

Teil IV: Analyse von Daten

In diesem Teil IV wird auf die Analyse von Daten eingegangen. Wie bereits in Kapitel 5.1 „Forschungszugänge“ und Kapitel 5.3 „Forschungsdesigns“ ausgeführt wurde, hängt die Art von (erhobenen) Daten und somit die Analyse dieser Daten von der Fragestellung, der gewählten Methode sowie dem Forschungsinstrument ab. Dabei kann zwischen qualitativen und quantitativen Daten, Analysen und Ergebnissen unterschieden werden. Qualitative und quantitative Forschungsansätze lassen sich nicht immer streng trennen, vielmehr gibt es häufig Kombinationen beider Forschungszugänge: Mixed Methods-Ansätze kombinieren resp. integrieren gezielt und systematisch qualitative und quantitative Forschungszugänge, um Forschungsfragen zu beantworten (siehe Kap. 5.1.1 „Qualitative und quantitative Forschung“). Dies bedeutet, dass in ein und demselben Forschungsprojekt oft sowohl quantitative als auch qualitative Daten vorhanden sind und weiterverarbeitet werden. In diesem Teil IV sind qualitative und quantitative Ansätze zur Analyse von Daten aufgeführt:

In Kapitel 12 „Auswertung qualitativer Daten“ wird auf drei verschiedene Ansätze der qualitativen Datenauswertung eingegangen: Qualitative Inhaltsanalyse, Grounded Theory sowie Objektive Hermeneutik. Zu jedem Verfahren werden die spezifischen Zielsetzungen, die theoretischen Hintergründe und das methodische Vorgehen (die „Techniken“) beschrieben sowie eine kritische Würdigung der Verfahren vorgenommen.

In Kapitel 13 „Auswertung quantitativer Daten“ wird erläutert, wie quantitative Daten ausgewertet werden. Dabei wird zuerst genauer auf den Messvorgang und die Datenaufbereitung eingegangen. Danach wird auf die Deskriptive (Beschreibende) Statistik eingegangen, bevor prüfstatistische (hypothesenprüfende) Verfahren der Datenanalyse vorgestellt werden.

12 Auswertung Qualitativer Daten

von Christina Huber und Lukas Lehmann

Qualitative Forschungsmethoden sind ein wichtiger Bestandteil der empirischen Forschung. Im Gegensatz zu quantitativen Methoden, die sich auf numerische Daten und Statistiken konzentrieren, zielt die qualitative Forschung darauf ab, tiefere Einblicke in menschliches Verhalten, Erfahrungen und Perspektiven zu gewinnen. Entsprechend gibt es mehrere Gründe, warum man sich für qualitative Methoden in der Forschung entscheidet. Zunächst einmal ermöglichen sie eine umfassende Analyse komplexer Sachverhalte, die schwierig zu quantifizieren sind. Bezüglich Datengrundlagen gilt der Grundsatz „all is data“ (Glaser, 2007, zit. nach Przyborski & Wohlrab-Sahr, 2014, S. 197), d. h. es können auch Daten, die bereits vorhanden sind und nicht eigens für eine Forschungsarbeit generiert werden müssen (sog. „natürliche“ Daten), ausgewertet werden. So können qualitative Methoden einfach und problemlos beispielsweise auch auf Unterrichtsmaterialien, SchülerInnen- und Schülerprodukte, auf administrative Dokumente und Rechtsgrundlagen, auf Internetseiten oder allgemein auch auf Beobachtungen oder gar auf Filmmaterial angewandt werden.

Ein Vorteil der Verwendung qualitativer Methoden ist ihre Flexibilität und Anpassungsfähigkeit an unterschiedliche Forschungsfragen, -ziele und -gegenstände. Darüber hinaus ermöglichen die in qualitativen Settings üblichen wenig oder nicht standardisierten Methoden, d. h. Vorgehensweisen, bei welchen Fragen und mögliche Antworten nicht im Vorhinein vorgegeben sind, den Forscherinnen und Forschern, die Perspektiven und Erklärungen der Forschungsteilnehmenden („Forschungsobjekte“) selbst einzubeziehen. Dies kann dazu beitragen, eine umfassendere und vielfältigere Sicht auf ein gewähltes Thema zu erhalten. Insgesamt bieten qualitative Methoden daher eine wertvolle Ergänzung zu quantitativen Methoden und tragen dazu bei, ein tieferes Verständnis komplexer sozialer Phänomene zu gewinnen.

In den Sozialwissenschaften existiert eine Vielzahl von unterschiedlichen qualitativen Auswertungsverfahren. Sie beziehen sich auf verschiedene Forschungs- und Theorietraditionen und haben unterschiedliche Ansprüche in Bezug auf die Geltung und Verallgemeinerbarkeit ihrer Resultate. Diese Vielfalt stellt gerade für Forschungsnovizinnen und -novizen eine wesentliche Eintrittshürde dar, da die unterschiedlichen „Methodenschulen“ nicht im Sinne von einfachen und komplexen Verfahren aufeinander aufbauen. Um den Einstieg zu erleichtern, wird in der folgenden Einleitung versucht, einige Unterschiede, aber auch Gemeinsamkeiten darzulegen. In einem weiteren Teil werden dann drei Verfahren detaillierter präsentiert und anhand von Beispielen erläutert.

12.1 Überblick

Allgemein und etwas plakativ ausgedrückt, besteht das Ziel der meisten qualitativen Auswertungsverfahren darin, soziale Wirklichkeit „von innen heraus“ (Flick et al., 2007, S. 14) zu beschreiben und zu verstehen. Wer qualitativ forscht, interessiert sich für die individuell-subjektiven Sichtweisen und das persönliche Wissen von Menschen. In der qualitativen Forschung werden deshalb keine vorformulierten Hypothesen getestet, sondern die Analysen folgen einer (relativ offenen) Fragestellung („Erkenntnisinteresse“) und den darauffolgenden Aussagen (von interviewten Personen oder anderen Text-, Ton- und Bildformen). Aus diesem Zusammenspiel von offener Frage und darauffolgenden Antworten wird versucht die entsprechenden Erkenntnisse abzuleiten. Im Unterschied zu quantitativ-statistischen Vorgehensweisen werden Zusammenhänge also nicht überprüft (deduktives Vorgehen), sondern sie werden aus dem Material heraus erarbeitet (induktives Vorgehen).

Prinzip der Offenheit. Das induktive Vorgehen verlangt von den Forschenden eine grosse Offenheit. Um diese Offenheit einzulösen, ist es zweckmässig, gegenüber dem Forschungsgegenstand (einer Person oder eines Sachverhalts) eine Grundhaltung des „Fremden“ (A. Schütz) einzunehmen, beobachtete Phänomene also konsequent und systematisch zu hinterfragen. Dies ist insbesondere dann wichtig, wenn man scheinbar klare, alltägliche und vermeintlich bekannte Sachverhalte untersucht.

Dahinter steckt eine Grundannahme: Auch das vermeintlich Alltägliche – etwa das routinierte Überqueren einer Strasse oder die beiläufige Begrüssung anwesender Personen beim Betreten eines Raumes – musste ursprünglich einmal gelernt und damit begründet werden. Diese Begründung einer Handlung ist durch die Routine zwar nicht mehr präsent – ich grüsse Freundinnen und Freunde herzlich, wenn ich ihnen begegne –, sie ist aber immer noch vorhanden und kann bei einer Nachfrage auch ausformuliert werden: „Ich grüsse meine Freundinnen und Freunde herzlich, um ihnen meine Aufmerksamkeit, meine Anerkennung als Freundinnen und Freunde dadurch kenntlich zu machen.“ So wird vieles, das im Prozess des Aufwachsens explizit gelernt wurde, später nur noch als automatisierte Routine angewendet. Solche Routinen – das können Handlungen sein, aber auch persönliche Überzeugungen – fassen auf einer ursprünglich bewusst getroffenen Entscheidung, etwas so und nicht anders zu deuten, zu bewerten oder auszuführen.

So verstanden hat jede Handlung also einen spezifischen Sinn. Diesen Sinn gilt es im Zuge des Analyseprozesses sichtbar zu machen. Mit qualitativen Zugängen kann man so aus vermeintlich Alltäglichem spannende und oft unvorhergesehene Erkenntnisse an den Tag bringen resp. Unbewusstes sichtbar und damit bewusst machen.

Das Prinzip der Offenheit in Bezug auf die zu analysierenden Daten meint keineswegs, dass qualitativ Forschende ohne Vorarbeit und Vorwissen an ein For-

schungsvorhaben herangehen. Im Gegenteil: Vorwissen hilft, die Aufmerksamkeit im Forschungsprozess auf bestimmte Fragen zu lenken, die Forschenden auf die Fülle möglicher Antworten vorzubereiten und für die genaue Analyse zu sensibilisieren. Vorwissen hilft auch, die mitunter grosse Fülle von Daten (z. B. in längeren Interviews) sinnvoll zu strukturieren.

Verstehen der sozialen Wirklichkeit als Forschungsziel. In der qualitativen Forschung geht es um das Interpretieren und das Verstehen sozialer Wirklichkeit, kurz: um die Rekonstruktion von Sinn. Was sich auf den ersten Blick kompliziert anhört, basiert auf einer einfachen Grundannahme (vgl. Abbildung 12-1): Während Tiere nach einem Reiz-Reaktion-Schema funktionieren, handeln Menschen stets auf der Basis von Interpretationen. Dem menschlichen Handeln geht dabei sowohl ein Interpretieren der jeweiligen Handlungssituation voraus als auch der möglichen Handlungsoptionen. Um beim obigen Beispiel zu bleiben: Das herzliche Grüssen von Freundinnen und Freunden setzt voraus, dass die Person eine Vorstellung – ein Konzept – von Freundschaft hat (das sich von anderen Konzepten, etwa die Begrüssung von Vorgesetzten, unterscheidet) und die Begrüssungsbewegung (z. B. Umarmung, Wangenküsse) entsprechend deuten und von anderen Gesten (z. B. einem Angriff) unterscheiden und dann ihrerseits angemessen darauf reagieren kann. Nun kommt hinzu, dass die Welt nicht statisch ist, sondern sie verändert sich mit der Zeit, mit dem Raum und mit den Personen, die in ihr agieren. Begrüssungsrituale verändern sich je nach Kontext und Ort, Neues kommt dazu oder wandelt sich. Bis zu welchem Punkt Personen noch bereit oder fähig sind, die manchmal sehr abstrakt ausfallenden Ideen richtig zu interpretieren, hängt sowohl vom Hintergrund der Personen selbst, aber auch von der Zeit und der Lebenswelt ab, in der sie sich befinden.

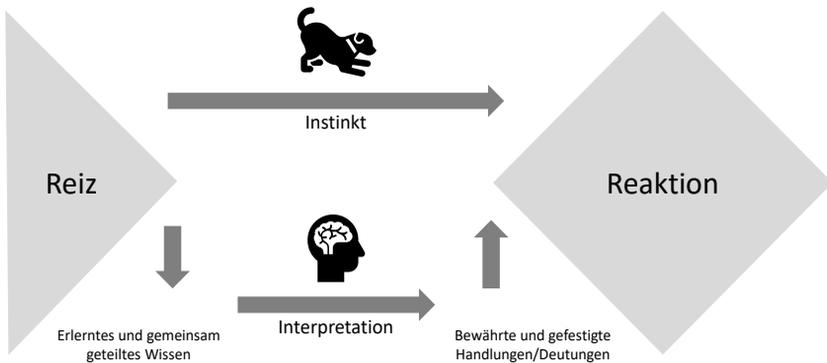


Abbildung 12-1: Ebenen des Verstehens: Reiz-(Interpretation)-Reaktion (eigene Darstellung)

Diese gesellschaftlichen Interpretationen und Sinngebungen in unterschiedlichen Kontexten versucht die qualitative Sozialforschung zu rekonstruieren (sichtbar und bewusst zu machen) und zu verstehen. Für die notwendige Grundhaltung gilt demnach:

Menschen, die immer schon wissen, was Sache ist, wo es lang geht und was als Nächstes zu tun ist, eignen sich für vieles – für die hermeneutische [i. e.: verstehende] Arbeit bringen sie allerdings keine guten Voraussetzungen mit. „!-Menschen“ haben in der Hermeneutik [i. e. verstehenden, qualitativen Forschung] nichts zu suchen (weil sie dort nichts finden können). Für „?-Menschen“ sind die Aussichten in der Hermeneutik ungleich besser. (Kurt, 2004, S. 31, zit. nach Breuer, 2010, S. 47).

Verstehen meint also nicht, dass wir als „!-Menschen“ durch die Welt gehen, sie mit Begriffen etikettieren und soziale Erscheinungen unter vorgefertigte Theorien subsumieren. Verstehen verlangt vielmehr, dass wir der Welt konsequent als „?-Menschen“ begegnen und dabei gesellschaftliche Ereignisse nicht einfach nur beobachten und in (mehr oder weniger plausible) Zusammenhänge stellen, sondern versuchen, den Sinn hinter diesen Ereignissen sichtbar zu machen. Dies korrespondiert in hohem Masse mit der geforderten Offenheit, denn es geht nachgerade nicht darum, Theorien zu beweisen, sondern interessierende Untersuchungsbereiche zu erforschen und im Forschungsprozess zu erkennen, was in diesem Bereich relevant ist (Strauss & Corbin, 1996, S. 8).

Dementsprechend werden qualitative Ansätze oft für die erforschende (explorative) Erschließung eines Forschungsfeldes eingesetzt. Denn sie erlauben Aussagen über Zusammenhänge oder über die Beschaffenheit eines Untersuchungsbereichs (Bsp.: welche Muster professionellen Handelns kommen bei Lehrpersonen vor oder welche Typen von Lehrpersonen gibt es?). Nicht zu qualitativen Ansätzen passend sind jedoch Aussagen über die Verteilungen von Ausprägungen (Bsp.: Wie viele Lehrpersonen gehören dem Typus „Unterrichtsvorbereitung in letzter Minute“ an?). Typische Fragestellungen beinhalten dementsprechend die Sichtweise von Personen auf spezifische Sachverhalte (Bsp.: „Wie planen Lehrpersonen ihren Unterricht und wie begründen sie ihre Planungsentscheidungen?“), das Verstehen einer Situation (Bsp.: „Wodurch zeichnet sich das Verhältnis von Lehrpersonen und Schülerinnen und Schülern aus?“) oder einer sozialen Ordnung (Bsp.: „Welche gesellschaftlichen Aufgaben soll die Schule gemäss der in ihr tätigen Fachpersonen erfüllen?“) (vgl. Flick et al., 2007, S. 18ff).

Das Spektrum qualitativer Verfahren. Qualitative Verfahren der Datengewinnung und Datenauswertung unterscheiden sich vor allem darin, unter welchen Voraussetzungen und Vorannahmen die Rekonstruktion und Interpretation sozialer Wirklichkeit zu erfolgen hat. Entsprechend herrscht auch innerhalb der qualitativen Ansätze gelegentlich Streit über das „richtige“ Vorgehen oder den „richtigen“ theoretisch-methodischen Zugang. Dies zeigt sich mitunter auch darin, dass für qualitative

Verfahren sehr verschiedene Bezeichnungen geführt werden: Die Rede ist u. a. von rekonstruktiven (gemäss Ralf Bohnsack) oder kommunikativen (Fritz Schütze) Verfahren, von sozialwissenschaftlicher oder wissenssoziologischer Hermeneutik (Hans-Georg Soeffner) oder von interpretativer Sozialforschung (Ronald Hitzler). Im Grundsatz lassen sich drei Verfahrensweisen qualitativer Forschung unterscheiden: inhaltsanalytische, kodierende und hermeneutische Verfahren (vgl. Tabelle 12-1).

Tabelle 12-1: Kategorisierung qualitativer Verfahren

	Inhaltsanalytische Verfahren	Kodierende Verfahren	Hermeneutische Verfahren
Erkenntnisziel	(theoriegeleitete) Erfassung und Beschreibung subjektiver Sichtweisen → Welchen Sinn schreiben Personen ihren Handlungen zu? Was wird (sinngemäss) mitgeteilt?	Rekonstruktion von Handlungsentscheiden in Prozessen → Welche Handlungs- bzw. Deutungsmuster zeigen sich?	Rekonstruktion von Sinn und Bildung von Typen → Welcher Sinn steckt hinter einer Handlung? Was bewegt die Handlung eines Falls?
Daten	stärker strukturierte Daten, meist (teil-)standardisierte Interviews	wenig strukturierte Daten, z. B. Interviews, Beobachtungsprotokolle, (bestehende) Dokumente	nicht oder wenig strukturierte Daten, z. B. narrative Interviews, Aufzeichnungen von natürlichen Gesprächen, bestehende Dokumente
Vorgehensweise	Linear: Daten erheben, auswerten und verdichten (jeweils abgeschlossene Zwischenschritte)	Zirkulär: Daten zu einem ersten Fall sammeln und auswerten, danach Daten zu einem zweiten Fall erheben und auswerten usw.	
Beispiel: Schulkultur	Befragung von Zeitzeugen (Schulleitung, Lehrpersonen) zum Thema Schulkultur, Zusammenfassung der subjektiven Sichtweisen und Deutungen	Analyse von Schulleitbildern, Beobachtungen vor Ort, Interviews mit Lehrpersonen, Schulleitungen etc. zur Bildung einer Theorie über Schulkultur	Analyse von Schulleitbildern, „natürliche“ Gespräche in Sitzungen oder Interviews mit Lehrpersonen, Eltern, Schulleitungen zur Entdeckung tiefer liegender Muster in der Ausbildung einer spezifischen Schulkultur

Datenerhebung. Allen Verfahren ist gemeinsam, dass sie meist auf nicht oder wenig standardisierte Daten zurückgreifen, also auf Daten, die wenig von aussen beeinflusst werden. Solche Daten können einerseits direkt von den Beforschten selbst erzeugt worden sein (Tagebücher, Kinderzeichnungen, Sitzungsprotokolle, Webauftritte usw.) oder aber durch die Forschenden generiert werden (Unterrichtsvideos, Beobachtungsprotokolle, Gesprächs- oder Interviewtranskripte etc.). Auch werden mit qualitativen Verfahren in der Regel eher kleinere Datensätze (Fallzahlen) verarbeitet, diese dafür aber umso intensiver. Um trotz niedriger Fallzahlen allgemeingültige Aussagen machen zu können, bedienen sich die qualitativen Methoden nicht der statistischen Repräsentativität (vgl. Kap. 5.4 „Stichproben“), sondern einer inhaltlichen Repräsentativität. Diese erreicht man dadurch, dass man erste zu untersuchende Fälle aufgrund der erarbeiteten Vorkenntnisse auswählt, erhebt und unmittelbar auswertet. Diese Erstauswertung eines Falls bestimmt alsdann den weiteren Verlauf und definiert die notwendige weitere Suche nach Fällen (zum Konzept des Theoretical Sampling vgl. nachfolgend sowie Kap. 12.3 „Grounded Theory“ und Kap. 5.4 „Stichproben“).

