdings ist eine genaue Kontrolle des zu erwartenden Verhaltens der Versuchspersonen weder möglich noch intendiert. Ein ebenso berühmtes wie kontroverses Feldexperiment war das Stanford-Gefängnis-Experiment, das der US-amerikanische Psychologe Philip Zimbardo 1971 an der Universität Stanford durchführte. 70 Personen wurden per Zeitungsannonce angeworben und nach einer ausführlichen Untersuchung ihrer psychischen Stabilität in zwei Gruppen eingeteilt: Aufseher und Gefängnisinsassen. Die Personen, denen die Insassenrolle zugedacht war, wurden einige Tage später unangekündigt von der Polizei verhaftet und in ein Gefängnis überstellt. Dort wurden sie von der Gruppe der Wärter erwartet. In den folgenden Tagen wurde der Tagesablauf in einem Gefängnis simuliert. Zimbardo sah seine Studie als Fortführung der Milgram-Experimente. Ziel war die Untersuchung von Gewalt und Unterordnung unter den Bedingungen von Anonymität und Uniformierung. Aufgrund des brutalen Verhaltens eines Teils der Aufseher wurde das Gefängnisexperiment nach wenigen Tagen abgebrochen. In dem Kinofilm „Das Experiment“ (Regie: Oliver Hirschbiegel, Hauptrolle: Moritz Bleibtreu), der 2001 in die deutschen Kinos kam, wurden Motive aus Zimbardos Experiment verarbeitet.

Quantitative Versuchspläne und Quasi-Experimente

Experimentelle Versuchspläne in der Psychologie sind dadurch gekennzeichnet, dass die experimentellen Bedingungen vorab definiert werden. Die Art der Variation der unabhängigen Variablen (UV), oft ist es nur eine einzige, steht vor der Datenerhebung fest. Die verschiedenen Ausprägungen der unabhängigen Variablen heißen **Faktorstufen**. Die untersuchten Personen werden vorab auf die Faktorstufen verteilt. Dieses Vorgehen ist bei Interventionsstudien in Medizin und Psychologie üblich. Die verschiedenen Formen der Intervention sind die Faktorstufen der unabhängigen Variablen. Die Anzahl der untersuchten Personen sollte in allen Fallgruppen gleich sein. Gemessen werden Wirkungen in der abhängigen Variablen (AV). Dies sind z. B. Beschwerderückgang oder Beschwerdefreiheit.

Faktorstufen

In Versuchsplänen werden die Ausprägungen der unabhängigen Variablen Faktorstufen genannt. Sie dienen der Gruppierung der untersuchten Fälle.

Tabelle 3: Fiktiver Versuchsplan in Medizin oder Psychologie mit drei Experimentalgruppen (Therapieformen) und einer Kontrollgruppe (keine Therapie)

UV: Formen der Therapie	Therapie A	Therapie B	Therapie C	Placebo (keine Therapie)
AV: Wirkung/Heilung bei Fällen	50 Fälle	50 Fälle	50 Fälle	50 Fälle

Quelle: erstellt im Auftrag der IU, 2022.

Explorative Experimente

Feldexperimente wie das Gefängnisexperiment sind oft explorativ angelegt. Es geht um die Untersuchung aller auftretenden Verhaltensmuster, nicht um die Messung von Effekten anhand numerischer Skalen oder die Ausschaltung von Störvariablen. Die Ergebnisse sind komplex und für die Versuchsleiter wenig vorhersehbar. Es handelt sich um **explorative Experimente**. Weitgehend synonym ist die Bezeichnung „qualitative Experimente“. Diese haben in der kognitiven Psychologie und der Entwicklungspsychologie eine lange Tradition (vgl. Burkart 2010, S. 253f.).

Explorative Experimente

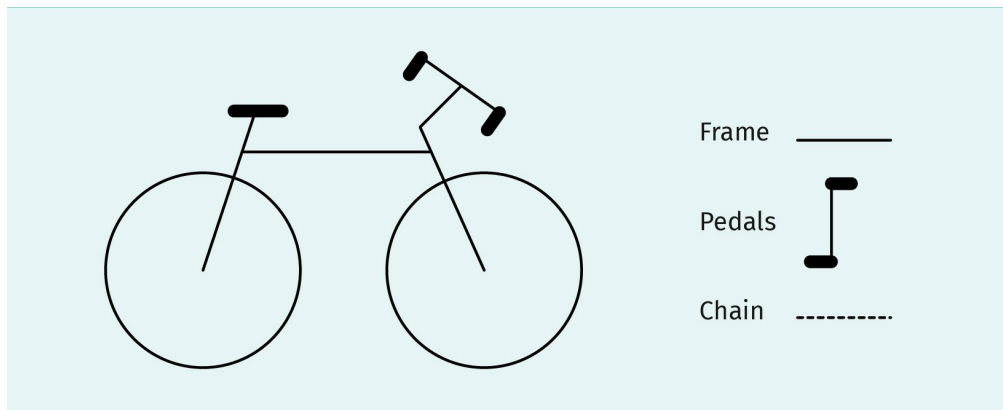
Bei explorativen Experimenten werden die unabhängigen Variablen in verschiedene Richtungen variiert. Sie dienen der Entdeckung (Exploration) von Sachverhalten.

Anders als in quantitativen Versuchsplänen werden in explorativen Experimenten nicht vorab Ursache-Wirkungs-Relationen definiert. Im Nachhinein, wenn die Ergebnisse vorliegen, lassen sich unter Umständen unabhängige und abhängige Variablen bestimmen. Die experimentellen Variationen gehen bei explorativen Experimenten in unterschiedliche Richtungen. Gleiches gilt für die Auswertung. Quantitative und qualitative Analysen sind bei explorativen Experimenten möglich und ergänzen sich.

Beispiel: Die Fahrradexperimente von Rebecca Lawson

Die in Cambridge lehrende britische Psychologin Rebecca Lawson hat explorative Experimente zur zeichnerischen Reproduktion von Fahrrädern durchgeführt (Lawson 2006, S. 1668f.). Ihre Versuchspersonen sollten Skizzen eines Fahrrads anfertigen, das alle wesentlichen Funktionselemente enthält. Eine Versuchsanordnung bestand darin, die drei Teile Rahmen, Kette und Pedale in eine vorgelegte Fahrradskizze einzufügen.

Abbildung 9: Vorlage für die Fahrradskizzen der Versuchspersonen von Rebecca Lawson



Quelle: Lawson 2006, S. 1670.

Lawson stellte den Versuchspersonen zusätzlich die Frage, an wie vielen Tagen in der Woche sie ein Fahrrad nutzen. Aus den Antworten auf diese Frage bildete sie zwei Gruppen: Gelegenheitsradler (weniger als einmal pro Woche Fahrradnutzung) und Vielradler (mehrere Male pro Woche Fahrradnutzung). Diese beiden Gruppen wurden nochmals unterteilt anhand der Versuchsinstruktion. Die Hälfte der Untersuchungspersonen wurde darüber in Kenntnis gesetzt, dass es in dem Experiment um ihr Wissen über den Antrieb von Fahrrädern geht. Der anderen Hälfte wurde das Untersuchungsziel nicht explizit mitgeteilt.

Außerdem variierte Lawson den Versuchsaufbau: Einer Gruppe wurde die obige Aufgabe zum Zeichnen vorgelegt. Eine andere Gruppe musste in vier Abbildungen die richtige Darstellung von drei fehlerhaften unterscheiden (Multiple Choice). Die numerischen Ergebnisse der Fehler sind wie folgt:

Tabelle 4: Numerische Ergebnisse der Fehler des Fahrradexperiments

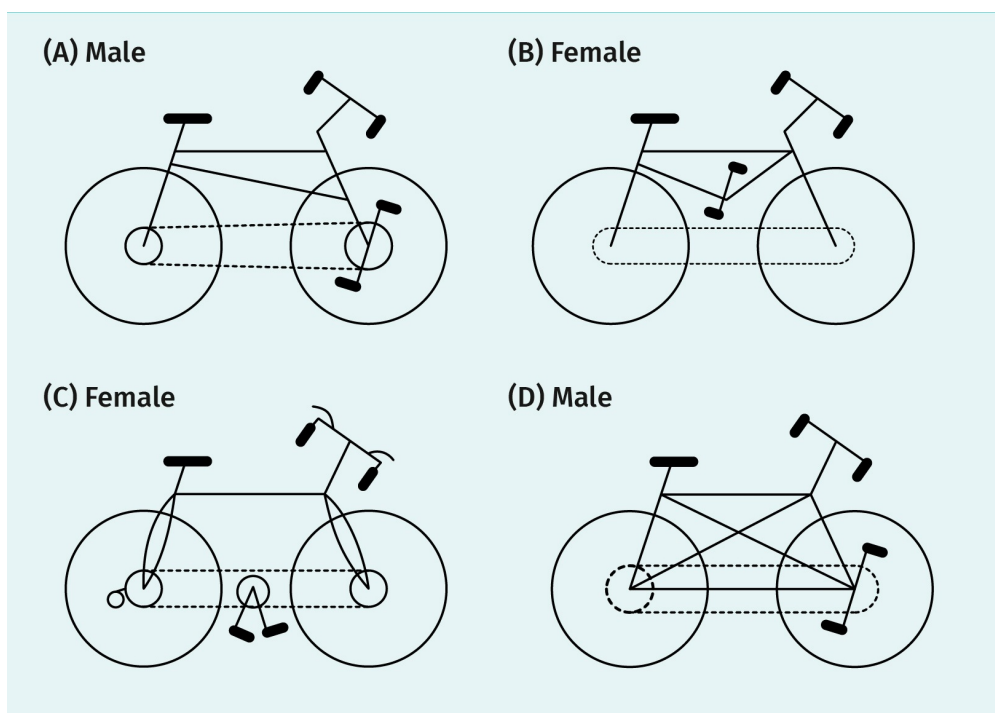
Varianten Experiment	Gelegenheitsradler (n = 81) Kenntnis Untersuchungsziel		Gelegenheitsradler (n = 81) Unkenntnis Untersuchungsziel		Vielradler (n = 61) Unkenntnis Untersuchungsziel	
	Zeichnung	Multiple Choice	Zeichnung	Multiple Choice	Zeichnung	Multiple Choice
Fehlerart						
1. Rahmen	12	26	23	46	3	7
2. Pedale	7	4	11	5	4	1
3. Kette	36	28	50	43	6	9
Einer der drei Fehler	44	40	60	55	9	3

Quelle: erstellt im Auftrag der IU, 2022 nach Lawson 2006.

Jede Person kann bis zu drei Fehler machen (Rahmen, Pedale, Kette). Deshalb fällt die Fehlerzahl in den Gruppen der Gelegenheitsradler größer aus als die Zahl der Fälle ($n = 81$).

In den beiden Gruppen der Gelegenheitsradler taucht damit bei über 40 % der Personen einer der drei Fehler auf. Aber auch bei den Vielradlern wiesen mehr als 10 % der Zeichnungen Fehler auf. Neben der numerischen Auswertung ist bei explorativen Experimenten auch die qualitative Analyse der Ergebnisse relevant. Rebecca Lawson entwickelte eine qualitative Fehlertypologie anhand der Zeichnungen der Versuchspersonen.

Abbildung 10: Typische Fehler in den Fahrradzeichnungen der Versuchspersonen, unterschieden nach deren Geschlecht



Quelle: Lawson 2006, S. 1670.

Der nächste Schritt der qualitativen Auswertung ist die Suche nach möglichen Gründen für die Fehler. In allen Beispielen wird der Hinterradantrieb mittels Kette und Pedale nicht richtig wiedergegeben. Die Bewegung des Nach-Vorne-Tretens und die Steuerung über die Lenkung des Vorderrads verführten offenbar die untersuchten Personen dazu, den Antrieb fälschlich zu weit nach vorne zu verlegen. Falsche Erklärungen zur Funktionsweise eines Fahrrads gehen in diesen Beispielen auf die Überbetonung des subjektiven Erlebens des Fahrradfahrens zurück, bei dem der Körperschwerpunkt vorne liegt.

Diese spekulative Erklärung könnte durch zusätzliche qualitative Befragungen direkt nach der Durchführung der Experimente belegt werden. Explorative Experimente können mit den im folgenden Lernzyklus dargestellten Methoden der kognitiven Psychologie kombi-

niert werden, indem die Personen retrospektiv zu ihrem Denkprozess befragt werden, oder indem sie vor den Experimenten zu ihrem Wissen zum Antrieb von Fahrrädern Auskunft geben.

2.6 Datenerhebungsverfahren in der kognitiven Psychologie

Die kognitive Psychologie bzw. Kognitionspsychologie gehört zur allgemeinen Psychologie und beschäftigt sich mit den Grundlagen von Wahrnehmung, Lernen, Denken und Gedächtnis. Diese vier Aspekte sind kognitive Prozesse im engeren Sinne. Im weiteren Sinne werden auch Aufmerksamkeit (Konzentration), Sprechen und Motivation zu den kognitiven Prozessen gezählt. Kognitive Vorgänge sind Prozesse der Informationsverarbeitung. Das Individuum nimmt Informationen aus der Umwelt (und Signale aus der psychophysischen Innenwelt) auf und verarbeitet diese im Bewusstsein. Durch Speicherung im Gedächtnis werden Informationen zu Bestandteilen des persönlichen Wissens, das abgerufen und erweitert werden kann.

Kognitive Prozesse eines Menschen sind nicht direkt von außen durch andere Personen zu beobachten. In der empirischen Forschung wird über äußerlich sichtbares Verhalten bzw. über Verhaltensresultate, z. B. Ergebnisse in einem Test, auf kognitive Zustände und Prozesse geschlossen. Für die betroffene Person selbst sind die eigenen kognitiven Zustände ebenfalls nicht ohne Weiteres beobachtbar. Wahrnehmung und Denken sind Vorgänge, die auf die Umwelt gerichtet sind, zur Bewältigung der täglichen Lebensaufgaben. Was in einer bestimmten Lebenssituation wahrgenommen und gedacht wurde, fällt rasch dem Vergessen anheim. Nur ein Bruchteil des täglich Wahrgenommenen gelangt ins Langzeitgedächtnis eines Menschen. Die eigenen kognitiven Prozesse innerlich zu beobachten und Erkenntnisse über eigene Denkformen zu gewinnen, ist im Alltag gewöhnlich nicht erforderlich. Zudem würde es eine Spaltung im Bewusstsein erfordern, also eine Teilung der Aufmerksamkeit in die Wahrnehmung der Umwelt und die innere Beobachtung der eigenen Informationsverarbeitung.

Trotz dieser methodischen Schwierigkeiten kommt der kognitiven Psychologie eine besondere epistemische Bedeutung für viele Wissenschaften zu, denn Wahrnehmung, Denken und Wissen sind die Grundlagen aller Formen wissenschaftlicher Erkenntnisgewinnung, die von Menschen ausgehen. Die Inhalte der kognitiven Psychologie haben überwiegend den Charakter **hypostasierender** Modelle, die die empirische Forschung leiten (z. B. Wissensmodelle, Gedächtnismodelle, Sprachproduktionsmodelle). Ihre empirischen Erhebungsmethoden sind Verfeinerungen der drei Grundformen psychologischer Methoden: Befragung, Beobachtung und Experiment. Die methodischen Verfeinerungen tragen den Besonderheiten der Untersuchung kognitiver Prozesse Rechnung. Außerdem bewegt sich die empirische Forschung in der kognitiven Psychologie im Schnittfeld mit Nachbarwissenschaften. Kognitionspsychologen kooperieren zum einen mit Informatikern bei der Erforschung und Anwendung der künstlichen Intelligenz und in der Softwareergonomie. Zum anderen sind Neurowissenschaftler und Mediziner Kooperationspartner bei der Erforschung von Bewusstsein und Gehirn.

Hypostasieren

Das Hypostasieren meint, etwas Unanschauliches zu veranschaulichen.

Kognitive Prozesse wie das Denken oder ein Gedächtnisabruf können nicht beobachtet werden. Hypostasierende Modelle fungieren als Hypothesen.

Sie dienen der Veranschaulichung kognitiver Prozesse.

Zur Orientierung über die Erhebungsmethoden in der kognitiven Psychologie passt als grobe Unterscheidung die zwischen quantitativen und qualitativen Methoden. Quantitative Methoden messen und registrieren äußerlich sichtbares Verhalten, wie z. B. die Zahl der richtig reproduzierten Worte oder Sätze oder die Reaktions- oder Blickgeschwindigkeit vor einem Computer. Von den gemessenen Leistungen wird zurückgeschlossen auf quantitative Kennzeichen von kognitiven Prozessen (Aufmerksamkeitsspanne, Gedächtniskapazität). Qualitative Verfahren untersuchen die subjektiven Wahrnehmungs-, Denk- und Gedächtniserlebnisse. Dazu müssen die untersuchten Personen diese Erlebnisse verbal mitteilen. Den Personen werden Aufgaben gestellt, die kognitive Prozesse (Denken, Lernen) erfordern. Nach Beendigung der Aufgabe teilen die Personen im Rückblick mit, wie sie die Aufgabe wahrgenommen haben und was sie während der Bearbeitung gedacht haben. In der qualitativen Wissensdiagnostik werden die Personen in Form eines Interviews zu ihrem Wissen zu einem bestimmten Thema befragt. In der Auswertung kommen in der kognitiven Psychologie nahezu alle Verfahren der quantitativen und qualitativen Datenauswertung zur Anwendung.

Quantitative Erhebungsverfahren der kognitiven Psychologie

Der Kognitionspsychologe Joachim Funke, Professor für Allgemeine Psychologie an der Universität Heidelberg, unterscheidet die folgenden quantitativen Methoden in der Kognitionspsychologie (Funke 1996):

1. Reaktions- und Entscheidungszeitmessung
2. Blickbewegungsmessung
3. Computersimulation kognitiver Prozesse
4. quantitative Modellierung kognitiver Strukturen
5. bildgebende Verfahren des Gehirns

Reaktions- und Entscheidungszeitmessung

Schnelligkeiten in der Reaktion geben Auskunft über die Geschwindigkeit der kognitiven Informationsverarbeitung. Entsprechende Experimente haben in der kognitiven Psychologie eine lange Tradition. Den Versuchspersonen werden bestimmte Reize (z. B. Worte, geometrische Figuren) präsentiert. Immer, wenn die Untersuchungsperson ein vorher definiertes Symbol sieht (z. B. ein bestimmtes Wort), drückt sie eine Taste eines Versuchsapparats. Heute werden solche Experimente oft vor dem Computerbildschirm durchgeführt. Eine andere Form ist, dass die Personen mehrere Reize auf ihre Gleichartigkeit oder Unterschiedlichkeit vergleichen (z. B. zwei Wortlisten). Nachdem sie sich in ihrem Urteil entschieden haben, drücken sie eine Taste oder nehmen eine Markierung auf dem Bildschirm vor.

Blickbewegungsmessung

Hier werden die mit den Augenbewegungen verbundenen Veränderungen des Blickfelds einer Person registriert. Beim Lesen von Texten auf einem Computerbildschirm oder der Bedienung von Software schweift der Blick auf dem Bildschirm umher. Dies kann heute mit Tracking-Tools leicht realisiert werden. Als Protokoll der Augenbewegungen entsteht ein Netz von Linien auf dem Bildschirm. Erkenntnisinteresse ist bei dieser Methode oft

Reaktions- und Entscheidungszeitmessung

Reaktions- und Entscheidungszeitmessungen sind zugleich quantitative Experimente. Die unabhängigen Variablen sind die visuellen Muster oder Worte, die auf dem Bildschirm erscheinen. Die abhängige Variable ist die Reaktionszeit.

eine Verbesserung des Layouts von Texten oder eine optimalere Gestaltung der Benutzeroberfläche eines Programms. Als gemessene quantitative Variable käme etwa infrage, wie oft und wie lange eine bestimmte Zeile fixiert wird, oder auch, wie oft auf vorhergehende Textpassagen zurückgegriffen wird.

Computersimulation kognitiver Prozesse

Dies ist keine Methode der Datensammlung an Menschen, vielmehr geht es um die Programmierung menschenähnlicher Denkstrukturen auf Computern. Dafür sind besondere Programmiersprachen erforderlich. Computer sollen, wenn sie entsprechend programmiert werden, menschliche Denkprozesse simulieren. Vor der Programmierung müssen unter Umständen Experten zu einer Domäne mit empirischen Methoden befragt werden, damit den Softwareentwicklern deutlich wird, was die Computersimulation leisten soll.

Quantitative Modellierung kognitiver Strukturen

Bestimmte kognitive Zustände oder Fertigkeiten von Menschen werden in formalen Modellen dargestellt. Beispielsweise kann man beim menschlichen Gedächtnis die Fertigkeiten Speichern (engl. „storage“) und Abrufen (engl. „retrieval“) unterscheiden. Beide Aspekte werden in unterschiedliche formale Modelle gefasst. Die Stimmigkeit dieser Modelle kann an Menschen über verschiedene Tests bzw. Experimente quantitativ überprüft werden. Unter Umständen werden die Modelle dann verändert und anschließend auf Computern programmiert. Das Ziel dieser Methode ist ebenfalls eine Computersimulation kognitiver Prozesse.

Bildgebende Verfahren des Gehirns

Hier werden kognitive Vorgänge, die Menschen bei der Bearbeitung von Aufgaben durchführen, als Stoffwechselfvorgänge bildlich dargestellt. Das ist heute die zentrale Methode bei der neurowissenschaftlichen Untersuchung des Gehirns.

Qualitative Erhebungsmethoden in der kognitiven Psychologie

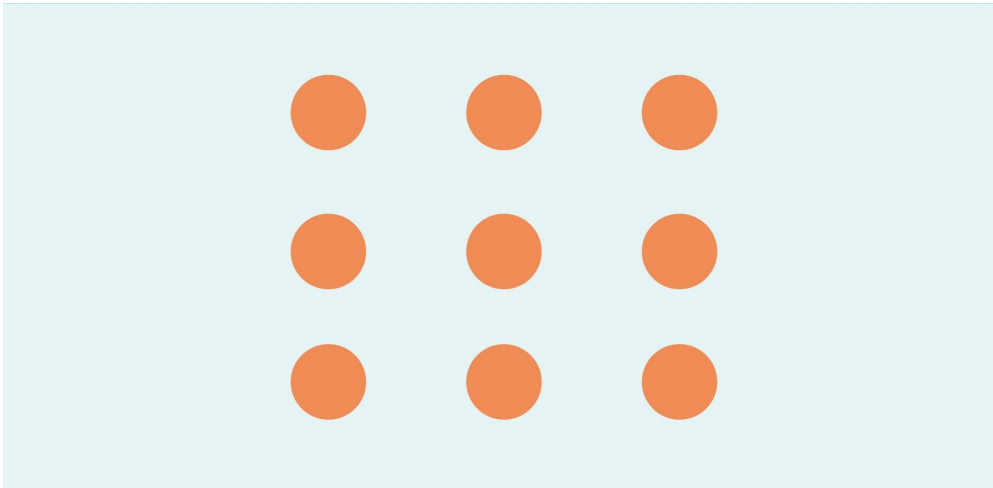
Qualitative Verfahren verzichten auf den Einsatz formaler Modelle bei der Datenerhebung und kommen ohne Messgeräte aus. Es geht um eine Ermittlung der Inhalte von Denkprozessen sowie von Wissen und Gedächtnis. Dazu geben die untersuchten Personen verbal Auskunft in Form von Interviews, in schriftlichen Berichten oder in grafischen Darstellungen. Die Untersuchungsanlage kann eher dynamisch oder eher statisch sein. Innerhalb der Problemlösungs- und Kreativitätsforschung ist der methodische Zugriff auf die Dynamik von Denkprozessen gerichtet. Bei der Prüfung des individuellen Wissensstands zu einem Thema handelt es sich dagegen um eine statische Momentaufnahme. Für beide Formen wird je ein Beispiel dargestellt.

Beispiel: Dynamische qualitative Denkprozessanalyse

Im Bereich der kognitionspsychologischen Forschung zu den Themen Problemlösung und Kreativität kann in einem einfachen explorativen Experiment untersucht werden, welche Richtungen des Denkens die Personen einschlagen, um ein Problem mit der Generierung

von neuen Ideen zu lösen. Ein Beispiel ist das sogenannte Neun-Punkte-Problem. Für dessen Lösung ist kein spezielles Expertenwissen vonnöten. Die Versuchsinstruktion lautet: „Verbinden Sie die folgenden neun Punkte mit vier Linien in einem Zug, also ohne zwischendurch abzusetzen und eine Linie neu zu beginnen“ (nach Maderthaler 2017, S. 250).

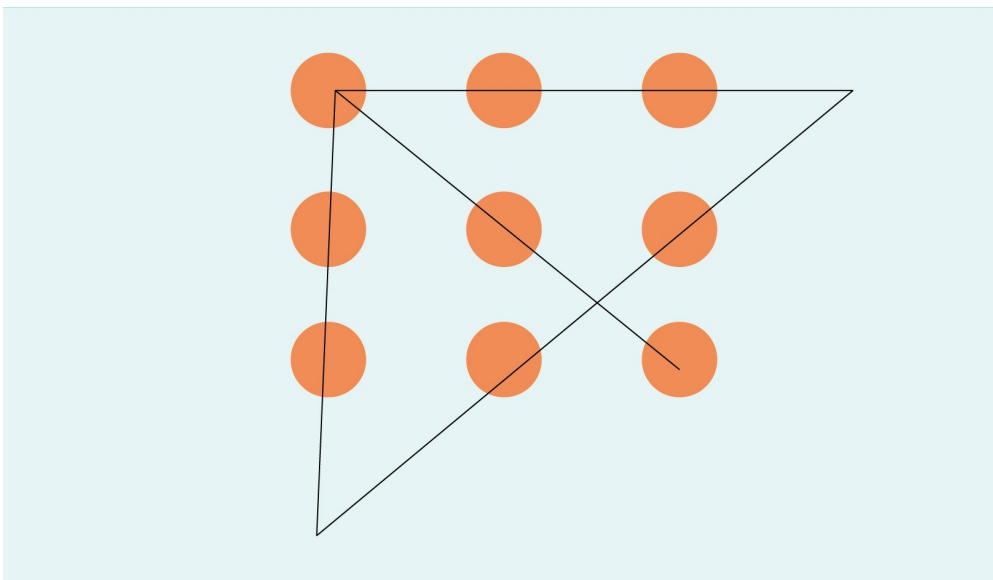
Abbildung 11: Neun-Punkte-Problem als Denkaufgabe



Quelle: erstellt im Auftrag der IU, 2022.

Die Lösung wird verhindert, wenn die Person die neun Punkte implizit als Quadrat mit Begrenzungslinien wahrnimmt. Zur Lösung ist ein Überschreiten der imaginierten Begrenzungen vonnöten.

Abbildung 12: Mögliche Lösung des Neun-Punkte-Problems



Quelle: erstellt im Auftrag der IU, 2022.

Zur Untersuchung der kognitiven Prozesse wird ein retrospektives Denkprotokoll angefertigt. Die Person schreibt auf oder teilt dem Versuchsleiter mit, welche Gedankenwege sie während der Bearbeitung eingeschlagen hat. Sie vergegenwärtigt sich im Rückblick ihre Denkprozesse. In einer größeren qualitativen Studie werden mehrere Personen mit dieser Aufgabe untersucht und die Denkprotokolle zusammengetragen. Denkprotokolle werden ähnlich behandelt wie Interviewprotokolle mit den entsprechenden Formen der Datenaufbereitung und qualitativen Auswertung.

Beispiel: Statische qualitative Wissensdiagnostik

Die qualitative Wissensdiagnostik hat das Ziel, dass die untersuchten Personen ihr persönliches Wissen zu einem Thema offenlegen, dies möglichst vollständig und in einem inneren logischen Zusammenhang. Dazu bedient sich die qualitative Wissensdiagnostik der Visualisierung mittels Karten und Verbindungslinien. Es wird von den untersuchten Personen jeweils ein Gebilde von Begriffskarten und Verbindungslinien angelegt oder gezeichnet. Die Karten enthalten Kategorien eines Wissensgebietes. Es sind die Knotenpunkte eines Wissensnetzwerkes. Diese werden durch Linien verbunden, welche die Beziehungen zwischen den Kategorien beschreiben.

Den meisten Lesern sind vermutlich in der Schule als Visualisierungstechniken von Wissen Mindmaps und Concept Maps begegnet. Das Verfahren der Struktur-lege-Technik (SLT) ist elaborierter und wurde in der kognitiven Psychologie in den 80er-Jahren in Deutschland entwickelt (Scheele/Groeben 2010, S. 507f.). Die Visualisierung des persönlichen Wissens wird mittels Karten auf einen Tisch gelegt. Die Beziehungen zwischen den Kategorien veranschaulichen spezielle Karten, die jeweils einen Typus formaler Beziehungen enthalten.

Abbildung 13: Beispiele von Beziehungskarten in der STL

<u>-</u>	ist identisch mit
<u>m</u>	manifestiert sich als Beispiel
<u>+</u>	je mehr X, desto stärker Y
<u>^</u>	logisches UND: beide, mehrere zugleich müssen zutreffen
<u>i</u>	ist Indikator/Beispiel für

Quelle: erstellt im Auftrag der IU, 2022 nach Groeben et al. 1988.

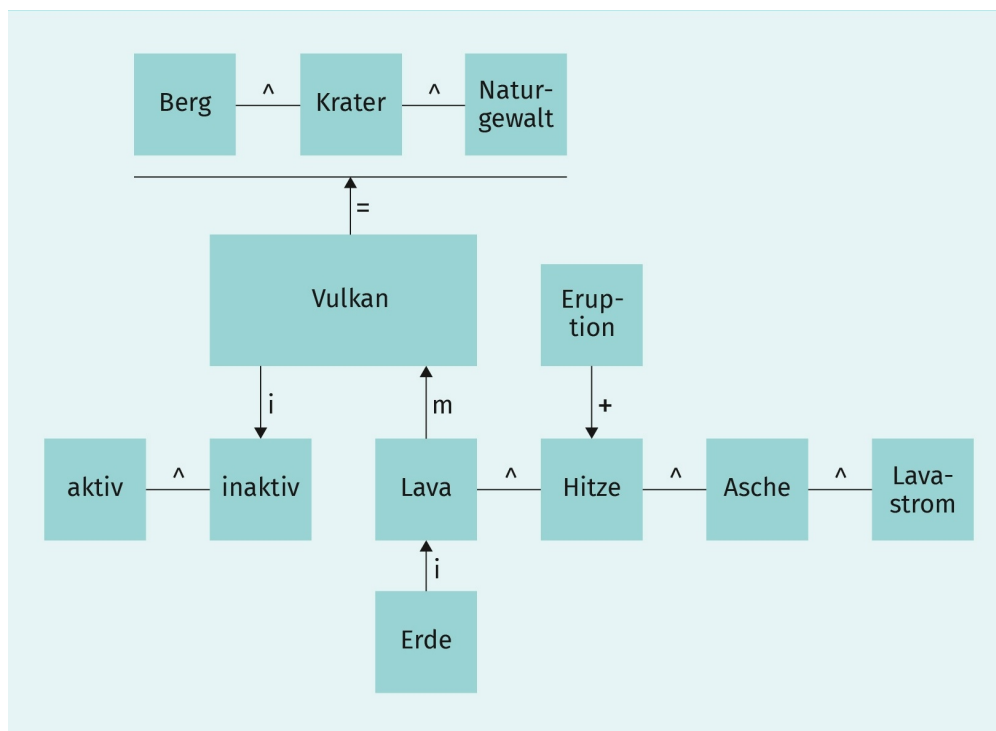
In **Lehr-Lern-Szenarien** können STL (wie Concept Maps und Mindmaps) zur Bestimmung des Vorwissens wie auch des durch Lernen neu gewonnenen Wissens herangezogen werden. Wenn das Vorwissen vor dem Lernprozess bei den Lernenden erhoben wurde, fungieren die grafischen Darstellungen zudem als eine Strukturierungs- und Gedächtnishilfe während der sich anschließenden Lernprozesse.

Lehr-Lern-Szenarien
Dies sind konkrete Situationen oder abstrakte Konzepte, in denen Lehr-Lern-Prozesse realisiert oder konzeptualisiert werden.

Zur Illustration der Anwendung von STL besprechen wir ein markantes Beispiel aus der Geologie bzw. Vulkanologie. Vermutlich haben alle Leser dieses Studienbriefs ein basales Wissen über Vulkanausbrüche und deren Folgen:

- Vulkane können aktiv oder inaktiv sein.
- Eine Eruption führt zu einem Ausbruch von Lava unter großer Hitze.
- Ein Lavastrom fließt talwärts.
- Asche wird in die Luft geschleudert.
- Ein Vulkan ist ein Berg, der einen unterirdischen Krater aufweist.
- Ein Vulkan symbolisiert die Kraft der Natur.

Abbildung 14: Mögliches Vorwissen von Menschen zu Vulkanen, visualisiert mittels STL



Quelle: erstellt im Auftrag der IU, 2022 nach Tergan 1988, S. 416.

Eine neue Lernphase zu Vulkanen verändert die Wissensstruktur. Eine öffentliche Lernphase waren die Medienberichte über den Ausbruch des Eyjafjallajökull auf Island. Er begann am 20. März 2010 und endete nach mehreren Eruptionsphasen endgültig am 9. Juli desselben Jahres. Menschen kamen zum Glück nicht zu Schaden, spektakulär waren jedoch die Auswirkungen auf den Flugverkehr, da eine Aschewolke nach Zentraleuropa

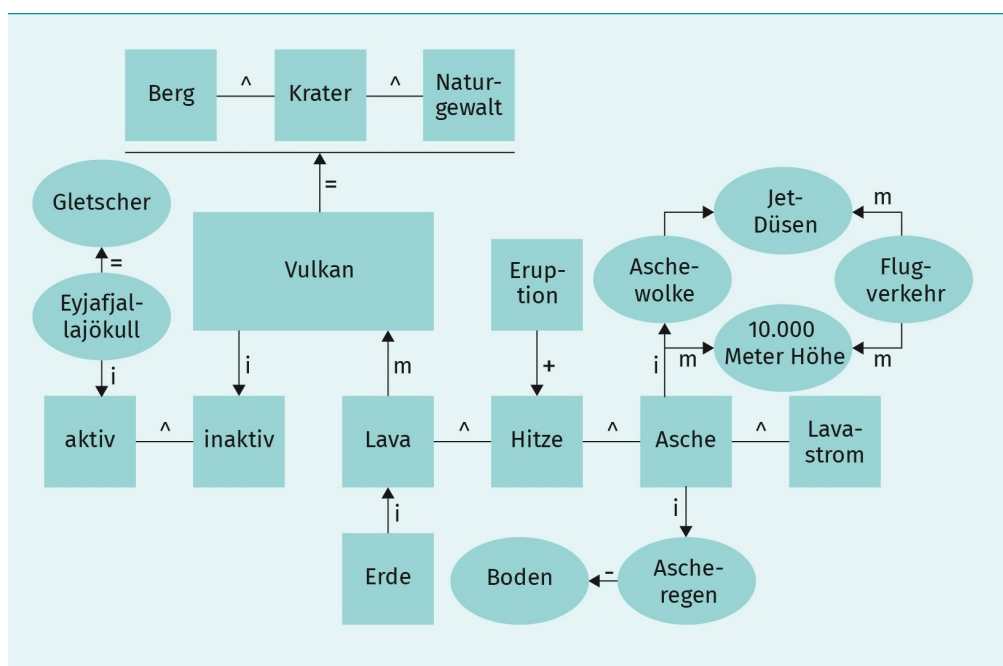
trieb. Ab dem 15. April wurde in vielen europäischen Ländern der Luftraum für Verkehrsflugzeuge gesperrt. Aufmerksame Medienrezipienten in Deutschland konnten aus den überregionalen Zeitungen und Nachrichtenmagazinen also Folgendes lernen:

- Eruptionen des Vulkans Eyjafjallajökull erfolgen in Phasen, nicht einmalig.
- Der Vulkan Eyjafjallajökull hat einen Gletscher.
- Schmelzendes Eis wird zu Wasserdampf und verstärkt die Eruptionen.
- Die Aschewolke steigt bis auf Höhen des zivilen Linienflugverkehrs.
- Der Linienflugverkehr mit düsengetriebenen Passagierjets fliegt 10.000–12.000 Meter hoch.
- Ein Jetflugzeug kann die Vulkanasche nicht als solche mit dem eigenen Wetterradar erkennen.
- Die Vulkanasche kann in die Triebwerke eines Düsenjets eindringen und diese stoppen.
- Die Vulkanasche kann in die Klimaanlage und den Frachtraum eindringen und die Manövrierfähigkeit eines Flugzeugs beeinträchtigen.
- Die Aschewolke kann zu Regen werden, der den Boden auf der Erde versiegelt.