► Beispiel „Theoretical Sampling“

Theoretical Sampling ist ein Begriff aus der qualitativen Forschungsmethodik, der eine systematische und iterative (d. h. sich schrittweise annähernde) Methode zur Datensammlung beschreibt. Diese Erhebungsmethode basiert auf der Idee, dass Forschende während des Forschungsprozesses fortlaufend neue Daten sammeln und analysieren, um die fortlaufende Erkenntnis resp. die Bildung von theoretischen Ansätzen und Modellen zu unterstützen. Im Gegensatz zur Stichprobenziehung bei quantitativen Methoden basiert die Auswahl der Fälle beim Theoretical Sampling auf ihrer Bedeutung für die Forschungsfragen und ihre erkenntnisbringende Leistung im Hinblick auf die im Forschungsprozess sich entwickelnden theoretischen Ansätze und Modelle.

In einer Studie zum Thema Burnout bei Lehrpersonen könnte das Theoretical Sampling folgendermassen aussehen: Die Forscherin oder der Forscher beginnt mit einer kleinen Stichprobe von Lehrpersonen, die sich für die Teilnahme an der Studie gemeldet haben. Aufgrund der Analyse der im Rahmen dieser Stichprobe gewonnenen Daten, vermuten die Forschenden, dass bestimmte Merkmale, wie zum Beispiel das Alter oder die Höhe des Arbeitspensums, einen Einfluss auf das Burnout-Risiko haben könnten. Die Forschenden entscheiden sich daraufhin, bei der Auswahl weiterer Lehrpersonen für die Studie auf diese Merkmale besonders zu achten (gleiche Ausprägung oder maximaler Kontrast, d. h. hier: hohes oder tiefes Alter; hohes oder tiefes Pensum), um die systematische Theoriebildung zu unterstützen.

Ein weiteres Beispiel wäre eine Studie zur Entstehung von Gruppendynamiken in einer Klasse. Die Forschenden beginnen mit der Beobachtung einer kleinen Grup-

pe von Schülerinnen und Schülern. Sie identifizieren bestimmte Verhaltensmuster. Durch die Analyse der Daten erkennen die Forschenden, dass diese Verhaltensmuster durch die Art des Classroom Managements der Lehrpersonen erheblich beeinflusst werden. Die Forschenden entscheiden daraufhin, weitere Gruppen unter verschiedenen Lehrpersonen zu beobachten. Sie suchen gezielt nach Lehrpersonen, die unterschiedliche Strategien des Classroom Managements anwenden. Auf diese Weise können sie systematisch vergleichen, wie Schülerinnen und Schüler unter verschiedenen Bedingungen reagieren, was die systematische Theoriebildung unterstützt.

In beiden Beispielen ist die Auswahl der Fälle nicht zufällig (wie bei quantitativen Vorgehensweisen), sondern basiert auf der Bedeutung für die Forschungsfrage und der Eignung für die Theoriebildung. Durch den Einsatz von Theoretical Sampling können Forschende eine umfassendere und tiefere Analyse durchführen und somit selbst bei kleinen Fallzahlen zu einer aussagekräftigeren Theoriebildung gelangen. Die Zahl, der am Ende zu analysierenden Fälle kann zu Beginn nicht festgelegt werden, wird sie doch durch den Forschungsprozess selbst bestimmt. Sie ist abhängig von der Fragestellung sowie von der Aussagekraft der Daten in Bezug auf die Fragestellung. Die Fallsuche wird im Idealfall so lange weitergeführt bis „nichts Neues“ mehr dazukommt. Auch lassen sich die Merkmale der Datenauswahl (z. B. Alter und Geschlecht von Befragten) erst im Anschluss an erste Analysen genauer bestimmen (induktives Vorgehen). Während die Ungewissheit über die Anzahl von zu untersuchenden Datenquellen teilweise schwierig auszuhalten ist, birgt die Methode dafür den Vorteil, dass keine „überschüssigen“ Daten, produziert werden, also kein „Datenfriedhof“ geschaffen wird (vgl. Legewie & Schervier-Legewie, 2004, Abs. 59).

Wahl der Auswertungsmethode. Vorab ist zu sagen: Es gibt keine klaren Entscheidungshilfen, welche die Wahl eines bestimmten qualitativen Auswertungsverfahrens vorgeben könnten. Ebenso wenig gibt es eine Hierarchie der Verfahren, sodass Forschungsnovizinnen und -novizen empfohlen werden könnte, mit dem einen oder anderen Verfahren einzusteigen und später dann mit komplexeren Verfahren weiterzufahren. Fälschlicherweise und der Sache kaum gerecht werdend, werden qualitative Verfahren oft in einer Art Negativselektion benutzt, d. h. Studierende und Forschende entscheiden sich primär und generell gegen quantitative Verfahren, aber nicht gezielt für ein qualitatives Vorgehen. Solche Arbeiten enden häufig in einem Durcheinander zwischen quantitativer Forschungslogik, paraphrasierendem Nachvollzug und einer unsystematischen Interpretation von Daten (vgl. Przyborski & Wohrab-Sah, 2014, S. 4). Deshalb ist für die Wahl der Auswertungsmethode empfehlenswert, dass man erstens die dem gewählten Verfahren zugrunde gelegte Sichtweisen auf die Welt (methodologischer Hintergrund) in groben Zügen versteht und zweitens, dass die Fragestellungen mit dem spezifischen Verfahren auch wirklich beantwortet werden können. Denn auch bei den qualitativen Ansätzen gibt es je nach Datenmaterial und Fragestellung mehr oder weniger geeignete Auswertungsverfahren (vgl. Tabelle 12-1). Und obwohl Vertretende der jeweiligen

methodischen Verfahren teils polemisch gegeneinander argumentieren, schliessen sich die Verfahren nicht aus, sondern können sich sogar gut ergänzen. Je nach Forschungsfrage kann es also ausgesprochen sinnvoll sein, verschiedene methodische Zugänge miteinander zu verknüpfen.

Mit der nachfolgenden zusammenfassenden Beschreibung von drei Verfahren soll die Festlegung eines Verfahrens bzw. einer Mischform von Verfahren erleichtert werden. Die Auswahl der vorzustellenden Ansätze wurde so getroffen, dass sich die wesentlichen Eckpunkte qualitativer Auswertungsansätze darin repräsentiert finden.

Qualitative Inhaltsanalyse. Die qualitative Inhaltsanalyse eignet sich für die systematische, theoriegeleitete Bearbeitung von Textmaterial; es sind mit diesen Verfahren auch grosse Textmengen bewältigbar. Anders als bei den anderen beiden Ansätzen werden die Daten eher klassifiziert und es wird weniger auf die Rekonstruktion tiefer liegender Sinnstrukturen abgestellt. Gerade für Forschungsvorhaben, die weniger explorativen Charakter haben, sondern sich auf bereits bekannte theoretische Überlegungen stützen, ist dieses Verfahren daher gut geeignet.

Grounded Theory. Grounded Theory ist ein Stil, analytisch über soziale Erscheinungen (Phänomene) nachzudenken, die man besser verstehen und erklären möchte. Ziel dieses Forschungsansatzes ist es, „in den Daten schlummernde Theorien zu entdecken“ (Strauss in Legewie & Schervier-Legewie, 2004, Abs. 51ff). Der Schwerpunkt liegt auf der Bildung von Arbeitshypothesen durch den ständigen Vergleich von Phänomenen, die in Prozessen des theoretischen Samplings beobachtet und dokumentiert werden. Die Arbeitshypothesen werden im Forschungsprozess zu theoretischen Modellen oder einer Theorie über den untersuchten Gegenstand verdichtet. Obwohl das Vorgehen eine grundsätzliche Offenheit gegenüber dem Forschungsgegenstand vorsieht – das theoretische Modell soll gleichsam aus dem Gegenstand heraus entstehen, in ihm verankert („grounded“) sein – setzt die Analyse der Daten auf Seiten der Forschenden „theoretische Sensibilität“ voraus, also die Fähigkeit, Wichtiges von Unwichtigem trennen zu können und ein Gespür für theoretisch relevante Aspekte in den Daten zu haben (vgl. Strauss & Corbin, 1996, S. 25). Diese Fähigkeit erlangt man durch Forschungserfahrung, aber auch durch die intensive Auseinandersetzung mit Fachwissen zum Thema.

Objektive Hermeneutik. Das ursprüngliche Anwendungsgebiet der Objektiven Hermeneutik war vor allem die Auswertung transkribierter Interviews im Rahmen soziologischer Studien zu Familien. Mit der Zeit wurde die Methode aber auf die Analyse weiterer Textdokumente ausgedehnt (auch auf Kunstwerke und Bildmaterial; vgl. Flick, 2002). Das Verfahren eignet sich gut für die Analyse „natürlicher“ Daten, die nicht zu Forschungszwecken künstlich erzeugt wurden (vgl. Schallberger, 2005). Bevorzugt wird in der Regel ein längerer Ausschnitt eines Gesprächs oder Dokuments.

Als Methode eignet sich die Objektive Hermeneutik speziell dort, wo „Forschungs-Neuland“ betreten wird. Denn das Vorgehen eröffnet nicht nur ein Verständnis für die spezifische Logik im Handeln einer Person, sondern stellt diese immer auch in einen allgemeinen Zusammenhang. Ein typischer Hinderungsgrund, sich mit den Methoden der Objektiven Hermeneutik einzulassen, ist die vermeintliche Ausführlichkeit und Langsamkeit der Analyse. Typischerweise werden aber nur die Eingangssequenzen intensiv analysiert, und bereits kurze Textausschnitte führen in der Regel zu einer präzisen Rekonstruktion. Das übrige Material wird anschliessend nur noch zur gezielten Falsifikation von am Material rekonstruierten Hypothesen verwendet.

So oder so gilt für alle Datenauswertungsverfahren und für die qualitativen Vorgehensweisen vielleicht besonders: Das „Sich-kümmern“ um methodische Zugänge sollte nicht erst mit dem Schreiben des Abschluss textes beginnen, denn das Gelingen und Scheitern einer wissenschaftlichen Arbeit hängt stark mit dem gewählten methodischen Vorgehen zusammen. Dabei lohnt es sich auch, sich Zeit zu lassen und unter Umständen Verfahren vorgängig auszuprobieren, um schliesslich ein der Fragestellung angemessenes Verfahren auswählen zu können. Die nachfolgende Darstellung der drei oben genannten Ansätze soll bei der Verfahrensauswahl helfen. Neben den Zielsetzungen sowie den theoretischen Fundamenten, auf denen die Ansätze gründen, werden insbesondere auch dazugehörige „Techniken“ der Datenanalyse skizziert und anhand von Beispielen verdeutlicht. Wer sich für das eine oder andere Verfahren entscheidet, kommt aber nicht umhin, sich im gewählten Verfahren weiter zu vertiefen (siehe dazu die entsprechenden weiterführenden Literaturempfehlungen am Ende dieses Kapitels).

12.2 Qualitative Inhaltsanalyse

Ziel der Inhaltsanalyse ist es, Material, das aus irgendeiner Form von Kommunikation bzw. Interaktion stammt, systematisch zu erfassen (vgl. Mayring, 1993, S. 11; 2007, S. 468). Sie eignet sich insbesondere zur Auswertung von transkribierten Interviews, aber auch von Beobachtungsprotokollen, Videoaufnahmen oder anderen Quellen (z. B. Förderberichte, Zeitungsartikel) und stellt deshalb bei Studierenden eine beliebte Standardmethode dar. Während die klassische Inhaltsanalyse primär auf die verdichtende Beschreibung von grossen Textmengen zielt, verfolgt die qualitative Inhaltsanalyse, dem qualitativen Forschungsparadigma verpflichtet, einen semantisch-hermeneutischen Anspruch.

12.2.1 Hintergrund des Ansatzes

Die Vorläufer der qualitativen Inhaltsanalyse wurden in den 1920er Jahren durch Forschende aus dem Bereich der Kommunikationswissenschaften entwickelt. Im

Vordergrund stand die systematische Auswertung grosser, massenmedialer Textmengen (z. B. Zeitungsartikel). Die Analyse war dabei vor allem quantitativ orientiert, das heisst man zählte bestimmte Textbestandteile aus (z. B. wie oft in einer Zeitung der Lehrberuf thematisiert wird). Seit den 1950er-Jahren wurde dieser inhaltsanalytische Ansatz in verschiedenen Disziplinen (u. a. Psychologie, Geschichtswissenschaft, Soziologie) weiterentwickelt. Dabei wollte man auch der Kritik Rechnung tragen, dass der quantitativ orientierte Ansatz zu oberflächlich sei. Man ergänzte diesen deshalb um eine qualitative Dimension, mit der auch Sinnstrukturen erfasst werden können (vgl. Kuckartz, 2016, Mayring, 2000).

Das in den 1980er-Jahren erstmals publizierte Buch „Qualitative Inhaltsanalyse“ von Mayring (1993) stellte den ersten als Methodenlehrbuch geschriebenen Text zu diesem Verfahren dar, der bis dato in vielen Neuauflagen erschienen ist (Kuckartz, 2016, S. 23, Mayring, 2010, S. 601). Im deutschsprachigen Raum wird der Begriff der Qualitativen Inhaltsanalyse häufig gleichgesetzt mit den von Mayring präsentierten Vorgehensweisen und dabei verkannt, dass es zahlreiche andere Formen qualitativ-inhaltsanalytischer Vorgehensweisen gibt (Kuckartz, 2016, S. 27).

12.2.2 Vorgehensweisen

Charakteristika. Qualitativ-inhaltsanalytische Verfahren (QIA) folgen erstens einem vergleichsweise strikten Ablaufmodell, bei welchem Analyseeinheiten (Kategorien) definiert und das Datenmaterial auf dieser Basis codiert und analysiert wird (vgl. Abbildung 12-2). Die zentrale Stellung der Kategorien im Forschungsprozess ist ein zweites Charakteristikum von QIA. Eine kategorienbasierte Vorgehensweise läuft in der Forschungspraxis aber selten so linear ab, wie dies Abbildung 12-2 suggeriert, sondern weist durchaus auch iterative Momente und Feedbackschritte auf. Dies ist dem Umstand geschuldet, dass die Gewinnung der Kategorien sowohl deduktiv (auf der Basis theoretischen Vorwissens) als auch induktiv (aus dem Datenmaterial heraus) erfolgen kann. Kuckartz (2016, S. 97) schlägt ein mehrstufiges Vorgehen vor, bei dem in einer ersten Phase grob entlang von deduktiv gewonnenen Hauptkategorien codiert wird. In einem Folgeschritt können diese Kategorien am Material induktiv weiterentwickelt, ausdifferenziert und das gesamte Datenmaterial anschliessend mit dem verfeinerten Kategoriensystem erneut codiert werden. Das im Vergleich – etwa zum Forschungsstil der Grounded Theory (siehe nachfolgend) – stärker systematisierte und regelgeleitete Vorgehen unterstützt drittens die intersubjektive Nachvollziehbarkeit des Verfahrens (siehe Gütekriterien). Viertens wird bei QIA die Entstehung des Datenmaterials und ihre Verortung in einem Kommunikationsmodell reflektiert (Kuckartz, 2016, S. 26, Mayring, 2000, Abs. 7). Dies meint, dass Informationen zu den Textproduzentinnen oder Befragten, zur Entstehungssituation des Textmaterials sowie zu dessen Wirkung beschrieben und reflektiert werden (vgl. Ramsenthaler, 2013, S. 24).

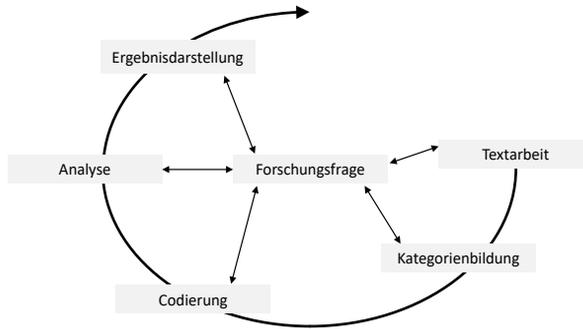


Abbildung 12-2: Generelles Ablaufschema qualitativ-inhaltsanalytischer Verfahren (Kuckartz, 2016, S. 45).

In der Praxis der Sozialforschung existiert eine Vielzahl von QIA-Methoden, die Kuckartz (2016) in drei grundlegende Verfahren differenziert:

- Inhaltlich-strukturierende QIA: Ziel ist es, mittels Kategorien und Subkategorien eine inhaltliche Strukturierung des Datenmaterials entstehen zu lassen, d. h. zentrale Inhalte und Themen zu identifizieren, zusammenzufassen und zu systematisieren.
- Evaluative QIA: Ziel ist es, zu einer Einschätzung und Bewertung von Inhalten zu kommen, d. h. das Datenmaterial auf einer Skala (in der Regel Ordinalskala, vgl. hierzu Kap. 13.1 „Schätzen bzw. Messen und Skalenniveau“) einzuschätzen.
- Typenbildende QIA: Ziel ist es, ausgehend von Fallvergleichen Muster bzw. Typen ausfindig zu machen, d. h. Gruppen oder Cluster von Fällen zu erkennen, die sich in Bezug auf bestimmte Merkmale ähnlich sind.

Nachfolgend wird ausschliesslich das erstgenannte Verfahren präsentiert, zumal dieses in zahlreichen Forschungsprojekten erprobt und daher auch in der Methodenliteratur ausgiebig beschrieben wird, und sich nachgerade als Einstiegsmethode und damit für studentische Arbeiten eignet.

Ablaufmodell. Der Ablauf einer inhaltlich-strukturierenden QIA lässt sich in folgenden Schritten beschreiben (vgl. Kuckartz, 2016, S. 97ff, Lamnek, 1989, S. 202ff, Mayring, 1993):

- 1) *Festlegung des Textmaterials:* Es muss nicht zwingend alles erhobene Datenmaterial vollständig ausgewertet werden, sondern es ist durchaus erlaubt, nur die für die Beantwortung der Forschungsfrage relevanten Ausschnitte zu verwenden. Eine Auswahl ist aber in jedem Fall zu begründen.
- 2) *Reflektierte Beschreibung der Entstehungssituation und formale Charakterisierung des Materials:* Informationen über den Zusammenhang, in welchem Daten entstanden sind, müssen für den Forschungsbericht festgehalten werden. Ebenso wird beschrieben, in welcher Form das Datenmaterial vorliegt. In Bezug auf Interviews bedeutet

dies etwa, dass auch Transkriptionsregeln, Interviewleitfäden sowie die methodologischen Überlegungen zur Gestaltung der Interviews offengelegt werden.

- 3) *Richtung der Analyse bestimmen*: Es gilt, sich vor der Analyse bewusst zu machen, was genau man anhand der Daten herausfinden möchte, welche Inhalte den Kern des eigenen Erkenntnisinteresses ausmachen, aber auch, welche subjektiven Vorannahmen man allenfalls hegt.
- 4) *Initiiierende Textarbeit*: In einem ersten Schritt wird das Textmaterial dann sorgfältig gelesen. Das Ziel dieser initiiierenden Textarbeit ist es, vor dem Hintergrund des Erkenntnisinteresses (der Analyserichtung), ein erstes Verständnis für die Texte zu entwickeln. Es lohnt sich, zentrale Begriffe und wichtige Abschnitte zu markieren, die inhaltliche Struktur sowie Argumentationslinien zu analysieren und hierzu Randnotizen zu machen. Ebenso ist es zielführend, wenn nach dem ersten Durcharbeiten eines Textes, eine Fallzusammenfassung oder ein Forschungsmemo erstellt wird, in dem erste Vermutungen und Erkenntnisse festgehalten werden (für Beispiele siehe Kuckartz, 2016, Kap. 3.3).
- 5) *Kategorienbasiertes Codieren des Materials*: Kuckartz (2016, S. 97) schlägt ein mehrstufiges Vorgehen in Bezug auf das Codieren des Datenmaterials vor. In einer ersten Phase wird eher grob entlang von deduktiv gewonnenen Hauptkategorien codiert, die am Material induktiv weiterentwickelt und ausdifferenziert werden. Der Analyseprozess wird dabei durch einen Kodierleitfaden gesteuert, in welchem die Kategorien definiert und mittels Ankerbeispielen (typische Textpassagen) veranschaulicht wird (vgl. Beispiel in Tabelle 12-2)

Tabelle 12-2: Beispiel Kodierleitfaden zum Thema (Motivations-) Ziele

Kategorie	Definition	Ankerbeispiele (aus dem Datenmaterial)
Anspruchsniveau	Aussagen dazu, welches Anspruchsniveau, verstanden als „das, was sich eine Person zu schaffen vornimmt“ (Wild et al., 2006, S. 212) als motivierend empfunden wird.	„Diese Ziele sollten mit den mir zur Verfügung stehenden Ressourcen erreichbar sein.“
Zieltypen	Aussagen, in denen verschiedene Zieltypen charakterisiert werden.	„Ist das Ziel in weiter Ferne, also nicht innerhalb eines halben Jahres zu erreichen, muss ich Teilziele festlegen.“

Sofern notwendig, werden ebenso Kodierregeln formuliert, um sicher zu stellen, dass alle Personen, welche einen Text mittels dieser Regeln analysieren, zum selben Resultat kommen (sog. Inter-coderreliabilität). Mit diesem Vorgehen soll

die Objektivität der Analyse sichergestellt werden (vgl. Mayring, 1993; 2007). Ein sorgfältig erstelltes Kategoriensystem bildet auch eine gute Struktur für den Forschungsbericht (Ergebnisteil).

Das kategorienbasierte Codieren kann entlang folgender Schritte strukturiert werden:

- (a) *Entwickeln thematischer Hauptkategorien:* Die Hauptkategorien werden deduktiv aus der Forschungsfrage sowie der Fachliteratur abgeleitet und sind in der Regel auch bei der Datenerhebung leitend (z. B. Themenblöcke im Interviewleitfaden).
- (b) *Codieren des gesamten Textmaterials mit den Hauptkategorien:* Der Text wird Zeile-für-Zeile oder Satz-für-Satz durchgearbeitet und Textabschnitte den Kategorien zugeordnet (z. B. in dem für jede Kategorie eine andere Farbe verwendet wird, mit welcher die entsprechenden Textabschnitte markiert werden).
- (c) *Zusammenstellen aller Textstellen einer Hauptkategorie und Induktives Bestimmen von Subkategorien am Material:* Für jede Hauptkategorie wird eine Zusammenstellung der codierten Textstellen erstellt, welche dann die induktive Gewinnung weiterer (Sub-)Kategorien am Datenmaterial ermöglicht. Dies kann bspw. mittels einer zusammenfassenden Inhaltsanalyse geschehen, bei der die vorliegenden Textstellen zunächst so paraphrasiert werden, dass sie kurz und bündig den Inhalt wiedergeben. In einem Folgeschritt werden die Paraphrasen generalisiert, d. h. alle auf dasselbe, allgemeine Abstraktionsniveau gebracht. Und schliesslich werden Aussagen mit gleichem oder ähnlichem Gegenstand gebündelt und zusammengefasst.
Im Beispiel (Tabelle 12-3) wird die Aussage einer Studentin zur Hauptkategorie Lernmotivation auf diese Weise analysiert, dabei wurde zunächst Satz-für-Satz paraphrasiert und die Paraphrasen anschliessend generalisiert und reduziert. So zeigt sich, dass die Studentin Zieltypen unterscheidet (Nah- und Fernziele). Diese können nun als induktiv gewonnenen Subkategorien in das bestehende Kategoriensystem aufgenommen werden.
- (d) *Erneutes Codieren des kompletten Materials:* Das Material wird mit dem ausdifferenzierten Kategoriensystem ein zweites Mal codiert, d. h. das Textmaterial mithilfe des erweiterten Kategoriensystems komplett systematisiert und strukturiert.
- (e) *Analyse und Interpretation des Materials:* Abschliessend werden die Textstellen zu den verschiedenen (Sub-)Kategorien erneut zusammengestellt und auf dieser Basis die finalisierende Analyse- und Interpretationsarbeit begonnen. Es lohnt sich, hier thematische und/oder fallbezogene Zusammenfassungen zu erstellen, um das analytische Nachdenken über das Material anzuregen. Ebenfalls gilt es in diesem Schritt, Zusammenhänge zwischen den Kategorien herauszuarbeiten und sichtbar zu machen. Dabei können auch Verfahren und Techniken ande-

rer Ansätze (z. B. das Kodier-Schema nach Strauss & Corbin, 1996, vgl. Abbildung 12-5) hilfreich sein.

Tabelle 12-3: Beispiel einer zusammenfassenden Inhaltsanalyse zur induktiven Kategoriengewinnung

Textdaten	Paraphrasierung	Generalisierung	Reduktion
Um eine Lern- und Leistungsmotivation hervorzubringen, braucht es die Grundvoraussetzung eines Zieles.	Motivation braucht Voraussetzung: Ziel	Motivationsvoraussetzung	Motivationsvoraussetzungen: Vorhandensein von erreichbaren Zielen
Ist das Ziel in weiter Ferne, also nicht innerhalb eines halben Jahres zu erreichen,	Zielerreichung erst nach halbem Jahr = Fernziel	Zieltyp	Zieltypen und ihre Merkmale: Fernziele (Zeithorizont mehr als ½ Jahr), Nahziele (= zerlegte Fernziele, sind mit aktuellen Ressourcen erreichbar)
muss ich Teilziele festlegen.	Fernziele müssen in Teilziele zerteilt werden, Teilziele = Subform von Fernzielen	Zieltyp	
Diese Ziele sollten mit den mir zur Verfügung stehenden Ressourcen erreichbar sein.	mit aktuellen Ressourcen erreichbar = Teilziel	Merkmal von Teilzielen (Zieltyp)	
Sind diese Grundvoraussetzungen gegeben, kann ich mich zum Lernen motivieren	Zielerreichbarkeit = Grundvoraussetzung für Lernmotivation	Motivationsvoraussetzung	Konsequenz von Zielerreichung = Leistungserfolg
und dadurch meinen Erfolg in der Leistung sichern, indem ich Ziele oder Teilziele erreiche.	Zielerreichung = sicherer Leistungserfolg	Konsequenz von Zielerreichung	

12.2.3 Kritische Würdigung

Die Systematik der QIA-Auswertungsverfahren ist vergleichsweise leicht erlernbar. Die Verfahren eignen sich deshalb auch für Forschungsnovizinnen und -novizen. Zudem erlaubt das systematische Vorgehen ein gutes Mass an Nachvollziehbarkeit. Das Verfahren bietet sich auch für grössere Stichproben an, weil sich relativ leicht quantitative Schritte einbauen lassen und damit eine Kombination qualitativer und quantitativer Vorgehensweisen ermöglicht wird (vgl. Mayring, 2007, S. 474). We-

Teil IV: Analyse von Daten, 978382561689, 2023
 wurde mit IP-Adresse 134.100.172.044 aus dem Netz der SUB Hamburg am Mai 20, 2025 um 08:22:42 (UTC) heruntergeladen.
 Das Weitergeben und Kopieren dieses Dokuments ist nicht zulässig.

niger geeignet ist das Verfahren bei Fragestellungen, die sehr offen sind oder explorativen Charakter haben.

12.3 Grounded Theory

Der Name Grounded Theory bezeichnet weniger die methodische Vorgehensweise als vielmehr das Ziel des Forschungsprozesses, das darin besteht, Theorien oder theoretische Ansätze und Modelle zu entwickeln, die in den Daten gründen (vgl. Strauss & Corbin, 1996, S. 7). Weiter bezeichnet Grounded Theory auch einen Forschungsstil zur Erarbeitung solcher Theorien. Dieser Forschungsstil unterscheidet sich vom „klassischen“ Forschungsprozess darin, dass er keinen linearen Ablauf des Forschungsprozesses vorgibt (vgl. Kap. 2.4 „Der idealtypische Ablauf eines empirischen Forschungsprozesses“), sondern von einer Zirkularität des Forschungsprozesses ausgeht (vgl. Abbildung 12-3). Bei Grounded Theory wird dabei die Gleichzeitigkeit von Datengewinnung, Datenanalyse und Theoriebildung sowie die Tatsache, dass sich diese Schritte gegenseitig beeinflussen, betont (Strauss, 1998, S. 44ff; vgl. auch Strübing, 2008, S. 13ff).

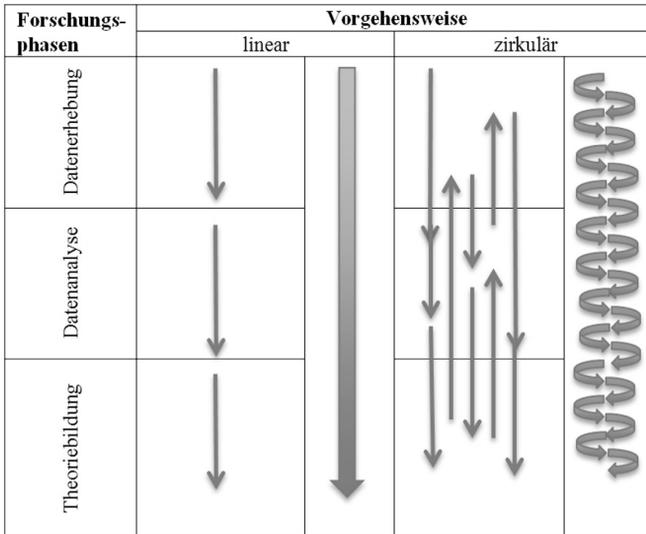


Abbildung 12-3: Zirkularität und Gleichzeitigkeit der Forschungsschritte in Abgrenzung zum linearen Vorgehen (modifiziert nach Strauss, 1998, S. 46 und Strübing, 2008, S. 15)

Im Vordergrund steht das kontinuierliche analytische Nachdenken über soziale Phänomene. Grounded Theory eignet sich deshalb auch zur systematischen

Teil IV: Analyse von Daten, 9783825261689, 2023
 wurde mit IP-Adresse 134.100.172.044 aus dem Netz der SUB Hamburg am Mai 20, 2025 um 08:22:42 (UTC) heruntergeladen.
 Das Weitergeben und Kopieren dieses Dokuments ist nicht zulässig.

Reflexion des eigenen berufspraktischen Handelns (vgl. auch Kap. 3 „Forschungskompetenzen für Lehrpersonen“).

12.3.1 Hintergrund des Ansatzes

„Grounded Theory“ wurde in den 1960er-Jahren durch die beiden amerikanischen Soziologen Barney Glaser und Anselm Strauss als Alternative zu den damals vorherrschenden deduktiven Vorgehensweisen begründet. In ihrem 1967 erstmals publizierten Buch „The Discovery of Grounded Theory“ (Glaser & Strauss, 2008) kritisierten sie, dass viele sozialwissenschaftlichen Methoden ausschliesslich auf die (deduktive) Überprüfung und Reproduktion bestehender Theorien abzielten, jedoch kaum zum Gewinn neuer Theorien beitragen. Von ihrem Forschungsstil erhofften sie sich neue Erkenntnisse und Theorien (vgl. dazu untenstehenden Interviewauszug).

► Anselm Strauss in einem Interview über das gemeinsam mit Barney Glaser publizierte Buch „The Discovery of Grounded Theory“:

„Wir entschieden Mitte [19]60, ein Buch über Methoden zu schreiben. Wir spürten schon, dass Veränderungen in der Luft lagen, denn wir wollten für die ‚Kids‘ schreiben – Leute über 30 schienen uns schon zu festgelegt. Barney hatte das bessere Gefühl, dass ein solches Buch ankommen würde, ich war skeptischer, weil ich älter war. Der Titel, ‚The Discovery of Grounded Theory‘ (1967; dt.: 1998), zeigt schon, worauf es uns ankam: nicht wie in den üblichen Methodenbücher die Überprüfung von Theorie, sondern deren Entdeckung ‚aus den Daten heraus‘. Grounded Theory ist keine Theorie, sondern eine Methodologie, um in den Daten schlummernde Theorien zu entdecken.

Wir verbanden drei Zielsetzungen mit dem Buch: Erstens versuchten wir, qualitative Forschung, die damals nicht anerkannt wurde, zu legitimieren. In vielen Departements ist es ja immer noch so, dass qualitative Forschung nicht als wissenschaftlich gilt! So wurde das Buch auch benutzt. – Studenten konnten damit ihre qualitativen Studien vor den Prüfungsausschüssen besser rechtfertigen. Zweitens wollten wir Funktionalisten wie Parsons und Mer-ton attackieren. Damals wurden deren Theorien von den Studenten und jungen Soziologen umstandslos übernommen und alles andere in Frage gestellt. Wir wandten uns gegen diese ‚überlieferten Theorien‘, und das war auch wirkungsvoll. Deswegen hat das Buch diese aktive und scharfe Diktion. Der Lektor hat zunächst alle aktiven Verben ins Passiv verkehrt, um das abzumildern, doch wir haben dagegen protestiert und uns durchgesetzt. Der dritte Grund war die Darstellung der Möglichkeit von Theoriebildung aus den Daten heraus.“

(Interviewauszug aus Legewie & Schwervier-Legewie, 2004, Abs. 51f)

Bei der Entwicklung waren Strauss und Glaser stark geprägt durch die Denkschulen des Pragmatismus sowie des Symbolischen Interaktionismus. Zentrale Themen dieser Denkschulen sind Prozesse des Wandels und soziale Interaktionen (vgl. Hildenbrand, 2007, S. 32ff). Vor diesem Hintergrund erstaunt es kaum, dass sich Grounded Theory vor allem für Forschungsvorhaben eignet, mit denen ein vertief-

tes Verständnis über (wenig bekannte) soziale Prozesse und Interaktionen gewonnen werden soll. Dabei interessiert insbesondere auch die Frage nach dem Sinn, den Menschen ihrem Handeln verleihen.