Die Befragung eines aufmerksamen Medienrezipienten mittels STL könnte zum folgenden **Strukturdiagramm** führen.

Strukturdiagramm
Als Strukturdiagramm bezeichnet man ein Netzwerk, das aus mindestens zwei Typen von ineinander verschachtelten Elementen besteht: Knoten und Beziehungen zwischen den Knoten.

Abbildung 15: Mögliches Wissensdiagramm nach Lernphase zum Thema Vulkane



Quelle: erstellt im Auftrag der IU, 2022.

In der Auswertung derartiger **Strukturdiagramme des Wissens** kommen sowohl qualitative als auch quantitative Verfahren zur Anwendung. Quantitativ könnten z. B. die numerischen Unterschiede in der Anzahl der Kategorien zwischen dem ersten Diagramm (vor der Lernphase) und dem zweiten Diagramm (nach der Lernphase) ermittelt werden. Als qualitative Inhaltsanalysen bieten sich Zusammenfassungen übereinander gelegter Diagr-

amme an. Besonders in der psychologischen Lehr-Lern-Forschung sowie bei der Befragung von Fachexperten zu einem Thema ist die grafisch orientierte qualitative Wissensdiagnostik eine Alternative zu qualitativen Interviews mit einem Leitfaden, da die Ergebnisse im Vergleich zu Interviewprotokollen prägnanter sind.

Wissensdiagramm
Grafiken der Wissensstruktur einer Person zu einem Thema ergeben ein Wissensdiagramm. Sie veranschaulichen das neu erworbene Wissen nach Lernphasen sowie das Wissen davor.

2.7 Psychophysiologische und neurowissenschaftliche Methoden

Psychophysiologische Forschungsmethoden kommen im Schnittpunkt von Medizin und Psychologie zum Einsatz. Gemessen werden körperliche Vorgänge mittels Apparaten, die auf den Körper bzw. die Haut von außen zugreifen. Viele psychophysiologische Studien werden mit Vorrichtungen durchgeführt, die in der Medizin als Standardinstrumente der Diagnostik fungieren (EKG, EEG und Röntgengeräte). Die Apparaturen liefern Normparameter in der medizinischen Diagnostik, die nach standardisierten Regeln quantitativ ausgewertet werden. Bei der empirischen Untersuchung größerer Stichproben können quantitative Auswertungsverfahren eingesetzt werden. Qualitative Verfahren kommen in der Psychophysiologie in der Regel nicht zum Einsatz.

Psychophysiologische Methoden
Als psychophysiologische Methoden werden solche bezeichnet, die anhand der Messung von Körperfunktionen Rückschlüsse über den Zusammenhang von Körper und Psyche zu ziehen versuchen.

Die von den Apparaten registrierten Kennwerte gelten als Indikatoren für körperliche Gesundheit, dienen aber auch dem Rückschluss auf psychische Zustände. Der Zusammenhang zwischen psychischen und körperlichen (somatischen, physiologischen) Prozessen ist kompliziert. Unstrittig ist, dass psychische Prozesse somatische beeinflussen und umgekehrt. In der Medizin widmet sich vor allem die Teildisziplin der Psychosomatik diesen Zusammenhängen. Deren quantitative Forschungsmethoden sind weitgehend identisch mit denen der Psychophysiologie.

Die Psychophysiologie versteht sich als naturwissenschaftliche Disziplin, die in der Forschung experimentell-messend verfährt (Gramann/Schandry 2009, S. 47f.). Messungen finden in psychologischen bzw. medizinischen Laboren statt. Leitmethode ist das Laborexperiment. Die die Experimente durchführenden Forscher müssen sich mit der Funktionsweise der eingesetzten Geräte gut auskennen, um Bedienungs- und Auswertungsfehler zu vermeiden. Wer als Student in einer Bachelor- oder Masterarbeit psychophysiologische Forschung betreibt, sollte vorher entsprechende Laborpraktika absolviert haben.

Die beiden Professoren für Psychophysiologie, Klaus Gramann und Rainer Schandry, stellen in ihrem Lehrbuch (2009) alle Forschungsthemen und Methoden der Psychophysiologie ausführlich dar. Sie unterscheiden folgende Bereiche psychophysiologischer Forschung:

1. elektrodermale Aktivität,
2. elektrische Muskelaktivität,
3. kardiovaskuläre Aktivität,
4. hirnelektrische Aktivität,

5. bildgebende Verfahren und
6. weitere Körperfunktionen.

Elektrodermale Aktivität

Hier wird die elektrische Hautleitfähigkeit untersucht und mit psychischen Prozessen in Zusammenhang gebracht. Erhöhte Schweißsekretion steigert die Hautleitfähigkeit. Eine Steigerung des allgemeinen physiologischen Erregungsniveaus, z. B. durch Aufregung oder Stress, fördert bei fast allen Menschen die Schweißsekretion. Ein populäres Messgerät ist der Lügendetektor. Er misst die Hautleitfähigkeit von Personen, während diese reden.

Elektrische Muskelaktivität

Messapparate sind Elektromyografen, die elektrische Signale der Muskeln registrieren. Dazu werden Elektroden auf die Haut der Personen in der Nähe des interessierenden Muskelgewebes aufgesetzt. Im Untersuchungsfokus stehen die Herzmuskulatur, die autonom innervierte glatte Muskulatur der Eingeweide und Gefäße sowie die willkürlich innervierbare, quer gestreifte Muskulatur des Bewegungsapparates.

Kardiovaskuläre Aktivität

Das kardiovaskuläre oder Herz-Kreislauf-System liefert eine Reihe psychophysiologisch relevanter Variablen. Zu diesen zählen Herzfrequenz, Blutdruck und periphere Gefäßmotorik. Das kardiovaskuläre System interagiert mit vielen psychischen Prozessen. Dazu gehören u. a. motorische Aktivierung, Emotionen, Stress und Konzentrationsprozesse. Die wichtigsten Messgeräte in der Forschung sind das Elektrokardiogramm (EKG) und Blutdruckmessgeräte.

Hirnelektrische Aktivität

Das gängigste Verfahren zur Registrierung hirnelektrischer Aktivitäten an der Schädeloberfläche ist die Elektroenzephalografie. Die Aufzeichnung führt zu einem **Elektroenzephalogramm (EEG)**. Höhere kognitive Prozesse (Denken, Lernen) werden in der Forschung mit Hirnaktivitäten in Zusammenhang gebracht, indem die Untersuchungspersonen Testaufgaben bearbeiten, während das EEG abgeleitet wird.

Bildgebende Verfahren

Bildgebende Verfahren leisten eine Visualisierung von Körper und Gehirn. Dazu werden Strahlen durch den Körper gesendet (Röntgen-Computertomografie, CT) oder es werden die von verschiedenen Körperteilen ausgehenden Strahlungen außerhalb des Körpers registriert (Positron-Emissions-Tomografie, PET). Ein weiteres Verfahren ist die funktionelle Magnet-Resonanz-Tomografie (fMRT). Diese beruht auf der Eigenschaft von Atomen, sich wie eine Kompassnadel auszurichten, wenn sie in ein Magnetfeld gesetzt werden. Die Messgeräte setzen Körper und/oder Gehirn einem Magnetfeld aus. Sobald keine Energieimpulse mehr eingestrahlt werden, schwingen die Atome wieder zurück und richten sich parallel zu den Feldlinien des Magnetfeldes aus. Beim Zurückschwingen emittieren sie ein

Elektroenzephalogramm (EEG)

Das EEG misst elektrische Spannungsschwankungen an der Kopfoberfläche. Je nachdem, wo am Kopf die Elektroden aufgelegt werden, sind Aussagen über elektrische Aktivitäten in Hirnarealen möglich.

Bildgebende Verfahren

In Medizin, Psychologie und Neurowissenschaften sind bildgebende Verfahren diejenigen Methoden der Diagnostik, die ein Bild des Körperinneren von Menschen oder Tieren generieren.

messbares, hochfrequentes Signal. Die Form des Signals hängt von der Anzahl der Atome und deren Zustand im Gewebe ab. Durch Variation des magnetischen Feldes kann man Erkenntnisse über die Atome im untersuchten Körpergewebe gewinnen.

Weitere Körperfunktionen

Dazu rechnen die beiden Autoren u. a. die Messung von Körpertemperatur und Atmung. Die Körpertemperatur kann mit Thermistoren gemessen werden. Ihr temperaturempfindlicher Teil kann in verschiedenen Formen, z. B. als Plättchen oder Kügelchen, ausgebildet sein, deren Durchmesser weniger als einen Millimeter beträgt. Für Messungen an der Körperoberfläche werden zumeist Plättchen eingesetzt. Die Atmung kann mit einem Atemgürtel registriert werden. Dieser wird von den untersuchten Personen umgebunden und registriert Zeit und Durchmesser von Brustumfangsänderungen. Die Messung von Körpertemperatur und Atmung spielt eine wichtige Rolle bei der Erforschung circadianer Rhythmen.

Psychophysiologische Forschung hat in Psychologie und Medizin eine lange Tradition. Ihre Ursprünge liegen im 19. Jahrhundert. Ihre Fortschritte im 20. Jahrhundert beruhen v. a. auf der Entwicklung moderner Messgeräte (Blutdruckmessgeräte, EKG, EEG). Die Psychophysiologie setzt am ganzen Körper an und untersucht das zentrale und vegetative Nervensystem. Sie ist abzugrenzen von der neurowissenschaftlichen Hirnforschung, die sich auf das Gehirn von Menschen und anderen Säugetieren konzentriert. Mit der Optimierung der bildgebenden Verfahren (PET und fMRT) ist die Kombination dieser Verfahren mit Lernaufgaben und psychologischen Tests möglich. Psychologische Daten aus Leistungstests können mit den Aufzeichnungen aus PET und fMRT in der quantitativen Auswertung kombiniert werden.

2.8 Der Einsatz standardisierter psychologischer Tests

Psychologische Tests sind Instrumente der quantitativen Datenerhebung. Ein Test ist ein mehr oder weniger standardisiertes Routineinstrument, das zur quantitativen Messung bestimmter Merkmale von Personen eingesetzt wird und bestimmten statistischen Gütekriterien genügt. Die zentralen Gütekriterien sind Objektivität, Reliabilität und Validität. Psychologische Tests werden nach wissenschaftlich verbindlichen formalen und statistischen Regeln konstruiert. Dazu gehört auch eine Anwendung des Tests in einer Normstichprobe. Damit wird die Verteilung der bei Anwendern zu erwartenden Testwerte bestimmt.

Die Konstruktionsregeln und -schritte werden in Lehrbüchern der psychologischen Diagnostik ausführlich beschrieben (z. B. im klassischen Lehrbuch von Lienert und Raatz 1998). Nach der Konstruktion, wenn die Gütekriterien für den betreffenden Test in Vorstudien ermittelt wurden und für die Anwendung ausreichen, werden psychologische Tests in Fachverlagen publiziert. Damit stehen sie für die Anwendung durch psychologisch ausgebildete Fachkräfte zur Verfügung. Für die Nutzung eines Tests müssen die Anwender die-

sen beim Verlag kaufen. Die Preise für einen publizierten psychologischen Test mit allen Materialien (Handbuch, Testbögen, Auswertungsschablone, Auswertungsprogramm) beginnen bei über 100 Euro und liegen oft weit darüber.

Tests sind empirische Analyseinstrumente für verallgemeinerte Zwecke. Sie sind in Hinsicht auf Forschungsfragestellungen unspezifisch. Inhaltlich beziehen sie sich auf übergeordnete psychologische Konstrukte (Intelligenz, soziale Kompetenzen u. a.). Damit sind sie bei vielen Forschungsfragestellungen anwendbar, soweit es um quantitative Analysen geht. Dass Tests in den meisten Lehrbüchern zur empirischen Sozialforschung gar nicht behandelt werden, ist kein Zufall. Tests sind primär diagnostische Instrumente. Sie werden für die psychologische Individualdiagnostik konstruiert. Diese hat vor allem im klinischen und pädagogischen Bereich ihren festen Platz, aber auch in der Wirtschaft (Bewerbersauswahl, Personalentwicklung). Dennoch sind psychologische Tests für die Untersuchung größerer Stichproben in der quantitativen empirischen Forschung prinzipiell geeignet.

Einteilung psychologischer Tests

Leistungstests

Zu den Leistungstests gehören u. a. Intelligenztests, Entwicklungstests und Tests in Kulturtechniken (Rechnen, Lesen, Rechtschreibung).

Üblich ist die Einteilung in Leistungs- und Persönlichkeitstests. **Leistungstests** verlangen von den Prüflingen kognitive, sprachliche oder motorische Leistungen. Die Aufgabenform ist bevorzugt Multiple Choice, oft mit nur einer richtigen Antwort. Die Anzahl aller richtigen Antworten ergibt den Endwert eines Leistungstests.

Persönlichkeitstests sind Fragebögen zur Selbsteinschätzung. Die getestete Person geht eine Liste von Aussagen durch und kreuzt auf einer numerischen Skala an, inwieweit eine Aussage auf sie selbst zutrifft oder nicht. Persönlichkeitstests bestehen aus verschiedenen Faktoren, z. B. dem heute in der psychologischen Fachwelt anerkannten Fünf-Faktoren-Modell der Persönlichkeit. Zu jedem Faktor erhält eine getestete Person einen quantitativen Kennwert.

Psychologische Tests in der Wirtschaft

Es werden fortlaufend neue Tests entwickelt und publiziert. Tests in den beiden gängigsten Anwendungsfeldern, der klinischen und der pädagogischen Psychologie, sind in den Katalogen der deutschsprachigen Testverlage zahlenmäßig immer noch am stärksten vertreten. In den beiden letzten Jahrzehnten hat der arbeits- und wirtschaftspsychologische Bereich bei der Anzahl verfügbarer Tests jedoch aufgeholt.

Einen guten Überblick über die für wirtschaftspsychologische Studien verfügbaren Tests gibt das Handbuch der Kompetenzmessung (Erpenbeck et al. 2017). **Kompetenzmessung** meint die quantitative Bestimmung der verschiedenen Aspekte der beruflichen Handlungskompetenz bei Fach- und Führungskräften. Zur beruflichen Handlungskompetenz gehören Fachkompetenzen, Methodenkompetenzen, Selbstkompetenzen und Sozialkompetenzen.

Kompetenzmessung

Verfahren der Kompetenzmessungen bzw. Kompetenzdiagnostik messen die Ausprägungen von beruflichen Handlungskompetenzen bei den getesteten Personen.

- **Fachkompetenzen:** Dies sind die für eine berufliche Tätigkeit spezifischen Kompetenzen, die durch Ausbildung, Studium und Berufstätigkeit erlernt worden sind und überwiegend routinemäßig angewendet werden. Fachkompetenzen sind größtenteils identisch mit Fachwissen.
- **Methodenkompetenzen:** Das sind alle Aspekte der effektiven Selbstorganisation am Arbeitsplatz, vor allem Zeit- und Selbstmanagement.
- **Selbstkompetenzen:** Sie betreffen die Fähigkeit zur Selbstbewertung der eigenen Arbeitsleistungen und der beruflich relevanten Kompetenzen. Soweit eine Person Vergleiche mit anderen Personen (Kollegen) vornimmt, sollten diese zu realistischen Ergebnissen führen.
- **Sozialkompetenzen:** Dazu gehören alle Facetten der gelingenden Kommunikation und Kooperation am Arbeitsplatz mit Kunden und Kollegen (z. B. verbaler Ausdruck, Konfliktfähigkeit, Führungsstil).

Critical Incident Technique

Leistungsprüfungen und Selbsteinschätzungen sind die beiden klassischen Formen der Aufgabenstellung in psychologischen Tests. In den letzten Jahren hat sich im Rahmen der psychologischen und pädagogischen Diagnostik im Kontext von Beruf und Wirtschaft eine neue Aufgabenform etabliert. Die entsprechenden Testaufgaben bestehen aus der Schilderung kritischer, aber durchaus typischer Konstellationen im Beruf. Diese Testformen werden unter die Kategorie **Critical Incident Technique (CIT)** subsumiert. Als Texterzählung oder in einem Video wird eine Problemsituation aus der Berufspraxis präsentiert, als Lösung muss die untersuchte Person einen Text formulieren oder aus mehreren vorgegebenen Lösungen die aus ihrer Sicht beste aussuchen. Die schriftlichen Protokolle werden anhand eines Klassifikationssystems von den Testauswertern mit Punkten versehen. Deutschsprachige Tests mit der Methode des CIT zielen bevorzugt auf die Diagnostik sozialer Kompetenzen von Fach- und Führungskräften ab (Vollmers 2015).

Critical Incident Technique (CIT)

In Critical Incident Tests werden die kritischen beruflichen Situationen entweder als Text bzw. Geschichte präsentiert oder als Video in Form einer kurzen Spielsequenz mit Schauspielern.

Abschließende Bewertung: Standardisierte Tests oder selbst entwickelte Fragebögen?

Psychologische Tests, die als Publikation vorliegen, haben den Vorteil, dass sie vor ihrer Veröffentlichung einer umfangreichen inhaltlichen und statistischen Qualitätskontrolle unterzogen wurden – schlecht formulierte und unpassende Aufgaben wurden in Vorstudien herausgefiltert. Außerdem ist sichergestellt, dass sich die Testaufgaben mit den theoretischen Konstrukten bzw. Faktoren decken, für die die Tests jeweils stehen.

Schließlich ist auch die statistische Vergleichbarkeit der von den Testanwendern erhaltenen Ergebnisse gewährleistet, da im Testhandbuch in der Regel die Kennwerte der Normstichproben angegeben sind. Standardisierte Tests stellen deshalb oft die bessere Alternative zu selbst entwickelten quantitativen Fragebögen dar, die in der Regel ohne systematische Pretests auskommen. Empirische Ergebnisse mithilfe anerkannter psychologischer Tests sind deshalb im Allgemeinen objektiver, zuverlässiger und valider als die Ergebnisse von quantitativen Fragebögen aus eigener Entwicklung.



ZUSAMMENFASSUNG

Die drei Grundmethoden der Datengewinnung in Psychologie und Sozialwissenschaften sind Befragungen, Beobachtungen und Experimente. Es gibt sie jeweils in qualitativen und quantitativen Varianten. Kombinationen von zwei oder drei Grundmethoden, wie auch eine Verknüpfung von qualitativen und quantitativen Ansätzen, sind in der Datenerhebung bei etlichen Fragestellungen möglich und sinnvoll. Komplexere Erhebungsmethoden, wie Gruppendiskussionen oder Grounded Theory, basieren auf einer oder mehrerer der Grundmethoden.

Spezifisch für die Psychologie sind als weitere Datenerhebungsverfahren die Methoden der Psychophysiologie, der kognitiven Psychologie und standardisierte Tests. Psychophysiologische Methoden und standardisierte Tests sind rein quantitativ-messende Methoden. In der kognitiven Psychologie eröffnet sich dagegen mit der qualitativen Untersuchung von Denkvorgängen und Wissen ein bedeutendes Forschungsfeld, besonders wenn es um Forschungen zu den Bereichen Lernen und Gedächtnis geht.

Die Rolle der Forscher bzw. derjenigen, die die Datenerhebungen im Kontakt mit den untersuchten Personen durchführen, ist unterschiedlich. In Interviews und Befragungen gilt das Primat der Zurückhaltung, um Suggestionen zu vermeiden. In Experimenten sind die Forschungspersonen aktiv. Sie nehmen zielgerichtet Einfluss auf die Versuchspersonen. In Beobachtungen können sich die Forschenden zwischen einer aktiven Rolle (teilnehmend) oder einer passiven (nicht-teilnehmend) entscheiden.

LEKTION 3

DATENAUFBEREITUNG: QUALITÄTSKONTROLLE UND OPTIMIERUNG DER GEWONNENEN DATEN FÜR DIE AUSWERTUNG

LERNZIELE

Nach der Bearbeitung dieser Lektion werden Sie wissen, ...

- wie qualitative und quantitative Daten nach deren Erhebung jeweils für die Auswertung in spezifischer Weise vorbereitet werden.
- wie die Aufbereitung quantitativer Daten konkret mit dem Statistikprogramm SPSS erfolgt.
- was bei der Transkription von Protokollen verbaler Daten (Interviews) zu bedenken ist.
- inwieweit spezielle Software die Datenaufbereitung und Auswertung qualitativer Daten unterstützt.

3. DATENAUFBEREITUNG: QUALITÄTSKONTROLLE UND OPTIMIERUNG DER GEWONNENEN DATEN FÜR DIE AUSWERTUNG

Einführung

Angenommen, die aus der Befragung von zehn Studierenden gewonnene Datentabelle des Beispiels zur Studienmotivation hätte folgendes Aussehen:

Abbildung 16: Mögliche Daten in SPSS aus der Befragung zur Studienmotivation

	ID	Alter	Geschl	Semz	Motiv 1	Motiv 2	Motiv 3	Motiv 4	Motiv 5	Motiv 6	Motiv 7	Motiv 8	Motiv 9
1	1	20	w	3	3	3	3	2	2	3	3	4	4
2	2	180	w	1	3	1	2	1	3	3	3	2	2
3	3	24	m	6	3	4	3	3	4	3	4	3	3
4	4	40	m	6	2	2	3	4	4	4	1	1	2
5	5	24	w	6	3	3	4	2	3	1	.	.	.
6	6	22	w	2	2	2	3	2	2	2	3	4	3
7	7	21	w	3	2	3	3	3	3	3	3	3	3
8	8	23	m	3	2	3	2	4	2	2	3	3	3
9	9	20	w	1	3	3	3	3	3	3	2	2	2
10	10	28	w	6	2	3	2	3	4	4	4	3	4

Quelle: erstellt im Auftrag der IU, 2022.

Es gibt ganz offensichtlich fehlende (Spalten Motive 7, 8 und 9) und falsche Angaben (Spalte Alter). Wie ist damit umzugehen, bevor die Auswertung beginnt? Wissenschaftler sind bestrebt, möglichst mit allen Werten für die untersuchten Fälle zu arbeiten. Nichtsdestoweniger kann es immer wieder zu Lücken und fehlenden Angaben kommen. Mögliche Gründe hierfür sind schlichtweg falsche Angaben durch die Befragten oder elektronische Übermittlungsfehler während einer Online-Erhebung. Durch den Austausch einer Datentabelle zwischen verschiedenen Statistikprogrammen kann es zu weiteren Unstimmigkeiten kommen. Die Datenaufbereitung, das Thema dieser Lektion, hat zum Ziel, diese Unstimmigkeiten im Rahmen des Erlaubten und Üblichen zu glätten. Fälschungen dürfen nicht vorgenommen werden. Es gibt wissenschaftliche Konventionen, wie mit fehlenden und falschen Daten vor der statistischen Analyse umgegangen werden darf. Ähnliche Regeln gibt es für die Behandlung dieser Probleme bei qualitativen Daten.

3.1 Bedeutung und Ziele der Datenaufbereitung

Die Datenaufbereitung hat zum Ziel, die erhobenen Daten für die Auswertung auf mögliche Fehler zu prüfen sowie technisch zu optimieren. Die Datenaufbereitung findet nach der Datenerhebung und vor der Datenauswertung statt. In dieser Lektion werden ihre wesentlichen Schritte besprochen. Das Vorgehen ist bei qualitativen und quantitativen Daten unterschiedlich. Am Ende der Aufbereitung quantitativer Daten steht eine **Datentabelle** als Datei, die auf Fehler geprüft und ggf. verändert wurde. Diese Datei sollte von allen Statistikprogrammen gelesen werden können. Zwischen verschiedenen Statistikprogrammen und Computern muss ein problemloser Austausch der jeweiligen Datentabelle möglich sein.

Die qualitative Datenaufbereitung führt dagegen zu einer Textdatei, welche die schriftlichen Protokolle der qualitativen Befragungen, Beobachtungen oder Experimente enthält. Die Textdatei liegt der folgenden Text- bzw. Inhaltsanalyse der Datenprotokolle zugrunde. Das Dateiformat ist ein gängiges für Textverarbeitungsprogramme, was die Kompatibilität der Datei zwischen verschiedenen Programmen und Betriebssystemen sichert. Für die Auswertung qualitativer Daten in größeren Projekten wird die Textdatei unter Umständen in Programme der qualitativen Datenanalyse (**QDA-Programme**) eingespeist und weiterverarbeitet.

Die qualitative wie die quantitative Datenaufbereitung haben beide zum Ziel, die erhobenen Daten möglichst vollständig und fehlerfrei für die Datenauswertung nutzbar zu machen. In der Phase der Datenaufbereitung muss entschieden werden, wie mit falschen und unvollständigen Daten umgegangen wird.

Die Datenaufbereitung wird in vielen, aber nicht allen Lehrbüchern zur empirischen Forschung in einem kurzen Buchkapitel beschrieben. Döring und Bortz (2016, S. 579–595) widmen diesem Thema in ihrem 900 Seiten starken Lehrbuch ganze 16 Seiten und behandeln sowohl die quantitative als auch die qualitative Form. Im Handbuch von Baur und Blasius (2014) wird die Aufbereitung qualitativer und quantitativer Daten in zwei verschiedenen Kapiteln behandelt. Der quantitativen Datenaufbereitung werden zwölf Seiten gewidmet (Lück/Landrock 2014). Der qualitativen Variante ist im gleichen Handbuch ein kaum längeres Kapitel zugeordnet, das ausführlich auf die Funktionen von QDA-Programmen eingeht (Kuckartz/Rädiker 2014).

Eine korrekte Datenaufbereitung stellt keine besondere intellektuelle Herausforderung dar. Forscher mit Erfahrung in empirischen Studien erledigen die entsprechenden Schritte schnell und routiniert. Studierende, die erstmalig empirische Studien durchführen, sollten sich vor der Auswertung mit den Schritten der Datenaufbereitung beschäftigen, um Anfängerfehler zu vermeiden.

Datentabelle

Statistikprogramme benutzen Datentabellen. Jede Datentabelle hat die gleiche Struktur. Zeilen listen die Fälle auf, Spalten die Variablen, zumeist in der Reihenfolge der Fragen des verwendeten Fragebogens.

Qualitative Data Analysis (QDA)

QDA-Programme sind speziell für die qualitative Sozialforschung entwickelte Software, die den qualitativen Forschungsprozess in allen Phasen unterstützt: von der Datenerhebung bis zum Verfassen des Abschlussberichts.

3.2 Schritte der Aufbereitung von quantitativen Daten

Ziel der Aufbereitung der quantitativen Daten ist es, eine Datei zu erstellen, in der alle untersuchten Fälle vollständig, d. h. mit allen Werten der Variablen, enthalten sind. Die Variablen sind die Fragen bzw. Aussagen eines Tests, Frage- oder Beobachtungsbogens.

Die Hauptschritte bei der Aufbereitung quantitativer Daten sind:

1. Übertragung der Angaben aus den Tests, Frage- oder Beobachtungsbögen in eine Tabelle als Datei,
2. Erstellung eines **Codebooks**,
3. Speicherung der Datei in einem Format, das von allen gängigen Statistikprogrammen gelesen werden kann,
4. Import der Datei in das Statistikprogramm,
5. Prüfung auf Fehler, Unvollständigkeiten (fehlende Werte) sowie Inkonsistenzen und deren Korrektur,
6. Neudefinition von Gruppierungsvariablen für die quantitative Auswertung sowie
7. abschließende Speicherung der veränderten Datei im Statistikprogramm für die Auswertung.

Codebook

Ein Codebook sollte schon vor der Datenerhebung erstellt werden. In ihm wird für alle Fragen bzw. Variablen definiert, welcher Wert der Variablen in der Datentabelle zu einem Kreuz bzw. einer Aussage auf dem Frage- oder Beobachtungsbogen gehört.

Liegen die ausgefüllten Fragebögen, Tests oder Beobachtungsbögen in Papierform vor, werden die Daten sukzessive per Hand in ein Tabellenkalkulationsprogramm oder direkt in das für die Auswertung benutzte Statistikprogramm eingegeben. Bei großen Fallzahlen ist das eine mühevoll und monotone Arbeit. Zum Glück werden viele Umfragen heute über das Internet durchgeführt. Die verwendete Online-Software sorgt dafür, dass die Werte direkt in eine Datentabelle bzw. Datenbank geschrieben werden. Jede Tabelle bzw. Datenbank ist in gleicher Weise organisiert. Jede Zeile umfasst einen Fall, jede Spalte markiert eine Variable.

In einem Codebook wird die Zuordnung der Werte der Datentabelle zu den Originalfragen des Fragebogens beschrieben. Jede Frage ist eine Variable. Ihr wird ein relativ kurzer Variablenname, oft ein Akronym der Frage im Fragebogen, zugewiesen. Viele Statistikprogramme akzeptieren sehr lange Variablennamen, in SPSS etwa liegt die Obergrenze bei 65 Zeichen. Kurze Variablennamen sind aber praktisch für die Auswertung und Ergebnisdarstellung.

Außerdem steht im Codebook, welchen Wertebereich die Variablen des Fragebogens umfassen und was die Werte bedeuten. Bei Skalenwerten auf dem Fragebogen ergeben sich diese Angaben von selbst. Bei nicht-numerischen Variablen ist die entsprechende Angabe dagegen alles andere als trivial. Beispielsweise kann das Geschlecht männlich bzw. weiblich als 0 und 1, 1 und 2 oder M und W für die Datentabelle kodiert werden. Bei Bildungsabschlüssen muss festgelegt werden, welche Formen von Abschlüssen in der Stichprobe auftauchen und welche Zahlen oder Buchstaben in der Datentabelle für die verschiedenen Bildungsniveaus stehen. Codebooks sind notwendig, wenn die Datensätze von vielen Nutzern ausgewertet werden. Für kleine Studien, die Studierende im Rahmen einer Bachelor- oder Masterarbeit durchführen, sind sie zumeist entbehrlich.

Die vollständige Datentabelle sollte in einem Format gespeichert werden, das von jedem Statistikprogramm erkannt wird. Üblich sind die Formate Character Separated Variables (Dateiname.csv), Text (Dateiname.txt) sowie Rich-Text-Format (Dateiname.rtf). Diese internationalen Dateiformate bestehen aus Zeilen und Spalten, die durch ein einheitliches Trennzeichen getrennt sind. Üblich sind die Trennzeichen Semikolon, Komma oder Punkt. Steuerzeichen, wie z. B. Zeilenumbruch, Leerzeichen oder Absatzende, finden sich in diesen Datentabellen nicht.

Für das Einlesen der Datentabelle in das Statistikprogramm stehen meist spezielle **Importbefehle** zur Verfügung. Alternativ können die Dateien auch mit dem Befehl „Datei öffnen“ eingelesen werden. Die drei genannten Dateiformate können von jedem Statistikprogramm problemlos erkannt werden, wenn die Trennzeichen den Konventionen entsprechen. Nach dem Import wird die Tabelle im Statistikprogramm dargestellt und steht der statistischen Auswertung zur Verfügung.

Im jeweiligen Statistikprogramm findet dann eine Prüfung auf Fehler, Inkonsistenzen und Unvollständigkeiten durch die Bearbeiter statt. Dies ist eine Form der internen Qualitätssicherung. Bei kleinen Stichproben kann der Qualitätscheck per Augenschein erfolgen. Die Wissenschaftler gehen die Datentabelle spaltenweise durch und prüfen, ob in den Zellen einzelne fehlende oder offensichtlich falsche Werte stehen.

Dieses Vorgehen ist bei großen Datensätzen nicht möglich. Hier werden zur Datenprüfung Häufigkeitstabellen, Grafiken und einfache statistische Kennwerte als Befehle aufgerufen. Fehlende Werte tauchen in den Ergebnissen auf, offensichtlich falsche Werte ebenfalls. Ein offenbar falscher Wert wäre z. B., wenn nur Erwachsene befragt wurden, bei der Variable Alter der Wert 11. Grundsätzlich ist der Verdacht auf eine fehlerhafte Eingabe bei nicht plausiblen **Ausreißern** in den Werten einer Variablen angebracht.

Fehlerhafte Werte werden anschließend per Hand im Datensatz gelöscht. An dieser Stelle erscheint im Programm ein fehlender Wert (engl. „missing value“). Leere Zellen erkennt jedes Statistikprogramm als fehlende Werte. Für die Darstellung der Ergebnisse einer Variablen in Tabellen und Grafiken ist es aber besser, einen speziellen Wert für fehlende Werte einzusetzen, der empirisch nicht gemessen wurde und nicht gemessen werden kann, z. B. den Wert -99. In SPSS ist dafür in der Variablenansicht, der Datendefinitionstabelle, ein entsprechender Wert zu kodieren. Eine weitere Möglichkeit bei fehlenden Werten ist das Einsetzen des Mittelwerts, der mit den anderen Werten der Variablen errechnet wurde. Der Mittelwert ist statistisch der plausibelste Schätzwert für fehlende Werte.

Neben fehlenden Werten gibt es noch zwei weitere häufig auftretende Unkorrektheiten bzw. Inkonsistenzen, die durch den Import einer Tabelle bzw. durch den Austausch der Datentabelle zwischen verschiedenen Programmen entstehen können. Erstens werden häufig Dezimalstellen nicht richtig erkannt, was u. a. in der unterschiedlichen Dezimaltrennung im Englischen (Punkt) und Deutschen (Komma) begründet sein kann. In SPSS können zwar in der Variablenansicht Einstellungen zur Dezimaltrennung vorgenommen werden, unter Umständen ist aber dennoch eine Korrektur per Hand für jeden einzelnen Fall notwendig. Zweitens kommt es vor, dass das Skalenniveau (Nominalskala, Ordinalskala, Intervallskala, Verhältnisskala) nicht richtig erkannt wurde. In SPSS sollte für jede Variable in der Variablensicht das Skalenniveau überprüft und ggf. geändert werden.

Importbefehl

In Statistikprogrammen gibt es einen Importbefehl für das erstmalige Einlesen einer Datentabelle. Dieser zeigt in der Vorschau, ob die Datentabelle richtig erkannt wird.

Ausreißer

Als Ausreißer werden in der Statistik Extremwerte bezeichnet. Es sind die Werte am Rande einer Verteilung, die die Kennwerte (z. B. den Mittelwert) einer Verteilung beeinflussen.