Da Grounded Theory neue Erkenntnisse gewinnen will, steht am Anfang des Forschungsprozesses nicht wie bei anderen Verfahren eine präzise und stark eingrenzen- de Fragestellung, sondern eine offene Beschreibung. Die Fragestellung entwickelt sich also erst im Verlaufe des Forschungsprozesses. Zu Beginn dient sie vor allem als „Wegweiser“ (Strauss & Corbin, 1996, S. 24), der die Forschungsrichtung vorgibt und den Forschenden einen Einstieg ins Thema ermöglicht. Mit dem Fortschreiten des Forschungsprozesses, d. h. nach der Erhebung und Analyse erster Daten, wird die Fragestellung verfeinert und spezifiziert.

Wichtig ist dabei, dass Forschende möglichst unvoreingenommen an das Forschungsfeld herangehen, d. h. sich auf die soziale Wirklichkeit einlassen und genau beobachten, was passiert. Umgekehrt formuliert: Es geht nicht darum, Hypothesen, die man bspw. aufgrund der Lektüre von Fachliteratur entwickelt hat, zu überprüfen, sondern vielmehr darum, solche Hypothesen aus den Daten heraus zu entwickeln und zu erklären.

In Untersuchungen mit der Grounded Theory möchten Sie Phänomene im Licht eines theoretischen Rahmens erklären, der erst im Forschungsverlauf selbst entsteht. Sie möchten nicht dadurch eingeengt werden, dass Sie an einer vorab entwickelten Theorie festhalten müssen, die sich auf den untersuchten Wirklichkeitsbereich anwenden lässt oder auch nicht. (Strauss & Corbin, 1996, S. 32)

Das bedeutet keinesfalls, dass bereits bekannte Theorien und Erklärungsansätze ignoriert werden (müssen). Ganz im Gegenteil: Jegliches Fachwissen, das bereits vorgängig bekannt ist, kann bei der Analyse und Interpretation der Daten hilfreich sein (vgl. Strauss & Corbin, 1996, S. 33ff). Zentral ist dabei aber, dass Forschende nicht „Gefangene der Literatur“ (Strauss & Corbin, 1996, S. 38) werden, sondern offen bleiben für die Erkenntnisse, die sich aus den Daten heraus ergeben.

12.3.2 Vorgehensweisen

Grounded Theory bietet kein starres Set an Verfahrensweisen und Methoden, die strikt befolgt werden müssen. Grounded Theory zeichnet sich vielmehr durch drei Schlüsselkonzepte aus, die das geforderte systematische Nachdenken über den interessierenden Untersuchungsbereich leiten. Es sind dies erstens das theoretische Kodieren, zweitens das theoretische Sampling und drittens die Idee des kontinuierlichen Vergleichs.

Theoretisches Kodieren. Das Kodierverfahren, welches gleichzeitig das Herzstück dieses Forschungsansatzes ist, besteht darin, dass Daten kodiert (aufgebrochen), interpretiert und auf neue Art zusammengesetzt werden (vgl. Strauss & Corbin, 1996, S. 39ff). Konkret bedeutet dies, dass man den Daten (oft sind dies Beobach-

tungsprotokolle oder transkribierte Interviews) „Namensetiketten“ (Kodes) zuordnet. Diese werden zunächst nahe am Text und später immer abstrakter formuliert. So entstehen Kategorien. Es geht dabei nicht ausschliesslich um die Klassifikation oder Beschreibung von Phänomenen, sondern vielmehr darum, theoretische Konzepte zu bilden, welche das interessierende Phänomen zu erklären vermögen. Hintergrund dieser Art des Kodierens bildet das sogenannte Phänomen-Indikator-Konzept-Modell (vgl. Abbildung 12-4). Darstellungen der sozialen Wirklichkeit (Phänomene) werden dabei als Anzeichen (Indikatoren) für etwas Allgemeineres (Konzepte, Kategorien) verstanden, das im Verlaufe des Forschungsprozesses entdeckt wird.

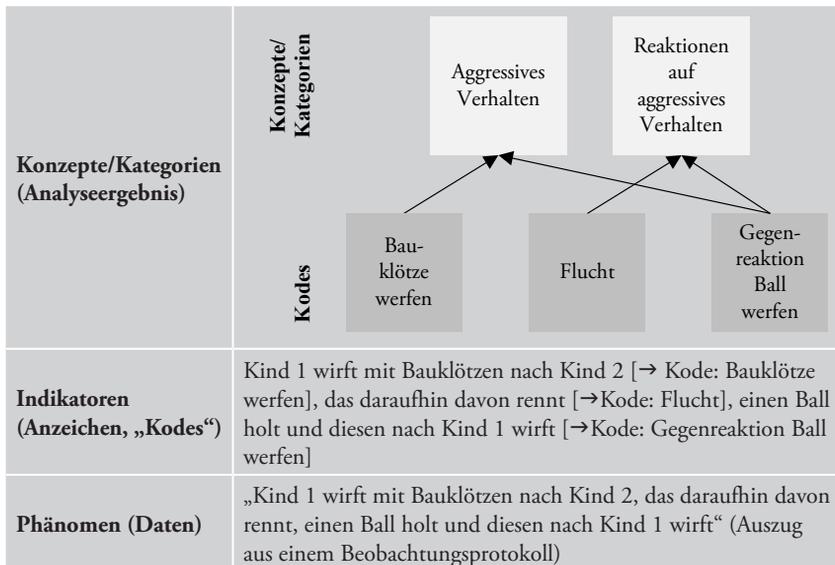


Abbildung 12-4: Phänomen-Indikator-Konzept-Modell

Die Entdeckung von Konzepten und Kategorien erfolgt dabei mit Hilfe unterschiedlicher Kodierverfahren:

- Offenes Kodieren
- Axiales Kodieren
- Selektives Kodieren

Diese drei Kodierverfahren sind nicht klar voneinander abzugrenzen und finden auch nicht notwendigerweise in einer bestimmten Abfolge statt, sondern man pendelt zwischen den verschiedenen Verfahren hin und her, was es letztlich nicht leicht macht, den Analyseprozess so zu dokumentieren, dass er für andere Forschende nachvollziehbar ist (vgl. Kap. 12.6 „Gütekriterien qualitativer Forschung“).

Nichtsdestotrotz zeigt sich in Bezug auf die Anwendung der Kodierarten eine gewisse Reihenfolge. Das offene und axiale Kodieren finden eher zu Beginn des Forschungsprozesses statt, das selektive Kodieren eher gegen Ende, wenn auch die Forschungsfrage spezifizierter ist und das Datenmaterial zunehmend gezielter und selektiver bearbeitet werden kann (vgl. Flick, 2002, S. 259).

Offenes Kodieren. Ziel des offenen Kodierens ist es, Phänomene zu benennen, die Daten zu untersuchen, zu vergleichen und Fragen an sie zu stellen. Es geht also um „das Herausgreifen einer Beobachtung, eines Satzes, eines Abschnittes und das Vergeben von Namen für jeden einzelnen darin enthaltenen Vorfall, jede Idee oder jedes Ereignis – für etwas, das für ein Phänomen steht oder es repräsentiert“ (Strauss & Corbin, 1996, S. 45).

Es ist normal und üblich, dass in diesem Prozess sehr viele Codes generiert werden. Ebenso kann es sein, dass ein Text – je nach Fragestellung und Stil der interpretierenden Person – ganz unterschiedlich kodiert wird. Dies ist unproblematisch, da die Codes nur vorläufigen Charakter haben. Im Verlaufe des Forschungsprozesses wird sich zeigen, ob diese ersten Codes brauchbar sind oder nicht, oder ob sie weiter spezifiziert werden müssen (vgl. Strauss, 1998, S. 100f). Der Prozess des offenen Kodierens kann unterschiedlich angegangen werden: Man kann Zeile-für-Zeile, Wort-für-Wort oder auch Abschnitt-für-Abschnitt kodieren (vgl. Strauss & Corbin, 1996, S. 53ff). In der Regel ist es hilfreich, die ersten Daten Wort-für-Wort zu analysieren, da es dadurch möglich wird, in die Daten einzutauchen und sich mit ihnen auseinanderzusetzen. Im späteren Verlauf können dann ganze Satzteile oder Abschnitte kodiert werden. Wichtig ist in jedem Fall, dass die Daten nicht einfach zusammengefasst, sondern vielmehr konzeptualisiert werden. Das bedeutet, dass sie auf eine abstraktere Ebene gebracht werden. Im Folgenden werden die Schritte anhand eines Interviewausschnittes mit einer Sozialpädagogin, die Menschen mit geistiger Behinderung betreut, verdeutlicht (vgl. Huber, 2006).

► Beispiel für offenes Kodieren

Da kamen wir zurück vom Spazieren, es ging mega gut mit dieser Person, 1/der kam zur Tür rein, 2/da sitzt jemand anders, eine Therapeutin auf der Treppe, 3/er geht gleich auf sie los und schlägt sie 4/und es ist die Situation für mich, dann solltest du ruhig bleiben, 5/dann solltest du pädagogisch gesehen, solltest du schauen, wo ist das Problem, 6/aber du kannst es nicht nachvollziehen, also du weisst nicht wieso und warum 7/und ich denke, es ist schwierig, 8/in diesen Situationen kannst du dir fast nicht mehr überlegen, was ist jetzt richtig, was ist jetzt falsch, wie handelst du, 9/sondern du handelst einfach mal aus der Situation raus. 10/Und ich denke im Nachhinein ist es für, finde ich es wichtig, dass du es anschaut, was war, was können Möglichkeiten sein, 11/und manchmal findet man auch gar nicht raus, was es gewesen ist. 12/

- 1/ „Normale“ (gute) Interaktion mit dem Klienten
- 2/ Wechsel der (räumlichen) Umwelt
- 3/ Wechsel in der (personalen) Umwelt

- 4/ (plötzliches) aggressives Verhalten des Klienten: unerwartete Situation
- 5/ normative Anforderung (ruhig bleiben sollen)
- 6/ normative (fachliche) Anforderung (Problemanalyse)
- 7/ Unwissenheit, fehlende Erklärung
- 8/ Belastungserleben
- 9/ keine Reflexionsmöglichkeit, in Situation nicht reflexionsfähig
- 10/ Routinehandeln
- 11/ normative Anforderung (Problemanalyse)
- 12/ fehlende Erklärung

Im nächsten Schritt werden relevante Phänomene ausgemacht und die Codes geordnet. Diesen Schritt nennt man Kategorienbildung. Dabei empfiehlt es sich, die Kategorien prägnant zu benennen, sodass man dann beginnen kann, diese analytisch weiterzuentwickeln. Der Name der Kategorie soll abstrakter sein als die Namen der um das Phänomen gruppierten Konzepte (vgl. Strauss & Corbin, 1996, S. 47ff).

Im vorliegenden Fallbeispiel könnte man beispielsweise den Code „Routinehandeln“ (Kode 10) näher untersuchen und fragen: Weshalb kommt es zu diesem Routinehandeln? Der Interviewausschnitt zeigt, dass das Routinehandeln in einer unerwarteten Situation notwendig wird, einer Situation, die ein rasches Eingreifen erfordert, weil jemand bedroht wird. Auch unterscheidet sich Routinehandeln offenbar vom fachlich geforderten, reflektierten Handeln. Dem Routinehandeln geht keine Reflexion voran, weil in dieser Situation keine Zeit zur Reflexion blieb (man muss sofort reagieren). Weiter zeigt sich, dass Routinehandeln eine nachträgliche Reflexion erfordert, die jedoch nicht immer klärend wirkt. Zusammenfassend kann festgehalten werden, dass normativ (fachlich) reflektiertes Handeln gefordert wäre, die Sozialpädagogin in der Praxis aber die Erfahrung macht, dass keine Zeit zur Reflexion bleibt. Auch wirkt Reflexion nicht immer (er-)klärend. Dies könnten wir als „Graben zwischen (normativer) Anforderung und Praxis“ bezeichnen. Im Anschluss stellt sich die Folgefrage, ob dieser Graben irgendetwas zu bedeuten hat, ob sich daraus irgendwelche Konsequenzen ergeben. Deshalb könnte man sich in einem nächsten Forschungs- und Analyseschritt – im Sinne eines systematischen Vergleichs – auf die Suche nach Handlungsbeschreibungen machen, in denen es keinen Graben zwischen (normativen) Anforderungen und der Praxis gibt, um zu vergleichen und zu untersuchen, ob es Gemeinsamkeiten und Unterschiede zwischen den beiden Handlungssituationen gibt.

Axiales Kodieren. Das axiale Kodieren zielt darauf ab, Kategorien (von Phänomenen) zu verfeinern. Dabei wird nach den Ursachen des Phänomens gefragt. Ebenso wird untersucht, in welchen Kontext das Phänomen eingebettet ist, wie mit dem Phänomen umgegangen wurde oder was die Konsequenzen daraus sind (vgl. Strauss & Corbin, 1996, S. 76; vgl. Abbildung 12-5).

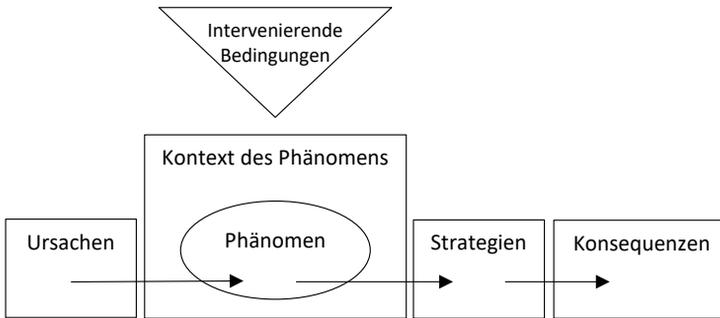


Abbildung 12-5: Kodierschema (modifiziert nach Strauss & Corbin, 1996, S. 78)

Gemäss dem Beispiel oben gilt: Für die beim offenen Codieren gewonnene (vorläufige) Kategorie „Graben zwischen (normativen) Anforderungen und Praxis“ ist noch nicht klar, ob es sich um die Ursache für etwas anderes oder um eine Konsequenz aus etwas handelt. Untersucht man dies auf der Basis des vorherigen Interviewausschnittes, stellt man fest: Die Sozialpädagogin musste rasch auf die Aggression des Klienten (= Ursache) reagieren (Phänomen: Handlungsdruck). Dies tat sie mit Routinehandeln (= Strategie) und die Konsequenz daraus ist der Graben zwischen den ihr bekannten Handlungsanforderungen und ihrem tatsächlichen Handeln. Daraufhin kann eine Hypothese formuliert werden, dass dieser „Graben“ zu Konflikten (Belastungserleben) führt. Diese Hypothese kann durch weiteres Datenmaterial überprüft werden.

Durch diesen Prozess bewegt man sich immer stärker zwischen induktivem und deduktivem Denken hin und her. Wie beim offenen Kodieren ist es wichtig, Fragen an den Text zu stellen und Vergleiche zu machen. Anders als beim offenen Kodieren geht man bei diesem Schritt aber fokussierter vor.

Selektives Kodieren. Der Prozess des selektiven Kodierens verdichtet schliesslich die im Vorfeld gewonnenen Kategorien zu einer „Grounded Theory“. Grundsätzlich unterscheidet sich dieser Integrationsprozess nicht sehr stark vom axialen Kodieren. Der Fokus des Kodierprozesses liegt nun aber auf einer Schlüsselkategorie, beachtet werden nur noch diejenigen Kategorien, welche einen bedeutenden Bezug zur ausgewählten Schlüsselkategorie haben. Der konkrete Ablauf dieses letzten Schrittes lässt sich folgendermassen beschreiben: Der erste Schritt des selektiven Kodierens

besteht im Offenlegen des *roten Fadens der Geschichte*. Der zweite besteht aus dem *Verbinden der ergänzenden Kategorien* rund um die *Kernkategorie* mit Hilfe des *Paradigmas*. Der dritte umfasst das *Verbinden der Kategorien auf der dimensional Ebene*. Der vierte beinhaltet das *Validieren dieser Beziehungen* durch die Daten. Der fünfte und letzte Schritt besteht im *Auffüllen der Kategorien*, die einer weiteren Verfeinerung und/oder Entwicklung bedürfen. (Strauss & Corbin, 1996, S. 95, Hervorhebungen im Original)

In Bezug auf das Beispiel kann man eine Theorie über das professionelle Handeln von Sozialpädagoginnen und -pädagogen entwickeln, mit der Routinehandeln erklärt wird. Da es im Rahmen vieler Forschungsarbeiten aus Zeit- und forschungsökonomischen Gründen nur selten um die Entwicklung einer neuen Theorie geht, dient dieser letzte selektive Schritt meistens dazu, die im Verfahren gewonnenen Hypothesen zu untermauern und in Form theoretischer Modelle zu präsentieren.

Memos. Der Prozess des Kodierens wird begleitet durch so genannte Memos. Hierbei handelt es sich um Einträge in ein „Forschungstagebuch“, in dem Ideen und Überlegungen zu Codes, Konzepten und Kategorien schriftlich ausformuliert und festgehalten werden (vgl. Flick, 2002, S. 259). Das Verfassen von Memos hilft den Forschenden Zusammenhänge zu erkennen, indem sie sich auch von den Daten distanzieren und bewusst eine analytische Perspektive einnehmen (vgl. Böhm, 2007, S. 477). Die in Form von Memos gewonnenen Textbausteine, ihre fortlaufende Überarbeitung, Systematisierung und Sortierung sind gerade auch im Hinblick auf das spätere Niederschreiben der Arbeit sehr hilfreich.

Theoretisches Sampling. Neben dem theoretischen Kodieren bildet das Theoretische Sampling ein weiteres Herzstück der Grounded Theory-Methodologie. Wie eingangs bereits formuliert, steht dabei die Frage im Zentrum, welcher nächste Fall das bereits gewonnene Wissen theoretisch sinnvoll erweitern und bereichern könnte.

Ich habe immer wieder diese Leute ... getroffen, die Berge von Interviews und Felddaten erhoben hatten und erst hinterher darüber nachdachten, was man mit den Daten machen sollte. Ich habe sehr früh begriffen, dass es darauf ankommt, schon nach dem ersten Interview mit der Auswertung zu beginnen, Memos zu schreiben und Hypothesen zu formulieren, die dann die Auswahl der nächsten Interviewpartner nahelegen. (Strauss in Legewie & Schervier-Legewie, 2004, Abs. 59).

Man geht ins Feld, erhebt erste Daten (z. B. Beobachtungen, Befragungen, Dokumente), analysiert diese und bildet erste Hypothesen. Dann geht man erneut ins Feld, um weitere Daten zu gewinnen, auf deren Basis man die Hypothesen weiterentwickeln kann (vgl. Hildenbrand, 2007, S. 476; Przyborski & Wohlrab-Sahr, 2014, S. 199ff). Die Stichprobe wird also laufend angepasst und erweitert. Dabei wird gezielt nach Daten und Fällen gesucht, welche zur Weiterentwicklung der Hypothesen und letztlich zur Sättigung der zu entwickelnden Theorie oder theoretischen Modelle beitragen (vgl. auch Kap. 5.4 „Stichproben“).

Kontinuierlicher Vergleich. Ein weiteres zentrales Konzept der Grounded Theory bildet die Idee des kontinuierlichen Vergleichens. Dabei kann es hilfreich sein, weit-hergeholte Vergleiche zu ziehen, d. h. Fälle stark zu kontrastieren, um Strukturen, Eigenheiten und Dynamiken erkennen zu können. Hughes (1970, zit. nach Star, 2010, S. 81) macht diese am Vergleich über berufliche Gemeinsamkeiten von Priestern und Prostituierten deutlich. Dieser Vergleich kann im ersten Moment vielleicht als fragwürdig erscheinen. Mit Fortschreiten der Arbeit wird aber erkennbar, dass die

Tätigkeiten einiges gemeinsam haben: „They both listen to people’s confessions; they work with people one-on-one (usually) in a setting that invites these sorts of intimacies, they listen rather than reveal their own lives, and so forth“ (Star, 2010, S. 81). In unserem Beispiel kann man Vergleiche über Situationen anstellen, bei denen es ebenfalls zu einem Graben zwischen bekannten Anforderungen und dem tatsächlichen Handeln kommt. So gibt es beispielsweise viele Menschen, die wissen, dass sie sich regelmässig körperlich bewegen sollten, sie aber dennoch lieber abends auf dem Sofa sitzen und fernsehen. Im Nachdenken hierüber fällt auf, dass – im Gegensatz zur Situation der Sozialpädagogin – der Entscheid fernzusehen meist bewusst und ohne Zeit- oder Handlungsdruck gefällt wird, dafür aber vor einem schlechten Gewissen begleitet sein kann. Diese Überlegung kann nun auf das Handeln der Sozialpädagogin übertragen und dabei festgestellt werden, dass wir es im ersten Fall mit Routinehandeln, als automatisiert ausgelöstes Handeln zu tun haben, während dessen das allabendliche Fernsehen eher als bewusst ausgelöstes Handeln bezeichnet werden kann, mit dem eine bestimmte Handlungspräferenz („lieber Fernsehen als Sport treiben“) verfolgt wird. Damit haben wir eine erste (vorläufige) Erkenntnis über die Beschaffenheit von Routinehandeln gewonnen.

12.3.3 Kritische Würdigung

Forschen im Stil der Grounded Theory erfordert Kreativität auf Seiten der Forschenden. Strauss und Corbin (1996, S. 39ff) betonen, dass nicht rigide an den von ihnen vorgestellten Verfahren und Techniken festgehalten werden soll, sondern dass man die Verfahren dem Forschungsgegenstand anpassen müsse. Dies kann gerade bei Forschungsanfängerinnen und -anfängern zu Verunsicherung führen. Die Begründer des Ansatzes machen Novizinnen und Novizen jedoch auch Mut, indem sie betonen, dass „die Geschicklichkeit in der Anwendung der Grounded Theory mit ihrer kontinuierlichen Übung und Praxis“ (Strauss & Corbin, 1996, S. 10) entsteht. Strauss (1998, S. 100) plädiert gar für eine gewisse Gelassenheit im Forschungsprozess: „Machen Sie sich keine Sorgen, fast jeder Weg bringt brauchbare Ergebnisse“ (Strauss, 1998, S. 100).

Auch die Frage nach dem Ende des Forschungsprozesses kann Unsicherheiten erzeugen. Im Grundsatz wird der Prozess der Datensammlung und -analyse nur so lange weitergeführt, bis eine theoretische Sättigung erreicht ist. Das bedeutet, dass irgendwann ein Punkt erreicht wird, an welchem der Beizug weiterer Fälle keine neuen Erkenntnisse mehr bringt (vgl. Corbin & Strauss, 1996, S. 159). Breuer (2010, S. 110) betont richtigerweise, dass „ein solches finales Urteil mit Gewissheit von keinem Forschenden gefällt werden“ kann. Es existiert kein objektiver Abschluss-Zeitpunkt, sondern dieser muss auf der Basis plausibler Argumente legitimiert und damit nachvollziehbar gemacht werden (vgl. Strübing, 2008, S. 34). Dies gilt insbesondere auch im Kontext von Abschlussarbeiten, welche innerhalb einer bestimmten Zeitdauer verfasst werden müssen.

12.4 Objektive Hermeneutik

Die objektive Hermeneutik (OH) ist eine Methode der qualitativen Sozialforschung, die sich auf die Interpretation von Texten konzentriert. Die Methode wurde von Ulrich Oevermann und seinen Mitarbeitenden in den 1970er Jahren entwickelt und zielt darauf ab, die Bedeutung eines Textes nicht allein auf Basis der Intention einer Person zu rekonstruieren, sondern es wird versucht, die impliziten Bedeutungen und Strukturen zu identifizieren, die in dem Text verankert sind. Der Begriff Hermeneutik stammt aus dem Griechischen und bedeutet die „Kunst der Auslegung, der Deutung“. Vom Begriff und Inhalt der konventionellen Hermeneutik unterscheidet sich die OH besonders bezüglich der im Namen enthaltenen Objektivität. Die OH bezieht sich nämlich nicht auf die (subjektive) Absicht hinter einer Aussage, sondern auf die tatsächliche und in diesem Sinn objektive Bedeutung eines Textes. Dieser Anspruch auf Objektivität mutet auf den ersten Blick reichlich überheblich an und führt bisweilen zu Missverständnissen. Ein kurzes Beispiel kann aber verdeutlichen, von welcher Art Objektivität die Rede ist: So ist ein Schraubenzieher primär dafür da, Schrauben einzudrehen. Seine objektive Bedeutung liegt also in seiner Funktion. Für diese Bedeutung spielt es keine Rolle, wer das Werkzeug verwendet und wie es der Handwerkerin bei der Arbeit erging. Der objektiven Hermeneutik geht es also nicht um die Rekonstruktion des subjektiven Gehalts einer Aussage, sie interessiert sich nicht für Absichten, Meinungen oder Wünsche, sondern vielmehr für die faktische Bedeutung einer Aussage.

12.4.1 Begriffe und konzeptuelle Hintergründe

Der Unterschied zwischen der objektiven Bedeutung und den subjektiven Absichten und Meinungen ist für die Analyse mittels OH entscheidend: Ein Text erzeugt Bedeutung, „die jenseits des Selbstverständnisses und Selbstbildes einer sozialen Praxis [bspw. Handlungen oder Äusserungen einer Person] liegen und die sich nicht in den Meinungen, Intentionen oder Wertorientierungen dieser Praxis erschöpfen“ (Wernet, 2009, S. 18). Anders als bei anderen qualitativen Verfahren, steht bei der objektiven Hermeneutik deshalb nicht der oberflächliche Informationsgehalt eines Textes oder einer Aussage im Zentrum, sondern die dahinterliegende Bedeutung. Diese gilt es zu rekonstruieren.

► Beispiel „Interaktion Lehrerin – Schüler“ (Teil I)

Die Lehrerin K. tritt vor die Klasse und sagt: „Guten Tag miteinander!“ Für die OH ist anfänglich nicht interessant, was K. mit ihrer Äusserung tatsächlich bezwecken wollte. Wichtiger ist vielmehr, welche Funktionen eine solche Aussage im Kommunikationszusammenhang – im sogenannten Interaktionssetting – haben kann. Denkbar und plausibel sind: (1) Die Lehrerin K. signalisiert damit, dass die Schulstunde offiziell begonnen hat, oder (2) sie will das allgemeine Geplauder unterbrechen und damit die Aufmerksamkeit der Schülerin-

nen und Schüler auf sich ziehen. Als Leseart (3) wäre auch plausibel, dass die Lehrerin K. damit eine rituelle Begrüssung zu Tagesbeginn durchführt. Die Reaktion der Schülerinnen und Schüler wird zeigen, welche der möglichen Bedeutungen verstanden und aufgenommen wurden.

Historie und Idee. Der Begründer dieser Forschungsmethode (zur Geschichte der OH vgl. u. a. Reichertz, 1997) entwickelte die Methode in den 1960er-Jahren explizit im Bereich der Lehrerbildungsforschung, und zwar im Zuge eines Forschungsprojekts zu „Elternhaus und Schule“ (Krappmann et al., 1968). Oevermann und Mitarbeitende entwickelten dabei ein Verfahren, das die Interaktionen zwischen Eltern und Kindern besser erfassen sollte. Im Zuge gründlicher Textinterpretationen stellten sie fest, dass Personen zwar immer in bestimmte Muster eingebunden sind und Konventionen folgen, dass sie sich aber dieser Zusammenhänge nicht stets bewusst sind. Aus dieser Beobachtung entwickelte sich eine erste wichtige theoretische Grundlage der objektiven Hermeneutik: Unser soziales Handeln richtet sich immer nach Regeln, die jedoch nicht immer bewusst wahrgenommen werden. Wer beispielsweise stimmig etwas zum Ausdruck bringen will, bedient sich den grammatikalischen, semantischen und pragmatischen Regeln der Sprache, ohne sich diese immer bewusst zu machen. Handelnde können sich diesen Regeln weder entziehen, noch können sie die Regeln ausser Kraft setzen, sie können die regelgeleitete Welt nicht verlassen. Denn auch wer gegen die Regeln handelt, nimmt sie trotzdem als Basis. Im Sinne des oben eingeführten Beispiels hiesse dies:

► Beispiel „Interaktion Lehrerin – Schüler“ (Teil II)

Die Lehrerin K. sagt: „Guten Tag miteinander!“. Diese Aussage wird von den Schülerinnen und Schülern als Begrüssung identifiziert und erfordert im Sinne der oben erklärten Regelgeleitetheit sozialen Handelns deshalb eine Reaktion („Guten Tag“ oder „Guten Tag Frau K.“, oder „Salut“). Als eine plausible und mögliche Reaktion zählt dabei auch, wenn (einige oder alle) Schülerinnen und Schüler sich dazu entschliessen, nichts zu sagen. Denn die Regel gilt auch dann, wenn man die geltenden Regeln – eine Begrüssung verlangt immer nach einer wie auch immer gearteten Erwidrung – verletzt. Eine Nicht-Erwidrung des Grusses gilt somit auch als eine (wie auch immer bewusste oder unbewusste) Reaktion, und kann entsprechend gedeutet werden: Wird die Lehrerin K. bewusst geschnitten? Herrscht ein Klassenklima, das dazu führt, dass eintretende erwachsene Personen nicht zur Kenntnis genommen werden?

Das Konzept der Regel ist nicht zu verwechseln mit den sozialen Normen, denn „das Konzept der Regelgeleitetheit formuliert, anders als soziale Normen, nicht, was zu tun ist“ (Wernet, 2000, S. 134), sondern nur, was es bedeutet etwas (nicht) zu tun. Regelgeleitetheit meint damit zuerst einmal, dass menschliche Handlungen nicht beliebig aneinandergereiht sind, sondern sie stehen immer in einer Abfolge. Oder um es mit einer Metapher zu erklären: Wenn ich mich für den Weg A ent-

scheide, gehe ich diesen weiter. Wenn ich ihn später verlasse oder anhalte, ist diese erste Wahl dennoch ein erster Referenzpunkt und alles weitere nimmt darauf Bezug. Insofern haben Handlungen immer eine innere Logik (ähnlich der Logik im Sprachgebrauch). Regelgeleitetheit bedeutet auch, dass es nichts Bedeutungsloses in der sozialen Welt gibt. Denn unsere Handlungen sind im wörtlichen Sinn immer „sinnvoll“. Drittens schaffen es erst die Regeln, dass sich die Möglichkeit von Alternativen bildet: So ergeben sich beispielsweise aus der Regel „eine Begrüssung verlangt nach einer Reaktion“ etliche Wahlmöglichkeiten: Ich erwidere den Gruss freundlich oder grimmig, förmlich oder freundschaftlich oder lasse den Gegengruss aus (denn auch die Verweigerung stellt eine Möglichkeit der Erwidern dar).

Angenommen wird dabei schliesslich, dass die vom Handelnden vorgenommenen Entscheidungen – ich reagiere auf die Begrüssung in dem Moment so und nicht anders – nicht willkürlich sind, sondern dass sie in spezifische Kontexte eingebettet sind und dabei jeweils einer bestimmten Logik folgen, welche als Identität oder als Habitus bezeichnet wird. Die Rekonstruktion dieser Logik nennt Oevermann „Fallstruktur“. Bei der Fallstruktur handelt es sich um das Kernprinzip eines Falls, auf dessen Suche der oder die Forschende ist.

Text als Ausgangsbasis. Um für eine Analyse einen Geltungsanspruch einzulösen (vgl. Gütekriterium „intersubjektive Überprüfbarkeit“, Kap. 12.6) muss sich die Wissenschaft allgemein an Protokolle halten. Denn nur so können Resultate für verschiedene Personen nachvollziehbar gemacht werden. Für die OH stellt der Ausgangspunkt der Analyse stets eine wie auch immer geartete Form von Text dar: „Jenseits von Texten hat die Wissenschaft ihr Recht verloren, da wissenschaftliche Aussagen erst dann formuliert werden können, wenn und insoweit Ereignisse Niederschlag bzw. eine Spur hinterlassen und diese wieder eine Interpretation erfahren haben“ (Garz & Kraimer, 1994, S. 8). Die OH geht also davon aus, dass die Welt durch Sprache konstituiert ist und demnach in Texten fassbar gemacht werden kann (Oevermann, 1986, S. 45). Eine Beobachtung selbst ist nach Oevermann wissenschaftlich nicht relevant, sondern sie wird es erst durch ihr Festhalten in irgendeiner Form, vorzugsweise als Text. Das heisst nun nicht, dass nicht-sprachliches Material (z. B. ein Bild, Tonaufnahmen oder eine Beobachtung) nicht bearbeitet werden kann. Es meint aber, dass diese Informationen zuerst versprachlicht werden müssen, damit sie für die wissenschaftliche Interpretation zugänglich gemacht werden können.

Sequenzanalyse, Wörtlichkeit und Kontextfreiheit. Es gibt verschiedene Verfahren der objektiv-hermeneutischen Textinterpretation (Hitzler & Honer, 1997, S. 38ff), welche aber trotz Differenzen ein gemeinsames Grundverständnis haben: Den eigentlichen Kern bildet die Sequenzanalyse. Hierbei ist jeder einzelne aufeinander folgende Interaktionsbeitrag wichtig und man arbeitet sich kontinuierlich im Text vor, und dies jeweils ohne Kontextwissen über den Gegenstand beizuziehen: „Der Fall lehrt den Wissenschaftler und nicht umgekehrt“ (Reichert, 1997, S. 42). Dies bedeutet nicht, dass die Kontexte einer Aussage oder Handlung nicht von Bedeutung

sind – im Gegenteil, sie spielen eine tragende Rolle bei der Einbettung. Das Kontextwissen soll jedoch erst nach der erfolgten Interpretation beigezogen werden.

Als weiteres Prinzip bei der Analyse gilt die Wörtlichkeit. Das heisst, dass die Interpretation auch dann nicht vom Text abweichen darf, wenn dieser in sich widersprüchlich wird: „Das Prinzip der Wörtlichkeit verpflichtet die Interpretation, den Text ‚auf die Goldwaage zu legen‘ in einer Weise, die uns in alltäglichen Verstehenskontexten als inadäquat und kleinlich erscheinen würde“ (Wernet, 2009, S. 23). Die beiden Prinzipien, Sequenzialität und Wörtlichkeit, hängen damit zusammen. Bei der Textanalyse mittels OH springt man zuerst nicht zwischen Textstellen hin und her, um nach geeigneten Passagen zur Interpretation zu suchen, sondern man folgt dem Text Schritt für Schritt. Die Devise lautet, den vorliegenden Text ernst zu nehmen – bezüglich der Wortwahl ebenso wie bezüglich der Reihenfolge der zu Text gebrachten Gedanken (Argumentationskette) – und ihn nicht nach brauchbaren Stellen „auszuschlachten“. Diese minutiöse Analyse zieht mit sich, dass beim Protokollieren von Gesprächen auf jedes Detail geachtet werden muss. Dadurch soll gewährleistet werden, dass keine Selektion oder Vorinterpretation seitens der Forschenden vorgenommen wird. Ein Wandern im Text ist erst dann unproblematisch, wenn durch die erste Sequenzanalyse eine erste Fallstruktur formuliert worden ist. Das Gebot der Sparsamkeit schliesslich besagt, dass alle gedankenexperimentellen Interpretationen (so genannte Lesarten) auszuschliessen sind, welche nicht von vorneherein mit dem Text kompatibel sind. Es sollen also nur solche Lesarten gebildet werden, die plausibel sind und sich ohne weiteres in den Zusammenhang des Textes stellen lassen. Damit werden der Interpretation Grenzen gesetzt.