Zur Datenaufbereitung kann auch die Neudefinition von Variablen gehören, die aus anderen abgeleitet bzw. berechnet werden. Oft ist es vor der Auswertung sinnvoll, neue Variablen zur Gegenüberstellung von Gruppen von Fällen bei anderen Variablen (z. B. für Mittelwertvergleiche) zu definieren. Beispielsweise wird die Variable Alter in statistischen Studien oft in Gruppen (unter 18, 18–30, 31–45, 46–65, über 65 Jahre) unterteilt. In unserem Beispiel der Untersuchung der Studienmotivation könnten die Bachelorstudierenden vor der Auswertung in die beiden Gruppen Anfänger (Semester 1–3) und Fortgeschrittene (Semester 4–6) unterteilt werden. Die beiden Werte der neuen Gruppierungsvariablen sind mit 1 und 2 zu kodieren. Das Skalenniveau wäre eine Ordinalskala. „Neukodieren“ (die alte Variable ist neben der neuen weiter in der Tabelle enthalten) oder „Umkodieren“ (die alten Werte der Variablen werden überschrieben) bestehender Werte heißt das in der Fachsprache. Die meisten Statistikprogramme bieten dazu Prozeduren an.

Nach Abschluss der Plausibilitäts- und Fehlerprüfung ist der letzte Schritt der Datenaufbereitung das Speichern der vorgenommenen Veränderungen in einem Dateiformat, das zum jeweiligen Statistikprogramm gehört. Bei SPSS ist es das Standardformat Dateiname.sav.

Anwendung auf das Beispiel Studienmotivation

Eine Untersuchung zur Studienmotivation beispielsweise hätte bei zehn Fällen als Datentabelle folgendes Aussehen. Die Reihenfolge der Spalten entspräche den Fragen des Fragebogens. Zusätzlich gibt es eine Gruppierungsvariable für die Semesterzahl (Anfänger oder Fortgeschrittene). Sie hat das Akronym Studierendenstatus.

Abbildung 17: Datentabelle in SPSS für das Beispiel zur Studienmotivation

	ID	Alter	Geschl	Semz	Motiv 1	Motiv 2	Motiv 3	Motiv 4	Motiv 5	Motiv 6	Motiv 7	Motiv 8	Motiv 9	Studierendenstatus
1	1	20	w	3	3	3	3	2	2	3	3	4	4	1
2	2	18	w	1	3	1	2	1	3	3	3	2	2	1
3	3	24	m	6	3	4	3	3	4	3	4	3	3	2
4	4	40	m	6	2	2	3	4	4	4	1	1	2	2
5	5	24	w	6	3	3	4	2	3	1	3	3	3	2
6	6	22	w	2	2	2	3	2	2	2	3	4	3	1
7	7	21	w	3	2	3	3	3	3	3	3	3	3	1
8	8	23	m	3	2	3	2	4	2	2	3	3	3	1
9	9	20	w	1	3	3	3	3	3	3	2	2	2	1
10	10	28	w	6	2	3	2	3	4	4	4	3	4	2

Quelle: erstellt im Auftrag der IU, 2022.

Die Variablenansicht in SPSS

In SPSS wird in der internen Darstellung zwischen der Datensicht und der Variablenansicht unterschieden. Die Datensicht ist die Datentabelle, anhand derer die Auswertungen vorgenommen werden. Die **Variablenansicht** enthält die Datendefinitionen, die maßgeblich darüber entscheiden, welche Auswertungen möglich und sinnvoll sind sowie welche grafischen Darstellungen am besten zu einer Variable passen.

Variablenansicht

Die Variablenansicht dient der Beschreibung der Variablen, damit die Programmbeefehle exakt durchgeführt werden können. Außerdem kann dort die Ausgabe von Tabellen und Grafiken für eine Variable beeinflusst werden.

Abbildung 18: Variablenansicht in SPSS im Datensatz zur Studienmotivation

	Name	Typ	Breite	Dezimal	Beschriftung	Werte	Fehlend	Spalten	Ausrichtung	Messniveau	Rolle
1	ID	Numerisch	2	0	Fallnummer	Keine	Keine	2	Rechts	Nominal	Eingabe
2	Alter	Numerisch	2	0	Alter	Keine	Keine	5	Rechts	Metrisch	Eingabe
3	Geschl	Zeichenfolge	1	0	Geschlecht	Keine	Keine	7	Rechts	Nominal	Eingabe
4	Semz	Numerisch	2	0	Semesterzahl	Keine	Keine	5	Rechts	Metrisch	Eingabe
5	Motiv1	Numerisch	2	0	Kompetenzen ...	Keine	Keine	5	Rechts	Metrisch	Eingabe
6	Motiv2	Numerisch	2	0	Eigene Vorstel...	Keine	Keine	5	Rechts	Metrisch	Eingabe
7	Motiv3	Numerisch	2	0	Neue Anregun...	Keine	Keine	5	Rechts	Metrisch	Eingabe
8	Motiv4	Numerisch	2	0	Austausch mit ...	Keine	Keine	5	Rechts	Metrisch	Eingabe
9	Motiv5	Numerisch	2	0	Positive Rück...	Keine	Keine	5	Rechts	Metrisch	Eingabe
10	Motiv6	Numerisch	2	0	Kooperation m...	Keine	Keine	5	Rechts	Metrisch	Eingabe
11	Motiv7	Numerisch	2	0	Freies Lernen	Keine	Keine	5	Rechts	Metrisch	Eingabe
12	Motiv8	Numerisch	2	0	Partner der Le...	Keine	Keine	5	Rechts	Metrisch	Eingabe
13	Motiv9	Numerisch	2	0	Interessen verf...	Keine	Keine	5	Rechts	Metrisch	Eingabe
14	Studierendenstatus	Numerisch	1	0	Anfänger/in od...	{1, Anfäng...	Keine	8	Rechts	Ordinal	Eingabe

Quelle: erstellt im Auftrag der IU, 2022.

Die Variablen sind in der Variablenansicht zeilenweise angeordnet in der Reihenfolge der Spalten in der Datenansicht. Die verschiedenen Spalten der Variablenansicht sind durch Klick auf den Spaltenkopf bearbeitbar. In den erscheinenden Menüs nehmen die Nutzer Eingaben vor. Die Bedeutung der verschiedenen Spalten ist wie folgt (von links nach rechts in obiger Abbildung):

- **Name:** Gemeint ist der Name der Variablen, wie er intern von SPSS-Befehlen gelesen wird.
- **Typ:** Das ist der interne Datentyp der Variablen in SPSS, entweder numerisch oder als Zeichenkette.
- **Breite:** Hier geht es um die Anzahl der Zeichen, die bei der Darstellung der Werte einer Variablen in der Datensicht maximal verwendet werden können.
- **Dezimalstellen:** Dies steht für die maximale Anzahl der Dezimalstellen, die bei der Darstellung der Werte einer Variablen in der Datensicht verwendet werden können. Bei ganzzahligen Werten von Variablen – in unserer Studie zur Studienmotivation wären das alle Variablen – würde hier der Wert 0 stehen.
- **Beschriftung:** Sie dient der verbalen Beschreibung der Variablenamen in der Datentabelle.
- **Werte:** Sie dienen der Beschreibung der verschiedenen Werte einer Variablen und werden bei Ergebnisausdrucken angezeigt. Bei der Variable Geschlecht würde z. B. bei 1 „weiblich“ eingegeben, bei 2 „männlich“.

Messniveau
Das metrische Skalenniveau in SPSS umfasst Intervall- und Verhältnisskala. Metrische Skalen beinhalten bei den Daten immer numerische Werte.

- **Fehlend:** Hier werden numerische Werte für fehlende Werte deklariert. Dabei sind immer Werte zu wählen, die empirisch nicht möglich sind. Für alle Variablen einer Tabelle sollten die fehlenden Werte einheitlich deklariert werden, beispielsweise immer als -99.
- **Spalten:** Das ist die Breite der Spalten in Zeichen in der Datensicht. Sie umfasst bei numerischen Variablen nur wenige Zeichen.
- **Ausrichtung:** Das ist die Ausrichtung der Variablendarstellung in der Datensicht. Zumeist passt links als Voreinstellung.
- **Messniveau:** Hier wird eingestellt, ob es sich um eine Nominal-, eine Ordinal- oder eine metrische Skala handelt.
- **Rolle:** Hier können Angaben dazu gemacht werden, welche Rolle eine Variable für die Auswertung haben soll, z. B. ob es sich um eine Gruppierungsvariable handelt. Werden in diesem Dialogfeld Angaben für die Variable gemacht, sind unter Umständen bestimmte statistische Prozeduren für diese Variablen in SPSS nicht mehr aufrufbar.

In anderen Statistikprogrammen wird in der Regel ebenso zwischen einer Datentabelle, die die Werte enthält, und einer Datendefinitionstabelle, die den formalen Rahmen für die Auswertung der Variablen setzt, unterschieden. Die entsprechenden Einstellungen sind ähnlich wie in SPSS.

3.3 Schritte der Aufbereitung von qualitativen Daten

Interviews werden in der Regel mitgeschnitten und liegen als Audiodateien vor. Diese Mitschnitte enthalten sowohl die Fragen der Interviewer als auch die Antworten der Befragten. Bei Beobachtungen und Experimenten ist heute ebenfalls der Einsatz technischer Hilfsmittel der Registrierung Usus. Bild- und Videodateien fungieren als Beobachtungsprotokolle. Bis in die 80er-Jahre des letzten Jahrhunderts hatten Beobachtungsprotokolle die Form handschriftlicher Aufzeichnungen auf Papier, erstellt von den Beobachtern. Die sozialwissenschaftlichen und psychologischen Methoden der Auswertung – das sind alle Formen der qualitativen Text- und Inhaltsanalyse – setzen jedoch in der Regel nicht direkt an multimedialen Dateiformaten an, sondern haben Texte zur Datengrundlage. Die Mediendateien werden deshalb für die Auswertung in Texte überführt. Dieser Vorgang heißt **Transkription**. Audio- und Videodateien werden transkribiert (abgetippt) zu einem Transkript. Dieses liegt den die Auswertung vornehmenden Personen vollständig vor.

Transkription
Die Überführung von Audio- oder Videomitschnitten der Datenerhebung in einen Text, der als Basis der qualitativen Auswertung fungiert, wird Transkription genannt.

„Vor einer Transkription sind mehrere Entscheidungen zu treffen. Diese beziehen sich darauf, wie vollständig, in welchem Umfang und auf welche Weise die Daten wiedergegeben werden sollen“ (Schreier 2013c, S. 246).

Ein vollständiges Transkript für Fremde zu erstellen, die nicht die Erhebung durchgeführt haben, ist ein zu hohes Ziel. Wer als Wissenschaftler an sich selbst den Anspruch stellt, alle Daten aus Interviews, Beobachtungen und Experimenten zu 100 % zu transkribieren,

sollte sich vor Augen führen, dass dieses Ideal niemals zu erreichen ist. Die Transkriptionsexperten Dresing und Pehl (2018) schreiben in ihrer Handreichung zur Transkription qualitativer Daten:

Jeder, der transkribiert oder mit Transkripten arbeitet, sollte sich im Vorfeld bewusst sein, dass eine Transkription nie die Gesprächssituation vollständig festhalten kann. Dafür spielen während der Kommunikation zu viele Faktoren eine Rolle, die unmöglich alle erfasst werden können. Selbst ein Transkript, das sich sehr eng an der Lautsprache orientiert, ‚vergisst‘ beispielsweise nonverbale Aspekte wie Geruch, Raumsituation, zeitliche Verortung, Optik, Mimik und Gestik. Es muss eine Fokussierung auf bestimmte Aspekte stattfinden. Diese Aspekte variieren je nach Forschungsziel bzw. intendierter Verwendung des Transkripts und der konkreten Situation. Ein Beispiel: Man befragt einen Geschäftsführer, ob er denn im nächsten Jahr Stellen streichen werde und er überlegt zehn Sekunden, schaut auf den Boden und sagt leise ‚Nöö‘. Wenn im Transkript nun ‚Nein‘ steht, kommt man sehr wahrscheinlich zu einer anderen Deutung, als wenn die Pausenlänge und Gestik mit transkribiert worden wären (Dresing/Pehl 2018, S. 16f.).

Kodierungsregeln legen fest, wie bei der Transkription mit den nonverbalen Aspekten von Interviews umzugehen ist. Nonverbale Aspekte der Kommunikation sind z. B. das Schweigen der Befragten bei einer Frage, Verzögerungen bei der Antwort oder ein Lächeln während des Sprechens. Weitere Aspekte sind die Betonung von Wörtern (Lautstärke) oder emotionale Ausbrüche (Trauer, Wut, Freude). Kuckartz und Rädiker (2014, S. 391) geben einen kurzen Überblick zu den wichtigsten bei qualitativen Forschungsprojekten relevanten Transkriptionsregeln. Ein leicht umsetzbares Transkriptionssystem präsentieren Dresing und Pehl (2018, S. 20f.).

Es hängt von der Forschungsfragestellung und den Erkenntniszielen ab, in welchem Ausmaß die Notation der non-verbalen Begleitumstände der Gesprächs- oder Beobachtungssituation erforderlich ist. Möglicherweise zielt die Hauptforschungsfrage gar nicht auf die von den untersuchten Personen explizit verbalisierten Aussagen und Meinungen. Stattdessen geht es vielleicht um verborgene Persönlichkeitsmerkmale oder unterschwellige politische Einstellungen oder Vorurteile. Bei diesen Erkenntniszielen sind nonverbale Aspekte, wie z. B. das Heben oder Senken der Stimme, eine feste oder zittrige Aussprache u. Ä. von Relevanz für die Auswertung. Sie gelten als Manifestationen eines Untersuchungsgegenstandes, der hinter den sprachlichen Aussagen liegt und psychologisch oder soziologisch zu interpretieren ist. Die Interpretation des nicht Gesagten ist der eigentliche Fokus der Auswertung. Passende Auswertungsverfahren sind überwiegend tiefenpsychologischer und hermeneutischer Natur (vgl. zu den Auswertungsdetails Kurt/Herbik 2014).

Die intendierten Auswertungsverfahren sind vorab bei der Transkription zu bedenken. Manche Auswertungsverfahren sehen Regeln für die Transkription nonverbaler Gesprächs- und Beobachtungsanteile als Teil ihrer spezifischen Methodik vor. Eine Beschäftigung mit den infrage kommenden Auswertungsverfahren bei einer qualitativen Untersuchung ist deshalb immer vor der Transkription erforderlich und sollte vor oder spätestens während der Erhebung stattfinden. Bei den zusammenfassenden, nicht interpretierenden Auswertungsverfahren der qualitativen Inhaltsanalyse nach Mayring und bei der Grounded Theory ist eine detaillierte Kodifizierung und Transkription nonverbaler Aspekte der Erhe-

bungssituation im Allgemeinen entbehrlich. Diese beiden Auswertungsverfahren beziehen sich allein auf das gesprochene Wort oder das direkt in einer Situation beobachtete Verhalten von Menschen.

Der Einsatz von QDA-Software in der Auswertung

Im Handbuch von Baur und Blasius (2014) widmet sich Kapitel 27 der Datenaufbereitung und -bereinigung in der qualitativen Sozialforschung (Kuckartz/Rädiker 2014). Einer der Autoren des Kapitels ist der Soziologe Udo Kuckartz (geb. 1951). Er ist der geistige Vater des Softwareprogramms MAXQDA, das in Deutschland vermutlich am häufigsten eingesetzt wird, wenn es um die Aufbereitung und Analyse qualitativer Daten geht. Weitere beliebte Programme sind ATLAS.ti und NVivo. Spezifische Software für die qualitative Sozialforschung heißt QDA-Software. QDA steht für „Qualitative Data Analysis“. Diese Software unterstützt den qualitativen Forschungsprozess in allen Phasen, also während der Datenerhebung, bei der Transkription, bei der inhaltsanalytischen Auswertung, bei der Literatursuche und -verarbeitung sowie beim Verfassen des Abschlussberichts. Im Unterschied zur Statistiksoftware gibt es in der QDA-Software allerdings keine aufrufbaren Befehlsroutinen für die Datenauswertung.

Was die QDA-Software nicht leistet

Die immensen Fortschritte der Computertechnologie in den letzten drei Jahrzehnten haben die Forschungspraxis in Psychologie und Sozialwissenschaften nachhaltig beeinflusst. Die QDA-Software soll die Forschungsarbeit in der qualitativen Sozialforschung unterstützen. Um den Nutzen von QDA-Software bewerten zu können, sind grundsätzlich drei Punkte zu bedenken.

Erstens übernimmt die Software nicht die Auswertung. Bisher ist keine QDA-Software erhältlich, die standardisierte Auswertungsprozeduren anbietet (im Unterschied zur Statistiksoftware in der quantitativen Forschung). Die Auswertung bleibt zu fast 100 % die kreative gedankliche, kommunikative und schriftliche Arbeit der beteiligten Wissenschaftler.

Die Software übernimmt zweitens nicht die Transkription. Um Audio- und Videodateien in verbale Protokolle zu verwandeln, ist immer die Eingabe von Text durch Menschen erforderlich. In den letzten Jahren hat, da sich die entsprechenden Programme verbessert haben, der Einsatz von **Spracherkennungssoftware** zwar deutlich zugenommen, die so gewonnenen Texte müssen für die Auswertung aber weiter aufbereitet werden.

Der Nutzen von QDA-Software besteht drittens vor allem darin, dass sich große Textmengen für die qualitative Forschungspraxis darin besser organisieren lassen als mit Textverarbeitungsprogrammen. In größeren Projekten, an denen mehrere Wissenschaftler beteiligt sind, lässt sich so der Datenaustausch zwischen den Computersystemen besser durchführen. Auswertungsschritte können geplant und dokumentiert werden. Für Studierende lohnt sich die Anschaffung und Benutzung der QDA-Software für ein Projekt zumeist nicht. Die am weitesten verbreiteten QDA-Programme sind im Übrigen alle kostenpflichtig.

Spracherkennungssoftware

Bei Programmen zur Spracherkennung handelt es sich um Diktiersoftware: Statt zu schreiben bzw. zu tippen, sprechen die Anwender in das Mikrofon des Computers.

Die Nichtprogrammierbarkeit der Schritte von Text- und Inhaltsanalysen

Die Durchführung von Textanalysen ist nicht als Prozedur eines Computerprogramms formalisierbar. Es wird in absehbarer Zukunft keine Software geben, die Menschen die Verstehens- und Analysearbeit bei wissenschaftlich auszuwertenden Texten abnehmen wird. Computer sind nicht fähig, die Bedeutungsnuancen einzelner Textpassagen zu erkennen. Es ist ihnen unmöglich, verbale Kategorien zu bilden, die längere Textpassagen zusammenfassen, verdichten oder interpretieren. Das muss der Nutzer der QDA-Software tun. Dies sind die Kernprozesse der qualitativen Text- und Inhaltsanalyse.

In den letzten Jahren kam es zu einer deutlichen Verbesserung der Spracherkennungs- und Übersetzungssoftware. Bei Computerenthusiasten unter den Novizen im Feld der wissenschaftlichen Textanalyse führt dies unter Umständen zu der irrigen Annahme, dass die Inhaltsanalysen schon bald eine Domäne der Computer sein werden. Bei der Spracherkennung geht es um die Umwandlung von akustischen Signalen in einzelne Worte, bei Übersetzungen um Zuordnungen von Worten aus Wortlisten für verschiedene Sprachen. Diese beiden programmierbaren Routinen spielen jedoch für die qualitative Inhaltsanalyse nur eine untergeordnete Rolle.

Wissenschaftliche Text- und Inhaltsanalysen beinhalten die Überführung von vorliegenden Texten in Texte höherer Ordnung, die Ergebnisse der Inhalts- bzw. Textanalyse. Diese fungieren als Erklärung für die Texte niederer Ordnung, die Datenprotokolle. Dazu sind komplexe Rekonstruktions-, Reorganisations-, Verdichtungs- und Ergänzungsprozesse der Ausgangstexte notwendig.

Formalisierte Varianten einfacher Textanalysen werden von Computerprogrammen dagegen sehr gut bewältigt. Die Ermittlung der Zahl der Zeichen oder Worte sowie die Feststellung, welche Zeichenketten in einem Text vorhanden sind oder nicht, das gelingt jedem Textverarbeitungsprogramm. Die Verbindung einfacher numerischer Zeichenanalysen mit komplexeren Verfahren der deskriptiven und prüfenden Statistik führt zu einer Software, die quantitative Inhaltsanalysen durchführen kann. Quantitative Analysen qualitativer Datenprotokolle sind in der empirischen Forschung möglich und sehr schnell von Computern abzuarbeiten. Im Ergebnis erhalten die Wissenschaftler Aussagen darüber, welche Begriffe wie oft in Texten vorkommen oder wie variantenreich Texte in Bezug auf ihre Wortwahl sind. Unterschiedliche linguistische Stile von Texten können so vergleichend zwischen verschiedenen Autoren oder Rednern analysiert werden.

Die Software für quantitative Inhaltsanalysen ist in etlichen QDA-Programmen enthalten. So hat z. B. das Paket MAXQDA als Erweiterungsmodul das Programm MAXDictio, das unterschiedliche Funktionen zur Ermittlung und Visualisierung (Stammbäume, Netzwerke) von Worthäufigkeiten umfasst. Im Internet sind zahlreiche Programme der quantitativen Inhaltsanalyse kostenlos verfügbar, zumeist allerdings nur in englischsprachigen Versionen. Das Programm AQUAD 7 ist dagegen kostenlos auf Deutsch erhältlich. Günter Huber, ein pensionierter Professor für Pädagogische Psychologie an der Universität Tübingen, hat es gemeinsam mit dem Erziehungswissenschaftler Leo Gürtler entwickelt. AQUAD hat Funktionen eines QDA-Programms und führt quantitative Inhaltsanalysen durch.

Spezielle Software allein für quantitative Inhaltsanalysen gibt es heute kaum noch. Häufigkeitsanalysen von Kategorien und Worten sind mit QDA-Software möglich. In Statistikprogramme werden zudem Text- bzw. Data-Mining-Verfahren integriert. Dies sind programmierbare Algorithmen, um statistische und grafische Verfahren auf Texte anzuwenden.



ZUSAMMENFASSUNG

Die Datenaufbereitung ist ein Zwischenschritt nach Abschluss der Datenerhebung und vor der Auswertung. Ziel ist es, die Daten auf Richtigkeit und Vollständigkeit zu prüfen und ggf. zu korrigieren und zu glätten. Bei qualitativen und quantitativen Auswertungen liegen unterschiedliche Datenformen vor, deshalb verläuft auch die Datenaufbereitung jeweils anders.

Quantitative Daten werden direkt im Statistikprogramm (z. B. SPSS), mit dem die Auswertung erfolgen soll, aufbereitet. Die Tabelle des Datensatzes der untersuchten Stichprobe wird auf fehlende und falsche Werte geprüft und die Daten- bzw. Variablendefinitionstabelle wird vollständig erstellt.

Die Datenaufbereitung qualitativer Daten ist Teil des Transkriptionsprozesses, der zuvor festgelegten Transkriptionsregeln folgt. Zusätzlich kann QDA-Software zum Einsatz kommen. Dies ist sinnvoll, wenn die QDA-Software die anschließenden Auswertungen unterstützen soll.

LEKTION 4

DATENANALYSE: AUSGEWÄHLTE AUSWERTUNGSMETHODEN

LERNZIELE

Nach der Bearbeitung dieser Lektion werden Sie wissen, ...

- aus welchen spezifischen Auswertungsschritten quantitative und qualitative Datenanalysen bestehen.
- wie Berechnungen in der beschreibenden Statistik durchzuführen und in SPSS als Befehl aufzurufen sind.
- wie Berechnungen in der prüfenden Statistik durchzuführen und in SPSS als Befehl aufzurufen sind.
- wie multivariate statistische Analysen von großen Datenmengen im Allgemeinen funktionieren.
- was die Erkenntnisziele von qualitativen Text- und Inhaltsanalysen sind und aus welchen Analyseschritten sie bestehen.

4. DATENANALYSE: AUSGEWÄHLTE AUSWERTUNGSMETHODEN

Einführung

Die deutsche Journalistin Alena Schröder (geb. 1979) berichtet in ZEIT WISSEN über ihre Erfahrungen als Teilnehmerin von Selbstbehauptungstrainings (Schröder 2019). Sie beschreibt sich dabei als Person, die dazu neigt, Konflikten aus dem Weg zu gehen und zurückzustecken, bis hin zur Verleugnung berechtigter eigener Ansprüche und Interessen. Für 2019 hat sie sich vorgenommen, öfter einmal Nein zu sagen und mutiger den Ansprüchen und Angriffen anderer entgegenzutreten. Sie fragt sich: „Kann ich trainieren, im passenden Moment eine Seite von mir zu zeigen, die sonst nichts mit meinem Selbstbild zu tun hat?“ (Schröder 2019, S. 19).

Sie beantwortet diese Frage mit Ja und nimmt an zwei Selbstbehauptungstrainings teil. Im Krav-Maga-Training lernt sie, bei körperlichen Angriffen zuzuschlagen und zuzutreten, auch jenseits der Regeln des Fair Play. Krav Maga ist ein aus Israel stammendes eklektizistisches Selbstverteidigungstraining, das ohne psychologische oder andere theoretische Fundierungen auskommt. Das zweite Training ist ein Kommunikationstraining. Es verspricht das Erlernen von Strategien, um persönliche Interessen gegenüber Widersachern rhetorisch zu behaupten und stammt aus der Schauspielausbildung. Konzipiert hat es der 2001 verstorbene Schauspieler und Clown Frieder Nögge mit Bezug auf antike Temperamentslehren (Fischer 2004). Es wird heute von der Kommunikationstrainerin Isabel García (geb. 1969) vertreten (García 2016) und basiert auf einer psychologischen Persönlichkeitstheorie, die vier Typen unterscheidet:

Es gibt den Erde-Typ, der ruhig, überzeugend, solide, dafür wenig emotional ist. Den Wasser-Typ, der tieferschürfend und mitfühlend ist, sich schwer abgrenzen kann und oft von seinen Emotionen geleitet wird. Den Feuer-Typ, der leidenschaftlich und begeisternd, dafür aber auch aufbrausend und arrogant sein kann. Sowie den Luft-Typ, der neugierig und lustig ist, aber auch oberflächlich und wankelmütig wirkt. Wir alle tragen diese Typen in uns, in unterschiedlicher Ausprägung, keiner ist ausschließlich das eine oder das andere, sagt Isabel García. Und wir können lernen, uns diese Typen wie Superhelden an unsere Seite zu holen, wenn es erforderlich ist (Schröder 2019, S. 23).

Die Persönlichkeit eines Menschen zu erkennen und zu beeinflussen ist ein Ziel, das eine Person sich selbst setzt, das aber auch von ihrer Umgebung, den Personen und Organisationen in ihrem Umfeld verfolgt wird. Dazu ist eine Theorie über die Formen menschlicher Persönlichkeiten erforderlich. Persönlichkeitstheorien haben in Philosophie und Psychologie eine lange Tradition. Sie wurden vor dem Aufkommen der wissenschaftlichen Psychologie am Ende des 19. Jahrhunderts durch Alltagsbeobachtungen und philosophische Reflexionen gewonnen. Heute ist dafür die empirische Psychologie zuständig. Moderne Persönlichkeitstheorien basieren auf der Analyse empirischer Daten. Ein Verfahren ist zu besonderer Prominenz aufgestiegen. Es ist die Faktorenanalyse, die später noch einmal kurz umrissen wird. Aber auch andere Analyseverfahren liefern Beiträge zur Persönlichkeitsforschung. Die beschreibende Statistik mit ihren Kennwerten, insbesondere der Kor-

relation, ist eine Voraussetzung der Faktorenanalyse. Qualitative Verfahren bieten ebenfalls entsprechende Möglichkeiten, da sie das Personentypische aus verbalen Protokollen (Aussagen über sich selbst und andere) herausarbeiten können.

4.1 Überblick zu den Verfahren der quantitativen Datenanalyse

Datenauswertungen zielen darauf ab, die zentralen Tendenzen der erhobenen Daten herauszuarbeiten. Eine Datenauswertung ist zugleich eine Zusammenfassung der Daten und eine Reduzierung auf das Wesentliche. Idealerweise endet sie mit wissenschaftlich fundierten Erklärungen für die gefundenen Ergebnisse.

Quantitative Daten sind allein mit statistischen Auswertungsverfahren kompatibel. Sie können nicht mittels qualitativer Text- oder Inhaltsanalyse bearbeitet werden. In dieser Lektion werden zunächst Grundzüge der statistischen Datenanalyse behandelt, anschließend geht es dann um die Auswertung qualitativer Daten.

Die Nutzung von Statistikprogrammen in der quantitativen Datenanalyse

Quantitative Datenanalysen werden mit **Statistikprogrammen** durchgeführt. Anwender rufen die entsprechenden Befehle in einem Programm auf. Das Statistikprogramm arbeitet mit der für die Stichprobe erstellten Datentabelle und errechnet statistische Kennwerte. In der wissenschaftlichen Psychologie, aber auch in der Wirtschaft, wird häufig das Statistikprogramm SPSS eingesetzt. Es ist das älteste kommerzielle Statistikprogramm und wurde erstmals 1968 auf den Markt gebracht. Die ersten Versionen liefen auf Großrechnern und arbeiteten bei der Befehlseingabe mit Lochkarten. Die aktuelle Version hat die Versionsnummer 25 und ist seit 2017 für verschiedene Geräte und Betriebssysteme erhältlich. SPSS ist ein modular aufgebautes Softwarepaket. Verschiedene Befehle und Prozeduren sind in separaten Modulen zusammengefasst. Die Befehle werden über Menüs aufgerufen, sodass Syntax- und Programmierkenntnisse für Anwender nicht unbedingt erforderlich sind. SPSS ist relativ leicht zu erlernen. Dazu gibt es Lehrbücher (z. B. Brosius 2018) und viele Handreichungen im Internet. SPSS ist allerdings kostenpflichtig. Jeder Anwender erwirbt eine zeitlich begrenzte Lizenz. Als kostenfreie Alternativen bieten sich Statistikprogramme an, die auf die SPSS-Syntax zurückgreifen, aber weniger umfangreich sind, wie z. B. das Programm PSPP.

Statistikprogramme
Zur quantitativen Auswertung größerer Datenmengen werden Statistikprogramme genutzt. Die Software enthält die in Lehrbüchern der Statistik dargestellten Formeln und Tests als Befehle, um die Daten auszuwerten.

Im wissenschaftlichen Bereich erfreut sich das Statistikprogramm R immer größerer Beliebtheit. R ist kein Programm mit Menü und Befehlen, sondern eine Programmiersprache für statistische Zwecke, die seit 1992 existiert. Anwender laden verschiedene Bibliotheken von Befehlen aus dem Internet herunter und können dann auf dem eigenen Gerät statistische Analysen durchführen. Dazu müssen die Befehle mit korrekter Syntax in einem Programmfenster (Interpreter) eingegeben werden. Die Syntax ist leicht zu verstehen, und verschiedene Befehle können kombiniert und verschachtelt werden. Dies macht R zu einer sehr flexiblen Statistiksoftware. Für ungeübte Nutzer ist allerdings aufgrund der fehlenden

Menüsteuerung eine gewisse Lern- bzw. Einarbeitungszeit einzuplanen. Mittlerweile sind die ersten deutschen Lehrbücher zur Anwendung von R in statistischen Auswertungen auf dem Markt (Wollschläger 2017; Sauer 2019).

Einteilung der statistischen Auswertungsverfahren

Bei statistischen Verfahren der Auswertung, die von Statistikprogrammen durchgeführt werden, sind vereinfacht drei Gruppen zu unterscheiden:

- Verfahren der deskriptiven bzw. explorativen Datenanalyse,
- Verfahren der Inferenzstatistik,
- Verfahren der multivariaten Datenanalyse.

In der deskriptiven bzw. explorativen Statistik werden statistische Kennwerte für die Variablen der untersuchten Stichprobe berechnet. Analysen mit Kennwerten können eindimensional und zweidimensional durchgeführt werden. Bei einer eindimensionalen Analyse müssen sich die Anwender überlegen, wie sich die Verteilung einer Variablen (z. B. Alter, Semesterzahl oder Geschlecht) am besten zu einem oder mehreren Kennwerten verdichten lässt. Eine zweidimensionale Analyse bildet den numerischen Zusammenhang zwischen zwei Variablen (z. B. Alter und Semesterzahl) ab.

Die Inferenzstatistik unternimmt wahrscheinlichkeitstheoretische Schlüsse von den in einer Studie ermittelten Kennwerten auf eine definierte **Grundgesamtheit**. Sie verwendet die Kennwerte als Ausgangspunkt. Dann wird abgeschätzt, wie wahrscheinlich der in einer Stichprobe gemessene Kennwert ist, wenn man unendlich viele Stichproben aus der Grundgesamtheit ziehen würde. Die Verteilung analoger Kennwerte aus unendlich vielen Stichproben lässt sich statistisch konstruieren, sodass die Anwender sehen, wo in dieser Verteilung der von ihnen gemessene Kennwert liegt und wie wahrscheinlich dieser ist, wenn alle Werte eine Gesamtwahrscheinlichkeit von 1 bzw. 100 % haben. Die Wahrscheinlichkeiten werden von den Statistikprogrammen ausgegeben.

Grundgesamtheit
Die Grundgesamtheit, auch Population genannt, ist die Summe aller Fälle, aus denen eine Stichprobe gezogen wird. Eine Grundgesamtheit kann empirisch und endlich sein (z. B. alle Wahlberechtigten in Deutschland) oder theoretisch und unendlich.

Multivariate Analyseverfahren ermitteln die quantitativen Zusammenhänge zwischen einer größeren Menge von Variablen. Sie sind zu unterscheiden nach strukturentdeckenden und strukturprüfenden Verfahren (vgl. Backhaus et al. 2018, S. 13f.). Entdeckende Verfahren suchen Gemeinsamkeiten in den quantitativen Daten. Sie ordnen diese nach Ähnlichkeiten und verdichten sie zu abstrakten Gruppen. Strukturprüfende Verfahren gehen von einem bestimmten Modell aus und bestimmen den quantitativen Wirkungsgrad von Variablen in Bezug auf andere des verwendeten Modells.

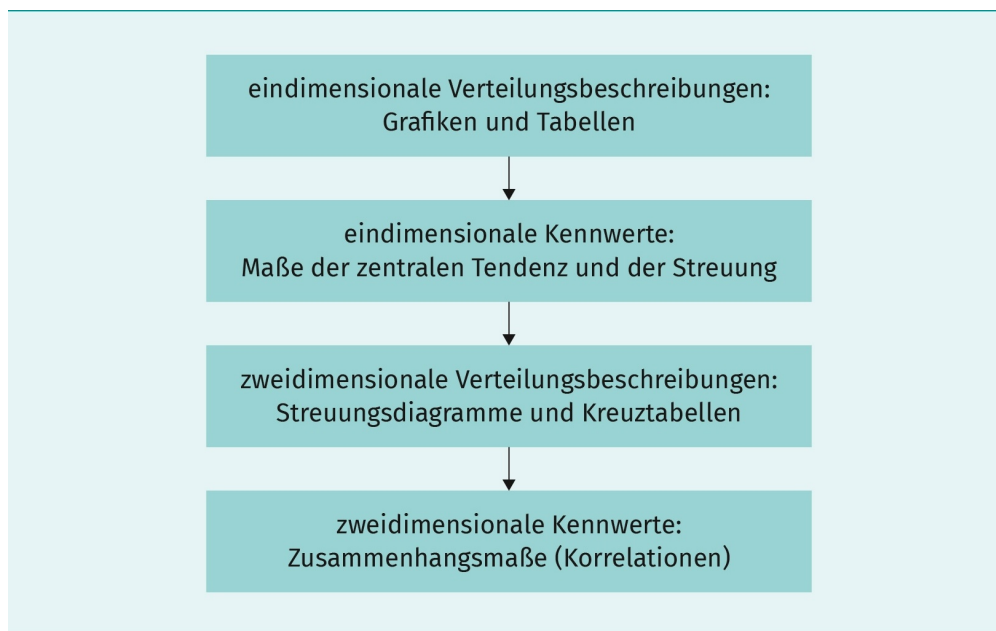
Im Rahmen dieses Studienbriefs werden die verschiedenen Auswertungsmethoden nur angerissen. Zu jedem Kapitel ist für den Nachvollzug der Auswertungsschritte die Lektüre von ausführlicheren Darstellungen in Lehrbüchern erforderlich. Ein gut verständliches Lehrbuch zur deskriptiven Statistik und zu den einfachen Tests in der Inferenzstatistik ist das Werk von Thomas Schäfer (2016).

4.2 Deskriptive Statistik und explorative Datenanalyse

Die Verfahren der deskriptiven Statistik liefern Beschreibungen (Deskriptionen) des erhobenen Datensatzes. In der Auswertung findet die deskriptive Analyse vor der Inferenzstatistik statt. Bis etwa 1990 wurde in statistischen Lehrbüchern allein zwischen diesen beiden Formen der Statistik unterschieden. In den letzten Jahren hat sich als neue Bezeichnung der Begriff der explorativen Statistik etabliert. Diesen prägte der US-amerikanische Statistiker John Wilder Tukey (1915–2000). Er plädierte in seinen Schriften für einen offenen und entdeckenden (explorativen) Umgang mit größeren Datenmengen. Die Forschenden sollten nicht exakt definierte statistische Hypothesen testen, sondern sich in den Analysen von den erhobenen Daten leiten lassen und in diesen nach unbekanntem Zusammenhängen suchen.

In diesem Studienbrief werden die Begriffe der deskriptiven und der explorativen Statistik synonym gebraucht. Beide beziehen sich auf die Auswertungsphase vor der Inferenzstatistik. Der erste Schritt der deskriptiven bzw. explorativen Statistik besteht darin, sich einen Überblick über die Verteilungen der Werte bei den gemessenen Variablen zu verschaffen. Dazu werden in den Statistikprogrammen Grafiken erstellt und/oder Tabellen von Häufigkeiten der Werte einer Variablen. Der nächste Auswertungsschritt ist die Berechnung statistischer Kennwerte, zunächst eindimensionale, danach zweidimensionale.

Abbildung 19: Auswertungsschritte in der deskriptiven bzw. explorativen Statistik

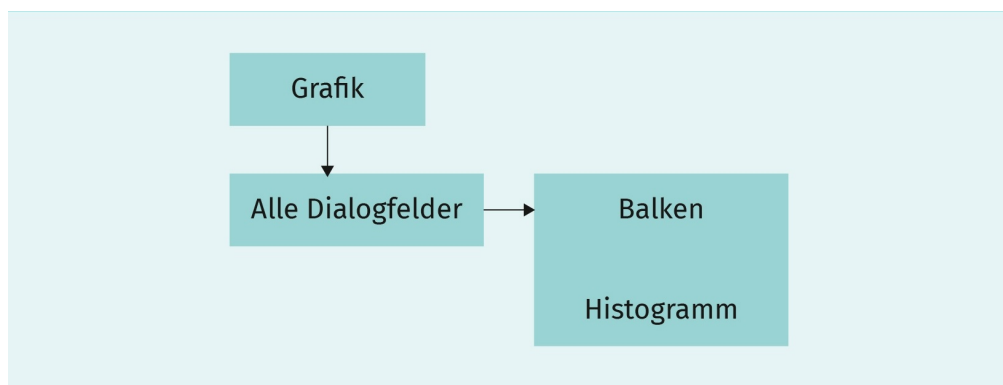


Quelle: erstellt im Auftrag der IU, 2022.

Grafiken zur Darstellung von Häufigkeiten

Balkendiagramme und Histogramme sind die am häufigsten verwendeten Grafiken zur Darstellung der Häufigkeitsverteilung der Werte einer Variablen. In beiden Diagrammformen werden die Werte der Variablen auf der x-Achse abgetragen, die Häufigkeiten auf der y-Achse. Bei Histogrammen sind die Balken zusammengezogen, bei Balkendiagrammen gibt es zwischen jedem Balken einen Zwischenraum. Die grafische Häufigkeitsverteilung vermittelt Anwendern einen summarischen Eindruck zur Tendenz einer Verteilung. Zu beiden Diagrammen gibt es einen leicht aufzurufenden Befehl in SPSS.

Abbildung 20: SPSS-Befehle für Balkendiagramme und Histogramme

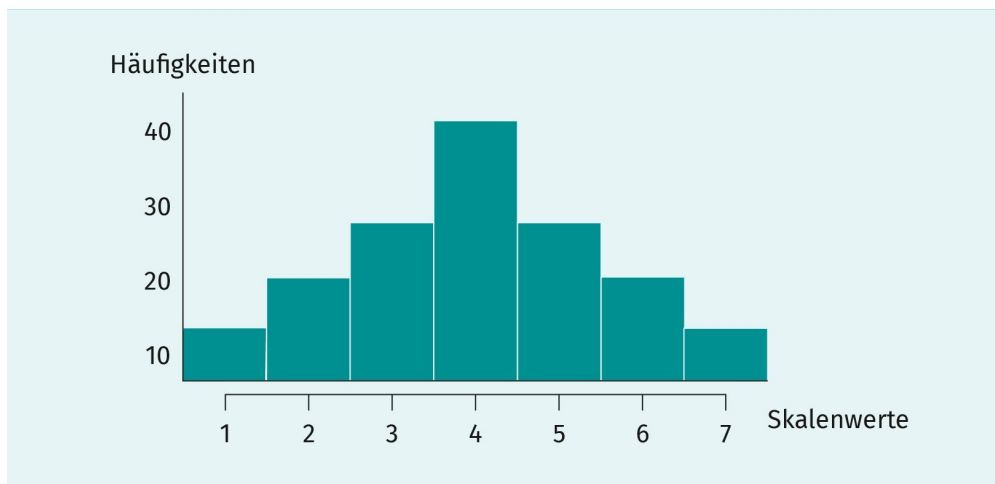


Quelle: erstellt im Auftrag der IU, 2022.

Kontinuierliche vs. diskrete Variable
Eine diskrete Variable kann nur endlich viele oder abzählbar unendlich viele Werte eines Intervalls der reellen Zahlen annehmen (z. B. Anzahl der Schüler in einer Klasse). Das Gegenteil sind stetige bzw. kontinuierliche Variablen. Bei ihnen kann die Differenz zwischen zwei benachbarten Werten unendlich klein werden (z. B. Gewicht).

Histogramme sind einzusetzen bei **kontinuierlichen** und metrischen Variablen (z. B. Intelligenz oder Gewicht), Balkendiagramme bei **diskreten** metrischen Variablen (z. B. Anzahl der Kinder) sowie bei Variablen auf dem Skalenniveau einer Nominal- (z. B. Geschlecht) und Ordinalskala (z. B. Bildungsabschluss). Bei LIKERT-Skalen können sowohl Histogramme als auch Balkendiagramme zur Darstellung der Häufigkeitsverteilung verwendet werden.

Abbildung 21: Histogramm einer siebenstufigen LIKERT-Skala



Quelle: erstellt im Auftrag der IU, 2022.

Die Abbildung zeigt eine fiktive symmetrische Häufigkeitsverteilung von sieben Werten.

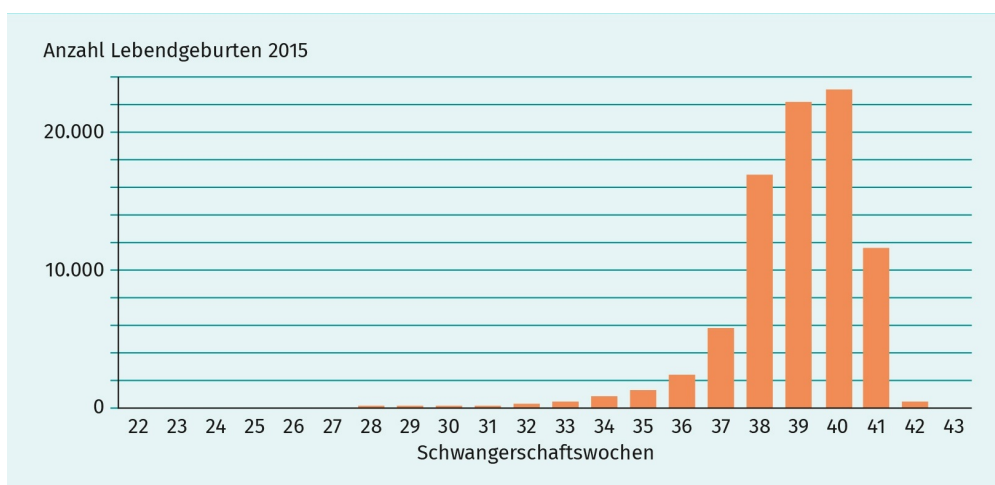
Normalverteilungen haben in der Statistik wie in der Psychologie eine große Bedeutung. Der Mathematiker Carl Friedrich Gauß (1777–1855) hat Normalverteilungen als Erster mathematisch untersucht. Auf ihn geht die Bezeichnung „Gaußsche Glockenkurve“ zurück. Wie eine Glocke spannt sich die symmetrische Verteilung links und rechts um ihren häufigsten Wert auf, der zugleich ihr mittlerer Wert ist. In der Psychologie gibt es Verteilungen von Variablen, die per Definition normalverteilt sind und auch in der Stichprobe so ausfallen sollten. Dies sind z. B. Werte in Intelligenztests. Diese sind so konstruiert, dass der Gipfel der Verteilung, der häufigste Wert, bei einem Intelligenzquotienten (IQ) von 100 liegt. 50 % der Bevölkerung haben einen höheren und 50 % einen niedrigeren IQ.

Es gibt unendlich viele Normalverteilungen. Diese lassen sich alle in die **Standardnormalverteilung** transformieren. Deren Mittelwert ist 0, die Abweichungen davon werden in standardisierten Streuungen von 1 ausgedrückt. Ein IQ-Wert von 100 entspricht dem Wert 0 in der Standardnormalverteilung. In der IQ-Normalverteilung ist die Streuung 15. Einem IQ-Wert von 115 entspricht deshalb der Wert 1 in der Standardnormalverteilung. In Normalverteilungen lässt sich abschätzen, wie viel Prozent der Messwerte in welchen Bereichen liegen. Beispielsweise liegen 68 % zwischen je einer Standardabweichung unterhalb und oberhalb der Mitte.

Standardnormalverteilung

Die Standardnormalverteilung ist eine Normalverteilung mit dem Mittelwert 0 und der Standardabweichung 1. Alle Werte aus anderen Normalverteilungen lassen sich leicht in die Werte der Standardnormalverteilung umrechnen.

Abbildung 23: Balkendiagramm: Häufigkeitsverteilung des Gestationsalters (Schwangerschaftswochen bei Entbindung)



Quelle: Schweizerisches Bundesamt für Statistik 2016.

Grafiken veranschaulichen Häufigkeitsverteilungen. Alternativ können in Statistikprogrammen wie SPSS auch Tabellen aufgerufen werden. Grafische wie tabellarische Häufigkeitsanalysen geben Hinweise, ob Verteilungen normal und erwartet oder unnormal und unerwartet ausfallen. Sie sind ein Teil der Datenaufbereitung, um fehlende und falsche Werte zu identifizieren. Außerdem geben sie Hinweise auf einzelne extrem hohe oder niedrige Werte, die unter Umständen die statistischen Analysen verzerren. Häufigkeitsanalysen mittels Grafiken und Tabellen bereiten die folgende Berechnung von statistischen Kennzahlen vor (vgl. ausführlich Schäfer 2016, S. 83f.).

Kennziffern bei eindimensionalen Häufigkeitsverteilungen

Die Kennwerte der deskriptiven Statistik (Maßzahlen) dienen der numerischen Charakterisierung der Tendenzen der Verteilung eines Merkmals in der Stichprobe.

Tabelle 5: Übersicht zu den Kennwerten der deskriptiven bzw. explorativen Statistik

Statistische Kennwerte: Wie lassen sich die Eigenschaften von Verteilungen in Maßzahlen abbilden?	
Maße der zentralen Tendenz: Wo liegt das Zentrum der Verteilung eines Merkmals?	Streuungsmaße: Wie stark unterscheiden sich die gemessenen Werte innerhalb der Verteilung eines Merkmals?
<p>Die zu berechnenden Kennwerte sind:</p> <ul style="list-style-type: none"> • Modus • Median • arithmetisches Mittel • geometrisches Mittel • Minimum bzw. Maximum • Quantil 	<p>Die zu berechnenden Maße sind:</p> <ul style="list-style-type: none"> • Spannweite • Interquartilsabstand • Varianz • Standardabweichung

Quelle: Wirtz/Nachtigall 2012, S. 39.

Maße der zentralen Tendenz

Zentrale Tendenz

Die zentrale Tendenz charakterisiert das Typische einer Häufigkeitsverteilung und somit auch die Mitte bzw. das Zentrum der Verteilung.

Die Maße der **zentralen Tendenz** repräsentieren alle Messwerte einer Verteilung zusammenfassend. Der Modus ist exakt der häufigste Wert, also derjenige, der die größte Anzahl von Fällen in der Stichprobe auf sich vereinigt. Der Median ist derjenige Wert, der die Verteilung aller Werte in zwei gleich große Hälften teilt. 50 % der Werte liegen oberhalb des Medians, 50 % darunter. Das arithmetische Mittel ist der sich ergebende Wert, wenn alle gemessenen Werte summiert und durch die Anzahl der Fälle geteilt werden. Das arithmetische Mittel wird umgangssprachlich als Durchschnitt bezeichnet, in der statistischen Fachsprache als Mittelwert. Das geometrische Mittel ist bei n Fällen die n -te Wurzel aus dem Produkt aller Werte. Es gibt die durchschnittliche Zuwachsrate von Wert zu Wert an, wenn die gemessenen Werte ansteigend geordnet sind. Das Minimum bzw. Maximum ist exakt der kleinste bzw. größte gemessene Wert. Das p -Quantil gibt an, welcher Wert die Verteilung in zwei komplementäre Teile aufspaltet. Beispielsweise ist das 25-Prozent-Quantil der Wert, über dem drei Viertel der Fälle liegen und ein Viertel darunter. Der Median ist zugleich das 50-Prozent-Quantil.

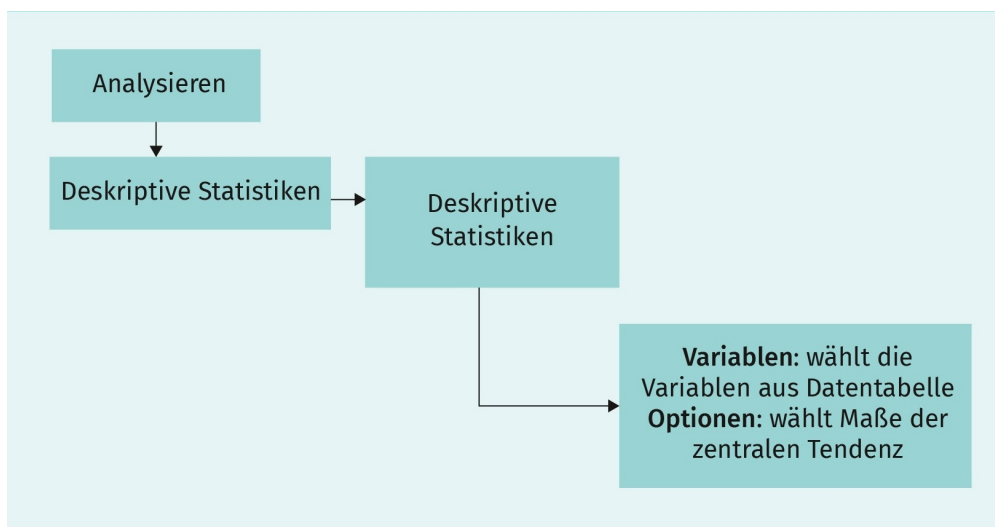
Zur Veranschaulichung werden die Maße der zentralen Tendenz für die Variable Alter des Beispiels Studienmotivation berechnet.

Tabelle 6

Berechnung der Maße der zentralen Tendenz:
Daten für 10 Fälle ($n = 10$): 20; 18; 24; 40; 24; 22; 21; 23; 20; 28.
Modus: Je zweimal kommen 20 und 24 vor. $x_{\text{mod}} = \{20, 24\}$
Median: Der Median ist 22,5.
Mittelwert: $20 + 18 + 24 + 40 + 24 + 22 + 21 + 23 + 20 + 28 = 240$ $\bar{x} = 240 : 10 = 24$
geometrisches Mittel: 10-te Wurzel aus $(20 \cdot 18 \cdot 24 \cdot 40 \cdot 24 \cdot 22 \cdot 21 \cdot 23 \cdot 20 \cdot 28)$ $x_{\text{geo}} = 1,4$
Maximum: $x_{\text{max}} = 40$
Minimum: $x_{\text{min}} = 18$
25-Prozent-Quantil: $x_{0,25} = 20$

Quelle: erstellt im Auftrag der IU, 2022.

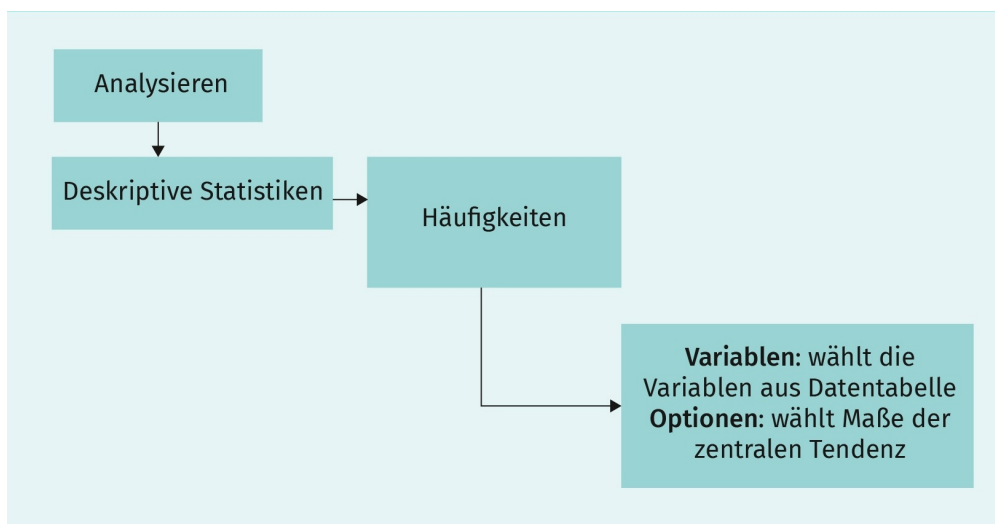
Abbildung 24: Erster SPSS-Befehl für die Maße der zentralen Tendenz und für Streuungsmaße



Quelle: erstellt im Auftrag der IU, 2022.

Welche Kennwerte bei Variablen sinnvoll zu berechnen sind, hängt im Wesentlichen von zwei Faktoren ab: dem Skalenniveau der betreffenden Variablen und der jeweiligen Häufigkeitsverteilung. Beim Skalenniveau gilt die Faustregel: Je niedriger dieses ist, desto eingeschränkter ist die Auswahl der statistischen Kennwerte. Beispielsweise können bei einer Nominalskala (z. B. Geschlecht) Median, arithmetisches und geometrisches Mittel nicht berechnet werden, da die Skala keine Größer-Kleiner-Relation enthält und die Abstände zwischen den Skalenwerten nicht definiert sind.

Abbildung 25: Zweiter SPSS-Befehl für die Maße der zentralen Tendenz und für Streuungsmaße



Quelle: erstellt im Auftrag der IU, 2022.

Tabelle 7: Skalenniveaus und jeweils anwendbare Maße der zentralen Tendenz

Nominalskala	Modus, Maximum, Minimum
Ordinalskala	Modus, Median, Maximum, Minimum, Quantile
Intervallskala	Modus, Median, arithmetisches Mittel, Maximum, Minimum, Quantile
Verhältnisskala	Modus, Median, arithmetisches Mittel, geometrisches Mittel, Maximum, Minimum, Quantile

Quelle: erstellt im Auftrag der IU, 2022.

Bei Verteilungen gilt als Faustregel, dass bei symmetrischen Verteilungen Modus, Median und der Mittelwert identisch sind. Beispielsweise ist in der symmetrischen Verteilung der siebenstufigen LIKERT-Skala der Wert 4 zugleich Modus, Median und Mittelwert. Bei annähernd symmetrischen Verteilungen, deren Variablen auf dem Niveau einer Intervall- oder Verhältnisskala vorliegen, wird im Allgemeinen in den Forschungsberichten nur das arithmetische Mittel mitgeteilt.

Bei schiefen Verteilungen, die entweder einen linken oder rechten Gipfel haben, fallen Modus, Median und Mittelwert auseinander. Bei Verteilungen mit einem rechten Gipfel bei den hohen Werten (siehe das Beispiel zur Schwangerschaftswoche) ist das arithmetische Mittel < Median < Modus. Bei linksgipfligen Verteilungen verhält es sich umgekehrt. Linksgipflige Verteilungen werden in der Fachsprache als „rechtsschief“ bezeichnet und rechtsgipflige als „linksschief“.

Streuungsmaße

Neben den Maßen der zentralen Tendenz werden zur Charakterisierung einer Häufigkeitsverteilung immer auch **Streuungsmaße** angegeben. Sie vermitteln einen Eindruck von der Unterschiedlichkeit bzw. Homogenität der gemessenen Werte. Haben alle untersuchten Fälle der Stichprobe ähnliche Messwerte, dann ist die Streuung klein. Sind die gemessenen Werte sehr unterschiedlich, so ist die Streuung groß. Grafisch zeigt sich dieser Gegensatz als enge oder breite Verteilung um den mittleren Wert.

Die gängigsten Streuungsmaße sind Spannweite, Interquartilsabstand, Varianz und Standardabweichung. Als Spannweite bezeichnet man die Differenz zwischen dem jeweils größten und kleinsten Wert einer Verteilung, als Interquartilsabstand jene zwischen dem unteren 25-Prozent-Quantil und dem oberen 75-Prozent-Quantil. Die Varianz ist die durchschnittliche quadratische Abweichung aller gemessenen Werte vom Mittelwert der Verteilung. Die Standardabweichung ist die Wurzel aus der Varianz. Zur Veranschaulichung werden die vier Streuungsmaße für die Variable Alter des Beispiels Studienmotivation berechnet.

Tabelle 8

Berechnung der Streuungsmaße für die Variable Alter
Daten für n = 10 Fälle: 20; 18; 24; 40; 24; 22; 21; 23; 20; 28. Mittelwert = 24.

Streuungsmaße
Liefen einen Wert für die Breite bzw. Homogenität der Verteilung einer Variablen

Spannweite: $40 - 18 = 22$

Interquartilsabstand (IQR): $24 - 20 = 4$

Varianz: $s^2 = ((20 - 24)^2 + (18 - 24)^2 + (24 - 20)^2 + (40 - 24)^2 + (24 - 24)^2 + (22 - 24)^2 + (21 - 24)^2 + (23 - 24)^2 + (20 - 24)^2 + (28 - 24)^2) = 354 : (n - 1) = 354 : 9 = 39,33$

Standardabweichung: $s = \sqrt{39,33} = 6,27$

Quelle: erstellt im Auftrag der IU, 2022.

Die Varianz errechnet sich als Summe aller quadrierten Abweichungen der gemessenen Werte einer Variablen von ihrem arithmetischen Mittel, dividiert durch die Anzahl der Werte bzw. Fälle der Stichprobe vermindert um einen Fall ($n-1$). Die entsprechende Summenformel lautet:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (4.1)$$

Die Standardabweichung (s) ist die Wurzel aus der Varianz (s^2). Nachdem zur Berechnung der Varianz die Abstände der Werte vom Mittelwert quadriert wurden, muss aus deren Summe die Wurzel gezogen werden, damit der mittlere Abstand der Werte vom Mittelwert ausgedrückt wird. Umgangssprachlich formuliert: Im Durchschnitt haben die zehn gemessenen Werte der Variablen Alter einen Abstand von 6,27 Jahren von ihrem Mittelwert (24 Jahre). Für diesen relativ hohen Wert gibt es zwei Gründe: Zum einen ist der vierte Messwert mit 40 Jahren ein Ausreißer, der die Streuung erhöht. Zum anderen ist die Stichprobe klein. Da bei der Varianz durch $n - 1$ (9 Fälle) geteilt wird, fällt die Streuung höher aus, als wenn durch die gesamte Fallzahl ($n = 10$) geteilt würde. Bei der Berechnung der Varianz wird bei der Anwendung in Statistikprogrammen in der Regel durch $n - 1$ dividiert. Der Grund dafür ist, dass bei inferenzstatistischen Analysen die mit $n - 1$ berechnete Varianz ein besserer Schätzwert für die Varianz in der Grundgesamtheit ist. Nähere Erläuterungen dazu sind Statistiklehrbüchern zu entnehmen (z. B. Nachtigall/Wirtz 2013, S. 177f.).

Welches Streuungsmaß berechnet wird und später in einem Forschungsbericht auftaucht, hängt, wie bei den Maßen der zentralen Tendenz, vor allem vom Skalenniveau und der Häufigkeitsverteilung der Variablen ab. Varianz und Standardabweichung können nur bei metrischen Niveaus (Intervall- und Verhältnisskala) berechnet werden.

Tabelle 9: Übersicht zu Skalenniveaus und den jeweils möglichen Streuungsmaßen

Nominalskala	keine Streuungsmaße
Ordinalskala	keine Streuungsmaße
Intervallskala	Spannweite, Interquartilsabstand, Varianz, Standardabweichung
Verhältnisskala	Spannweite, Interquartilsabstand, Varianz, Standardabweichung

Quelle: erstellt im Auftrag der IU, 2022.

Bei Verteilungen mit Extremwerten (Ausreißern) haben besonders in kleinen Stichproben ($n < 100$) die Extremwerte einen großen Einfluss auf Varianz und Standardabweichung. Deshalb sollten bei metrischen Variablen (z. B. Alter) sowohl Varianz bzw. Standardabweichung als auch der Interquartilsabstand in Forschungsberichten angegeben werden. Gleiches gilt für schiefe Verteilungen mit einem rechten oder linken Gipfel. Die SPSS-Befehle für die Streuungsmaße sind identisch mit denen für die Maße der zentralen Tendenz.

Zweidimensionale Analysen

Zweidimensionale Kennwerte beschreiben den numerischen Zusammenhang zwischen zwei Variablen. Welcher statistische Kennwert in zweidimensionalen Analysen berechnet wird, hängt von den Skalenniveaus der beiden Variablen ab. Errechnet wird zumeist ein Korrelationskoeffizient. Für die Skalenniveaus gibt es verschiedene Koeffizienten.

Tabelle 10: Auswahl gängiger Zusammenhangsmaße in Abhängigkeit vom Skalenniveau der Merkmale

	nominal	ordinal	intervall
nominal	Phi, Cramer's V	biseriale Rangkorrelation	punktbiseriale Korrelation
ordinal		Spearman	(Spearman)
intervall			Pearson

Quelle: erstellt im Auftrag der IU, 2022.

Zusammenhänge zwischen zwei metrischen Variablen (z. B. Alter und Semesteranzahl im Beispiel Studienmotivation) werden mittels Produkt-Moment-Korrelation errechnet. Dieser Korrelationskoeffizient ist ein Maß für den linearen Zusammenhang zweier Variablen, auch **Covarianz** genannt. Untersucht wird, ob die Streuung beider Variablen gleichsinnig ist. Wenn hohe Werte der ersten Variablen mit hohen Werten der zweiten einhergehen (in unserem Beispiel: hohes Alter mit hoher Semesteranzahl), und niedrige Werte der ersten mit niedrigen der zweiten, dann ist die Korrelation hoch. Maximal kann der Wert des Korrelationskoeffizienten den Wert 1 erreichen. Der Grund ist, dass die Covarianz der beiden Variablen in der Formel des Korrelationskoeffizienten durch das Produkt der Standardabweichungen beider Variablen geteilt wird. Das hat den Effekt einer Standardisierung. So werden Covarianzen von Variablen mit unterschiedlichen Wertebereichen vergleichbar. Die Formel für den Korrelationskoeffizienten lautet:

$$r = \frac{s_{xy}}{s_x \cdot s_y} \quad (4.2)$$

Im Nenner stehen die beiden Standardabweichungen (s_x und s_y) der beiden Variablen.

Die Berechnung der Covarianz (s_{xy}) ist relativ aufwendig. Ermittelt wird die Covarianz für jedes Wertepaar pro Fall in der Stichprobe. Über alle Fälle hinweg werden die Covarianzen addiert und durch die Anzahl der Fälle, vermindert um 1, geteilt. Die Summenformel lautet:

Covarianz
Die Covarianz beschreibt den Grad des „Miteinander-Variierens“ (Streuens) der Werte von X und Y. Es ist die Summe der gemittelten Abweichungsprodukte zweier Variablen von ihren Mittelwerten.

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (4.3)$$

Korrelationskoeffizienten von $r > 0$ zeigen einen positiven linearen Zusammenhang an, $r < 0$ einen negativen. Negative Zusammenhänge liegen vor, wenn hohe Werte einer Variablen systematisch mit niedrigen der anderen einhergehen. Korrelationen um 0 zeigen, dass es keinen Zusammenhang zwischen den beiden Variablen gibt, d. h., in einigen Fällen kombinieren sich niedrige und hohe Werte, in anderen ist es umgekehrt. Je größer der absolute Betrag von r , desto größer ist der lineare Zusammenhang. Korrelationen sind ein beschreibender quantitativer Zusammenhang. Sie enthalten keine Aussagen über den Grund des Zusammenhangs und implizieren keine kausalen Zusammenhänge.

Die Schritte der Darstellung und Berechnung des Korrelationskoeffizienten werden im Folgenden auf das Beispiel Studienmotivation angewendet. Für beide Variablen Alter und Semesterzahl sind die Werte für die Fälle 1 bis 10 wie folgt:

Tabelle 11: Liste der zehn Fälle aus dem Beispiel Studienmotivation für Alter und Semesterzahl

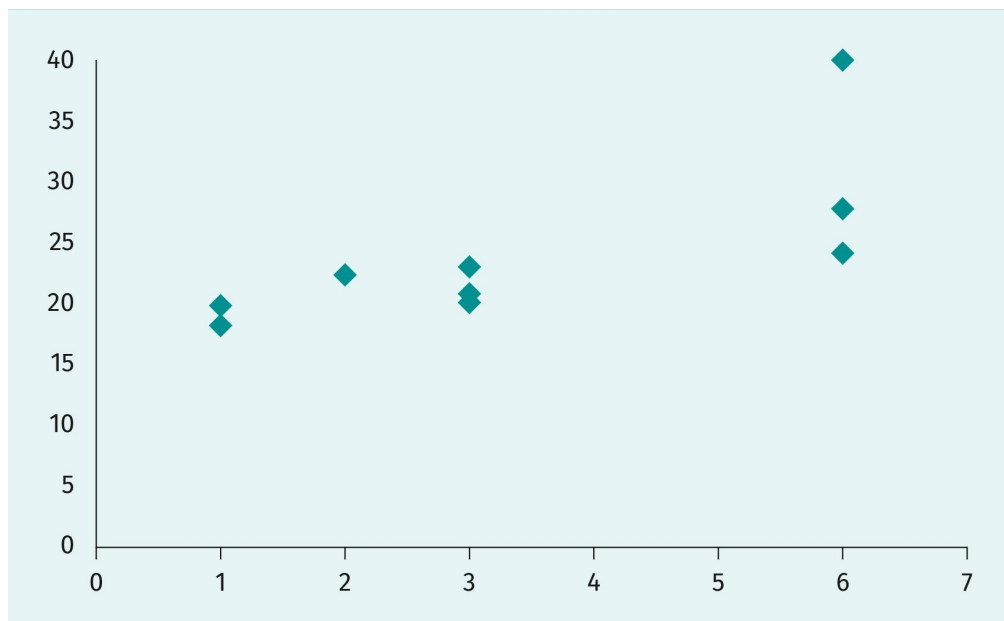
Fallnummer	1	2	3	4	5	6	7	8	9	10
Alter	20	18	24	40	24	22	21	23	20	28
Semesterzahl	3	1	6	6	6	2	3	3	1	6

Quelle: erstellt im Auftrag der IU, 2022.

Grafische Analyse im Streudiagramm

Einen ersten Eindruck, welcher Zusammenhang zwischen zwei metrischen Variablen vorliegt, vermittelt ein Streu- bzw. Punktdiagramm. Bei einem Streudiagramm werden die Werte der ersten Variablen auf der x-Achse abgetragen, die der zweiten auf der y-Achse. Jeder Punkt repräsentiert einen Fall. In SPSS ist das Streudiagramm im gleichen Menü als Befehl aufzurufen wie Balkendiagramme und Histogramme.

Abbildung 26: Streudiagramm der Variablen Semesterzahl (X) und Alter (Y)

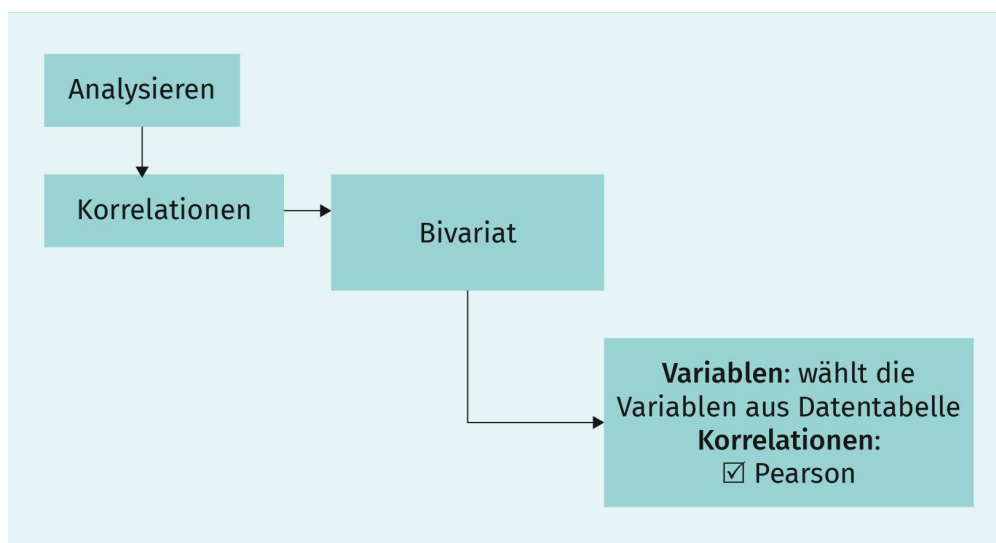


Quelle: erstellt im Auftrag der IU, 2022.

Aus einem Streudiagramm ist abzulesen, ob sich die Werte einer Variablen systematisch mit der anderen verändern. Im obigen Diagramm ist ein leichter Anstieg von Semesteranzahl und Alter von links nach rechts erkennbar. Gedanklich lässt sich eine Gerade konstruieren, die von links leicht ansteigend nach rechts verläuft und neun der zehn Punkte näherungsweise erfasst. Eine Ausnahme bildet der Fall 4 mit dem hohen Alter von 40 Jahren. Er liegt deutlich oberhalb der imaginären Linie.