12.4.2 Vorgehensweise

Im Folgenden wird eine auf drei Schritte reduzierte Vorgehensweise vorgestellt, welche den gemeinsamen Kern der verschiedenen Verfahren repräsentiert. Das konkrete Analysevorgehen wird zudem am Beispiel einer Interviewpassage aus Wernet (2009, S. 47ff) dargestellt.

Die drei wesentlichen Schritte umfassen:

- 1) Die Interpretation der objektiven Daten,
- 2) die Feinanalyse, und
- 3) die Bildung einer Strukturhypothese.

„Qualitative Interpretationen lassen sich nicht nach der Logik von Computerprogrammen organisieren. Die Analyseschritte sollen primär zur Sorgfalt der Explikation anleiten“ (Przyborski & Wohlrab-Sahr, 2014, S. 269). Insofern handelt es sich hierbei nicht um eine strikt festgelegte Abfolge von Schritten, die mechanisch befolgt und immer in der gleichen Weise praktiziert werden.

Interpretation der objektiven Daten. Beim ersten Schritt geht es darum, den Rahmen eines Textes (Interview oder andere Textsorte) zu beschreiben, um damit

den Kontext einer Interaktion oder Aussage in ihren korrekten Zusammenhang einzubetten. Mit „objektiven Daten“ sind bspw. Angaben über die soziale Herkunft (Bildungsbiografie, Beruf, Einkommen etc.) und die familiäre Einbettung (Beruf der Eltern, Geschwister) einer Person oder die „Beschaffenheit“ des Milieus (bspw. Anzahl Klassen in einer Schule, Anteile mehrsprachiger Schülerinnen und Schüler, sozioökonomische Daten der Schule) gemeint. Je nach Art des Materials, das untersucht werden soll, gehören dazu auch Vorstellungen über die Beschaffenheit eines Textes. Dies in der begründeten Annahme, dass ein Gesetzestext, ein Lehrplan oder ein Schreiben der Schulleitung nicht dieselben Regeln befolgen müssen, resp. sich nicht derselben Sprache und Konventionen bedienen.

Die Interpretation solcher Kontextdaten wird in der Regel zur Erzeugung von ersten, sehr allgemeinen Hypothesen gebraucht und dient im Sinne von „Normalitätsfolien“ zur Einbettung möglicher Interpretationen sowie zur Spezifizierung der Fragestellung. Voraussetzung für diesen Schritt ist natürlich, dass die objektiven Daten erhoben werden bzw. die Informationen über ein Forschungsfeld vorhanden sind.

► Beispiel „Klassenarbeiten“

Schüler (S): Wann geben Sie uns die Klassenarbeiten wieder?

Lehrperson (L): Nächste Woche.

S: Oh, Sie haben sie doch schon drei Wochen.

L: Und wenn ich sie fünf Wochen hätte.

S: Meine Mutter denkt schon, ich hätt die weggeschmissen

Die objektiven Daten, die hier interessieren, sind nicht die biographischen Angaben zur Lehrperson und zum Schüler, sondern das Interaktionssetting „Lehrperson-Schüler“. Folgendes kann für die Analyse festgehalten werden:

- 1) Die protokollierte Interaktion findet in einem Schulzimmer und nicht im privaten Rahmen statt. Diese an sich triviale Beschreibung ist für die spätere Deutung insofern wichtig, als dass beispielsweise ein Interview zwischen Vater und Sohn, zwischen Anwalt und Klient etc. in der Regel andere Interaktionen erwarten lassen würden.
- 2) Das zu erwartende Kommunikationsmuster, die Art wie die Beteiligten miteinander sprechen, ist – ungeachtet des subjektiven Erziehungsstils – nicht ein gleichberechtigtes (symmetrisches). Im schulischen Kontext gibt grundsätzlich die Lehrperson die Rahmenbedingungen vor – und dies gilt selbst dann, wenn sie sich für partizipativen Unterricht entscheidet. Diese Asymmetrie im Schulalltag (Hierarchie) typisiert den Lehrberuf insofern, als dass sie ein typisches Spannungsfeld von Unterrichtsarbeit darstellt (prototypisch: Meister-Schüler-Verhältnis).
- 3) Für die Einbettung der Interaktion ist weiter wichtig, dass sich die Kommunikation in einem Klassenraum abspielt, also in einem (teil-)öffentlichen Raum.

Mit diesen ersten Ergebnissen kann man sich nun der Textsequenz zuwenden, in der Erwartung, etwas über dieses Spannungsfeld im Lehrberuf zu erfahren. Wichtig ist, dass die Interpretation der objektiven Daten die weiteren Interpretationen nicht determinieren, d. h. nicht zu eng begrenzen sollte, weil man ansonsten gegen das Prinzip der Kontextfreiheit (siehe oben) verstossen würde.

Natürlich ist es auch möglich, eine Interpretation vorzunehmen, ohne diese vorbereitenden Operationen auszuführen. Die Interpretation der objektiven Daten hilft jedoch, sich den Kontext des Datenmaterials zu verdeutlichen und in dieser Hinsicht forschungsökonomisch sowie forschungslogisch vorzugehen: Schliesslich sollen nicht alle möglichen Interpretationen bis ins Detail durchgearbeitet werden, sondern nur diejenigen, die für die Fragestellungen und die konkrete Untersuchung von Relevanz sind.

Feinanalyse. Die Feinanalyse beginnt grundsätzlich mit dem Anfang eines Texts (vgl. oben, Sequenzialitätsprinzip). Dies begründet sich darin, dass anfangs Weichen gestellt werden, die den weiteren Verlauf einer Argumentation massgeblich beeinflussen: „In lebensgeschichtlichen Interviews z. B. präsentieren sich die befragten Personen zu Beginn häufig in besonders verdichteter Form, so dass hier geradezu das ‚Motto‘ einer Lebensgeschichte erkennbar wird“ (Przyborski & Wohlrab-Sahr, 2014, S. 268). Auch bei anderen Texten (Geschichten, Werbebotschaften etc.) werden die zentralen Argumente und Ideen oft zu Beginn dargelegt.

Die Feinanalyse durchläuft in der Regel zwei Schritte: Erstens Geschichten erzählen und Lesarten bilden sowie zweitens der Vergleich mit real existierendem Text. Beim „Geschichten erzählen“ geht es um die gedankenexperimentelle Produktion von Interpretationsmöglichkeiten: Um herauszufinden, welche Bedeutungen in einer Aussage stecken, erfinden die Forschenden Geschichten, in denen dieselbe Aussage vorkommen könnte. Dies hilft dabei, den Fall unvoreingenommen zu analysieren. Die zentrale Frage, die das Geschichten erzählen leitet, ist also: Was kann das Gesagte alles bedeuten? Dieser Schritt verlangt ein Verständnis von allgemeinen Bedeutungen einzelner Begriffe sowie von möglichen Verwendungsweisen einzelner Wendungen und Formulierungen. Ein Wörterbuch, vor allem aber das gemeinsame Analysieren des Textes in einer Gruppe kann hier grosse Hilfe leisten.

Auf der Grundlage solch konstruierter Geschichten erfolgt die Lesartenbildung, indem die Geschichten auf ihre Gemeinsamkeiten hin untersucht werden. Bei diesem Schritt der Interpretation, bei dem es um die Bildung von adäquaten Lesarten (meistens kann ein Textausschnitt verschieden verstanden und gedeutet werden) geht, soll dabei gerade nicht interessieren, welchen Sinn und welche Intention die sprechende Person selbst mit seiner Aussage verbindet. Denn wie bereits mehrfach erwähnt, rekonstruiert die objektive Hermeneutik den „objektiven“ Gehalt einer Aussage und nicht den subjektiv gemeinten oder intendierten Sinn (deshalb „objektive“ Hermeneutik).

Hilfreich bei solchen Gedankenexperimenten sind folgende Leitfragen (vgl. Klee-
mann et al., 2009, S. 128ff):

- Welche Bedeutung steckt in einem Wort, in einem Satz? (evtl. unter Zuhilfenahme eines Wörterbuchs)
- Welche unterschiedlichen Bedeutungen (Konnotationen) hat ein Begriff?
- Gibt es Auffälligkeiten in der Wortwahl?
- In welchen Kontexten verwendet man normalerweise solche Begriffe und Formulierungen?

► Beispiel „Klassenarbeiten“

Welche sinnmachenden Geschichten und Lesarten lassen sich für die Aussage des Schülers: „Wann geben Sie uns die Klassenarbeiten wieder?“ also erfinden?

Geschichte 1: Die Arbeit wurde gerade geschrieben und ein Schüler erkundigt sich nach dem Rückgabetermin.

Geschichte 2: Die Arbeit wurde noch nicht geschrieben und der Schüler erkundigt sich nach dem Terminplan.

Lesart 1: Bei beiden Geschichten handelt es sich um eine reine Informationsfrage, und die Antwort darauf müsste die Information wiedergeben.

Geschichte 3: Die Arbeit wurde bereits geschrieben. Der Rückgabetermin ist bekannt und die Lehrperson hat sich nicht darangehalten.

Geschichte 4: Die Arbeit wurde bereits geschrieben. Üblicherweise werden Klassenarbeiten von der Lehrperson in einer bestimmten Frist bearbeitet, was diesmal aber nicht der Fall ist.

Lesart 2: In beiden Fällen würde ein Versäumnis vorliegen und der Schüler unternimmt mit seiner Frage eine Kritik im Sinne von „Wann geben sie uns die Arbeit endlich zurück?“ Falls die Kritik berechtigt wäre, müsste die Antwort auf die Beschwerde eingehen und unter Umständen eine Entschuldigung beinhalten.

Geschichte 5: Die Arbeit wurde bereits geschrieben, ein Termin für die Rückgabe ist bekannt und es liegt kein Versäumnis seitens der Lehrperson vor.

Lesart 3: Die Frage würde eine Gegenfrage aufwerfen: Warum fragt der Schüler? Will er sich beschweren (Kritik) oder ist etwas eingetreten, das den Rückgabetermin verändert (Information)?

Die dargestellten Lesarten (1–3) lassen sich auf zwei Bedeutungen reduzieren: Entweder die Frage zielt auf einen Informationsaustausch oder sie zielt auf die Frage der Verbindlichkeit, die durch ein mögliches Versäumnis der Lehrperson verletzt wurde. Bei Letzterem kann die Lehrperson durch eine Entschuldigung oder eine Erklärung die verletzte Verbindlichkeit zumindest anerkennen und reparieren.

Ein Dauerproblem der OH besteht in der Frage, wann die Lesarten erschöpft sind. Diese Frage lässt sich nicht absolut bestimmen, sondern entscheidet sich letztlich an der Analyse. Grundsätzlich zeigt sich, dass die Feinanalyse zu Beginn viel Zeit erfordert, wenn man sicher gehen will, dass man alle möglichen Lesarten entwickelt

hat. Als weitere „Absicherung“ empfiehlt es sich zudem, die Analysearbeiten in einer Gruppe vorzunehmen. Letztlich begrenzt das Sparsamkeitsprinzip (siehe oben) aber die Anzahl möglicher Geschichten und schliesst bspw. folgende Lesart von vornherein aus:

Geschichte 6: Der Schüler hat ein angespanntes Verhältnis zur Lehrperson und will sie ärgern, indem er eine unsinnige Frage stellt.

Diese Geschichte ist zwar möglich, jedoch insofern unpassend, weil sie auf jede denkbare Interaktion anwendbar ist und deshalb kaum etwas über die hier vorliegende Situation auszusagen vermag. Würde man die Lesart beibehalten wollen, dass der Schüler ein notorischer Querulant sei, müsste sie zumindest durch weitere Textstellen belegt werden können.

Im zweiten Schritt geht es darum, die formulierten Lesarten mit dem tatsächlichen Text zu vergleichen. Aufschlussreich ist dabei nicht nur, welche Aussage im Anschluss tatsächlich realisiert wird, sondern analysiert werden die nicht realisierten Lesarten. Denn informativ und beachtenswert ist sowohl das, was gesagt wird, als auch das, was eben nicht gesagt wurde. So oder so steht bei diesem Schritt die Frage im Zentrum: Was kann es bedeuten, dass gerade diese Antwort gewählt wurde und nicht eine andere, ebenfalls plausible?

► Beispiel „Klassenarbeiten“

Schüler (S): Wann geben Sie uns die Klassenarbeiten wieder?

Lehrperson (L): Nächste Woche.

Die Antwort der Lehrperson lässt vorerst darauf schliessen, dass es sich um Informationsaustausch handelt: Für die Rückgabe nennt sie den Termin „nächste Woche“.

S: Oh, Sie haben sie doch schon drei Wochen.

L: Und wenn ich sie fünf Wochen hätte.

S: Meine Mutter denkt schon, ich hätt die weggeschmissen

Die dritte Zeile und die folgende Reaktion der Lehrperson zeigen dann aber, dass es sich bei der Frage des Schülers eben nicht um einen Wunsch nach Information handelt, sondern dass beide, Lehrperson und Schüler, davon ausgehen, dass die Lehrperson etwas versäumt hat. Die erste Antwort der Lehrperson, die Information, stellt also eine „Regelverletzung“ dar, weil sie die in der Frage formulierte Kritik nicht beachtet hatte. Hätte sie mit „ja, ich weiss, ich bin spät dran“ geantwortet, wäre der Dialog wohl früher beendet gewesen. Im vorliegenden Fall wird der Schüler in diesem Sinn „regelkonform“ fast dazu gedrängt, seinen Unmut nochmals zu formulieren, da er davon ausgeht, dass die Lehrperson diesen als solchen nicht verstanden hat oder nicht verstehen will.

Hypothesenbildung. Aus der Interpretation einer ganzen Sequenz ergibt sich nun ein erstes Muster in Bezug auf das Lehrpersonenhandeln. Die Hypothese, die so genannte Fallstruktur, formuliert die Zusammenhänge und setzt sie in Bezug zur Fragestellung.

► Beispiel „Klassenarbeiten“ – Hypothesenbildung

Zum Spannungsfeld im Lehrpersonenhandeln kann aufgrund dieses kleinen Textausschnitts folgendes festgehalten werden: Die Lehrperson, welche die Frage des Schülers beantwortet, erzeugt ein Interaktionsproblem, und zwar ohne, dass es einen zwingenden Grund dafür gäbe.

Ein solcher erster Befund kann nun weiterverwendet werden und es kann überprüft werden, ob sich die Hypothese auch in nachfolgenden Interaktionssequenzen bestätigt. Die Fortsetzung der Feinanalyse an den jeweiligen Folgesequenzen dauert so lange, bis eine schlüssige Hypothese über den jeweils vorliegenden Fall (Fallstrukturhypothese) ausformuliert werden kann. Schlüssigkeit bedeutet demnach auch, dass alternative, bis dahin ebenfalls textkompatible Lesarten, ausgeschlossen werden können.

In der Praxis zeigt sich, dass drei weitere Segmente in der Regel reichen, um die Hypothese zu festigen. Der Rest eines Textes oder Interviews dient dann nunmehr der punktuellen Überprüfung der Fallstrukturhypothese.

12.4.3 Kritische Würdigung

Das Verfahren der OH eignet sich für die Ausarbeitung von Typen (bspw. von Lehrpersonenhandeln oder von typischen Interaktionsmustern zwischen Eltern und Lehrpersonen etc.). Mit Hilfe des Theoretical Samplings (siehe oben) können weitere Fälle nach dem Kriterium des maximalen Kontrasts gesucht werden. Bei einem solchen Vorgehen sind auch generalisierende Aussagen möglich, sofern nicht nach der Häufigkeit des rekonstruierten Typs gefragt wird, sondern eine vollständige Typologie gesucht wird (bspw. es gibt fünf Typen von Lehrpersonen).

Das Markenzeichen der OH ist die detaillierte Rekonstruktion von Sinn. Dabei werden nicht nur die tatsächlich vorgefundenen Bedeutungen herausgearbeitet, sondern es werden wie oben am Beispiel der Lesarten gezeigt, auch andere mögliche Situation mitgedacht und beleuchtet. Angebracht ist ein solch ganzheitliches Vorgehen dort, wo wenig über ein Phänomen bekannt ist und ein Feld explorativ erschlossen werden soll.

Die OH findet man deshalb oft in pädagogischen Forschungsarbeiten (vgl. Kutschke, 2011), aber auch in der Biographie- und der Milieuforschung. Weiter gibt es gibt aber auch eine Reihe von klinischen Anwendungsfeldern (Polizeiarbeit, klinische Psychotherapie, Unternehmensberatung, Sozialarbeit etc.), in denen mit OH gearbeitet wird. Die Schwierigkeiten, und damit auch die Hauptkritik, beziehen sich auf die Genauigkeit beim Vorgehen: Bei der OH handelt es sich um eine „Kunstlehre“, da es keine präzisierten methodischen Standards gibt. Für eine verlässliche Bildung von Bedeutungen eines Textes (Lesarten) empfiehlt es sich daher, in einer Gruppe zu arbeiten. Zudem braucht es, um das Prinzip der Wörtlichkeit ernst zu nehmen, vertiefte Kenntnisse der Sprache oder des Dialekts, in dem der Text produziert wurde. Die Analyse eines fremdsprachlichen Textes schliesst sich damit eigentlich aus. Was

für die Sprachkenntnisse gilt, zählt auch für den untersuchten Gegenstand: Um die in einem spezifischen Feld gültigen Kontexte als Kontrastfolie nutzen zu können, braucht es unter Umständen einige Zeit und Mühe, sich dieses Wissen anzueignen. Denn: Viele (subjektive) Meinungen machen zusammen noch keine objektive.