Nach dem Streudiagramm wird in zweidimensionalen Analysen metrischer Variablen der Korrelationskoeffizient für die beiden Variablen X und Y berechnet. Dieser trägt nach den beiden Erfindern Auguste Bravais (1811–1863) und Karl Pearson (1857–1936) den Namen Pearson-Bravais-Korrelationskoeffizient. In SPSS wird der folgende Befehl aufgerufen:

Abbildung 27: SPSS-Befehl für den Korrelationskoeffizienten von Pearson und Bravais



Quelle: erstellt im Auftrag der IU, 2022.

In der folgenden Tabelle sind zur Veranschaulichung die Rechenschritte zur Berechnung des Korrelationskoeffizienten für die Variablen Semesterzahl (X) und Alter (Y) aufgeführt. Diese Schritte führt das jeweils verwendete Statistikprogramm beim Aufruf des Befehls zum Korrelationskoeffizienten von Pearson und Bravais durch, ohne diese anzuzeigen. Es gibt nur das Ergebnis für den Korrelationskoeffizienten (r) aus.

Tabelle 12: Berechnungsschritte der Korrelation für die Variablen Semesteranzahl und Alter

Fallnummer X_i	x-Werte	y-Werte	$X_i - X_M$	$Y_i - Y_M$	$(X_i - X_M)^2$	$(Y_i - Y_M)^2$	$(X_i - X_M) \cdot (Y_i - Y_M)$
1	3	20	-0,7	-4	0,49	16	2,8
2	1	18	-2,7	-6	7,29	36	16,2
3	6	24	2,3	0	5,29	0	0
4	6	40	2,3	16	5,29	256	36,8
5	6	24	2,3	0	5,29	0	0
6	2	22	-1,7	-2	2,89	4	3,4
7	3	21	-0,7	-3	0,49	9	2,1
8	3	3	-0,7	-1	0,49	1	0,7
9	1	20	-2,7	-4	7,29	16	10,8

Fallnummer Xi	x-Werte	y-Werte	$X_i - X_M$	$Y_i - Y_M$	$(X_i - X_M)^2$	$(Y_i - Y_M)^2$	$(X_i - X_M) \cdot (Y_i - Y_M)$ *
10	6	28	2,3	4	5,29	16	9,2
Summe	37	240	0	00	40,1	354	82
Mittelwerte: X_M und Y_M	$X_m = 3,7$	$Y_m = 24$					
Varianzen S^2_X und S^2_Y	$S^2_X = 40,1 : 9 = 4,45$	$S^2_Y = 354 : 9 = 39,33$					
Streuungen S_X und S_Y	$S_X = 2,11$	$S_Y = 6,27$					
Covarianz S_{xy}	$82 : 9 = 9,11$						
Korrelation r	$\frac{9,11}{2,11 \cdot 6,27}$ $r = 0,69$						

Quelle: erstellt im Auftrag der IU, 2022.

Der Korrelationskoeffizient von Alter und Semesterzahl hat eine Höhe von 0,69. Somit liegt eine mittlere positive lineare Korrelation vor. Für die Stichprobe gilt: Je höher das Lebensalter, desto höher fällt im Mittel die Semesteranzahl aus. Das ist ein zu erwartendes Ergebnis. Ältere Studierende sollten in höheren Semestern sein, jüngere in unteren.

Determinationskoeffizient

Determinationskoeffizienten

Ein Determinationskoeffizient gibt in Prozenten an, wie hoch der Anteil der gemeinsamen Varianz zweier Variablen X und Y ist, wenn die gesamte Varianz beider Variablen 100 ist.

Der quadrierte Korrelationskoeffizient (r^2) ist der **Determinationskoeffizient**. In unserem Beispiel ist $r^2 = (0,69)^2 = 0,48$. Der Determinationskoeffizient wird in der Statistik auch als Bestimmtheitsmaß bezeichnet. Er drückt aus, welcher Anteil der Varianz der einen Variablen durch die andere bestimmt (determiniert) wird. Entsprechend ist die Umkehrung $1 - r^2$ das Unbestimmtheitsmaß. Häufig werden die beiden Maße in Prozent ausgedrückt. In unserem Beispiel wird die Variable Alter (Y) zu 48 % in ihrer Varianz bzw. Streuung durch die Variable Semesterzahl (X) bestimmt. Das Bestimmtheitsmaß gibt Anwendern Aufschluss darüber, ob sich die beiden Variablen für Prognosen mittels linearer Regressionen eignen. Die lineare Regressionsanalyse ist ein Verfahren, mittels dessen die Ausprägungen der zweiten Variablen (Y) durch die erste (X) vorhergesagt werden sollen (vgl. zur einfachen linearen Regressionsanalyse Schäfer 2016, S. 102f.).

Weitere Korrelationskoeffizienten

Die punktbiseriale Korrelation wird eingesetzt, wenn eine der beiden Variablen metrisch ist und mehrere Werte umfasst, die andere Variable dagegen nur in zwei Ausprägungen vorliegt. Solche Variablen heißen dichotome Variablen. Bis vor Kurzem war z. B. die Variable Geschlecht in Deutschland eine dichotome Variable. Seit dem Urteil des Bundesverfassungsgerichts von 2017 ist in Deutschland ein drittes Geschlecht (Bezeichnung: divers) anerkannt und sollte als solches auch in empirischen Studien abgefragt werden. In unserem Beispiel ist die zusätzlich definierte Variable Studierendenstatus (Anfänger oder Fortgeschrittene) dichotom.

Haben beide Variablen das Niveau der Ordinalskala, werden die Koeffizienten Spearman's Rho und Kendall's Tau verwendet. Diese Namen haben sie von ihren beiden Erfindern Charles Spearman (1863–1943) und Maurice Kendall (1907–1983). Beide Berechnungsweisen beruhen darauf, dass die gemessenen Werte in eine aufsteigende Rangreihe gebracht werden. Bei zehn Fällen gibt es zehn Rangplätze von 1. bis 10.; korreliert werden nicht die gemessenen Werte, sondern die Rangplätze. Beide Rangreihen werden systematisch miteinander verglichen. Ist die Rangreihe der Fälle in beiden Variablen identisch, so ergibt sich für den Korrelationskoeffizienten der Wert 1. Steigt die Rangreihe der Werte bei X an und fällt in der gleichen Reihenfolge bei Y ab, ist die Korrelation -1 .

Bei den Koeffizienten von Spearman und Kendall resultieren bei gleichen Daten bzw. Rangreihen leicht unterschiedliche Werte. Spearman's Rho berechnet die Mittelwerte, Varianzen und Kovarianzen von Rangplätzen. Die Berechnung ist analog zu der Produkt-Moment-Korrelation für metrische Daten. Der Unterschied besteht allein darin, dass mit Rangplätzen und nicht mit Datenwerten gearbeitet wird. Dagegen konzentriert sich Kendall's Tau allein auf die Unterschiede zwischen den Rangplätzen in beiden Datenreihen. Details der Berechnung von Rangkorrelationen sind in allen Lehrbüchern zur deskriptiven Statistik dargestellt. Im Internet gibt es zahlreiche Einführungen. Beide Koeffizienten können in allen Statistikprogrammen aufgerufen werden.

Bei Nominalskalen (z. B. Geschlecht) werden andere Korrelationskoeffizienten berechnet (tetrachorischer Korrelationskoeffizient, Phi-Koeffizient, Odds Ratio, Yule's Y). Diese Koeffizienten sind Abwandlungen der Korrelationskoeffizienten für Ordinalskalen bzw. metrische Skalen. In der Forschungspraxis werden diese Korrelationskoeffizienten eher selten verwendet. Stattdessen werden die Häufigkeiten von zwei Variablen „gekreuzt“, was zu einer Schnittmenge der Häufigkeiten beider Variablen führt. Diese Häufigkeitstabelle heißt Kreuztabelle (engl. „crosstabs“) und ist in jedem Statistikprogramm aufzurufen. Werden zwei Variablen miteinander gekreuzt, die jeweils nur zwei Ausprägungen haben (z. B. männlich vs. weiblich für Geschlecht und Ja vs. Nein für Staatsangehörigkeit Deutsch), entsteht eine Vierfeldertafel. Das ist die einfachste Form einer Kreuztabelle. Die folgende Vierfeldertafel ist ein imaginäres Beispiel aus der Medizin zur Messung des Behandlungserfolges von Rückenpatienten. Es ist üblich, die Randsummen der Spalten und Zeilen in der Tabelle mit anzugeben.

Tabelle 13: Fiktive Vierfeldertafel zum Zustand der behandelten Patienten nach Rückenoperation und konventioneller Behandlung ohne Operation (n = 400)

	Rücken-OP	konventionell	Zeilensumme
ohne Beschwerden	125	150	n = 275
mit Beschwerden	75	50	n = 125
Spaltensumme	n = 200	n = 200	n = 400

Quelle: erstellt im Auftrag der IU, 2022.

Die Kreuztabelle vermittelt einen ersten Eindruck über den Zusammenhang zwischen den beiden Variablen. Sie zeigt, ob eine Variable von der anderen statistisch abhängig ist oder nicht. Offensichtlich liegt in diesem Beispiel eine Abhängigkeit vor, denn die konventionelle Therapie führt zu einem statistisch besseren Ergebnis als die Operation. 150 Personen sind bei der konventionellen Therapie ohne Beschwerden, bei Operationen nur 125. Im nächsten Schritt wird mit dem Chi-Quadrat-Test, einem Verfahren der Inferenzstatistik, geprüft, ob dieses Ergebnis statistisch signifikant ist.

4.3 Inferenzstatistik

Die Inferenzstatistik wird auch als prüfende oder schließende Statistik bezeichnet. Sie prüft, wie wahrscheinlich die gemessenen Kennwerte der Stichprobe (Mittelwerte, Korrelationen) unter Zugrundelegung einiger wahrscheinlichkeitstheoretischer Annahmen zu Ereignissen (z. B. Ergebnis eines Münzwurfs) und Verteilungen (z. B. Verteilung von Kopf und Zahl bei 10.000 Münzwürfen) sind. Die Inferenzstatistik schließt dabei von den empirischen Ergebnissen der Stichprobe auf die Grundgesamtheit, also die Gesamtheit aller Fälle, aus denen die Stichprobe gezogen wurde.

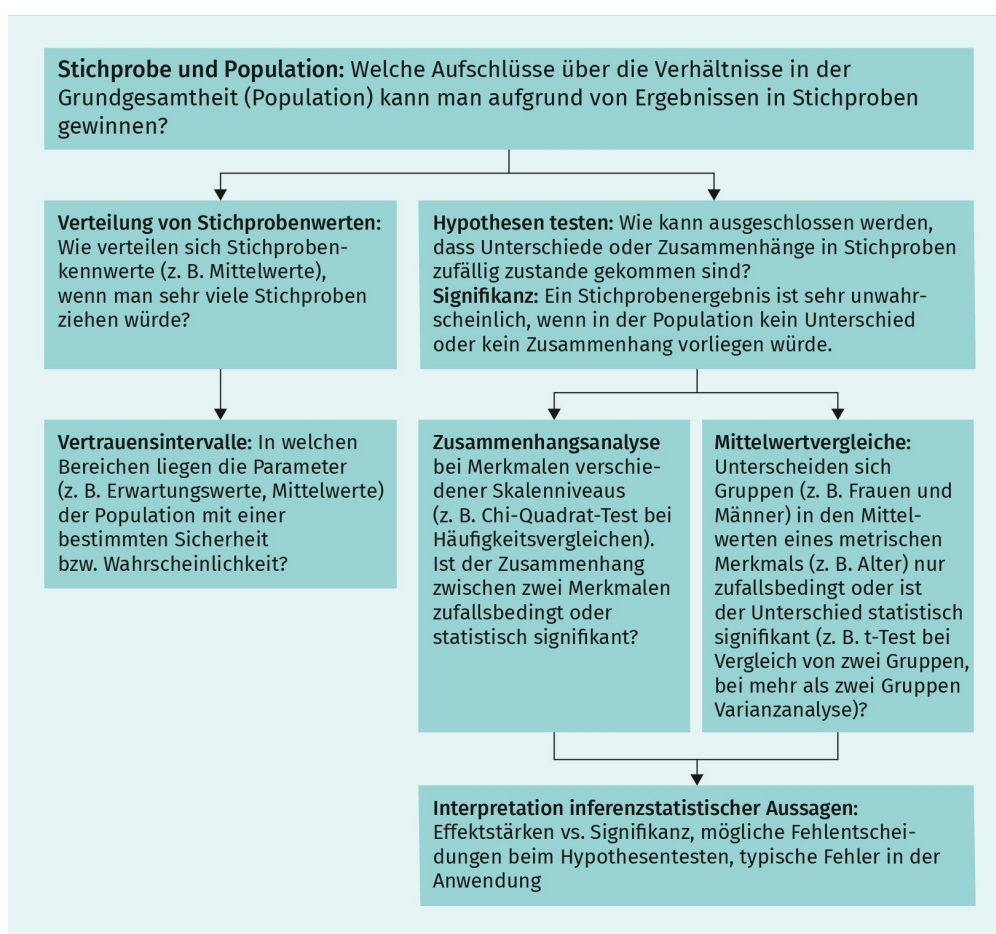
In Ausnahmefällen kann die Verteilung der in der Stichprobe gemessenen Kennwerte für die Grundgesamtheit bekannt sein oder anhand anderer Daten ermittelt werden. Für die Variable Alter in unserem Beispiel Studienmotivation könnte als Grundgesamtheit die Gesamtheit aller Altersverteilungen von Studierenden an deutschen Hochschulen ermittelt werden. Dies wäre allerdings sehr aufwendig und würde den Rahmen einer konventionellen Studie sprengen. Deshalb geht man in der Inferenzstatistik von theoretisch konstruierten Verteilungen aus und ermittelt die Lage des empirisch in der untersuchten Stichprobe gemessenen Kennwertes in der theoretischen Verteilung.

Neulinge können sich diese theoretische Verteilung vorstellen als die Verteilung eines Kennwertes, der aus unendlich vielen Stichproben gezogen wurde unter Zugrundelegung einer bestimmten Hypothese über die numerische Ausprägung des Kennwertes (z. B. dass der Mittelwert des Alters = 24 sein soll). Wenn der empirisch in einer Studie gemessene Kennwert in dieser theoretisch konstruierten Verteilung sehr weit am Rand liegt, wird die zuvor aufgestellte Hypothese verworfen. Das in der Studie gemessene Ergebnis gilt als unwahrscheinlich bzw. als statistisch auffällig (signifikant). Diese Richtung der Inferenzstatistik wird als Signifikanztest bezeichnet.

Die zweite Richtung der Inferenzstatistik will dagegen von den in der Stichprobe gemessenen Kennwerten auf die Ausprägung des wahren, aber unbekanntes Mittelwertes in der Grundgesamtheit schließen. Ausgehend vom Mittelwert einer Variablen in einer Stichprobe (z. B. des Mittelwertes 24 beim Alter) wird abgeschätzt, in welchem Bereich (Intervall) der Mittelwert aller Mittelwerte von unendlich vielen Verteilungen liegt. Dieses Intervall kann mit einer bestimmten Wahrscheinlichkeit als ein Wertebereich (z. B. zwischen 20 und 28 bei der Variable Alter mit der Wahrscheinlichkeit von x Prozent) angegeben werden. Die zweite Richtung der Inferenzstatistik ist die Schätzung von **Vertrauensintervallen** (Konfidenzintervallen).

Vertrauensintervall
 Der Vertrauensintervall ist jener Bereich eines Merkmals, in dem sich 95 % bzw. 99 % aller Populationsparameter befinden, die den empirisch gemessenen Kennwert erzeugt haben können. So gehört z. B. der in der Stichprobe gemessene Mittelwert mit 95 %-iger Wahrscheinlichkeit zu einer Population, deren Mittelwert sich im 95 %-igen Vertrauensintervall befindet.

Abbildung 28: Übersicht: Erkenntnisschritte in der Inferenzstatistik



Quelle: Nachtigall/Wirtz 2013, S. 15.

Diese Lektion konzentriert sich auf den rechten Bereich der Abbildung. Die Schätzung von Vertrauensintervallen wird in diesem Studienbrief nicht erläutert, findet sich aber in nahezu jedem Lehrbuch zur Inferenzstatistik (vgl. z. B. Wirtz/Nachtigall 2012, S. 115).

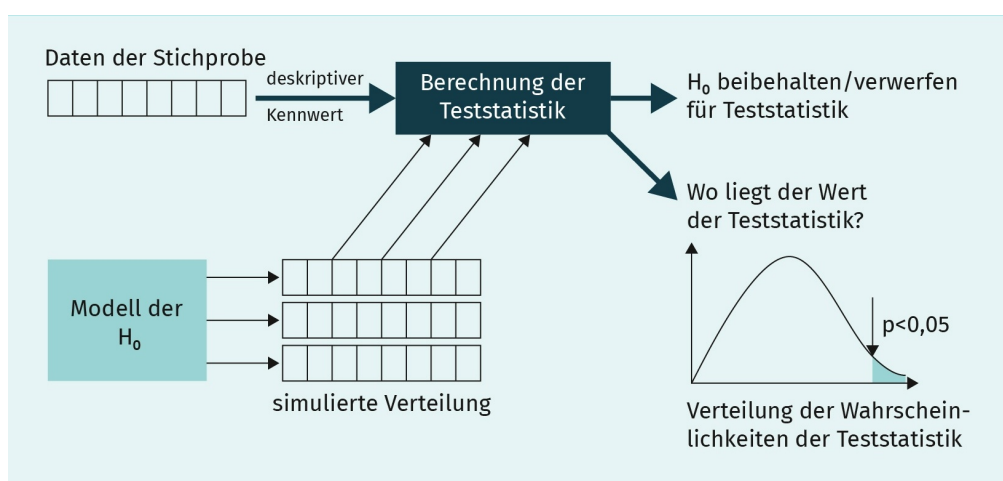
In der Praxis der empirischen Datenanalyse kommt das Hypothesentesten häufiger vor als das Schätzen von Vertrauensintervallen. Das Testen von statistischen Hypothesen ist immer der erste Schritt nach der Berechnung der deskriptiven Kennwerte. Getestet wird, wie wahrscheinlich die in der Stichprobe gemessenen Kennwerte (z. B. Mittelwerte bzw. deren Differenz zwischen zwei Gruppen) sind.

Dass ein gemessener Unterschied zwischen Gruppen in der Häufigkeit von Werten oder in den Mittelwerten nur Zufall ist und nicht an den Besonderheiten der Stichprobe liegt, das behauptet die sogenannte Nullhypothese (H_0) in der Inferenzstatistik. Diese wird vor der Durchführung des Signifikanztests formuliert. Sie kommt inhaltlich häufig in Form der Behauptung daher, die Mittelwerte oder Häufigkeiten für verschiedene Fallgruppen seien gleich. Die Nullhypothese gilt es zu verwerfen. Wenn dies geschieht, weil die Wahrscheinlichkeit des gemessenen empirischen Ergebnisses sehr gering ist, ist das Ergebnis in der untersuchten Stichprobe statistisch bedeutsam (signifikant). Als Grenze für das Verwerfen der Nullhypothese gilt oft eine Wahrscheinlichkeit von 0,05 bzw. 5 %. Liegt die Wahrscheinlichkeit für den in der Stichprobe gemessenen Kennwert darunter, so wird H_0 verworfen. Ganz präzise ausgedrückt: Es wird nicht der direkt gemessene Kennwert, sondern dessen inferenzstatistische Transformation, die sogenannte Teststatistik, für die Hypothesenprüfung verwendet.

Der Signifikanztest in der Inferenzstatistik – allgemeine Form

Die Logik des statistischen Testens von Nullhypothesen ist für alle statistischen Kennwerte (Häufigkeiten, Mittelwerte, Varianzen, Korrelationen) ähnlich. Für ein tieferes Verständnis der verschiedenen Verfahren und Verteilungen der Inferenzstatistik ist es hilfreich, sich das folgende Schema des Statistikers und Computerwissenschaftlers Allen Downey (geb. 1967) vom Olin College of Engineering in Needham, Massachusetts, einzuprägen.

Abbildung 29: Schema des Testens von Nullhypothesen in der Inferenzstatistik



Quelle: eigene Übersetzung von Downey 2016.

Die Daten der Stichprobe (links in der Abbildung) stellen die Grundlage für die Berechnung der Teststatistik dar. Die Teststatistik ist ein inferenzstatistischer Kennwert, welcher sich aus der Berechnung der Kennwerte in der Stichprobe ergibt, wenn man die Nullhypothese zugrunde legt. In die Berechnung einer Teststatistik gehen also zwei Aspekte ein: die empirisch gemessenen Häufigkeiten bzw. Kennwerte (Mittelwerte, Varianzen) sowie die Daten aus einer theoretisch konstruierten Verteilung, die sich aus einer Simulation aus unendlich vielen Stichproben ergeben. Das Ergebnis dieser Berechnung ist die sogenannte **Teststatistik**, der Kennwert für den angewendeten inferenzstatistischen Test.

Zu jedem deskriptiven Kennwert (Mittelwert, Varianzen, Häufigkeiten) gibt es einen spezifischen inferenzstatistischen Test. Für Häufigkeiten ist es der Chi-Quadrat-Test, für Mittelwerte ist es der t-Test, für Varianzen der F-Test. Die entsprechenden Teststatistiken heißen Chi-Quadrat-Wert, t-Wert und F-Wert und werden als die in der Stichprobe beobachteten Effekte bezeichnet. Deren Berechnungsweise ist unterschiedlich, das Anwendungsprinzip aber immer gleich. Im letzten Schritt wird geprüft, wie wahrscheinlich der mit Bezug auf die untersuchte Stichprobe errechnete Wert der Teststatistik ist, wenn deren Verteilung unter der Nullhypothese als Bezug genommen wird. Die Verteilung aller möglichen Teststatistiken stellt die Kurve dar (rechts unten). Je weiter der ermittelte Wert der Teststatistik am Rand dieser Verteilung liegt, desto unwahrscheinlicher ist er. Unter Umständen fällt er in den sogenannten Ablehnungsbereich der Nullhypothese (schraffierter Bereich rechts). Dann wird die H_0 verworfen und die Alternativhypothese angenommen.

Die Auswertungsschritte des Hypothesentestens werden im Folgenden an zwei Beispielen erläutert. Es sind der Chi-Quadrat-Test, der Häufigkeiten prüft, und der t-Test, der Mittelwertdifferenzen zwischen zwei Gruppen prüft. In Statistikprogrammen können beide Verfahren durch die betreffenden Befehle aufgerufen werden. Die verschiedenen Rechenschritte, die das Statistikprogramm durchführt, um theoretische Verteilungen zu konstruieren und Teststatistiken zu ermitteln, bleiben für die Anwender weitgehend unsichtbar. Ausgegeben werden allein die Ergebnisse der Hypothesenprüfung. Entscheidende Bedeutung kommt dem ausgegebenen p-Wert zu. Dieser gibt an, wie wahrscheinlich die für die Stichprobe errechnete Teststatistik (Chi-Quadrat-Wert bzw. t-Wert) bei Geltung von H_0 ist. Ist der p-Wert $> 0,05$ (die Wahrscheinlichkeit ist höher als 5 %), wird die Nullhypothese beibehalten. Das Ergebnis ist statistisch nicht auffällig. Im gegenteiligen Fall ($p < 0,05$) wird die Nullhypothese verworfen. Die Definition von 5 % als Randwahrscheinlichkeit ist eine willkürliche Setzung. Es kommt vor, dass 1 % ($p < 0,01$) als Grenze der Annahme bzw. Ablehnung von H_0 definiert werden.

Teststatistik

Eine Teststatistik ist die statistische Umwandlung des Kennwertes der Stichprobe. Mit der Teststatistik wird die Nullhypothese geprüft.

Der Signifikanztest bei Häufigkeiten (Chi-Quadrat-Test)

Wir besprechen das Vorgehen anhand des eingeführten Beispiels aus der Medizin. Jeweils 200 Patienten wurden mit einer Operation oder mit konventionellen Methoden (Rückengymnastik, Physiotherapie) behandelt.

Tabelle 14: Fiktive Vierfelder-Häufigkeitsverteilung von zwei Gruppen und zwei Behandlungsformen

	Rücken-OP	konventionell	Zeilensumme

ohne Beschwerden	125	150	n = 275
mit Beschwerden	75	50	n = 125
Spaltensumme	n = 200	n = 200	n = 400

Quelle: erstellt im Auftrag der IU, 2022.

Diese Tabelle enthält die empirischen Häufigkeiten des Datensatzes. Die Daten weisen auf den Einfluss der Behandlungsmethode hin. Die Heilungsquote ist bei der konventionellen Behandlung um 25 Personen höher. In der Inferenzstatistik wird geprüft, ob der Einfluss einer Variablen auf die andere statistisch bedeutsam (signifikant) oder nur zufällig ist. Ist der Einfluss nur zufällig, stellt sich in unserem Beispiel der Behandlungserfolg unabhängig von der Behandlungsmethode ein. Die zweite Variable (Behandlungserfolg) ist nicht dem Einfluss der ersten (Behandlungsform) unterworfen. Der Chi-Quadrat-Test wird deshalb als Unabhängigkeitstest bezeichnet.

Nullhypothese/Alternativhypothese

Die Nullhypothese H_0 ist die Hypothese, die mittels der Teststatistik überprüft wird. Ist ihre Wahrscheinlichkeit geringer als 5 %, wird sie verworfen und die Alternativhypothese H_1 wird angenommen.

Für die Durchführung des Signifikanztests wird die **Nullhypothese** (H_0) aufgestellt. Diese behauptet, dass die Behandlungsmethode keinen statistisch signifikanten Einfluss auf den numerisch unterschiedlichen Behandlungserfolg hat. Beide Behandlungsmethoden seien in ihrer Wirkung gleich. Zudem wird vor der Durchführung des Signifikanztests auch die gegenläufige **Alternativhypothese** (H_1) formuliert.

- H_0 : Die Behandlungsformen haben keinen Einfluss auf die Erfolgsquote der Behandlung.
- H_1 : Die Behandlungsformen unterscheiden sich signifikant in ihrer Wirkung auf die Erfolgsquote der Behandlung.

Unter Zugrundelegung der Nullhypothese müsste die Verteilung von beiden Patientengruppen (Operation oder konventionelle Behandlung) identisch sein. Bei 400 untersuchten Personen in der Stichprobe ist der Anteil von Gesunden 275 und der von Patienten mit Beschwerden 125. Das ist ein Verhältnis von 68,8 % zu 31,2 %. Genauso sollten die relativen Anteile der Häufigkeiten in beiden Behandlungsgruppen sein, falls die H_0 gilt. Die folgende Tabelle zeigt bei unterstellter Geltung der Nullhypothese die zu erwartende Verteilung.

Tabelle 15: Theoretisch zu erwartende Vierfelder-Häufigkeitsverteilung der dichotomen Variablen Behandlungserfolg bei zwei Behandlungsformen

	Rücken-OP	konventionell	Zeilensumme
ohne Beschwerden	137,5	137,5	n = 275
mit Beschwerden	62,5	62,5	n = 125
Spaltensumme	n = 200	n = 200	n = 400

Quelle: erstellt im Auftrag der IU, 2022.

Die bei Geltung der H_0 zu errechnende Teststatistik für Häufigkeiten ist der **Chi-Quadrat-Wert**. Dieser errechnet sich immer aus der Differenz zwischen den gemessenen und den erwarteten Häufigkeiten in allen Feldern einer Mehrfeldertafel. Pro Zelle der Vierfeldertafel wird die Differenz aus beobachteten (observed) und erwarteten (expected) Häufigkeiten quadriert und durch die erwarteten Häufigkeiten dividiert. Letzteres ist notwendig, um unterschiedliche Gruppengrößen zu verrechnen. In unserem einfachen Beispiel sind allerdings beide Gruppengrößen gleich (jeweils 200 Personen). Die allgemeine Formel für den Chi-Quadrat-Wert bei Vierfeldertafeln lautet (f steht für Häufigkeiten):

$$x^2 = \frac{(f_{o,1} - f_{e,1})^2}{f_{e,1}} + \frac{(f_{o,2} - f_{e,2})^2}{f_{e,2}} + \frac{(f_{o,3} - f_{e,3})^2}{f_{e,3}} + \frac{(f_{o,4} - f_{e,4})^2}{f_{e,4}} \quad (4.4)$$

Chi-Quadrat-Wert
Der Chi-Quadrat-Wert als Teststatistik basiert auf der Abweichung der gemessenen Häufigkeiten von den erwarteten, wenn die Nullhypothese gilt.

In die folgende Formel sind die Werte unseres Beispiels eingesetzt:

$$x^2 = \frac{(125 - 137,5)^2}{137,5} + \frac{(150 - 137,5)^2}{137,5} + \frac{(75 - 62,5)^2}{62,5} + \frac{(50 - 62,5)^2}{62,5} \quad (4.5)$$

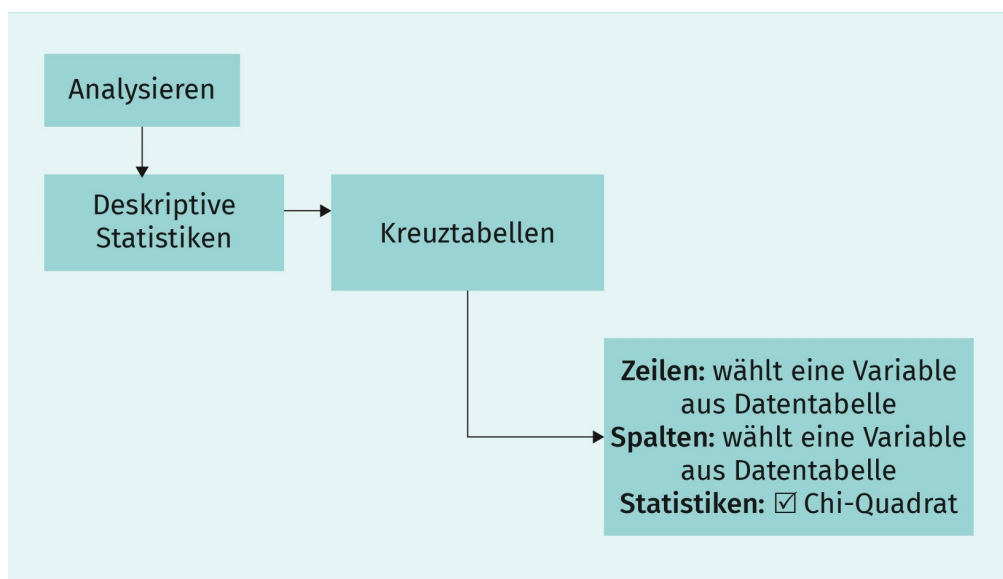
Daraus ergibt sich:

$$x^2 = 1,14 + 1,14 + 2,5 + 2,5 = 7,28 \quad (4.6)$$

Damit ist die Teststatistik für die empirischen Werte bestimmt. Der gemessene Chi-Quadrat-Wert beträgt bei Geltung der Nullhypothese 7,28. Der letzte Schritt der Signifikanzprüfung besteht darin, die Wahrscheinlichkeit für diesen Wert in einer Chi-Quadrat-Verteilung zu ermitteln (in der Abbildung die Verteilung unten rechts). Der p-Wert für den Chi-Quadrat-Wert 7,28 liegt bei 0,055. Die für die Stichprobe ermittelte Teststatistik ist damit zwar recht unwahrscheinlich, liegt aber noch knapp innerhalb des Annahmebereiches der Nullhypothese links vor der Grenze von 0,05. In der Chi-Quadrat-Verteilung haben alle Werte, die größer als 7,81 sind, eine Wahrscheinlichkeit von $p < 0,05$. Das Ergebnis ist für die Stichprobe der 400 untersuchten Patienten damit statistisch nicht signifikant. H_0 , die Behandlungsform hat keinen Einfluss auf die Heilungsquote, wird beibehalten, trotz der gemessenen Unterschiede zwischen den beiden Behandlungsformen.

Die Chi-Quadrat-Verteilung ist eine theoretisch konstruierte Verteilung. Diese liegt den Statistikprogrammen vor bzw. wird darin simuliert. Mit Aufruf des Chi-Quadrat-Befehls für eine Kreuztabelle von zwei Variablen und deren Häufigkeiten errechnet das Programm den Chi-Quadrat-Wert und zeigt die Wahrscheinlichkeit (p-Wert) für die gemessene Teststatistik. In der Regel wird auch angegeben, ob das Ergebnis signifikant ist oder nicht.

Abbildung 30: SPSS-Befehl für eine Mehrfelder-Tafel mit Chi-Quadrat-Test



Quelle: erstellt im Auftrag der IU, 2022.

Ein Vierfeldertest ist ein einfaches Beispiel für einen Mehrfelder-Chi-Quadrat-Test. Beim obigen Beispiel wird der Chi-Quadrat-Test mit $n - 1$ Freiheitsgraden angesetzt. Freiheitsgrade sind ein wichtiger Begriff in der Inferenzstatistik. Grob gesagt geben sie an, wie viele Werte in einem statistischen Ausdruck frei variieren können. Bei einer Vierfeldertafel können nur für drei von vier Zellen die Werte willkürlich ausfallen. Für die vierte Zelle bleibt immer nur der durch die drei anderen Zellen festgelegte Rest. Die Anzahl der Freiheitsgrade ist in unserem Beispiel deshalb $4 - 1 = 3$.

Die Wahrscheinlichkeit einer gemessenen Teststatistik hängt von verschiedenen Faktoren ab. Ein bestimmter Chi-Quadrat-Wert, wie z. B. 7,28, kann bei einer Stichprobe signifikant ausfallen, in einer anderen Stichprobe nicht. Ein wichtiger Einflussfaktor ist die Stichprobengröße. Je größer die Stichprobe, desto eher wird der Wert der Teststatistik signifikant und die H_0 verworfen. Außerdem sind die Freiheitsgrade ein relevanter Faktor. Für jeden Freiheitsgrad ergibt sich eine andere Verteilung der Teststatistik unter Geltung der Nullhypothese. Im Allgemeinen erkennt das Statistikprogramm, welchen Freiheitsgrad es für eine bestimmte Variablenkombination auswählen muss, vorausgesetzt, die Anwender wählen die richtigen Befehle. Was Freiheitsgrade genau sind und welche Bedeutung sie für die Inferenzstatistik haben, das wird in den Lehrbüchern ausführlich erläutert (z. B. Nachtigall/Wirtz 2013, S. 119f.).

Der Signifikanztest bei Mittelwertdifferenzen (t-Test)

Der **t-Test** wird angewendet, wenn die Abhängigkeit zwischen einer dichotomen und einer metrisch skalierten Variable untersucht werden soll. Die dichotome Variable definiert die zu vergleichenden Gruppen (z. B. Männer und Frauen, Raucher und Nichtraucher, Neulinge und Experten). Überprüft wird, ob die gemessenen Unterschiede zwischen den beiden Gruppen in der metrischen Variable statistisch signifikant sind oder nur zufällig in der

untersuchten Stichprobe gemessen wurden. Für die metrische Variable werden nicht die Häufigkeiten der Werte bzw. Fälle ermittelt, sondern Mittelwert und Streuung. Diese sind der Ausgangspunkt zur Berechnung der Teststatistik (t-Wert). Der t-Wert ermittelt die Differenz zwischen den Mittelwerten der beiden Gruppen, falls die Nullhypothese (H_0) gilt.

t-Test

Der t-Test überprüft die Mittelwertdifferenzen von zwei Gruppen (z. B. Frauen und Männer) bei einer gemessenen metrischen Variable (z. B. Gewicht).

Die Anwendung des t-Tests ist oft verknüpft mit der Definition einer Versuchs- und einer Kontrollgruppe in einem experimentellen Versuchsplan. Beispielsweise wird in der Medizin oder der Psychologie eine Gruppe von behandelten Patienten (Versuchsgruppe) verglichen mit einer Kontrollgruppe, die nicht behandelt wurde bzw. die ein Placebo-Medikament erhielt. Gemessen werden die Unterschiede in den Mittelwerten beider Gruppen in heilungsrelevanten Variablen (z. B. mittlerer Wert auf einer psychologischen Skala). Aber auch in nicht-experimentellen Untersuchungsplänen oder explorativen Studien werden t-Tests zur Überprüfung von Gruppenunterschieden durchgeführt. Dazu werden Gruppierungsvariablen als dichotome Variablen benutzt. Diese werden im Fragebogen abgefragt oder hinterher bei der Datenaufbereitung definiert. In unserem Beispiel wurde der Studierendenstatus (Anfänger oder Fortgeschrittene) nachträglich definiert. Für den t-Test müssen Gruppierungsvariablen zweistufig (dichotom) sein.

Anwendung des t-Tests auf das Beispiel zur Studienmotivation

Die Variable Studierendenstatus hat zwei Ausprägungen. Sie wurde aus der Variable Semesterzahl in SPSS berechnet. 1 bedeutet Anfänger und umfasst die Werte 1, 2 und 3 bei der Semesteranzahl. 2 bedeutet Fortgeschrittene und beinhaltet die Werte 4, 5 und 6. Die Verteilung über die zehn Fälle zeigt die folgende Tabelle.

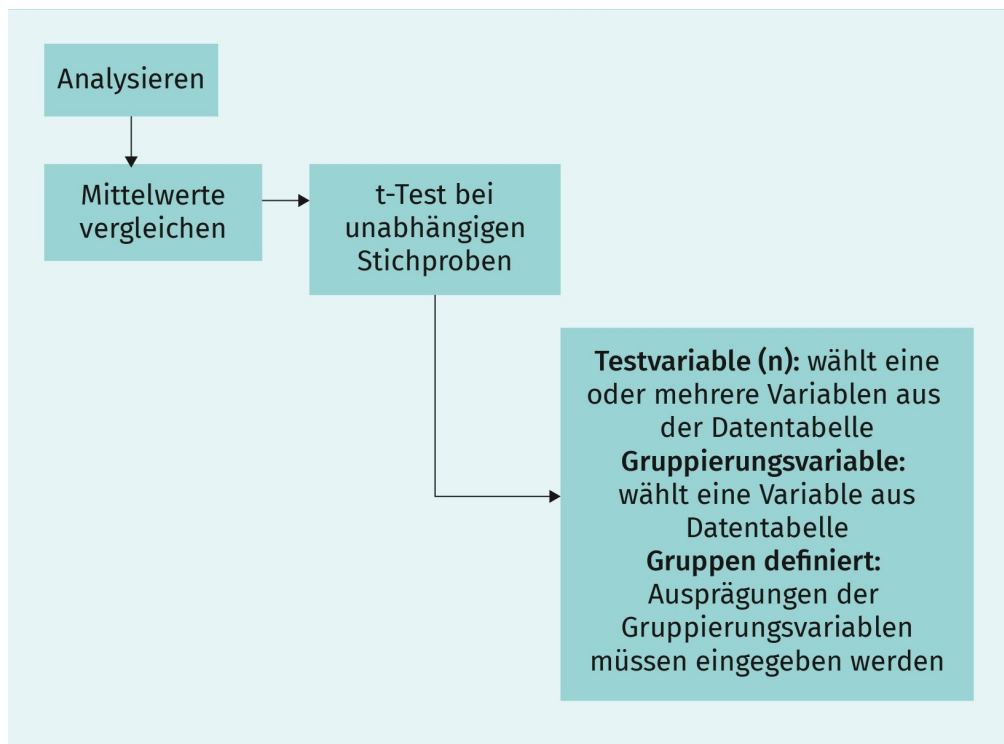
Tabelle 16: Verteilung der Gruppierungsvariablen Studierendenstatus im Beispiel Studienmotivation (n = 10)

Alter	20	18	24	40	24	22	21	23	20	28
Semesterzahl	3	1	6	6	6	2	3	3	1	6
Studierendenstatus	1	1	2	2	2	1	1	1	1	2

Quelle: erstellt im Auftrag der IU, 2022.

In der deskriptiven bzw. explorativen Datenanalyse wurden als Maße der zentralen Tendenz Mittelwerte und Streuungen für die Variable Alter in der Gesamtstichprobe ermittelt. Nun interessiert uns der genaue Altersunterschied zwischen den beiden Gruppen Anfänger (Semester 1–3) und Fortgeschrittene (Semester 4–6). Dazu berechnen wir in SPSS Mittelwert und Streuungen getrennt nach den beiden Gruppen der Variable Studierendenstatus.

Abbildung 31: SPSS-Befehl zum t-Test für zwei unabhängige Gruppen



Quelle: erstellt im Auftrag der IU, 2022.

SPSS gibt das folgende Ergebnis für die Gruppenmittelwerte aus:

Tabelle 17: Mittelwerte und Streuungen der Variablen Alter gruppiert nach Studierendenstatus

Anfänger oder Fortgeschrittene	Mittelwert	N	Standardabweichung
Anfänger	20,67	6	1,751
Fortgeschrittene	29,00	4	7,572
Insgesamt	24,00	10	6,272

Quelle: erstellt im Auftrag der IU, 2022.

Das Alter der fortgeschrittenen Studierenden ist im Mittel 8,33 Jahre höher. Es folgt die inferenzstatistische Analyse. Sie geht der Frage nach, ob diese zwischen beiden Gruppen in der untersuchten Stichprobe signifikant ausfällt. Der Ablauf des Signifikanztests ist ähnlich wie beim Chi-Quadrat-Test, aber die zu berechnende Teststatistik (t-Wert) ist eine andere:

1. Aufstellen der Nullhypothese H_0
2. Berechnung der Teststatistik unter Geltung von H_0

3. Ermittlung der Wahrscheinlichkeit des Wertes der Teststatistik in der Prüfverteilung aller möglichen Werte der Teststatistik
4. Verwerfen oder Annehmen von H_0

Die Nullhypothese besagt, dass es keinen Unterschied in den Mittelwerten gibt. Die Mittelwerte beider Gruppen sind gleich. In der formalisierten Schreibweise der Inferenzstatistik wird für Mittelwerte das griechische μ (Aussprache: müh) verwendet.

$$H_0: \mu_A = \mu_B \quad (4.7)$$

Die **Alternativhypothese** besagt, dass es einen Unterschied gibt.

$$H_1: \mu_A \neq \mu_B \quad (4.8)$$

Mit dieser Formulierung startet ein zweiseitiger (nicht gerichteter) Signifikanztest. Es wird nur behauptet, dass es einen Unterschied gibt, nicht, dass einer der beiden Mittelwerte größer ist als der andere. Letzteres wäre eine gerichtete Hypothese (vgl. ausführlich zur Anwendung einseitiger und zweiseitiger Tests Schäfer 2016, S. 163f.).

Alternativhypothese (H_1)
Die Alternativhypothese ist das Gegenteil der Nullhypothese (H_0). Zumeist handelt es sich dabei um eine nicht gerichtete, zweiseitige Hypothese, d. h., die Werte sind größer oder kleiner als in H_0 .

Der zweite Schritt ist die Berechnung der Teststatistik. Diese ist der t-Wert. Dessen allgemeine Formel beim Vergleich zweier unabhängiger Gruppen ist:

$$t = \frac{(29 - 20,67)}{\sqrt{\frac{1,75^2_A}{6_A} + \frac{sd_B^2}{4_B}}} = \frac{(8,33)}{\sqrt{\frac{3,06}{6} + \frac{57,3}{4}}} = 2,16 \quad (4.9)$$

Im Zähler steht die Differenz der Mittelwerte der beiden Gruppen. Der Nenner ist die Wurzel aus einer Summe. Deren beide Summanden sind jeweils die Varianzen ($sd^2 = \text{Standard Deviance zum Quadrat}$) in beiden Gruppen geteilt durch die Anzahl der Fälle (n) in der jeweiligen Gruppe. Für die Werte der untersuchten Stichprobe ergibt sich:

$$t = \frac{(29 - 20,67)}{\sqrt{\frac{1,75^2_A}{6_A} + \frac{sd_B^2}{4_B}}} = \frac{(8,33)}{\sqrt{\frac{3,06}{6} + \frac{57,3}{4}}} = 2,16 \quad (4.10)$$

Die berechnete Teststatistik (t-Wert) unter der Nullhypothese hat damit für die Stichprobe den Wert 2,16. Im nächsten Schritt ist zu prüfen, wo dieser konkrete t-Wert in der Verteilung aller t-Werte liegt. Das Statistikprogramm SPSS erledigt die beiden Schritte, t-Wert berechnen und Check in der Prüfverteilung, zusammen. Die Ergebnisse nach Aufruf des entsprechenden Befehls sind:

Tabelle 18: SPSS-Ergebnis des t-Tests für die Mittelwertdifferenzen der Variablen Alter im Studierendenstatus

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit		
		F	Signifikanz	T	df	Sig. (2-seitig)
Alter	Varianzen sind gleich	5,988	,040	-2,668	8	,028
	Varianzen sind nicht gleich			-2,163	3,215	,113

Quelle: erstellt im Auftrag der IU, 2022.

In der unteren Zeile ergibt sich der zuvor in der Formel errechnete Wert von 2,16, allerdings als negativer Wert. Der Grund ist, dass SPSS mit der umgekehrten Gruppenreihenfolge gerechnet hat. Gruppe_A hat den Mittelwert 20,33 und Gruppe_B 29. Dadurch wird der Zähler negativ und folglich der gesamte Bruch. Negative und positive t-Werte haben in der Inferenzstatistik die gleiche Bedeutung. Es interessiert nur der absolute Betrag beim t-Test.

Von SPSS werden zwei verschiedene Formen der Berechnung des t-Wertes durchgeführt (erste und zweite Zeile in der Tabelle). Die zweite Zeile (t-Wert = -2,16) geht von ungleichen Varianzen in beiden Gruppen aus, was für unser Beispiel gilt (Gruppe_A = 7,57², Gruppe_B = 1,75²). Die erste Zeile der von SPSS ausgegebenen Tabelle errechnet den t-Wert dagegen mit Schätzungen der Varianz für beide Gruppen in einer imaginären Grundgesamtheit, wenn unterstellt wird, die Varianzen beider Gruppen seien gleich. Die Berechnung führt zu einem höheren t-Wert als absoluter Betrag (2,67). Gleichzeitig wird ein inferenzstatistischer Test durchgeführt, der in einem F-Wert die reale und geschätzte Varianz ins Verhältnis setzt. Wie ein solcher F-Wert berechnet wird, wird später noch im Abschnitt zur Varianzanalyse erläutert.

Der F-Wert ist für unser Beispiel 5,99 und mit $p = 0,04$ statistisch signifikant ($p < 0,05$). Im Levene-Test auf Varianzgleichheit kommt es zu einer Zurückweisung von dessen Nullhypothese (H_0 : In beiden Gruppen sind die Varianzen gleich.). Die Varianzen sind in beiden unterschiedlich. Deshalb sollte für unser Beispiel der Mittelwertdifferenzen zwischen den beiden Gruppen Anfänger und Fortgeschrittene im Studium die zweite Zeile (die Varianzen sind nicht gleich) verwendet werden. In der zweiten Zeile führt die Verortung des gemessenen t-Wertes in der t-Verteilung zu einer Wahrscheinlichkeit von $p = 0,11$ (rechte Spalte in der Tabelle). Dieser Wert liegt oberhalb der Grenze von $p < 0,05$. $H_0 (\mu_A = \mu_B)$ muss mit den Daten unseres Beispiels beibehalten werden. Der Altersunterschied zwischen der Gruppe Anfänger und Fortgeschrittene ist statistisch nicht signifikant.

Effektstärke: Numerische Unterschiede zwischen zwei Gruppen

Beim Testen von Mittelwertdifferenzen in der Inferenzstatistik ist es üblich, im Forschungsbericht ein weiteres Maß anzugeben: die Effektstärke. Im t-Test wurde überprüft, ob es einen Effekt einer Gruppierungsvariable auf die metrische Variable gibt. Über die Größe des Effekts wird nichts ausgesagt. Deshalb wird mit der zusätzlichen Berechnung der Effektstärke abgeschätzt, ob die Mittelwertdifferenz in der metrischen Variable zwischen den beiden Gruppen vergleichsweise groß oder gering ist. Das Grundprinzip der Berechnung der Effektstärke ist, die Differenz der Mittelwerte durch die Streuung der metrischen Variablen in der Gesamtstichprobe zu teilen. Die Differenz der Mittelwerte wird quasi normiert oder standardisiert. Als Koeffizient der Effektstärke wird Cohens d verwendet. Der Koeffizient ist nach dessen Erfinder, dem Psychologen Jacob Cohen (1923–1998), benannt.

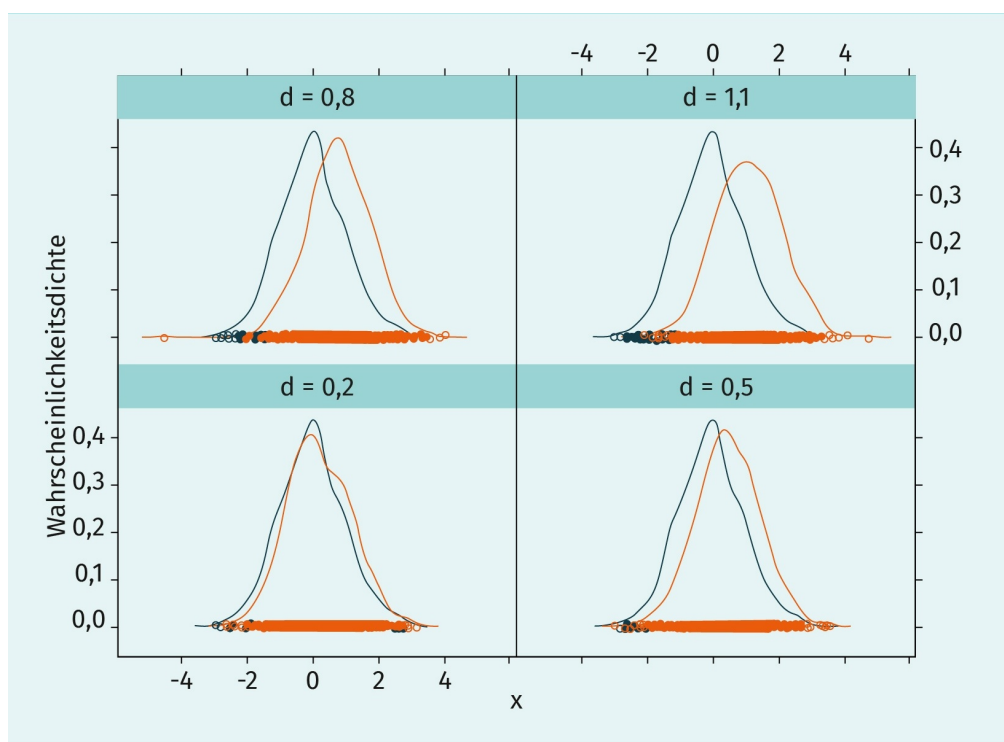
$$\text{Effektstärke: } d = \frac{(\bar{x}_A - \bar{x}_B)}{s} \quad (4.11)$$

In unserem Beispiel sind die Mittelwerte im Alter für A (Fortgeschrittene) 29 und für B (Anfänger) 20,67. Die Streuung der Variablen für alle zehn untersuchten Fälle beträgt 6,27. In die Formel eingesetzt:

$$d = \frac{(29 - 20,67)}{6,27} = 1,33 \quad (4.12)$$

Nach den von Cohen vorgegebenen Kriterien ist das ein außerordentlich hoher Effekt. Cohen definierte eine $d < 0,8$ als sehr hohen Effekt (Cohen 1988, S. 79f.). Die Wirkung der dichotomen Gruppenvariablen (Anfänger vs. Fortgeschrittene) auf die Variable Alter ist statistisch gesehen sehr groß. Anwender können sich den Koeffizienten d als ein Maß für die Überlappung zweier Verteilungen vorstellen. Sind die Mittelwerte, also die Gipfel der beiden Verteilungen, nahezu identisch und die Streuungen ebenso, überlappen sich die Verteilungen zu 100 % und $d = 0$. Die folgende Abbildung zeigt vier verschiedene imaginäre Verteilungen mit unterschiedlichen d -Werten. Auf der x -Achse ist jeweils eine Variable mit dem Mittelwert 0 abgetragen, auf der y -Achse Wahrscheinlichkeiten.

Abbildung 32: Überlappungsgrad zweier Verteilungen für variierende Werte des Koeffizienten d



Quelle: erstellt im Auftrag der IU, 2022.

In der Inferenzstatistik existieren diverse Formeln für die Berechnung von Cohens Koeffizienten d. Die oben genannte ist einfach und veranschaulicht das Berechnungsprinzip. Leider existiert in SPSS kein Befehl, um Cohens d direkt aufzurufen. Die Originalformel ist komplizierter, da diese nicht von der in der Stichprobe gemessenen Varianz bzw. Streuung ausging, sondern ein Schätzverfahren einsetzte, um die Varianz aus unendlich vielen Stichproben für eine imaginäre Grundgesamtheit zu berechnen. Basis dafür sind aber die gemessenen Stichprobenwerte. Cohens Formel lautet:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{sd_{\text{pool}}} \quad (4.13)$$

sd_{pool} steht für gepoolte Varianz. Die gepoolte Varianz wird verwendet, um die Varianz von zwei oder mehr Gruppen bzw. Grundgesamtheiten zu schätzen, wenn die Varianzen unbekannt sind, man aber davon ausgeht, dass sie in etwa gleich groß sind. Die Formel für die Schätzung der Varianzen, also für den Nenner, ist:

$$sd_{\text{pool}} = \sqrt{\frac{1}{n_A + n_B - 2} ((n_A - 1)sd_A^2 + (n_B - 1)sd_B^2)} \quad (4.14)$$

Neben Cohens d gibt es weitere Maße zum Vergleich der Effekte von dichotomen Gruppierungs- und/oder unabhängigen Variablen auf die Mittelwerte von metrischen Variablen. Cohens Koeffizient wird aber am häufigsten verwendet.

Andere Varianten des t-Tests

Wir haben in unserem Beispiel zum Alter von Studierenden einen t-Test für zwei unabhängige Stichproben mit abweichenden Streuungen durchgeführt. Diese Form des t-Tests wird nach seinem Namensgeber Bernard Lewis Welch (1911–1989) als Welchs-t-Test für ungleiche Varianzen bezeichnet. Die anderen Formen des t-Tests gehen auf den britischen Mathematiker William Sealy Gosset (1876–1936) zurück. Er benutzte in seinen Publikationen das Pseudonym „Student“, das als Namensgeber fungiert. Der Students-t-Test für zwei unabhängige Stichproben geht von der Varianzhomogenität in der Grundgesamtheit aus, aus der beide Gruppen stammen. Oben im SPSS-Ergebnis enthält die erste Zeile die Ergebnisse vom Students-t-Test. SPSS berechnet als Voreinstellung immer beide Varianten des t-Tests für unabhängige Stichproben. In anderen Statistikprogrammen (z. B. in R) gibt es dafür separate Befehle.

Weitere t-Tests für den Vergleich von Mittelwertdifferenzen, die auf Gosset bzw. Student zurückgehen, sind der t-Test für zwei **abhängige Stichproben** und der t-Test für eine Stichprobe (engl. „one sample t-test“). Zwei abhängige Stichproben liegen vor, wenn in den beiden Gruppen die gleichen Fälle vorkommen, z. B. alle Patienten vor und ebenfalls alle nach der Therapie. Der t-Test für eine Stichprobe ist anzuwenden, wenn der Mittelwert der gesamten Stichprobe für eine Variable mit einem feststehenden Mittelwert zu vergleichen ist. Beispielsweise könnte man den Mittelwert des Alters, der in unserem Beispiel für alle zehn untersuchten Fälle bei 24 Jahren liegt, mit dem bekannten Mittelwert einer größeren Studierendengruppe vergleichen, z. B. den Mittelwert aller eingeschriebenen Bachelorstudenten an der gleichen Universität. In SPSS sind dafür jeweils spezifische Befehle im gleichen Menü aufzurufen.

Abhängige Stichproben
Abhängige Stichproben sind solche mit identischen Fällen, d. h., an der gleichen Stichprobe werden zu zwei Zeitpunkten Messungen vorgenommen, z. B. vor einer Behandlung und danach.

4.4 Multivariate statistische Analysemethoden

Multivariate Analyseverfahren dienen der Auswertung mehrerer bzw. vieler (multi) Variablen. Die Anwendung multivariater Verfahren in der Auswertung muss bereits in der Datenerhebungsphase mitbedacht werden. Damit diese Verfahren zum Einsatz kommen, sollte eine größere Anzahl von Variablen in die Erhebung einfließen, zudem muss die Stichprobe groß genug sein. Beides trifft allerdings für studentische Bachelor- und Masterarbeiten oft nicht zu. Dennoch soll im Rahmen dieses Studienbriefs ein kurzer Einblick in diese fortgeschrittenen Analyseverfahren gegeben werden. Für ein intensives Studium multivariater Analysemethoden ist das gleichnamige Lehrbuch von Backhaus, Erichson, Plinke und Weiber besonders zu empfehlen (Backhaus et al. 2018).

Die Autoren unterscheiden zwischen strukturprüfenden und strukturentdeckenden Verfahren (Backhaus et al. 2018, S. 15f.). Strukturprüfende Verfahren zielen auf die exakte numerische Überprüfung von Abhängigkeiten zwischen verschiedenen Variablen ab. Die Richtung der Abhängigkeiten wird vor Untersuchungsbeginn in statistischen Modellen definiert. Es wird oft zwischen unabhängigen und abhängigen Variablen unterschieden. Ein typisches Beispiel ist die Varianzanalyse, die eng mit experimentellen Erhebungsplä-

nen verbunden ist. Bei ihr wird die Wirkung einer oder mehrerer unabhängiger Variablen (z. B. verschiedene Fortbildungen) auf eine abhängige (z. B. Lernerfolg von Seminarteilnehmern) geprüft. Die strukturprüfenden Methoden beinhalten inferenzstatistische Signifikanztests. Es wird überprüft, inwiefern die definierten statistischen Modelle auf die untersuchte Stichprobe passen.

Strukturentdeckende Verfahren suchen dagegen nach unbekanntem Mustern in den Daten. Zu Beginn der Untersuchung besitzen die Anwender nur vage Vorstellungen über die verdeckten Zusammenhänge der zu untersuchenden Fälle und Variablen. Strukturentdeckende Verfahren extrahieren Gemeinsamkeiten aus den Daten. Die Auswertung ist eine Verdichtung der Daten zu Oberkategorien (Personengruppen oder inhaltlich beschreibende Kategorien für Merkmale). Die Kategorien stellen eine latente Erklärungsstruktur für die Ausprägungen der Werte in einer Stichprobe dar. Dieses statistisch gewonnene Strukturmuster fungiert als ein Raster, um die untersuchten Fälle einzuordnen. Jeder untersuchte Fall kann am Ende der Analyse mindestens einer der Oberkategorien (z. B. einer Persönlichkeitsform oder einer Käufergruppe) zugeordnet werden. Strukturentdeckende Verfahren zählen zu den explorativen Verfahren der statistischen Datenanalyse. Ihre Ergebnisse können teilweise zusätzlich inferenzstatistisch überprüft werden. Diese Prüfung kann aber auch entfallen.

Tabelle 19: Übersicht zu den multivariaten Verfahren

Strukturen prüfen	multiple Regressionsanalyse, logistische Regressionsanalyse, Zeitreihenanalyse, Varianzanalyse, Diskriminanzanalyse, Kontingenzanalyse, Strukturgleichungsanalyse, Conjoint-Analyse
Strukturen entdecken	Faktorenanalyse, Clusteranalyse, multidimensionale Skalierung, Korrespondenzanalyse

Quelle: Backhaus et al. 2018, S. 15.

Welches der multivariaten Verfahren die Anwender einsetzen, das hängt in erster Linie von den Erkenntniszielen und in zweiter Linie von den Skalenniveaus der in die Analyse einfließenden Variablen ab. Bei der Varianzanalyse (Strukturen prüfen) müssen die unabhängigen Variablen in Form von Klassen oder Stufen vorliegen. Die passenden Skalenniveaus sind Nominal- oder Ordinalskala. Die abhängige Variable muss metrisch skaliert sein, denn zwischen den Stufen werden Mittelwert- und Varianzunterschiede betrachtet. Bei der Faktorenanalyse (Strukturen entdecken) sollten alle Variablen auf dem Niveau einer Intervall- oder Verhältnisskala vorliegen. Beide Verfahren werden im Folgenden überblicksartig in ihren wesentlichen Auswertungsschritten umrissen.

Die Varianzanalyse als Beispiel einer prüfenden multivariaten Analyse

Der Befehl für die **Varianzanalyse** lautet in den meisten englischen Statistikprogrammen ANOVA. Diese Abkürzung steht für „Analysis of Variance“. Analysiert werden die Varianzen in den verschiedenen Gruppen der gemessenen abhängigen Variablen. Die verschiedenen Ausprägungen der unabhängigen Variablen definieren die Gruppen. Sie werden auch als Faktorstufen bezeichnet. Es müssen mindestens drei verschiedene Stufen eines Faktors vorliegen, um die Varianzanalyse zu rechnen. Liegen nur zwei vor, wird der t-Test zur Überprüfung von Mittelwertdifferenzen eingesetzt.

Die Varianzanalyse prüft, ob die Lage der Filialen (unabhängige Variable) einen Einfluss auf die Verkaufszahlen (abhängige Variable) hat. Wirkungen zeigen sich als unterschiedliche Mittelwerte bzw. Streuungen in den unterschiedlichen Faktorstufen bzw. Gruppen. Es ist möglich, in einem Versuchsplan mehrere unabhängige Variablen aufzunehmen. Das führt zu Gruppen und Untergruppen. Dies wäre eine mehrfaktorielle Varianzanalyse. Im Folgenden wird die einfachste Variante einer Varianzanalyse besprochen. Es ist ein einfaktorieller Versuchsplan mit drei Ausprägungen der unabhängigen Variablen. Eine Erläuterung mehrfaktorieller Varianzanalysen findet sich in dem empfohlenen Lehrbuch (Backhaus et al. 2018, S. 175f.).

Varianzanalyse

Dies ist die Bezeichnung für eine Klasse multivariater Verfahren zur Überprüfung von gemessenen Unterschieden zwischen Gruppen von Fällen. Die einfachste Form ist die einfaktorielle Varianzanalyse. Bei ihr gibt es nur eine unabhängige Variable. Zugleich sind Varianzanalysen auch Versuchspläne mit einer oder mehreren unabhängigen und einer abhängigen Variable. Nur deren Mittelwerte und Varianzen werden gemessen.

Imaginäres Fallbeispiel für eine einfaktorielle Varianzanalyse

Angenommen, es gibt ein Handelsunternehmen mit drei Filialen in unterschiedlicher Lage in einer Region (im Stadtzentrum, am Stadtrand und in ländlicher Lage). An einem bestimmten Tag werden in allen drei Filialen die Verkaufszahlen von den gleichen fünf Produkten gemessen. Es soll untersucht werden, ob sich die Verkaufszahlen zwischen den Filialen signifikant unterscheiden.

Tabelle 20: Fiktive Verkaufszahlen von fünf Produkten in drei Filialen

	Filiale A (Stadt)	Filiale B (Rand)	Filiale C (Land)
Produkt A	18 Stück	12 Stück	17 Stück
Produkt B	14 Stück	16 Stück	18 Stück
Produkt C	14 Stück	13 Stück	16 Stück
Produkt D	18 Stück	15 Stück	19 Stück
Produkt E	16 Stück	14 Stück	20 Stück
Mittelwerte	16	14	18

Quelle: erstellt im Auftrag der IU, 2022.

Die Varianzanalyse prüft, ob die Filialen (unabhängige Variable) einen Einfluss auf die Verkaufszahlen (abhängige Variable) haben. Per Augenschein scheint es bei Betrachtung obiger Datentabelle so, denn die drei Mittelwerte unterscheiden sich. Eine zusätzliche Bedingung für die Anwendung der Varianzanalyse ist, dass alle anderen potenziellen Einflüsse auf die abhängige Variable in ihrer Wirkung aus- oder gleichgeschaltet werden. Andere Einflussfaktoren in unserem Beispiel wären z. B. Produktpreise, Werbemaßnahmen oder Öffnungszeiten der Filialen. Wir setzen für die folgenden Berechnungen voraus, dass diese für die untersuchten Filialen gleich sind.

Die Varianzanalyse ist ein statistischer Signifikanztest. Die Abfolge der Testschritte entspricht dem Vorgehen, das anhand des Schemas von Allen Downey (2016) erläutert wurde.

1. Aufstellen der Nullhypothese H_0
2. Berechnung der Teststatistik unter Geltung von H_0

3. Ermittlung der Wahrscheinlichkeit des Wertes der Teststatistik in der Prüfverteilung aller möglichen Werte der Teststatistik
4. Verwerfen oder Annehmen von H_0

Diese vier Schritte werden im Folgenden für die Varianzanalyse konkretisiert. Die entsprechende Nullhypothese besagt, dass die unabhängige Variable ohne Einfluss auf die abhängige Variable ist. Die Mittelwerte der abhängigen Variablen (mittlere verkaufte Stückzahl fünf verschiedener Produkte) wären dann in den drei Filialen identisch.

$$H_0: \mu_1 = \mu_2 = \mu_3 \quad (4.15)$$

Streuungs- bzw. Varianzzerlegung

Bei der Varianzanalyse wird die Summe aller Streuungen (Varianzen) von allen gemessenen Werten zerlegt (analysiert).

Der zweite Schritt, die Berechnung der Teststatistik, beruht in der Varianzanalyse auf dem Grundgedanken der **Streuungs- bzw. Varianzzerlegung**. Es gibt zwei verschiedene Formen der Varianz in der Gesamtstichprobe. In jeder untersuchten Gruppe bzw. Faktorstufe variieren die Messwerte um den Gruppenmittelwert. Das ist die Varianz in den Gruppen (engl. mss_w = „mean sums of squares within“). Außerdem gibt es die Streuung der Gruppenmittelwerte um den Gesamtmittelwert aller untersuchten Fälle (engl. mss_b = „mean sums of squares between“). In unserem Beispiel ist der Gesamtmittelwert die mittlere Stückzahl aller verkauften Produkte in den drei Filialen. Sie beträgt 16.

Die Varianzanalyse vergleicht Varianzen, nicht Mittelwerte. Indirekt kann über die Nullhypothese aber auf die Mittelwerte geschlossen werden. Genauer gesagt: Wenn H_0 gilt, dann sollten die mittleren Unterschiede zwischen den Gruppen gering sein. Die Gruppenmittelwerte liegen dann nahe beim Gesamtmittelwert und haben so eine geringe Varianz. Ist das nicht der Fall, sollten sich die Streuungen bzw. Varianzen von Gruppe zu Gruppe merklich unterscheiden. Die zu berechnende Teststatistik ist ein Bruch, der das Verhältnis zwischen der Streuung in den Gruppen (mss_w) ins Verhältnis setzt zur Streuung der Gruppenmittelwerte (mss_b).

$$\text{Teststatistik (F-Wert): } \frac{mss_b}{mss_w} \quad (4.16)$$

Wenn die Varianz zwischen den Faktorstufen der Varianz innerhalb der Faktorstufen entspricht, sind Zähler und Nenner gleich und es ergibt sich $F = 1$. H_0 wird dann verworfen. Damit F wächst, muss der Zähler (Varianz zwischen den Faktorstufen) größer sein als der Nenner (Varianz innerhalb der Gruppen). Als Faustregel gilt: Je größer F , desto eher wird die Nullhypothese verworfen. Allerdings hängt das bei jeder Untersuchung von den Freiheitsgraden ab, die für die **Berechnung von F** verwendet werden. Für die Berechnung von F sind mss_w und mss_b zu bestimmen. Dies sind die summierten Quadratsummen dividiert durch die Freiheitsgrade. Bei den Varianzen in den Gruppen sind die Freiheitsgrade bei je fünf Fällen in drei Gruppen: $15 - 3 = 12$. Bei der Varianz zwischen den drei Gruppenmittelwerten ergeben sich als Freiheitsgrade: $3 - 1 = 2$.

Tabelle 21: Varianzanalyse: Berechnung der Quadratsummen in drei Filialen (m_{ss_w})

	Filiale A (Stadt)	Filiale B (Rand)	Filiale C (Land)
Produkt A	$(18 - 16)^2$	$(12 - 14)^2$	$(17 - 18)^2$
Produkt B	$(14 - 16)^2$	$(16 - 14)^2$	$(18 - 18)^2$
Produkt C	$(14 - 16)^2$	$(13 - 14)^2$	$(16 - 18)^2$
Produkt D	$(18 - 16)^2$	$(15 - 14)^2$	$(19 - 18)^2$
Produkt E	$(16 - 16)^2$	$(14 - 14)^2$	$(20 - 18)^2$
Quadratsummen	16	10	10

Quelle: erstellt im Auftrag der IU, 2022.

$$m_{ss_w}: \frac{16 + 10 + 10}{12} = 3 \quad (4.17)$$

Die Quadratsumme der drei Gruppenmittelwerte ist:

$$(16 - 16)^2 \cdot 5 + (14 - 16)^2 \cdot 5 + (18 - 16)^2 \cdot 5 = 0 + 20 + 20 = 40 \quad (4.18)$$

$$m_{ss_b} = \frac{40}{2} = 20 \quad (4.19)$$

Im letzten Schritt zur Berechnung der Teststatistik für die Varianzanalyse wird die Varianz zwischen den Faktorstufen durch die Varianzen innerhalb der Gruppen geteilt.

$$m_{ss_b} = \frac{40}{2} = 2 \quad (4.20)$$

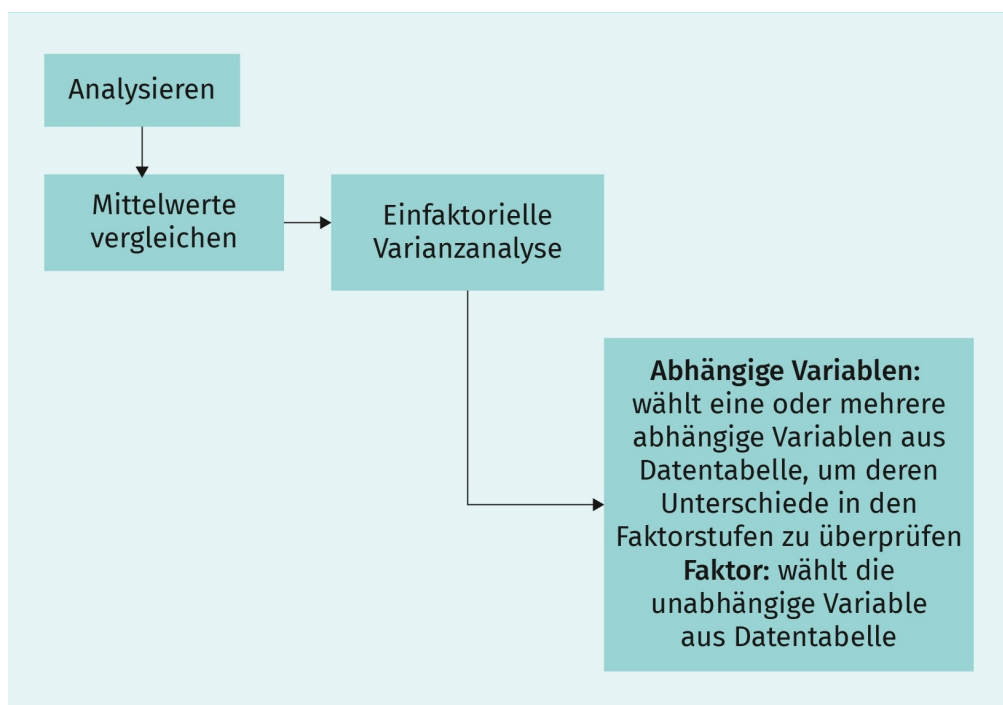
Der für unser Beispiel berechnete F-Wert sagt aus, dass die Varianzen zwischen den Gruppen bzw. Faktorstufen insgesamt höher sind als innerhalb der Gruppen. Für die Überprüfung von H_0 muss abschließend ermittelt werden, wo der empirisch gemessene F-Wert von 6,67 in der theoretisch konstruierten F-Verteilung liegt. Dazu kann in der F-Verteilung der Grenzwert für die Wahrscheinlichkeit $p < 0,05$ abgelesen werden. Bei zwei Freiheitsgraden für den Zähler und zwölf Freiheitsgraden für den Nenner des F-Wertes liegt dieser Grenzwert bei 3,81. 95 % der F-Werte sind bei unendlich vielen Stichproben kleiner, 5 % größer. Für diese 5 % wird H_0 verworfen. In unserem Beispiel trifft das zu. H_0 wird also verworfen. Die Lage der Filiale hat einen Einfluss auf die Verkaufszahlen der fünf Produkte.