12.5 Generalisierbarkeit qualitativer Analysen

Die Frage der Generalisierbarkeit qualitativer Analysen ist eine wichtige Diskussion in der qualitativen Forschung. Im Gegensatz zur quantitativen Forschung, die auf statistische Generalisierbarkeit abzielt, beabsichtigt die qualitative Forschung, die Tiefe und Komplexität von Phänomenen zu verstehen.

Klar ist, dass qualitative Auswertungsmethoden keine empirische Generalisierung im Sinne von Häufigkeitsverteilungen zulassen. Jedoch ist es sehr gut möglich und auch erwünscht, typische Muster und Strukturen im Hinblick auf einen bestimmten Forschungsgegenstand zu rekonstruieren. „An die Stelle von Gültigkeitskriterien wie ‚Repräsentativität‘ oder ‚Signifikanz‘ tritt das Gültigkeitskriterium der ‚Sättigung‘“ (Schallberger, 2005, S. 16f). Es geht also bei der spezifischen Art der Generalisierung nicht um die Frage, wie häufig ein bestimmtes (Handlungs-) Muster oder ein bestimmter (Handlungs-) Typus in der Wirklichkeit vorkommt, sondern darum, die Typen überhaupt erst einmal ausfindig zu machen und in ihrer Beschaffenheit zu beschreiben (Brüsemeister, 2008, S. 30).

Qualitative Forschung kann also dazu beitragen, allgemeine Erkenntnisse zu gewinnen, indem sie beispielsweise spezifische Kontexte untersucht, die für ein bestimmtes Phänomen (Interaktion, Prozess etc.) relevant sind. Durch die Untersuchung dieser Bedingungen und Kontexte können Forscherinnen und Forscher bestimmte Muster und Zusammenhänge identifizieren, welche für ein ganzes Feld ihre Gültigkeit entwickeln. So kann qualitative Forschung dazu beitragen, die Komplexität von Phänomenen zu verstehen und allgemeine Erkenntnisse zu gewinnen, indem sie spezifische Kontexte und Bedingungen untersucht und Theorien und Modelle entwickelt, die auf andere Kontexte und Gruppen übertragbar sind.

In jedem Fall kann die Übertragbarkeit von Befunden durch die Ausführlichkeit und Sorgfalt in der Beschreibung der Forschungsmethoden und -ergebnisse erhöht werden. Eine ausführliche Beschreibung der Forschungsmethoden und -ergebnisse ermöglicht es anderen Forscherinnen und Forschern, die Ergebnisse der Studie auf ihre eigenen Forschungsfragen und -kontexte anzuwenden.

12.6 Gütekriterien qualitativer Forschung

Die Frage danach, welche Qualität („Güte“) Forschungsergebnisse und -erkenntnisse aufweisen, wird mittels sogenannter Gütekriterien zu definieren versucht. Die Gütekriterien von qualitativer und quantitativer Forschung unterscheiden sich primär

aufgrund ihrer Zielsetzung und den davon abgeleiteten Verfahren der Datenerhebung und -analyse. Während quantitative Forschung darauf abzielt, durch systematische Erhebung und Analyse von Daten statistische Muster zu identifizieren und allgemeine Aussagen über eine Population zu treffen, zielen qualitative Verfahren auf die Entwicklung eines tieferen Verständnisses von Phänomenen und ihren Kontexten. Die für die quantitative Forschung geltenden drei Gütekriterien der Objektivität, Reliabilität (Verlässlichkeit) und Validität (Gültigkeit der Ergebnisse; vgl. Kap. 5.3.3 „Gütekriterien quantitativer Forschung“) können deshalb nicht einfach auf qualitative Verfahren übertragen werden.

Die Gütekriterien qualitativer Forschung zielen vor allem auf die „prozedurale Verlässlichkeit“ bzw. „Verfahrensrationaltät“ (Flick, 2007, S. 501) des Forschungsprozesses. Deshalb ist es wichtig, dass zentrale Entscheidungen vor allem in den Phasen der Datenerhebung und -auswertung im Forschungsbericht dokumentiert und begründet werden. Insgesamt sind dabei folgende Aspekte zu beachten:

- Intersubjektive Nachvollziehbarkeit
- Angemessenheit des Forschungsprozesses
- Empirische Verankerung
- Reflektierte Subjektivität (vgl. Steinke, 2007, S. 323ff).

Intersubjektive Nachvollziehbarkeit. Der Forschungsprozess muss so dokumentiert werden, dass er auch von anderen Personen nachvollzogen werden kann. Das heisst, sowohl das Vorverständnis der Forschenden als auch die Methoden der Datenerhebung, -aufbereitung und -auswertung müssen dargestellt und begründet werden. Die intersubjektive Nachvollziehbarkeit kann erhöht werden, indem Textinterpretationen nicht alleine, sondern in Forschungsteams durchgeführt werden (vgl. Steinke, 2007, S. 324ff).

Angemessenheit des Forschungsprozesses. Qualitative Forschung muss stets der Fragestellung und dem Forschungsgegenstand angemessen sein. Das heisst, es muss begründet werden, welche Fälle ausgewählt und untersucht werden (Theoretical Sampling), welche Methoden verwendet werden und wie diese zueinander passen (vgl. Steinke, 2007, S. 326ff).

Empirische Verankerung. Hypothesen sollen aus den Daten heraus entwickelt werden. Ihre Überprüfung erfolgt wiederum anhand von empirischen Daten, d. h. die Forschungsergebnisse sind in den Daten (empirisch) verankert (vgl. Steinke, 2007, S. 327f).

Reflektierte Subjektivität. Die Rolle der Forschenden als Subjekte (mit ihren Forschungsinteressen, Vorannahmen, Lebenserfahrungen) muss ebenfalls methodisch reflektiert werden. Das heisst, der Forschungsprozess sollte laufend durch Selbstbeobachtung und -reflexion begleitet werden, um bspw. festzustellen, ob man bestimmte Aspekte der Erkenntnisse verdrängt, weil sie einem nicht zusagen (vgl. Steinke, 2007, S. 330f).

Zentrales Element des Nachweises der Güte von Forschungsergebnissen ist schliesslich die Dokumentation der Ergebnisse: Hier gilt es jeweils zu entscheiden, ob und in welcher Ausführlichkeit die Verfahrensschritte, insbesondere die Dokumentation des Entscheidungswegs, dokumentiert werden kann. So sollte es den Lesenden zumindest ermöglicht werden, den Erschliessungsprozess der zentralen Hypothesen (auszugsweise) an Datenprotokollen nachprüfen zu können.

Weiterführende Literatur

— Allgemeine Übersichten

Bohnsack, R., Marotzki, W. & Meuser, M. (2006). *Hauptbegriffe qualitativer Sozialforschung*. Barbara Budrich.

Brüsemeister, T. (2008). *Qualitative Forschung. Ein Überblick*. VS Verlag für Sozialwissenschaften.

Kleemann, F., Krähnke, U. & Matuschek, I. (2009). *Interpretative Sozialforschung. Eine praxisorientierte Einführung*. VS Verlag für Sozialwissenschaften.

Przyborski, A. & Wohlrab-Sahr, M. (2014). *Qualitative Sozialforschung. Ein Arbeitsbuch* (4., erw. Aufl.). Oldenbourg.

— Links zu Qualitativer Sozialforschung

Schweizerisches Netzwerk für Qualitative Sozialforschung: <https://qualitative-research.ch>

Forum Qualitative Social Research (FQS): <http://www.qualitative-research.net>

<https://qualitativeinhaltsanalyse.de>

<https://www.audiotranskription.de>

<https://methodenzentrum.ruhr-uni-bochum.de/e-learning/qualitative-auswertungsmethoden/>

— Grounded Theory

Breuer, F. (2010). *Reflexive Grounded Theory eine Einführung für die Forschungspraxis*. VS Verlag für Sozialwissenschaften.

Strauss, A. (1998). *Grundlagen qualitativer Sozialforschung* (2. Aufl.). Fink.

Strauss, A. & Corbin, J. (1996). *Grounded Theory: Grundlagen qualitativer Sozialforschung*. Psychologie Verlags Union.

Strübing, J. (2008). *Grounded Theory. Zur sozialtheoretischen und epistemologischen Fundierung des Verfahrens der empirisch begründeten Theoriebildung* (2., überarb. u. erw. Aufl.). VS Verlag für Sozialwissenschaften.

— Qualitative Inhaltsanalyse

Kuckartz, U. (2016). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (3., überarb. Aufl.). Beltz Juventa.

Mayring, P. (1993). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (4., erw. Aufl.). Deutscher Studienverlag.

Mayring, P. (2007). Qualitative Inhaltsanalyse. In U. Flick, E. von Kardoff & I. Steinke (Hrsg.), *Qualitative Forschung. Ein Handbuch* (5. Aufl., S. 468–475). Rowohlt.

— Objektive Hermeneutik

Kutschke, M. (2011). *Objektive Hermeneutik im Schulalltag – Möglichkeiten und Grenzen eines hermeneutischen Vorgehens in der Unterrichtsforschung*. GRIN Verlag GmbH.

Wernet, A. (2009). *Einführung in die Interpretationstechnik der Objektiven Hermeneutik. Qualitative Sozialforschung, Bd. 11* (3. Aufl.). VS Verlag für Sozialwissenschaft.

13 Auswertung quantitativer Daten

von Jürg Aepli und Luciano Gasser

In diesem Kapitel 13 wird erläutert, wie quantitative Daten ausgewertet werden. Dabei wird zuerst genauer auf den Messvorgang und die Datenaufbereitung eingegangen, bevor Daten beschreibende (*deskriptive*) Verfahren und Hypothesen prüfende (*inferenzstatistische*) Verfahren der Datenanalyse dargestellt werden.

13.1 Grundlagen: Schätzen bzw. Messen und Skalenniveaus

Ausprägungen von Merkmalen. Sozialwissenschaftliche Forschung befasst sich mit dem Erleben und Verhalten von Menschen, welches sie näher untersucht. Werden Daten erhoben, um bestimmte Eigenschaften, Meinungen oder Einstellungen der Menschen in ihren jeweiligen Ausprägungen festzuhalten, wird dies als *Messvorgang* bezeichnet. So wird etwa in einem Fragebogen u. a. gefragt, wie gerne man Schokolade mag und es werden die Antwortmöglichkeiten „sehr gerne“, „eher gerne“, „eher nicht gerne“ und „gar nicht gerne“ vorgegeben. Somit hat das Merkmal „Schokolade mögen“ vier mögliche Ausprägungen.

Zuordnung von Zahlenwerten. Sollen nun diese jeweiligen Merkmalsausprägungen bei allen befragten Personen angeschaut und miteinander verglichen werden, ist es hilfreich, ihnen Zahlenwerte zuzuordnen. Das Vorhandensein von Zahlenwerten ermöglicht es beispielsweise, auszuzählen, wie oft eine bestimmte Merkmalsausprägung (z. B. „sehr gerne Schokolade mögen“) vorkommt. Dies ist vor allem bei Merkmalen mit vielen Ausprägungen und bei grösseren Stichproben von Nutzen. Die den Messobjekten zugeordneten Zahlen heissen *Skalenwerte* oder *Messwerte*. Die Zuordnung der Zahlen muss nach der Regel erfolgen, dass die *numerischen Relationen* (Verhältnisse zwischen den Zahlen) den *empirischen Relationen* (beobachtete Verhältnisse der Ausprägung eines Merkmals) entsprechen, d. h. dass die entsprechende Abbildung *homomorph*, also strukturerhaltend, ist (vgl. Bortz & Döring, 2006; Diekmann, 2007).

Skalenniveau (Messniveau, Massstabstypen)

Für die Beschreibung von Merkmalsausprägungen stehen vier verschiedene Massstabstypen oder Skalenniveaus zur Verfügung: (1) die Nominalskala, (2) die Ordinalskala (Rangskala), (3) die Intervallskala und (4) die Verhältnisskala (Rationalskala). Die Skalenniveaus legen fest, wie man die Zahlen interpretieren darf und welche Operationen mit den Zahlen sinnvoll sind.

Die vier Skalenniveaus. Für jedes z. B. mithilfe eines Fragebogens erhobene Merkmal mit seinen Ausprägungen muss zunächst das zutreffende Skalenniveau ermittelt werden, um die Skalen- oder Messwerte strukturerhaltend zuordnen zu können. Die vier wichtigsten Skalenniveaus (Skalenarten) sind: *Nominalskala*, *Ordinalskala* (Rangskala), *Intervallskala*, *Verhältnisskala* (Rationalskala). Diese Aufzählung entspricht bereits einer aufsteigenden Ordnung, wobei die Nominalskala dem einfachsten Skalenniveau entspricht. Im Folgenden werden die vier genannten Skalenniveaus vorgestellt (Übersicht siehe Tabelle 13-1).

Nominalskala. Von einem nominalskalierten Merkmal spricht man, wenn die Ausprägungen des Merkmals zwar unterschieden, aber nicht in eine Rangfolge gebracht werden können. Ein Merkmal wird durch die Beschreibung von Kategorien messbar gemacht, indem den Ausprägungen Zahlen („1“, „2“) oder auch Beschreibungen („ja“, „nein“) zugeordnet werden.

Um welche Zahlen oder Beschreibungen es sich dabei handelt, ist unwichtig. Die zugeordneten Zahlen (Symbole) haben nur eine kategoriale, also einordnende Funktion. Ihre Zuordnung erfolgt willkürlich. Damit lassen sich Merkmale kennzeichnen, die keine unterschiedlichen Ausprägungsgrade (Intensitätsgrade) bei den einzelnen Objekten besitzen, die aber verschiedenen Kategorien zugeordnet werden können. So liegt beispielsweise das Merkmal „Geschlecht“ traditionellerweise in zwei Ausprägungen vor, wobei Männer nicht „besser“ oder „höher“ als Frauen, sondern einfach anders sind.

Ordinalskala (Rangskala). Die Ordinalskala ordnet den Objekten die Zahlen so zu, dass das Objekt mit der grösseren/höheren Merkmalsausprägung die grössere Zahl erhält. Die Zahlen haben hier eine ordnende Funktion und erlauben die Bildung einer Rangfolge bezüglich der Ausprägung eines Merkmals. Aufgrund der Rangfolge kann jedoch nicht entschieden werden, wie weit die Merkmalsausprägungen auseinander liegen, da der „grösseren“ Ausprägung lediglich irgendeine höhere Zahl zugeordnet werden kann als der „kleineren“ Ausprägung. Es geht also um die Beschreibung von Relationen wie „grösser als“, „kleiner als“, „mehr als“, „weniger als“ usw.

Intervallskala. Die Intervallskala ordnet den Merkmalsausprägungen Zahlen in der Weise zu, dass Differenzen zwischen den Zahlen (numerisches Relativ) den empirischen Differenzen (empirisches Relativ) entsprechen. Dazu braucht es eine definierte Masseinheit, d. h. die Abstände müssen gleich sein. Beispiel Celsiusskala: Die Abstände auf der Skala (z. B. 1–4°C und 8–11°C) entsprechen gleichen Temperaturdifferenzen in der Realität. Bei der Intervallskala wird sowohl die Masseinheit als auch der Nullpunkt willkürlich festgelegt. Es können deshalb keine Aussagen über das Verhältnis von Skalenwerten gemacht werden. Es wäre beispielsweise falsch zu sagen, in Florida (36°C) sei es dreimal wärmer als in Luzern (12°C), weil sich die Temperaturangaben nicht auf den absoluten Nullpunkt beziehen (in Wahrheit ist

es in Florida bezogen auf den absoluten Nullpunkt von $-273,15^{\circ}\text{C}$ nur 1,08 mal wärmer als in Luzern). Das gleiche gilt bei vielen eingesetzten Tests: Wer in einem Leistungsmotivationstest 40 Punkte erreicht, ist nicht „doppelt so motiviert“, wie jemand, der 20 Punkte erreicht.

Verhältnisskala (Rationalskala). Im Gegensatz zur Intervallskala besitzt die Verhältnisskala einen absoluten Nullpunkt. Es können deshalb Aussagen über die Verhältnisse von Skalenwerten gemacht werden. Dies ist zum Beispiel bei der Körpergröße der Fall: Wenn Maria 140 cm gross ist und Martin 70 cm, so kann man sagen, Maria ist doppelt so gross wie Martin.

In der Literatur wird nicht immer zwischen Verhältnisskala und Intervallskala unterschieden, sondern diese beiden Skalenniveaus werden in einer Kategorie zusammengefasst: Man spricht dann von einem *metrischen* Skalenniveau.

Die Bedeutung der Skalenniveaus. Die Kenntnis dieser Skalenniveaus ist entscheidend, da die Zulässigkeit statistischer Verfahren vom jeweiligen Skalenniveau eines Merkmals resp. seiner Ausprägungen abhängt. Das heisst, dass bestimmte statistische Verfahren nur bei gewissen Skalenniveaus verwendet werden dürfen.

Ohne Kenntnis des Skalenniveaus der (quantitativen) Daten der eigenen Untersuchung sind weiterführende statistische Analysen nicht möglich. Bereits einfache deskriptiv-statistische Verfahren, wie die Ermittlung der zentralen Tendenz (Durchschnitt) der Antworten auf die Frage „Wie viele Stunden sehen Sie pro Tag fern?“ erfordern es, dass das Skalenniveau des Merkmals („täglicher Fernsehkonsum“) bekannt ist, um den geeigneten Kennwert auszuwählen (vgl. Kap. 13.3.4 „Masse der zentralen Tendenz“).