In SPSS gibt es verschiedene Befehle für die Durchführung von Varianzanalysen. Die einfaktorielle Varianzanalyse ist in SPSS im gleichen Menü wie der t-Test.

F-Wert

Der F-Wert ist der Quotient zweier Varianzschätzungen für die Grundgesamtheit. Bei der einfaktoriellen Varianzanalyse ist der F-Wert der Quotient aus den Varianzschätzungen auf der Basis der Quadratsummen zwischen den Faktorstufen und auf der Basis der Quadratsummen innerhalb der Faktorstufen. Bei Varianzanalysen ist die Teststatistik immer ein F-Wert. Dieser wird an der F-Verteilung auf Signifikanz geprüft.

Abbildung 33: SPSS-Befehl zur Durchführung der einfaktoriellen Varianzanalyse



Quelle: erstellt im Auftrag der IU, 2022.

Statistikprogramme geben in der Regel nicht den Fünf-Prozent-Grenzwert in der F-Verteilung an, sondern die Wahrscheinlichkeit für den empirisch gemessenen F-Wert. Ist der vom Statistikprogramm ausgegebene p-Wert $< 0,05$, wird H_0 abgelehnt. Das Verwerfen der H_0 bei einer Varianzanalyse bedeutet immer, dass der oder den unabhängigen Variablen eine statistisch signifikante Wirkung auf die abhängige Variable zugeschrieben wird. Es wird allerdings nichts darüber ausgesagt, wie groß die Wirkung der einzelnen Faktorstufen ist. Die Varianzanalyse liefert allein eine globale Antwort auf die Frage, ob überhaupt ein Einfluss besteht. Fällt diese Antwort positiv aus, wird H_0 (alle Mittelwerte der Faktorstufen sind gleich) verworfen. Im Anschluss können weitere multivariate Analysen mit strukturprüfenden Verfahren (Diskriminanzanalyse oder multiple Regressionsanalyse) durchgeführt werden, um die numerischen Einflüsse der unabhängigen Variablen auf die abhängige genauer zu spezifizieren.

Die Faktorenanalyse als Beispiel einer explorativen multivariaten Analyse

Faktorenanalyse
Die Faktorenanalyse setzt an einer größeren Menge von Korrelationen zwischen jeweils zwei Variablen an und führt im Ergebnis zu voneinander unabhängigen Faktoren, welche die Zusammenhänge zwischen vielen Variablen erklären.

Die **Faktorenanalyse** sucht wie andere strukturentdeckende Verfahren nach einer latenten Struktur, die nicht direkt in den Daten zu finden ist. Die zu Anfang verborgene Struktur ist bei allen explorativen Verfahren ein Gerüst von Kategorien, das als Erklärung für die in den Daten vorgefundenen Zusammenhänge fungiert. In verschiedenen explorativen Verfahren werden die Abhängigkeiten mit jeweils anderen Kennwerten ermittelt. Bei der Faktorenanalyse fungieren Korrelationen zwischen den Variablen als Ausgangsbasis für die Generierung einer erklärenden Tiefenstruktur. Alle Korrelationen zwischen den Variablen, die zu dem erforschten Thema gehören und in einer Stichprobe erhoben wurden, müssen in einer Faktorenanalyse berücksichtigt werden. In unserem Beispiel zur Studienmotiva-

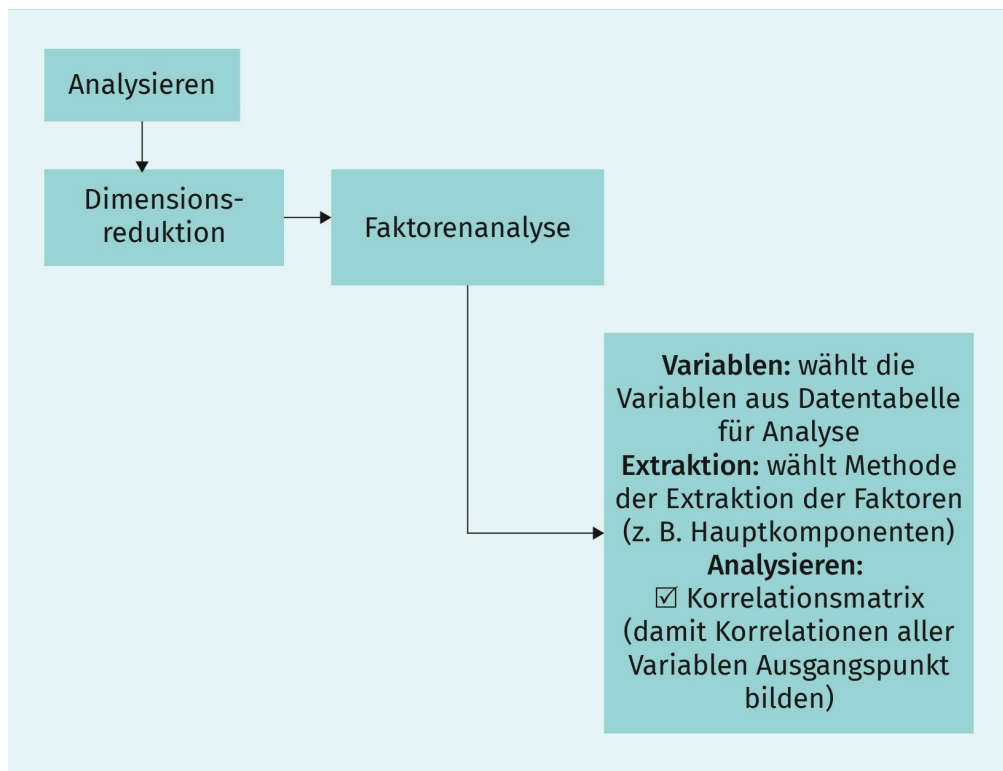
tion enthält der verwendete Fragebogen neun Variablen zur Motivation. Diese müssen also komplett für eine Faktorenanalyse zur Motivationsstruktur der Studierenden berücksichtigt werden.

Als Erfinder der explorativen Faktorenanalyse gilt der britische Psychologe Charles Spearman, der sich mit Intelligenztests beschäftigte. Er fand heraus, dass die Ergebnisse von Testaufgaben miteinander korrelierten, die sich inhaltlich auf ähnliche Intelligenzaspekte beziehen. Beispielsweise korrelierten die Ergebnisse von Rechenaufgaben untereinander hoch. Ihre Korrelation zu den Ergebnissen von verbalen Aufgaben war dagegen geringer. Dies führte Spearman zu der Idee eines Zweifaktorenmodells der Intelligenz. Er postulierte einen Generalfaktor der Intelligenz und einen bereichsspezifischen Faktor für jede Testaufgabe. In ähnlicher Weise beschrieben später Persönlichkeitsforscher Faktoren der Persönlichkeit aufgrund der Korrelation der Ergebnisse von Persönlichkeitsfragebögen. Die Geschichte der psychologischen Intelligenz- und Persönlichkeitsforschung ist aufs Engste mit der Entwicklung der Faktorenanalyse als statistischem Verfahren verknüpft. Intelligenz und Persönlichkeit sind Konstrukte, die aus verschiedenen Faktoren bestehen. Das Verfahren der Faktorenanalyse besteht darin, diese Faktoren in statistisch optimaler Weise zu generieren.

Faktoren (latente Merkmale) sind hypothetische Konstrukte, die für die Beantwortung der Items in einem Datensatz ausschlaggebend sind. Um die Faktoren zu schätzen, werden bei der Faktorenanalyse neue Variablen generiert, die die Informationen im Datensatz maximal erklären. Das Analyseergebnis kann zur Datenreduktion genutzt werden. Am häufigsten wird folgendes Vorgehen zur Datenreduktion gewählt: Variablen, die Indikatoren desselben Faktors sind, werden zu einem Summenwert zusammengefasst. Solche Summenwerte sind zuverlässigere und fehlerfreiere Messgrößen als die Einzelvariablen im Ursprungsdatensatz (Wirtz/Nachtigall 2012, S. 202).

Die Faktorenanalyse ist ein komplexes Verfahren, das auf vielen Rechenschritten beruht. Zudem gibt es verschiedene Varianten der Faktorenanalyse (vgl. Backhaus et al. 2018, S. 365f.). Die Anwendung in SPSS und anderen Statistikprogrammen ist dennoch relativ einfach. Die Anwender werden von allen Rechenschritten freigestellt. Sie wählen nur aus, welche Variablen eingelesen werden und welche Form der Faktorenanalyse angewendet werden soll.

Abbildung 34: SPSS-Befehl zur Durchführung der Faktorenanalyse



Quelle: erstellt im Auftrag der IU, 2022.

Korrelations- bzw. Interkorrelationsmatrix

Eine Korrelations- bzw. Interkorrelationsmatrix ist eine symmetrische Matrix, d. h., sie enthält genauso viele Zeilen wie Spalten. Sie gibt alle paarweisen Korrelationen der verwendeten Variablen wieder.

1. Erstellung der **Korrelations- bzw. Interkorrelationsmatrix** der verwendeten Variablen
2. Bestimmung der Faktorstruktur
3. Festlegung der Anzahl der Faktoren
4. Benennung und Interpretation der Faktoren

Für die Berechnungsschritte verwenden wir den Datensatz zur Studienmotivation. Eine Faktorenanalyse lässt sich für die neun Variablen zur Einschätzung der Studienmotivation durchführen. Verwendet wurde eine vierstufige LIKERT-Skala. Die Werte liegen damit im Bereich 1–4 für neun Variablen und zehn Fälle vor. Die Mindestanforderungen an Variablen- und Fallzahl werden damit in unserem Beispiel formal nicht erfüllt, denn es sollten mindestens 50 Personen und dreimal mehr Personen als Variablen Eingang in eine Faktorenanalyse finden (Wirtz/Nachtigall 2012, S. 202). Das Prinzip einer Faktorenanalyse kann aber auch mit geringeren Zahlen veranschaulicht und mit SPSS gerechnet werden.

Zunächst wird die Interkorrelationsmatrix der neun Variablen erstellt.

Abbildung 35: Interkorrelationsmatrix in SPSS der Variablen zur Studienmotivation

	Kompetenz-zunahme	eigene Vorstellungen	neue Anregungen	Austausch mit Studenten	positive Rückmeldungen	Kooperation mit anderen	freies Lernen	Partner der Lehrenden	Interessen verfolgen
Kompetenz-zunahme	1	,12	,33	-,55	,00	-,23	,12	,00	-,14
eigene Vorstellungen	,12	1	,29	,44	,17	-,08	,42	,35	,49
neue Anregungen	,33	,29	1	-,11	,00	-,46	-,24	,12	-,05
Austausch mit Studenten	-,55	,44	-,11	1	,29	,31	-,31	-,33	-,05
positive Rückmeldung	,00	,16	,00	,28	1	,59	,00	-,59	-,18
Kooperation mit anderen	-,23	-,09	-,46	,31	,59	1	-,16	-,48	-,03
freies Lernen	,12	,42	-,24	-,31	,00	-,17	1	,66	,67
Partner der Lehrenden	,00	,35	,12	-,33	-,59	-,45	,67	1	,79
Interessen verfolgen	-,14	,49	-,05	-,05	-,18	-,03	,67	,79	1

Quelle: erstellt im Auftrag der IU, 2022.

Die Interkorrelationsmatrix enthält in der Diagonale jeweils die Korrelation einer Variable mit sich selbst. Diese Korrelationen sind deshalb gleich 1. Alle anderen Korrelationen sind ohne die 0 vor dem Komma angegeben. Diese Darstellung ist in Statistikprogrammen üblich. Der Matrix der Korrelationen können Anwender entnehmen, welche Variablen aufgrund ihrer hohen positiven oder negativen Korrelationen zusammengehören und damit eine Gruppe bilden, die sich zu einem Faktor verdichten lässt. In der obigen Matrix fällt auf, dass es negative und positive Korrelationen gibt. Ausschließlich positive Korrelationen treten zwischen den drei Variablen auf, die jeweils zu einem der drei Aspekte der Selbstbestimmungstheorie der Motivation von Ryan und Deci gehören. Das sind die Variablen 1–3, 4–6 und 7–9. In der Abbildung oben ist das durch die zusätzlich eingezeichneten drei Quadrate angedeutet. Da die Formulierung dieser Variablen aus der Selbstbestimmungstheorie abgeleitet wurde, war zu erwarten, dass sie statistisch jeweils eine Gruppe bilden. Da die Faktorenanalyse von den gruppierten Ähnlichkeiten bei der Faktorenlösung ausgeht, ist ein Ergebnis mit drei Faktoren in unserem Beispiel wenig überraschend.

Als Form der Faktorenanalyse wenden wir auf unsere Daten die **Hauptkomponentenanalyse** an. Das ist die in der Psychologie gängigste Variante der Faktorenanalyse. Diese Variante der Faktorenanalyse generiert neue Variablen (Faktoren), die die gemessene Varianz aller Variablen sukzessive aufklären. Die Korrelation einer Variable mit einem Faktor ist die

Hauptkomponentenanalyse
 Hauptkomponentenanalysen sind spezielle Varianten der Faktorenanalyse. In der Psychologie werden sie gegenüber anderen Varianten bevorzugt.

sogenannte Faktorladung. Zunächst wird ein Faktor extrahiert und überprüft, wie viel Prozent der gesamten Varianz, die in allen Korrelationen der Variablen enthalten ist, von diesem erfasst werden. Dieser Faktor wird in der Auswertung wie eine neue Variable behandelt. Alle Variablen werden dann mit diesem Faktor korreliert. Danach wird ein zweiter Faktor aus der verbliebenen Gesamtvarianz aller Korrelationen extrahiert. Für diesen werden ebenfalls die Faktorladungen der Variablen ermittelt. Danach werden nach dem gleichen Prinzip weitere Faktoren extrahiert. Von Faktor zu Faktor wird weniger Varianz verarbeitet. Auf diese Weise werden von einem Statistikprogramm zunächst so viele Faktoren ermittelt, wie es Variablen gibt.

Tabelle 22: SPSS-Ergebnis der Hauptkomponentenanalyse für das Beispiel Studienmotivation

Komponente	anfängliche Eigenwerte		
	Gesamt	% der Varianz	kumulierte %
1	3,063	34,033	34,033
2	2,069	22,986	57,019
3	1,515	16,835	73,854
4	1,285	14,276	88,130
5	,518	5,756	93,885
6	,427	4,741	98,627
7	,109	1,206	99,833
8	,013	,143	99,975
9	,002	,025	100,00

Quelle: erstellt im Auftrag der IU, 2022.

SPSS liefert für eine sukzessive Extraktion der neun Faktoren obiges Ergebnis. Dabei nimmt die Bedeutung der Faktoren für die Varianzaufklärung in den Daten von Zeile zu Zeile ab. Der erste Faktor, der bei der Hauptkomponentenanalyse „erste Komponente“ heißt, extrahiert 34 % der Varianz, die Faktoren 8 und 9 aber nur noch jeweils weniger als 0,2 %. Es stellt sich nun die Frage, wie viele Faktoren in die am Ende resultierende Faktorenstruktur aufgenommen werden.

Das Statistikprogramm schlägt den Anwendern in unserem Beispiel eine Vier-Faktoren-Lösung vor. Dabei geht es von der Höhe der Eigenwerte (zweite Spalte) aus. Der Eigenwert eines Faktors ist die Summe der bivariaten Determinationskoeffizienten des Faktors mit allen Variablen des Datensatzes. In unserem Beispiel ergeben sich neun Variablen. Der Eigenwert des Faktors 1 ist 3,1.

Dies kann wie folgt interpretiert werden: Der erste Faktor erklärt so viel Varianz wie 3,1 Variablen des Ausgangsdatensatzes aller neun Variablen. Teilt man den Eigenwert durch die Anzahl der Ausgangsvariablen, so erhält man den durch den Faktor vorhersagbaren oder aufgeklärten Varianzanteil.

$$\text{Varianzanteil Faktor1} = \frac{3,063}{9} = 0,34 \quad (4.21)$$

Das entspricht 34 % der aufgeklärten Varianz (Spalte 3).

Aufgenommen in die endgültige Faktorenstruktur werden nur die Faktoren, deren Eigenwert größer als 1 ist. Diese Regel ist das sogenannte Kaiser-Guttman-Kriterium bzw. Kaiserkriterium als Kurzform. „Nach dem Kaiserkriterium gilt als Mindestanforderung für einen aussagekräftigen Faktor, dass dieser mehr Varianz aufklärt als eine der Ursprungsvariablen. Dieses Kriterium ist für alle Faktoren erfüllt, die einen Eigenwert größer 1 besitzen“ (Wirtz/Nachtigall 2012, S. 208).

SPSS gibt als zweite Tabelle eine Übersicht mit den vier Faktoren und den zugehörigen **Faktorladungen** aller neun Variablen aus.

Tabelle 23: Komponentenmatrix in SPSS für das Beispiel Studienmotivation

	Komponente			
	1	2	3	4
Kompetenzzunahme	,227	-,548	,398	,584
eigene Vorstellungen	,425	,531	,683	-,046
neue Anregungen	,200	-,418	,757	-,217
Austausch mit Studenten	-,420	,645	,299	-,511
positive Rückmeldungen	-,530	,402	,403	,522
Kooperation mit anderen	-,579	,541	-,155	,384
freies Lernen	,733	,374	-,120	,441
Partner der Lehrenden	,950	,109	-,148	-,136
Interessen verfolgen	,748	,537	-,071	,028

Extraktionsmethode: Hauptkomponentenanalyse

vier Komponenten extrahiert

Quelle: erstellt im Auftrag der IU, 2022.

Diese Komponentenmatrix ist ein Vorschlag für die Anwender. Sie entscheiden, ob sie die Lösung akzeptieren. In der Forschungspraxis würden Anwender weitere Formen der Faktorenanalyse ausprobieren, ebenso verschiedene Extraktionsmethoden und Abbruchkrite-

Faktorladungen

Mit Rotationsverfahren werden in einer Faktorenanalyse die Koordinatensysteme (und damit die extrahierten Faktoren) so lange gedreht, bis sie ein zuvor definiertes statistisches Kriterium erfüllen. Die Varimax-Rotation dreht die Faktoren schrittweise so lange, bis die Varianz der quadrierten Ladungen pro Faktor maximal ist. Mittelgroße Ladungen werden so entweder geringer oder stärker und können damit eindeutiger ihren jeweiligen Faktoren zugeordnet werden.

rien. Unter den verschiedenen vom Statistikprogramm vorgeschlagenen Faktorenlösungen suchen sie für ihren Forschungsbericht die optimale aus. Dies ist im Allgemeinen diejenige, bei der sich die verschiedenen Variablen mit ihren Faktorladungen optimal jeweils nur einem Faktor zuordnen lassen. Es gibt also für jede Variable eine hohe Ladung auf einem Faktor und für den Rest an Faktoren niedrige Ladungen. Dadurch werden die Faktoren trennscharf.

Im vierten und letzten Auswertungsschritt sind die Komponenten inhaltlich zu benennen und zu interpretieren. In unserem Beispiel könnte die abschließende Benennung der vier Faktoren wie folgt aussehen:

1. **Autonomiebedürfnis:** Die drei höchsten Faktorladungen haben die Variablen 7, 8 und 9. Diese drei Variablen wurden vor dem Hintergrund genau dieser Kategorie der Selbstbestimmungstheorie formuliert.
2. **Bedürfnis nach sozialer Eingebundenheit:** Diese Kategorie orientiert sich ebenfalls an der Selbstbestimmungstheorie. Die beiden höchsten positiven Ladungen haben die Variablen 4 (0,645) und 6 (0,541), die in Übereinstimmung mit dieser Kategorie formuliert wurden. Eine zusätzliche Begründung für die Benennung dieses Faktors ergibt sich aus der ersten Variable (Zunahme an persönlichen Kompetenzen). Diese hat eine negative Ladung (-0,548). Niedrige Werte dieser individuell ausgerichteten Variable gehen also mit höheren Werten der sozialen Eingebundenheit einher.
3. **Kompetenzerleben:** Der dritte Faktor ist ebenfalls ein Aspekt der Selbstbestimmungstheorie. Die Variablen 2 und 3 wurden vor diesem theoretischen Hintergrund konstruiert und haben hohe Ladungen (0,683 und 0,757).
4. **Kompetenzzunahme durch Feedback:** Dies ist der einzige neue Aspekt gegenüber der Selbstbestimmungstheorie. Die Daten weisen auf diesen vierten Motivationsfaktor hin. Er hat hohe Ladungszahlen bei der ersten Variable (Kompetenzzunahme) und der fünften (positive Rückmeldungen durch Lehrende). Rückmeldungen durch Lehrende haben positive Effekte für die studentische Motivation.

Die Faktorenanalyse unseres Datensatzes bestätigt demnach größtenteils die theoretischen Vorannahmen. Das ist wenig überraschend, denn die Fragen des Fragebogens wurden vor diesem Hintergrund formuliert und die Selbstbestimmungstheorie von Deci und Ryan gilt als empirisch gut abgesicherte Theorie. Hinzu kommt für unsere Stichprobe ein Faktor, der vor allem auf der Wirkung der Interaktion von Lehrenden und Lernenden beruht.

4.5 Überblick zu den qualitativen Formen der Datenanalyse

In diesem Lernzyklus geht es um die Auswertung qualitativer Daten. Die qualitative Datenauswertung hat die aufbereiteten, d. h. in der Regel transkribierten Datenprotokolle als Basis. Es gibt Auswertungsverfahren, in denen die Interpretation nonverbaler bzw. latenter Aspekte zentral ist. Andere fokussieren dagegen auf das Wortprotokoll und kommen weitgehend ohne nonverbale Hintergründe von Texten und Gesprächen aus. Gemeinsa-

mes Ziel der qualitativen Verfahren der **Text- und Inhaltsanalyse** sind fundierte theoretische Erklärungen für die empirisch beobachtete Verschiedenheit der in der Stichprobe untersuchten Fälle. Dies leisten Kategorien von Aussagen und von verschriftlichten Beobachtungen sowie Typisierungen von Fällen bzw. Personen. Der Fortgang der qualitativen Datenauswertung führt von der Oberflächenstruktur der transkribierten Datenprotokolle hin zu einer erklärenden Tiefenstruktur, einem System von Theorien, Kategorien und Typisierungen.

Text- und Inhaltsanalyse

Qualitative Text- und Inhaltsanalysen werten aufgezeichnete bzw. fixierte Kommunikationsvorgänge so aus, dass Textpassagen Oberkategorien zugeordnet werden.

Verschiedene qualitative Ansätze der Datenauswertungen unterscheiden sich in den Erkenntniszielen und Auswertungsschritten. Schreier (2013c) gibt einen informativen Überblick zu den unterschiedlichen methodischen Konzepten in der qualitativen Datenauswertung. Sie unterscheidet acht verschiedene Ansätze

1. **Hermeneutik** als allgemeine Form wissenschaftlichen Verstehens: Hier geht es um die Schritte des wissenschaftlichen Interpretierens fremder Texte. Die Interpretation vollzieht sich oft in einer dafür speziell ausgebildeten Personengruppe. Die Interpretation hat zum Ziel, den manifesten Inhalt des Textes in seinen wahren bzw. latenten zu überführen. Manifest ist derjenige Inhalt eines Textes, der sich bei einer ersten Lektüre den Lesern aufdrängt. Der latente Inhalt ist der verborgene Kerninhalt von Texten. Dieser ist nur durch zusätzliche Analysen bzw. die Hinzunahme von historischen oder persönlichen Kontexten der Textentstehung zu erschließen. Ein Beispiel für einen hermeneutischen Ansatz ist die Traumdeutung in der Psychoanalyse nach Sigmund Freud (Freud 1900). Der Traumtext gilt als Manifestation latenter Wünsche und Konflikte des Träumers, die gemeinsam von Therapeuten und Klienten in der Psychoanalyse erschlossen werden.
2. **Objektive Hermeneutik:** Dies ist eine spezielle Form der Hermeneutik, die der deutsche Soziologe Ulrich Oevermann (geb. 1940) entwickelt hat. Bei der objektiven Hermeneutik geht es darum, die objektive, den Lesern zunächst verborgene Sinnstruktur von Texten zu entschlüsseln. Zentral ist das sequenzanalytische Vorgehen. Die Interpretation beginnt mit dem ersten Satz bzw. der Überschrift eines Textes und schreitet Satz für Satz voran bis zum Textende (vgl. Schreier 2013c, S. 250–252).
3. **Codieren:** Diese Form der Textanalyse ist Teil der Auswertung in der Grounded Theory. Mittlerweile wird das Codieren aber aufgrund der einfachen Erlernbarkeit und Anwendung auch bei verbalen Daten verwendet, die mit anderen methodischen Ansätzen erhoben wurden. Codieren bedeutet, dass induktiv aus den Protokollen Kategorien (Codes) verallgemeinert werden, welche die Protokolle verdichten.
4. **Qualitative Inhaltsanalyse:** Dies ist eine Form der Analyse verbaler Daten, die Oberkategorien definiert, um Textpassagen zusammenzufassen. Es gibt verschiedene Formen, die aber im Grundprinzip – Datenreduktion durch Oberkategorien – übereinstimmen. Im deutschen Sprachraum dominiert das Verfahren der qualitativen Inhaltsanalyse nach Mayring, welches im folgenden Lernzyklus näher vorgestellt werden wird.
5. **Computergestützte Verfahren:** Diese sind keine Methoden der qualitativen Datenanalyse, sondern Programme, die die qualitative Auswertung von Daten unterstützen. Es sind in den letzten beiden Jahrzehnten eine Reihe von Programmen entwickelt worden, die jeweils zu einem Auswertungsverfahren optimal passen. Da heute viele Daten elektronisch erhoben werden, ist der Einsatz von Software eine Hilfe für die gedankliche und schriftliche Auswertung, ersetzt diese aber nicht.

Hermeneutik

Die Hermeneutik ist eine Methode der Auslegung bzw. Interpretation von Texten. In die Interpretation werden Kontexte einbezogen, die oft nicht unmittelbar den zu interpretierenden Texten zu entnehmen sind. Das sind z. B. die Biografie eines Autors oder kulturelle Hintergründe der Entstehung des Textes.

Diskursanalyse

Diskursanalysen untersuchen gesellschaftliche Diskurse, also Redeweisen, die Einfluss haben auf das Denken und Handeln aller Menschen in einer Gesellschaft. Institutionen, die Diskurse prägen, sind z. B. die Politik oder die Medien.

Ikonologie

Die Ikonologie als Methode der visuellen Analyse wird auch als Ikonografie bezeichnet. Ursprünglich ist sie eine Methode in der Kunstgeschichte bzw. Kunstwissenschaft, um Gegenstände der bildenden Kunst zu analysieren.

6. **Diskursanalyse:** Aussagen in einem Interview können als Teil eines gesellschaftlichen Diskurses begriffen werden, der jenseits des Interviews in den Medien und der allgemeinen Öffentlichkeit stattfindet. Wer sich z. B. in einem Interview abfällig über Migranten äußert, tut dies wie weitere Personen in unserer Gesellschaft in anderen Gesprächssituationen. In der Diskursanalyse bezieht die Auswertung eines Interviews den allgemeinen gesellschaftlichen Diskurs eines Themas mit ein. Dazu ist die Hinzuziehung entsprechender Äußerungen gesellschaftlich relevanter Gruppen in den Medien erforderlich (vgl. Schreier 2013c, S. 261–264).
7. **Semiotik:** Dies ist ein sprachwissenschaftlicher Zugang der Datenanalyse. In der Sprachwissenschaft ist die Semiotik die Zeichentheorie. Aussagen von Menschen bestehen aus sprachlichen Zeichen und deren Verweis auf etwas Dahinterstehendes. Spricht z. B. jemand von Attentaten durch Ausländer, so ist das semiotisch etwas anderes, als wenn er sagen würde, die Anschläge seien von Terroristen oder Islamisten begangen worden. Zeichen rufen bei Lesern Assoziationen hervor, und Sprecher verbinden mit ihren Aussagen bestimmte Assoziationen. Eine semiotische Auswertung von Datenprotokollen sucht nach bedeutungstragenden Zeichen, Symbolen, Metaphern und Sprachbildern (vgl. Schreier 2013c, S. 265–267).
8. **Visuelle Analyse:** Diese Auswertungsform ist relevant für visuelle Daten, wenn keine Transkripte, sondern Bilder vorliegen. Die Auswertung visueller Daten kann als eine Form der **Ikonologie** erfolgen. „Ikone“ ist das altgriechische Wort für „Bild“ und „Abbild“ und bezeichnet ursprünglich Kult- und Heiligenbilder der orthodoxen Christenheit. Schreier definiert Ikonologie als Auswertungsmethode wie folgt: „Im Mittelpunkt stehen die Analyse von Symbolen und die motivgeschichtliche Einbettung der Darstellung. Die Analyse erfolgt auf drei aufeinander aufbauenden Ebenen: der darstellenden, der ikonografischen und der ikonologischen Ebene“ (Schreier 2013c, S. 268).

Im Folgenden wird allein die qualitative Auswertungsmethode besprochen, die nach den Erfahrungen des Autors dieses Studienbriefs an deutschen Universitäten in den Fächern Psychologie, Soziologie und Erziehungswissenschaften den vermutlich größten Verbreitungsgrad in Examensarbeiten hat. Es ist die qualitative Inhaltsanalyse nach Mayring.

4.6 Qualitative Text- bzw. Inhaltsanalyse von verbalen Daten

Eine Inhaltsanalyse ist eine Form der Analyse verbaler Datenprotokolle, die Oberkategorien definiert, um dazu passende Textpassagen in den transkribierten Protokollen zusammenzufassen. Im deutschen Sprachraum findet das von Philipp Mayring (geb. 1952) entwickelte inhaltsanalytische Verfahren viel Zuspruch. Es ist in etlichen Lehrbüchern dokumentiert und auch im Internet gibt es viele einführende Darstellungen. Mayring war bis zu seiner Pensionierung 2017 Professor für Psychologische Methodenlehre an der Universität Klagenfurt. Er hat die neueren Formen der Inhaltsanalyse im deutschen Sprachraum konzeptionell mitbegründet, systematisiert und weiter ausdifferenziert (vgl. ausführlich zu verschiedenen Formen der Inhaltsanalyse Schreier 2014). Als methodisch

kontrolliertes Verfahren hat Mayring seine Form der Inhaltsanalyse erstmalig 1983 umfassend in einem Lehrtext dargestellt. Mittlerweile liegt dieses Lehrbuch mit deutlich veränderten Inhalt in der zwölften Auflage vor (Mayring 2015).

Formen der qualitativen Inhaltsanalyse nach Mayring

Im Handbuch von Baur und Blasius (2014) hat Mayring gemeinsam mit Thomas Fenzl einen Überblick zu seinen Formen der Inhaltsanalyse verfasst. Er unterscheidet in diesem Artikel drei Varianten: die zusammenfassende, die strukturierende und die explikative Inhaltsanalyse. Die tatsächliche Anzahl der Varianten einer Inhaltsanalyse nach Mayring ist höher, da der Autor für spezifische Forschungsfragen zu diesen drei Hauptformen weitere Unterformen definiert hat (Mayring 2015). Die Namen der drei Hauptformen verdeutlichen die allgemeinen Zielrichtungen qualitativer Inhaltsanalysen, die auf Mayrings Ideen basieren:

- **Zusammenfassung:** Große Mengen von Aussagen bzw. Passagen aus verschiedenen Texten sollen zusammengefasst und reduziert werden.
- **Struktur:** Im Ergebnis der Inhaltsanalyse werden die Aussagen bzw. Texte neu strukturiert, damit sie tiefer verstanden werden.
- **Explikation:** Diese Form der Inhaltsanalyse zielt vor allem auf Erklärungen. Das Ergebnis der Inhaltsanalyse sind Kategorien mit belegenden Zitaten aus den Texten. Die Kategorien fungieren zugleich als Erklärungen für Textpassagen.

Text oder Inhalt – was genau wird analysiert?

Texte enthalten Inhalte, z. B. Aussagen von Interviewten zu verschiedenen Themen. Inhaltsverzeichnisse organisieren Texte. Sie geben einen Überblick über Texte. Für Leser fungieren Inhaltsverzeichnisse als Orientierungspunkte. Vereinfacht können sich Anfänger die Methodik der qualitativen Inhaltsanalyse zu Beginn als Prozess der Erstellung eines Inhaltsverzeichnisses vorstellen. Sie geben den Daten, also den Aussagen bzw. Texten, eine übersichtliche Struktur. Die Gliederungspunkte des Inhaltsverzeichnisses sind die **Kategorien**. Die Arbeit mit Kategorien ist entscheidend bei der qualitativen Inhaltsanalyse von Mayring. Kategorien sind Oberbegriffe, die Textpassagen zusammenfassen. Kategorien können hierarchisch (Ober- und Unterkategorien) oder quantitativ (starke vs. schwache Ausprägung eines Merkmals) differenzieren.

Die Kategorien werden zu Beginn der Inhaltsanalyse definiert, z. B. anhand wissenschaftlicher Literatur, und in Form einer Liste notiert. In der eigentlichen Inhaltsanalyse, der Arbeit mit und an den erhobenen Daten, werden diese Kategorien auf die Daten bezogen. In den transkribierten Protokollen werden Formulierungen gesucht, die zu den Kategorien passen. Im Beitrag für das Handbuch von Baur und Blasius (2014) schreibt Mayring, die treffendste Kennzeichnung seines Verfahrens sei „qualitativ orientierte kategoriengeleitete Textanalyse“ (Mayring/Fenzl 2014, S. 544).

Kategorien

Kategorien sind grundlegende Allgemeinbegriffe. Sie dienen der Zusammenfassung von und der Orientierung über empirisch registrierte Daten. In Inhaltsanalysen besteht zwischen Daten und Kategorien immer ein explizit definierter Zusammenhang.

Der Ablauf der Inhaltsanalyse nach Mayring

Mayring entwickelte seine Version der Inhaltsanalyse gegen Ende der 70er-Jahre des vorigen Jahrhunderts in einem Forschungsprojekt, das sich dem Erleben von Arbeitslosigkeit unter Lehrern widmete. In diesem Projekt hatte er 75 qualitative Interviews mit betroffenen Lehrkräften auszuwerten. Bei der Analyse der Interviewprotokolle arbeitete er mit einem Kodierleitfaden. Das ist eine Tabelle mit vier Spalten. In verschiedenen Veröffentlichungen hat Mayring ein längeres Extrakt dieses Rasters publiziert (Mayring/Fenzl 2014, S. 549; Mayring 2000).

Tabelle 24: Ausschnitt aus einem Kodierleitfaden der qualitativen Inhaltsanalyse

1. Kategorien	2. Definition	3. Ankerbeispiele	4. Kodierregeln
K1: hohes Selbstvertrauen	Hohe subjektive Gewissheit, mit der Anforderung gut fertig geworden zu sein, d. h. Klarheit über die Art der Anforderung und deren Bewältigung, positives, hoffnungsvolles Gefühl beim Umgang mit der Anforderung, Überzeugung, die Bewältigung der Anforderung selbst in der Hand gehabt zu haben.	„Sicher hat’s mal ein Problemchen gegeben, aber das wurde dann halt ausgeräumt, entweder von mir die Einsicht, oder vom Schüler, je nachdem, wer den Fehler gemacht hat. Fehler macht ja ein jeder“ (17,23). „Ja klar, Probleme gab’s natürlich, aber zum Schluss hatten wir ein sehr gutes Verhältnis, hatten wir uns zusammengerauft“ (27,33).	Alle drei Aspekte der Definition müssen in Richtung „hoch“ weisen, es soll kein Aspekt auf nur mittleres Selbstvertrauen schließen lassen, sonst Kodierung „mittleres S.“
K2: mittleres Selbstvertrauen	Teilweise oder schwankende Gewissheit, mit der Anforderung gut fertig geworden zu sein.	„Ich hab mich da einigermaßen durchlaviert, aber es war oft eine Gratwanderung“ (3,55). „Mit der Zeit ist es etwas besser geworden, aber ob das an mir oder an den Umständen lag, weiß ich nicht“ (77,20).	Wenn nicht alle drei Definitionsaspekte auf „hoch“ oder „niedrig“ schließen lassen.
K3: niedriges Selbstvertrauen	Überzeugung, mit der Anforderung schlecht fertig geworden zu sein, d. h. wenig Klarheit über die Art der Anforderung, negatives, pessimistisches Gefühl beim Umgang mit der Anforderung, Überzeugung, den Umgang mit der Anforderung nicht selbst in der Hand gehabt zu haben.	„Das hat mein Selbstvertrauen getroffen; da hab ich gemeint, ich bin eine Null – oder ein Minus“ (5,34).	Alle drei Aspekte deuten auf niedriges Selbstvertrauen, auch keine Schwankungen erkennbar.

Quelle: Mayring 2000.

Das Raster für die Kodierung der Texte aus den Interviews besteht aus vier Spalten mit folgenden Funktionen:

1. **Kategorien:** Diese Spalte enthält die Liste der Kategorien der Inhaltsanalyse. In dem Forschungsprojekt zur Lehrerarbeitslosigkeit gab es z. B. die Kategorie „hohes Selbstvertrauen“.
2. **Definition:** In der zweiten Spalte wird zu jeder Kategorie eine festgelegte Definition eingetragen. Zu „hohes Selbstvertrauen“ gehörte: „Hohe subjektive Gewissheit, mit der Anforderung gut fertig geworden zu sein, d. h. Klarheit über die Art der Anforderung und deren Bewältigung, positives, hoffnungsvolles Gefühl beim Umgang mit der Anforderung, Überzeugung, die Bewältigung der Anforderung selbst in der Hand gehabt zu haben.“
3. **Ankerbeispiele:** In der dritten Spalte befinden sich Beispiele zu den Definitionen bzw. Kategorien. Das sind Zitate aus den erhobenen Daten. Zu der Kategorie „hohes Selbstvertrauen“ war dies: „Sicher hat’s mal ein Problemchen gegeben, aber das wurde dann halt ausgeräumt, entweder von mir die Einsicht, oder vom Schüler, je nachdem, wer den Fehler gemacht hat, Fehler macht ja ein jeder.“
4. **Kodierregeln:** Sie beziehen sich auf den Zusammenhang zwischen Ankerbeispielen und Kategorien. Sie dienen als Grundlage der Entscheidung beim Durchgehen der Datenprotokolle, ob Passagen zu der Kategorie passen oder nicht. Im vorliegenden Fall, bei der Kategorie „hohes Selbstvertrauen“, galt die Regel: „Alle drei Aspekte der Definition müssen in Richtung ‚hoch‘ weisen, es soll kein Aspekt auf nur mittleres Selbstvertrauen schließen lassen.“

Dieser Leitfaden fungiert in der qualitativen Inhaltsanalyse als ein analytisches Suchraster, um die Protokolle der Interviews systematisch zu durchdringen. In den transkribierten Datenprotokollen werden entsprechende **Textpassagen** markiert und den Kategorien über Markierungen zugeordnet. Das praktische Vorgehen ist iterativ. Immer wieder werden die Datenprotokolle durchgelesen und die entsprechenden Textstellen mit den zugeordneten Kategorien verglichen. Ziel ist es, alle Textpassagen optimal in Kategorien zu integrieren. Dabei kommt es zu Mehrfachzuordnungen. Eine Textpassage passt zu mehreren Kategorien und umgekehrt. Eine qualitative Textanalyse mit hoher Reliabilität sorgt dafür, dass alle Textpassagen aus allen Interviews der Stichprobe zu 100 % in den Kategorien aufgehen.

Textpassage

Als Textpassage gilt eine kurze oder längere Sequenz von Worten, die eine einheitliche Bedeutung in sich trägt. Eine Textpassage kann in einem Datenprotokoll auch aus einem einzelnen Wort (z. B. „Neiini!“ als Ausdruck einer Emotion oder einer Überraschung) bestehen.

Mayrings Verfahren im Bezug zu anderen Varianten der Inhaltsanalyse

Systematische Verfahren der Inhaltsanalyse stammen aus der amerikanischen Soziologie. Die ersten Anleitungen und Lehrtexte verfassten Soziologen, die in den USA die Inhalte von Massenmedien (Zeitungen und Zeitschriften) in den 40er-Jahren des letzten Jahrhunderts analysierten. Sie mussten große Mengen an Texten verarbeiten. Der Soziologe Bernard Berelson (1912–1979) publizierte 1952 ein Buch, in welchem er seine Arbeitsweise bei der inhaltlichen Analyse von Texten in Zeitungen und Illustrierten beschrieb (Berelson 1952). Er ging quantitativ vor und definierte für die Inhaltsanalyse vorab ein Kategorienraster. In einem zweiten Schritt ordnete er jeder Kategorie typische Worte und Kurzaussagen aus den Texten zu. Auf diese Weise ermittelte Berelson, wie häufig bestimmte Themen in den US-Massenmedien genannt wurden oder nicht.

Der deutsch-amerikanische Soziologe Siegfried Kracauer (1888–1996) ersann als Gegenentwurf zu Berelson das Konzept einer qualitativen Inhaltsanalyse. Kracauer sprach sich gegen die Quantifizierungen von Worten in Texten aus, da so der innere Sinnzusammen-

hang der analysierten Texte verloren gehen würde. Eine rein quantitative Inhaltsanalyse, so wie sie Berelson als Verfahren konzipierte, verharrt aus Kracauers Sicht zwangsläufig an der Oberfläche der analysierten Texte:

Overemphasis on quantification tends to lessen the accuracy of analysis. Content analysis is frequently obliged to isolate and process the more intricate characteristics of a sample; and whenever this happens it runs the risk of treating them inadequately (Kracauer 1952, zitiert nach Lobinger 2012, S. 245).

Mayring bezieht sich sowohl auf Berelson als auch auf Kracauer (vgl. Mayring 2000, Absatz 6). Auch wenn seine Variante den Namen „qualitative Inhaltsanalyse“ trägt, ist eine Ausdehnung auf quantitative Analysen möglich. Anhand des Kodierleitfadens können Anwender Häufigkeitsanalysen durchführen, die die Anzahl der Nennungen von Worten in den Datenprotokollen ermitteln (Mayring 2000, Absatz 26–28).

Im Unterschied zu Mayrings Ansatz verzichtet das Codieren in der Grounded Theory auf die Vorabdefinition eines festen Kodierrasters. Stattdessen sollen die Kategorien induktiv aus den Protokollen von Gesprächen und Beobachtungen gewonnen werden. Das Codieren dient der Entwicklung von Kategorien, die die verschiedenen Daten zusammenfassen. Dazu sehen sich die Untersucher die aus ihrer ganz persönlichen Sicht auffälligen bzw. interessanten Passagen in den Beobachtungsprotokollen an und stellen offene Fragen. Beispielsweise: Was ist das beobachtete Phänomen in der Alltagssprache der Beteiligten? Wo im Untersuchungsfeld tritt es auf? Welche Funktionen hat es? Welche Formen und Facetten hat es? Im späteren Verlauf der Analyse werden dazu passende Textstellen gesucht. Der Begründer des Auswertungsverfahrens in der Grounded Theory hat dazu ein mit vielen Beispielen versehenes, gut verständliches Lehrbuch geschrieben (Strauss 1998). Dieses sei zum Abschluss des Studienbriefs allen Lesern empfohlen, die sich auf qualitative Inhaltsanalysen vorbereiten.



ZUSAMMENFASSUNG

Qualitative und quantitative Auswertungsmethoden sind in der Abfolge ihrer Auswertungsschritte unterschiedlich. Quantitative Auswertungen beginnen mit explorativen Analysen von Verteilungen der gemessenen Variablen (Grafiken und Tabellen). Danach werden deskriptive Kennwerte für Variablen ermittelt (Mittelwerte, Streuungen). Dem schließen sich Untersuchungen von Zusammenhängen zwischen zwei Variablen an (Kreuztabellen, Korrelationen). Danach erfolgt die Anwendung der Inferenzstatistik. Sie prüft anhand theoretischer Verteilungen die Wahrscheinlichkeiten von Kennwerten (z. B. Chi-Quadrat-Verteilung oder t-Verteilung). Nicht erwartete Ergebnisse sind statistisch signifikant.

Multivariate statistische Verfahren sind bei großen Fall- und Variablenzahlen anzuwenden. Sie prüfen entweder die Zusammenhänge zwischen mehreren Variablen auf ihre statistische Signifikanz, wie z. B. die Varianzanalyse, oder sie zielen auf in den Daten verborgene Grundmus-

ter ab, die als eine Erklärung für die empirisch gemessenen Unterschiede zwischen den Fällen fungieren. Ein solches strukturentdeckendes Verfahren ist die Faktorenanalyse, deren Anwendung in der empirischen Psychologie eine lange Tradition hat.

Qualitative Verfahren der Text- und Inhaltsanalyse suchen ebenfalls oft nach gemeinsamen Grundmustern in den untersuchten Fällen. Verschiedene verbale Aussagen der befragten Personen werden zu Kategorien verdichtet. Ein wichtiges Instrument in der qualitativen Inhaltsanalyse nach Mayring ist ein Kategorienraster, das auf die qualitativen Datenprotokolle als Suchraster bezogen wird. Mayrings qualitative Inhaltsanalyse ist ein sehr populäres Verfahren, aber es gibt auch andere Verfahren der qualitativen Datenanalyse, deren Erkenntnisziele in verschiedene Richtungen gehen.

ANHANG

LITERATURVERZEICHNIS

- Backhaus, K. et al. (2018): *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. 15. Auflage, Springer Gabler, Berlin/Heidelberg.
- Baur, N./Blasius, J. (Hrsg.) (2014): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden.
- Baur, F./Krapp, M./Bamberg, G. (2017): *Statistik – eine Einführung für Wirtschafts- und Sozialwissenschaftler*. [E-Book] De Gruyter, Oldenbourg/Berlin/Boston.
- Berelson, S. (1952): *Content Analysis in Communication Research*. Free Press, Glencoe.
- Bierhoff, H.-W. (1999): *Attribution*. In: Asanger, R./Wenninger, G. (Hrsg.): *Handwörterbuch Psychologie*. Beltz, Weinheim, S. 60–66.
- Brosius, F. (2018): *SPSS. Umfassendes Handbuch zu Statistik und Datenanalyse*. 8. Auflage, mitp, Frechen.
- Burkart, T. (2010): *Qualitatives Experiment*. In: Mey, G./Mruck, K. (Hrsg.): *Handbuch Qualitative Forschung in der Psychologie*. Springer VS, Wiesbaden, S. 252–262.
- Burkart, T./Kleining, G./Witt, H. (2010): *Dialogische Introspektion. Ein gruppengestütztes Verfahren zur Erforschung des Erlebens*. Springer VS, Wiesbaden.
- Calaprice, A. (Hrsg.) (2007): *Albert Einstein. Einstein sagt: Zitate, Einfälle, Gedanken*. Piper, München/Heidelberg.
- Cohen, J. (1988): *Statistical Power Analysis for the Behavioral Sciences*. 2. Auflage, L. Erlbaum Associates, Hillsdale.
- DIW – Deutsches Institut für Wirtschaftsforschung (2018): *SOEP Survey Papers. Series A: Survey Instruments (Erhebungsinstrumente). SOEP-Core 2017. Personenfragebogen. Stichprobe A-L3*. (URL: https://www.diw.de/documents/dokumentenarchiv/17/diw_01_c.611290.de/diw_ssp0563.pdf [letzter Zugriff: 17.05.2019]).
- Döring, N. (2014): *Evaluationsforschung*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 167–181.
- Döring, N./Bortz, J. (2016): *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. 6. Auflage, Springer, Berlin/Heidelberg.
- Downey, A. (2016): *There is still only one test*. (URL: <http://alldowney.blogspot.de/2016/06/there-is-still-only-one-test.html> [letzter Zugriff: 17.05.2019]).

- Dresing, T./Pehl, T. (2018): *Praxisbuch Interview, Transkription & Analyse. Anleitungen und Regelsysteme für qualitativ Forschende*. 8. Auflage, Eigenverlag, Marburg. (URL: <https://www.audiotranskription.de/Praxisbuch-Transkription.pdf> [letzter Zugriff: 17.05.2019]).
- Ebbinghaus, H. (1885): *Über das Gedächtnis*. Unveränderter Neudruck 2016. Hansebooks, Norderstedt.
- Eifler, S. (2014): *Experiment*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 195–209.
- Ernst, H. (2014): *Psychologie ist sich selbst und andere besser verstehen. Heiko Ernst im Gespräch mit Martin Tschechne anlässlich seines Abschieds als Chefredakteur*. In: *Psychologie Heute*, 41. Jg., Heft 10, S. 44-48.
- Erpenbeck, J. et al. (Hrsg.) (2017): *Handbuch Kompetenzmessung. Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis*. 3. Auflage, Schäffer-Poeschel, Stuttgart.
- Fischer, A. (2004): *Frieder Nögge: Clown in vier Temperamenten*. In: *Erziehungskunst*. Heft 7/8, S. 836–840.
- Flick, U. (2014): *Gütekriterien qualitativer Sozialforschung*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 411–424.
- Freud, S. (1900): *Die Traumdeutung*. Deuticke, Leipzig/Wien.
- Friedrichs, J. (1990): *Methoden empirischer Sozialforschung*. 14. Auflage, Westdeutscher Verlag, Opladen.
- Friedrichs, J. (2014): *Forschungsethik*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 81–91.
- Funke, J. (1996): *Methoden der kognitiven Psychologie*. (URL: https://www.psychologie.uni-heidelberg.de/ae/allg/mitarb/jf/meth_kogn.html [letzter Zugriff: 17.05.2019]). Aktualisierte Fassung des Originals in: Erdfelder, E. et al. (Hrsg.) (1996): *Handbuch Quantitative Methoden*. Psychologie Verlags Union, Weinheim, S. 515–528.
- García, I. (2016): *Ich kann auch anders. Von freundlich bis unbarmherzig – wie Sie das Repertoire Ihrer Kommunikationsmuster wirksam erweitern*. Berlin, Econ.
- Geiger, S. M./Grossman, P./Schrader, U. (2019): *Mindfulness and Sustainability: Correlation or Causation?* In: *Current Opinion in Psychology*, Heft 28, S. 23–27
- Glaser, B. G./Strauss, A. L. (1974): *Interaktion mit Sterbenden : Beobachtungen für Ärzte, Schwestern, Seelsorger und Angehörige*. Vandenhoeck & Ruprecht, Göttingen.

- Goethe, J. W. (1989): *Der Versuch als Vermittler zwischen einem Subjekt und einem Objekt*. In: Richter, K. (Hrsg.): *Sämtliche Werke nach Epochen seines Schaffens. Band 12. Zur Naturwissenschaft überhaupt, besonders zur Morphologie*. Hanser, München S. 684–693.
- Goffman, E. (1973): *Asyle. Über die soziale Situation psychiatrischer Patienten und anderer Insassen*. Suhrkamp, Frankfurt am Main.
- Gramann, K./Schandry, R. (2009): *Psychophysiologie. Körperliche Indikatoren psychischen Geschehens*. 4. Auflage, Beltz, Weinheim/Basel.
- Groeben, N. et al. (1988): *Das Forschungsprogramm Subjektive Theorien: eine Einführung in die Psychologie des reflexiven Subjekts*. (URL: <https://www.ssoar.info/ssoar/handle/document/2765> [letzter Zugriff: 17.05.2019]).
- Hasselhorn, M./Gold, A. (2017): *Pädagogische Psychologie. Erfolgreiches Lernen und Lehren*. 3. Auflage, Kohlhammer, Köln.
- Helfferrich, C. (2014): *Leitfaden- und Experteninterviews*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 559–574.
- Hossiep, R./Schulz, R. (2013): *Persönlichkeitsinventare, -strukturtests*. In: Frey, A./Lissmann, U./Schwarz, B. (Hrsg.): *Handbuch Berufspädagogische Diagnostik*. Beltz, Weinheim, S. 196–215.
- Hussy, W./Schreier, M./Echterhoff, G. (2013): *Forschungsmethoden in Psychologie und Sozialwissenschaften*. 2. Auflage, Springer, Berlin/Heidelberg.
- Kleining, G. (1982): *Umriss zu einer Methodologie qualitativer Sozialforschung*. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 34. Jg., S. 224–253. (URL: <http://www.heuristik-hamburg.net/UmrissaufsatzKleining.PDF> [letzter Zugriff: 17.05.2019]).
- Knoblauch, G. (2014): *Ethnographie*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 521–528.
- Kopp, J./Lois, D. (2014): *Sozialwissenschaftliche Datenanalyse. Eine Einführung*. 2. Auflage, Springer VS, Wiesbaden.
- Kuckartz, U./Rädiker, S. (2014): *Datenaufbereitung und Datenbereinigung in der qualitativen Sozialforschung*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 383–396.
- Kurt, R./Herbik, R. (2014): *Sozialwissenschaftliche Hermeneutik und hermeneutische Wissenssoziologie*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 473–491.

- Lawson, R. (2006): *The science of cycology: Failures to understand how everyday objects work*. In: *Memory & Cognition*, 34. Jg., Heft 8, S. 1667–1675.
- Lienert, G. A./Raatz, H. (1998): *Testaufbau und Testanalyse*. 6. Auflage, Beltz/Psychologie Verlags Union, Weinheim.
- Lobinger, K. (2012): *Visuelle Kommunikationsforschung. Medienbilder als Herausforderung für die Kommunikations- und Medienwissenschaft*. Springer VS, Wiesbaden.
- Lück, D./Landrock, U. (2014): *Datenaufbereitung und Datenbereinigung in der quantitativen Sozialforschung*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 397–409.
- Maderthaner, R. (2017): *Psychologie*. 2. Auflage, facultas, Wien.
- Mausfeld, R. (2005): *Über die Bedingungen der Möglichkeit von Lernen*. In: Käsermann, M.-L./Altörfer, A. (Hrsg.): *Über Lernen. Ein Gedankenaustausch. Festschrift für Klaus Foppa zum 75. Geburtstag*. Edition Solo, Bern, S. 218–236. (URL: http://www.uni-kiel.de/psychologie/mausfeld/pubs/Voraussetzungen_von_Lernen.pdf [letzter Zugriff: 17.05.2019]).
- Mayring, P. (2000): *Qualitative Inhaltsanalyse*. In: *Forum Qualitative Sozialforschung*, 1. Jg, Heft 2. (URL: <http://www.qualitative-research.net/index.php/fqs/article/view/1089/2383> [letzter Zugriff: 17.05.2019]).
- Mayring, P. (2010): *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Beltz, Weinheim/Basel.
- Mayring, P. (2015): *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 12. Auflage, Beltz, Weinheim/Basel.
- Mayring, P./Fenzl, T. (2014): *Qualitative Inhaltsanalyse*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 543–556.
- Mayring, P./Gahleitner, S. (2010): *Qualitative Inhaltsanalyse*. In: Bock, K./Miethe, I. (Hrsg.): *Handbuch Qualitative Methoden in der Sozialen Arbeit*. Barbara Budrich, Opladen/Farmington Hills, S. 295–304.
- Medjedović, I. (2014): *Qualitative Daten für die Sekundäranalyse*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 223–232.
- Métraux, A. (1985): *Der Methodenstreit und die Amerikanisierung der Psychologie in der Bundesrepublik 1950–1970*. In: Ash, M./Geuter, U. (Hrsg.): *Geschichte der deutschen Psychologie im 20. Jahrhundert. Ein Überblick*. Westdeutscher Verlag, Wiesbaden, S. 225–251.

- Meyer, C./Meier zu Verl, C. (2014): *Ergebnispräsentation in der qualitativen Forschung*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 245–257.
- Meyer-Wellmann, J. (2017): *Nabu Hamburg räumt Fehler bei Volksinitiative ein*. In: *Hamburger Abendblatt* vom 11.12.2017. (URL: <https://www.abendblatt.de/incoming/article212804817/Nabu-raeumt-Fehler-bei-Volksinitiative-zum-Gruenschutz-ein.html>) [letzter Zugriff: 17.05.2019]).
- Meyer-Wellmann, J. (2019): *Die Angst von Rot-Grün vor den Naturschützern*. In: *Hamburger Abendblatt* vom 09.03.2019. (URL: <https://www.abendblatt.de/hamburg/article216624673/Die-Angst-von-Rot-Gruen-vor-den-Naturschuetzern.html>) [letzter Zugriff: 17.05.2019]).
- Mühlichen, A. (2014): *Informationelle Selbstbestimmung*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 93–102.
- Nachtigall, C./Wirtz, M. (2013): *Wahrscheinlichkeitsrechnung und Inferenzstatistik. Statistische Methoden für Psychologen, Teil 2*. 6. Auflage, Beltz Juventa, Weinheim/Basel.
- Paulus, J. (2017): *Versuch und Irrtum*. In: *Psychologie Heute*, 44. Jg., Heft 2, S. 72–76.
- Pohl, R. (2007): *Das autobiografische Gedächtnis. Die Psychologie unserer Lebensspanne*. Kohlhammer, Stuttgart.
- Reinecke, J. (2014): *Grundlagen der standardisierten Befragung*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 601–617.
- Rindermann, H. (2003): *Lehrevaluation an Hochschulen. Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und seine Evaluation*. In: *Zeitschrift für Evaluation*, 2. Jg., Heft 2, S. 233–256.
- Ryan, R. M./Deci, E. L. (2000): *Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development and Well-Being*. In: *American Psychologist*, 55. Jg., Heft 1, S. 66–78.
- Sauer, S. (2019): *Moderne Datenanalyse mit R: Daten einlesen, aufbereiten, visualisieren, modellieren und kommunizieren*. Springer Gabler, Wiesbaden.
- Schäfer, T. (2016): *Methodenlehre und Statistik. Einführung in Datenerhebung, deskriptive Statistik und Inferenzstatistik*. Springer, Wiesbaden.
- Scheele, B./Groeben, N. (2010): *Dialog-Konsens-Methoden*. In: Mey, G./Mruck, K. (Hrsg.): *Handbuch Qualitative Forschung in der Psychologie*. Springer VS, Wiesbaden, S. 506–523.

- Schreier, M. (2013a): *Qualitative Forschungsmethoden*. In: Hussy, W./Schreier, M./Echterhoff, G.: *Forschungsmethoden in Psychologie und Sozialwissenschaften*. 2. Auflage, Springer, Berlin/Heidelberg, S. 189–221.
- Schreier, M. (2013b): *Bewertung qualitativer Forschung*. In: Hussy, W./Schreier, M./Echterhoff, G.: *Forschungsmethoden in Psychologie und Sozialwissenschaften*. 2. Auflage, Springer, Berlin/Heidelberg, S. 274–286.
- Schreier, M. (2013c): *Qualitative Analyseverfahren*. In: Hussy, W./Schreier, M./Echterhoff, G.: *Forschungsmethoden in Psychologie und Sozialwissenschaften*. 2. Auflage, Springer, Berlin/Heidelberg, S. 245–275.
- Schreier, M. (2014): *Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten*. In: *Forum Qualitative Sozialforschung*, 14. Jg., Heft 1. (URL: <http://www.qualitative-research.net/index.php/fqs/article/view/2043> [letzter Zugriff: 17.05.2019]).
- Schröder, A. (2019): *Ich kann auch anders! Wie man seinen Willen stärkt*. In: *ZEIT WISSEN*, 5. Jg., Heft 2, S. 18–26.
- Schurz, G. (2014): *Einführung in die Wissenschaftstheorie*. 4. Auflage, Wissenschaftliche Buchgesellschaft, Darmstadt.
- Schweizerisches Bundesamt für Statistik (2016): *Statistik der natürlichen Bevölkerungsbewegung*. (URL: <https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken/grafiken.assetdetail.212271.html> [letzter Zugriff: 17.05.2019]).
- Strauss, A. (1998): *Grundlagen qualitativer Sozialforschung: Datenanalyse und Theoriebildung in der empirischen soziologischen Forschung*. 2. Auflage, Fink, München.
- Strübing, J. (2014): *Grounded Theory und Theoretical Sampling*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 457–472.
- Tergan, S. O. (1988): *Qualitative Wissensdiagnose. Methodologische Grundlagen*. In: Mandl, H./Spada H. (Hrsg.): *Wissenspsychologie*. Psychologie Verlags Union, München, S. 400–422.
- Thierbach, C./Petschick, G. (2014): *Beobachtung*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 855–866.
- Vogl, S. (2014): *Gruppendiskussion*. In: Baur, N./Blasius, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Springer VS, Wiesbaden, S. 581–586.
- Vollmers, B. (1992): *Kreatives Experimentieren. Die Methodik von Jean Piaget, den Gestaltpsychologen und der Würzburger Schule*. Deutscher-Universitäts-Verlag, Wiesbaden.

- Vollmers, B. (2009): *Das Rollenspiel Verkauf. Ein Instrument der pädagogischen Diagnostik sozialer Kompetenzen*. In: *Wirtschaft und Berufserziehung*, 61. Jg., Heft 10, S. 18–23.
- Vollmers, B. (2015): *The Challenge of Assessing Social Competencies in VET. A Review of all Actual Research Practices in German-speaking Countries*. In: Gessler, M./Freund, L. (Hrsg.): *Crossing Boundaries in Vocational Education and Training: Innovative Concepts for the 21st Century. Conference Proceedings. Evaluate Europe Handbook Series*, University of Bremen, S. 232–236. (URL: https://vetnetsite.files.wordpress.com/2015/10/crossing_boundaries_2015.pdf [letzter Zugriff: 17.05.2019]).
- Vollmers, B./Kindervater, A. (2009): *Evaluation der betrieblichen Praxis im Modellversuch VAmB: Kompetenzmessungen bei Auszubildenden mit Lernschwierigkeiten*. In: *bwp@Berufs- und Wirtschaftspädagogik – online*, Ausgabe 17, 1–23. (URL: http://www.bwpat.de/ausgabe17/vollmers_kindervater_bwpat17.pdf [letzter Zugriff: 17.05.2019]).
- Vollmers, B./Kindervater, A. (2010): *Sozialkompetenzen in simulierten Berufssituationen von Auszubildenden mit Lernschwierigkeiten: Ein empirischer Vergleich von Beobachterurteilen und Selbsteinschätzungen im Modellversuch VAmB*. In: *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 106. Jg., Heft 4, S. 517–533.
- Wallraff, G. (1985): *Ganz unten*. Kiepenheuer & Witsch, Köln.
- Westermann, R. (2016): *Methoden psychologischer Forschung und Evaluation*. Kohlhammer, Stuttgart.
- Wetzel, S. (2019): *Leben achtsame Menschen nachhaltiger?* In: *Psychologie Heute*, 46. Jg., Heft 2, S. 6.
- Wirtz, M./Nachtigall, C. (2012): *Deskriptive Statistik. Statistische Methoden für Psychologen Teil 1*. 6. Auflage, Beltz Juventa, Weinheim/Basel.
- Wollschläger, D. (2017): *Grundlagen der Datenanalyse mit R: Eine anwendungsorientierte Einführung*. 4. Auflage, Springer, Wiesbaden.

ABBILDUNGSVERZEICHNIS

Abbildung 1: Verknüpfung von Attributionstheorie, Selbstkonzept und Anreiztheorien zur Erklärung der Auswahl von Lernaufgaben von misserfolgsorientierten Personen	17
Abbildung 2: Multidimensionales Modell des Lehrerfolgs an Hochschulen	18
Tabelle 1: Unterschiede zwischen qualitativer und quantitativer Forschung	21
Tabelle 2: Aus der SDT abgeleitete Indikatoren für die empirische Forschung	25
Abbildung 3: Quantitativer Fragebogen zur Studienmotivation basierend auf der SDT ..	26
Abbildung 4: Vierstufige LIKERT-Skala	52
Abbildung 5: Beispiel einer Ja/Nein-Zuordnung	52
Abbildung 6: Beispiel einer Ja/Nein-Zuordnung mit neutraler Option	53
Abbildung 7: Beispiel: Wie oft in der Woche trinken Sie Alkohol?	53
Abbildung 8: Beobachtungsbogen im Rollenspiel Verkauf	61
Tabelle 3: Fiktiver Versuchsplan in Medizin oder Psychologie mit drei Experimentalgruppen (Therapieformen) und einer Kontrollgruppe (keine Therapie)	63
Abbildung 9: Vorlage für die Fahrradskizzen der Versuchspersonen von Rebecca Lawson	64
Tabelle 4: Numerische Ergebnisse der Fehler des Fahrradexperiments	64
Abbildung 10: Typische Fehler in den Fahrradzeichnungen der Versuchspersonen, unterschieden nach deren Geschlecht	65
Abbildung 11: Neun-Punkte-Problem als Denkaufgabe	69
Abbildung 12: Mögliche Lösung des Neun-Punkte-Problems	69
Abbildung 13: Beispiele von Beziehungskarten in der STL	70
Abbildung 14: Mögliches Vorwissen von Menschen zu Vulkanen, visualisiert mittels STL	71
Abbildung 15: Mögliches Wissensdiagramm nach Lernphase zum Thema Vulkane	72

Abbildung 16: Mögliche Daten in SPSS aus der Befragung zur Studienmotivation	80
Abbildung 17: Datentabelle in SPSS für das Beispiel zur Studienmotivation	84
Abbildung 18: Variablenansicht in SPSS im Datensatz zur Studienmotivation	85
Abbildung 19: Auswertungsschritte in der deskriptiven bzw. explorativen Statistik	95
Abbildung 20: SPSS-Befehle für Balkendiagramme und Histogramme	96
Abbildung 21: Histogramm einer siebenstufigen LIKERT-Skala	97
Abbildung 22: IQ-Normalverteilung und Standardnormalverteilung	98
Abbildung 23: Balkendiagramm: Häufigkeitsverteilung des Gestationsalters (Schwangerschaftswochen bei Entbindung)	99
Tabelle 5: Übersicht zu den Kennwerten der deskriptiven bzw. explorativen Statistik ...	99
Tabelle 6	100
Abbildung 24: Erster SPSS-Befehl für die Maße der zentralen Tendenz und für Streuungsmaße	101
Abbildung 25: Zweiter SPSS-Befehl für die Maße der zentralen Tendenz und für Streuungsmaße	101
Tabelle 7: Skalenniveaus und jeweils anwendbare Maße der zentralen Tendenz	102
Tabelle 8	102
Tabelle 9: Übersicht zu Skalenniveaus und den jeweils möglichen Streuungsmaßen ..	103
Tabelle 10: Auswahl gängiger Zusammenhangsmaße in Abhängigkeit vom Skalenniveau der Merkmale	104
Tabelle 11: Liste der zehn Fälle aus dem Beispiel Studienmotivation für Alter und Semesterzahl	105
Abbildung 26: Streudiagramm der Variablen Semesterzahl (X) und Alter (Y)	106
Abbildung 27: SPSS-Befehl für den Korrelationskoeffizienten von Pearson und Bravais	107
Tabelle 12: Berechnungsschritte der Korrelation für die Variablen Semesteranzahl und Alter	107

Tabelle 13: Fiktive Vierfeldertafel zum Zustand der behandelten Patienten nach Rückenoperation und konventioneller Behandlung ohne Operation (n = 400)	110
Abbildung 28: Übersicht: Erkenntnisschritte in der Inferenzstatistik	111
Abbildung 29: Schema des Testens von Nullhypothesen in der Inferenzstatistik	112
Tabelle 14: Fiktive Vierfelder-Häufigkeitsverteilung von zwei Gruppen und zwei Behandlungsformen	113
Tabelle 15: Theoretisch zu erwartende Vierfelder-Häufigkeitsverteilung der dichotomen Variablen Behandlungserfolg bei zwei Behandlungsformen	114
Abbildung 30: SPSS-Befehl für eine Mehrfelder-Tafel mit Chi-Quadrat-Test	116
Tabelle 16: Verteilung der Gruppierungsvariablen Studierendenstatus im Beispiel Studienmotivation (n = 10)	117
Abbildung 31: SPSS-Befehl zum t-Test für zwei unabhängige Gruppen	118
Tabelle 17: Mittelwerte und Streuungen der Variablen Alter gruppiert nach Studierendenstatus	118
Tabelle 18: SPSS-Ergebnis des t-Tests für die Mittelwertdifferenzen der Variablen Alter im Studierendenstatus	120
Abbildung 32: Überlappungsgrad zweier Verteilungen für variierende Werte des Koeffizienten d	122
Tabelle 19: Übersicht zu den multivariaten Verfahren	124
Tabelle 20: Fiktive Verkaufszahlen von fünf Produkten in drei Filialen	125
Tabelle 21: Varianzanalyse: Berechnung der Quadratsummen in drei Filialen (m_{ss_w}) ..	127
Abbildung 33: SPSS-Befehl zur Durchführung der einfaktoriellen Varianzanalyse	128
Abbildung 34: SPSS-Befehl zur Durchführung der Faktorenanalyse	130
Abbildung 35: Interkorrelationsmatrix in SPSS der Variablen zur Studienmotivation ..	131
Tabelle 22: SPSS-Ergebnis der Hauptkomponentenanalyse für das Beispiel Studienmotivation	132
Tabelle 23: Komponentenmatrix in SPSS für das Beispiel Studienmotivation	133

Tabelle 24: Ausschnitt aus einem Kodierleitfaden der qualitativen Inhaltsanalyse 138



IU Internationale Hochschule GmbH
IU International University of Applied Sciences
Juri-Gagarin-Ring 152
D-99084 Erfurt



Postanschrift
Albert-Proeller-Straße 15-19
D-86675 Buchdorf



media@iu.org
www.iu.org



Hilfe & Kontakt (FAQ)
Antworten rund um Dein Studium findest
Du jederzeit auf myCampus.