Tabelle 13-1: Übersicht über die vier Skalenniveaus

Skalenniveau	Eigenschaften	Beispiele
Nominalskala	<ul style="list-style-type: none"> • es gibt verschiedene Gruppen, die nebeneinander bestehen • keine Gruppe ist der anderen übergeordnet (oder besser) • es gibt keine Rangfolge 	<ul style="list-style-type: none"> • Geschlecht (männlich, weiblich) • Familienstand (verheiratet, ledig, geschieden etc.) • Lieblingsfarben • Parteien • Freizeitaktivitäten
Ordinalskala	<ul style="list-style-type: none"> • es gibt verschiedene Gruppen, die übereinander liegen • obwohl es eine Rangfolge gibt, kann nicht gesagt werden, wie weit die Gruppen auseinander liegen (die Abstände der Gruppen sind unklar) 	<ul style="list-style-type: none"> • Schulnoten • abgestufte Kategorien von Häufigkeiten („nie“, „manchmal“, „oft“, „immer“) • Wetter-Abstufungen • militärische Ränge

Skalenniveau	Eigenschaften	Beispiele
Intervallskala	<ul style="list-style-type: none"> • es liegt eine klar definierte, regelmässige Skala vor • der Abstand zwischen den Gruppen ist klar messbar • der Abstand zwischen den Gruppen ist immer gleich • der Nullpunkt ist willkürlich gewählt • zwischen den Gruppen können keine Verhältnisse gebildet werden 	<ul style="list-style-type: none"> • Temperatur (Celsius) • Kalenderzeit
Rationalskala	<ul style="list-style-type: none"> • es liegt eine klar definierte, regelmässige Skala vor • der Abstand zwischen den Gruppen ist immer gleich • der Nullpunkt ist sinnvoll definiert resp. absolut • zwischen den Gruppen können Verhältnisse gebildet werden 	<ul style="list-style-type: none"> • Alter • Dienstalster • Körpergrösse • Einkommen

Unterscheidung zwischen stetigen und diskreten Variablen. Schliesslich ist im Zusammenhang mit den Skalenniveaus und den damit möglichen statistischen Analysen auch die Unterscheidung zwischen *stetigen* (kontinuierlichen) und *diskreten* (diskontinuierlichen) Variablen von Belang.

Stetige Variablen zeichnen sich dadurch aus, dass sich „in jedem beliebigen Intervall unendlich viele Merkmalsausprägungen befinden (z. B. die Variablen Gewicht, Länge oder Zeit)“ (Bortz & Döring, 2003, S. 7).

Bei einer *diskreten Variable* finden sich in einem (begrenzten) Intervall nur eine endliche Anzahl Ausprägungen, z. B. die „Anzahl Bücher zu Hause“ oder die „Anzahl Wörter, welche in zehn Minuten auswendig gelernt werden können“ (vgl. Bortz & Döring, 2003).

Wird also in einem Experiment gemessen, wie schnell jemand auf einer Leinwand einen projizierten Buchstaben erkennt, so ist die Variable „Reaktionszeit“ als stetige Variable zu bezeichnen. Stetige Variablen kann man daran erkennen, dass bei Messungen resp. den mit ihr verbundenen Messeinheiten eine unendliche Anzahl Stellen nach dem Komma möglich wäre.

13.2 Aufbereitung von quantitativen Daten

Ausgangslage: Vorhandene Antworten bzw. Daten. Quantitative Daten können auf verschiedene Weise erhoben werden und liegen zunächst als Antworten bzw. Daten in Fragebögen, Beobachtungs-, Interview- oder Versuchsprotokollen vor. Die Daten sind damit zwar vorhanden, aber in einer noch ungeordneten und unübersichtlichen Form. Quantitative Daten werden aufbereitet, indem die Antworten bzw. Daten aus Fragebögen, Beobachtungs-, Interview- oder Versuchsprotokollen zunächst kodiert und danach in einer Urliste aufgelistet werden. Die Kodierung wird in einem Kodeplan festgehalten. Bevor mit der eigentlichen Auswertung begonnen wird, müssen die Daten selber einer Untersuchung unterzogen werden, um Eingabefehler entdecken und um z. B. nicht seriös ausgefüllte Fragebögen oder Items, die offensichtlich missverstanden wurden, ausschliessen zu können. Darauf wird nachfolgend eingegangen.

Wenn die Ausprägungen eines Merkmals (Variable) in tabellarischer Form dargestellt werden, spricht man von *Rohwerten*, die in einer Urliste dargestellt werden. In einer *Urliste* werden die erhobenen Ausprägungen eines Merkmals (Variable) in ihrer „ursprünglichsten“ Form erfasst. Die erhobenen Informationen sind damit vollständig dokumentiert.

Auflistung der Daten. Daten werden üblicherweise computergestützt (z. B. mittels Excel, SPSS) ausgewertet. Vor der Dateneingabe in eine entsprechende Datei müssen einige Vorüberlegungen angestellt werden, damit die Daten in eine für die Auswertung geeignete Form gebracht werden können.

► Fallbeispiel zum Vorgehen beim Erstellen einer Datentabelle

Studierende wurden mittels Fragebogen nach ihren Erfahrungen bei der Bearbeitung einer Online-Lerneinheit befragt. Die Daten liegen somit als Antworten zu den Fragen des Fragebogens vor. Die Daten aus allen Fragebögen müssen nun in einer Datei aufgelistet werden.

Kurzbezeichnung des Merkmals festlegen. Wie aus Abbildung 13-1 ersichtlich ist, wird für jede Frage mit ihren Antwortmöglichkeiten eine Variable gebildet. Die Bezeichnung für die Variable wird als Spaltentitel gesetzt. Häufig muss für die Variable eine Kurzbezeichnung gewählt werden, da bei computergestützter Auswertung gewisse, vor allem auch ältere Versionen von Statistikprogrammen für die Kennzeichnung der Merkmale nur eine beschränkte Anzahl Buchstaben und/oder Ziffern zulassen (in der Abbildung wurde für das Merkmal „Geschlecht“ die englische Kurzbezeichnung „sex“ gewählt).

Kodierung der Ausprägungen. Die Antworten der Studierenden werden pro Student/Studentin in je einer Zeile festgehalten. Man spricht bei den eingetragenen Ausprägungen eines Merkmals auch von *Rohwerten*. Bei den Rohwerten handelt es sich um *kodierte Daten*, d. h. jeder sprachlichen Antwortmöglichkeit muss vorgängig ein numerischer Wert zugeordnet werden (siehe weiter unten „Datenkodierung“). In Abbildung 13-1 wurde beispielsweise für das Merkmal „Geschlecht“ die Ausprägung „1“ für männlich und „2“ für weiblich gewählt.

Indexspalte zur Kennzeichnung der Befragten. Zur Kennzeichnung der Zeilen wird dem Datensatz eine *Indexspalte* zur Kennzeichnung der Befragten vorangestellt. In Abbildung 13-1 beinhaltet die Indexspalte die Fragebogen-Nummern (mit „nr“ bezeichnet). Dazu wurden vorgängig alle ausgefüllten Fragebögen durchnummeriert. Anstatt der Nummer des Fragebogens kann hier auch der Code der befragten Person eingetragen werden. Durch diese Massnahme können die in der Urliste eingetragenen Antworten bis auf den ausgefüllten Fragebogen zurückverfolgt werden. Nach der Auflistung liegen die Daten in Form einer so genannten *Urliste* vor. In dieser Urliste sind die Merkmale (z. B. „Geschlecht“, „Alter“) als Spaltentitel eingetragen und die Rohwerte dazu für jeden Fragebogen bzw. jede befragte Person jeweils in einer Zeile festgehalten. Für den in Abbildung 13-2 aufgeführten Fragebogen „Nummer 1“ werden folgende Werte eingetragen: 19 (= *alter*), 2 (= *sex*), 3 (= *lernstil*), 1,5 (= *dauer*). In Abbildung 13-2 ist ein Ausschnitt aus der Urliste, die mit der Software Excel erstellt wurde, dargestellt.

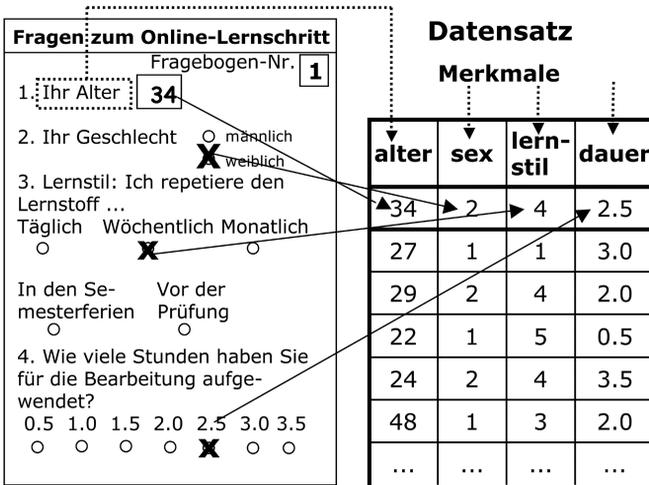


Abbildung 13-1: Auflistung von Antworten aus einem Fragebogen in einer Urliste

nr	alter	sex	lernstil	dauer
1	19	2	3	1.5
2	21	1	4	1
3	21	1	4	2
4	27	2	5	1.5
5	21		3	2
6	33	2	4	3
7	33	2	4	3
8	29	2	2	2.5
9	20	2	4	2.5
10	48	2	4	2.5
11	21	2	4	2.5
12	20	2	4	0.5

Abbildung 13-2: Urliste mit Indexspalte, welche die Fragebogen-Nummern enthält

Dokumentation der Datenkodierung in einem Kodeplan. Durch die Dokumentation in einem Kodeplan ist eine übersichtliche Darstellung von wichtigen Informationen zum Datensatz gegeben. Damit kann die Kodierung auch später noch nachvollzogen und die Bedeutung der Werte erschlossen werden. Ein Kodeplan sollte (mindestens) folgende Angaben enthalten:

- *Kurzbezeichnung* für das Merkmal in der Urliste,
- *Bedeutung des Merkmals* bzw. Wortlaut der Frage,
- *zugelassene Werte* (welche Ausprägungen darf ein Merkmal annehmen?) und *Bedeutung der Ausprägungen*,
- *Bezeichnung fehlender Werte* (wie sind fehlende Angaben bezeichnet und eingetragen?).

Anmerkung zur Bezeichnung fehlender Werte: Wenn in gewissen Fällen keine Antwortvorgaben vorliegen, weil die Antwort z. B. übersehen wurde oder weil aus Unwissenheit eine Frage nicht beantwortet werden konnte, kann bei gewissen Statistikprogrammen (z. B. SPSS) ein Zahlenwert bestimmt und dieser Zahlenwert als „fehlende Angabe“ deklariert werden. Dies wäre beispielsweise in Abbildung 13-2 bei der Versuchsperson Nummer 5 der Fall, die das Geschlecht nicht angegeben hat. Damit kann diese Person aus den weiteren statistischen Auswertungen ausgeschlossen werden. Werden Auswertungen mit der Software Excel gemacht, so ist darauf zu achten, dass solche Zahlenwerte für fehlende Angaben (*missing values*) allenfalls nicht für weitere Auswertungen verwendet werden. Fehlende Angaben können ausgelassen oder allenfalls auch mit „x“ bezeichnet werden.

- *Skalenniveau* des Merkmals (vgl. Kap 13.1).

In Tabelle 13-2 ist der Kodeplan zu der in Abbildung 13-2 dargestellten Urliste aufgeführt. Sobald die Urliste und der Kodeplan erstellt sind, kann mit der eigentlichen Datenauswertung begonnen werden.

Tabelle 13-2: Beispiel eines Kodeplans zur Urliste in Abbildung 13-2

Kurzbezeichnung	Bedeutung	Ausprägungen	Bezeichnung fehlender Werte	Skalenniveau
nr	Fragebogen-Nummer	1–120	leere Zelle	Nominal
alter	Alter	Alter in Jahren	leere Zelle	Metrisch
sex	Geschlecht	1 = männlich 2 = weiblich	leere Zelle	Nominal
lernstil	Repetition des Lehrstoffs	5 = täglich 4 = wöchentlich 3 = monatlich 2 = in den Semesterferien 1 = vor der Prüfung	leere Zelle	Ordinal
dauer	Bearbeitungsdauer für den Lernschritt	Bearbeitungsdauer in Stunden: 0,5/1,0/1,5/2,0/2,5/3,0/3,5	leere Zelle	Metrisch

Prüfung der eingegebenen Daten (Datenexploration). Wenn die Daten eingegeben worden sind, sollte danach nicht gleich mit der eigentlichen Auswertung begonnen werden, sondern die Daten selber sollten einer eingehenden Untersuchung unterzogen werden (Datenexploration).

Dabei sind folgende Fragen zu klären:

- 1) Sind die Dateneingaben korrekt vorgenommen worden oder können Eingabefehler entdeckt werden?
- 2) Sind die Items von den Personen wohl richtig verstanden worden? Sind Ausreißer und Extremwerte vorhanden? Müssen einzelne Personen oder einzelne Angaben wegen „Unstimmigkeit“ von weiteren Auswertungen ausgeschlossen werden?

Entdeckung von Eingabefehlern. Eingabefehler sind am besten zu erkennen, wenn eine Häufigkeitsauszählung der auftretenden Werte vorgenommen wird. Beispielsweise kann ein Fehler entdeckt werden, indem bei der Häufigkeitsverteilung zu „Momentanes Gewicht“ bei einer Schülerinnen- und Schülerbefragung die Ausprägung „645 kg“ vorkommt, was nicht stimmen kann. Eine solche Häufigkeitszählung kann tabellarisch oder grafisch dargestellt werden (siehe dazu die Kap. 13.3.1 und 13.3.2).

Neben der Darstellung von Häufigkeitsverteilungen können auch die Verteilungsform überprüft werden (dies interessiert vor allem bei intervall- und verhältnisskalierten Variablen, da die Wahl von passenden analytischen Tests von der Form der Verteilung – verlangt wird häufig eine Normalverteilung der Werte – abhängt).

Entdeckung von Widersprüchlichkeiten. Mit etwas Geschick und Einfallsreichtum lassen sich weitere Fehler entdecken. So könnte z. B. durch den Vergleich zweier Fragen mittels Kreuztabelle festgestellt werden, dass etwas nicht stimmen kann: Eine Schülerin hat auf die Frage, ob Tanzen eine Freizeitbeschäftigung von ihr sei, mit *nein* geantwortet hat; bei einer anderen Frage hat sie jedoch angegeben, dass sie in ihrer Freizeit pro Woche zwischen 4 und 6 Stunden tanze.

Wenn in einen Fragebogen Kontrollfragen eingebaut worden sind, kann herausgefunden werden, wie sorgfältig ein Fragebogen ausgefüllt worden ist. Wenn Fehler entdeckt werden oder „unstimmige“ Werte auftreten, sollten diese korrigiert bzw. die Daten dieser Personen von weiteren Auswertungen ausgeschlossen werden.

13.3 Deskriptive Statistik I: Auf ein Merkmal bezogene Auswertungen

Mit einer beschreibenden Statistik sollen Untersuchungsergebnisse so dargestellt werden, dass die im Rahmen einer Untersuchung erhobenen Aspekte klar und in verständlicher Form sichtbar und beschrieben werden.

Die *Deskriptive Statistik* befasst sich mit statistischen Methoden zur Beschreibung und Charakterisierung von Daten in Form von Tabellen, Grafiken und einzelnen Kennwerten.

Welche Auswertungen könnten nun interessieren? Beispielsweise interessiert die Frage, wie viele Stunden Studierende zur Bearbeitung eines Online-Lernschrittes aufgewendet haben. Bei diesem Beispiel geht es um die Darstellung der Antworten zum Merkmal „Bearbeitungsdauer für den Online-Lernschritt“.

Bei der *univariaten (d. h. auf eine Variable bezogenen) Datenanalyse* interessiert die Frage, was bei einem ausgewählten Merkmal über die erhobenen Ausprägungsgrade ausgesagt werden kann.

Bei *bivariaten Datenanalysen* interessiert der Zusammenhang zwischen zwei Merkmalen (z. B. ob es einen Zusammenhang zwischen „Alter“ und „Lernstil“ gibt).

Bei *multivariaten Auswertungen* wird das Zusammenwirken mehrerer Merkmale untersucht. Untersucht man z. B., wie „Alter“ und „Lernstil“ mit der „Bearbeitungsdauer für den Lernschritt“ zusammenhängen, so spricht man von einer multivariaten Datenanalyse.

Die folgenden Kapitel 13.3.1 bis 13.3.4 beziehen sich auf die univariate Datenanalyse. Auf bivariate Datenanalysen wird in Kapitel 13.4 „Deskriptive Statistik II: Beziehungen zwischen zwei Merkmalen“ genauer eingegangen.

13.3.1 Die tabellarische Darstellung

Eine Häufigkeitsverteilung einer univariaten Datenanalyse beschreibt, wie sich die untersuchten Personen auf die Merkmalsausprägungen eines Merkmals verteilen. Beispielsweise kann zum Merkmal „Bearbeitungsdauer für den Lernschritt“ folgende Häufigkeitsverteilung erstellt werden:

- 1 Person hat für die Bearbeitung des Lernschritts 0,5 Stunden aufgewendet,
- 6 Personen gaben an, 1 Stunde aufgewendet zu haben,
- 29 Personen gaben 1,5 Stunden an,
- 32 Personen gaben 2 Stunden an, etc.

Es können zu allen Merkmalen – im Fallbeispiel zu „Alter“, „Geschlecht“, „Lernstil“ und „Bearbeitungszeit für den Lernschritt“ – Häufigkeitsverteilungen berechnet werden. In Tabelle 13-3 ist eine solche Häufigkeitstabelle zum Merkmal „Bearbeitungsdauer“ enthalten. Um eine bessere Überschaubarkeit zu erzielen, kann die Anzahl der Ausprägungen reduziert werden, indem Kategorien gebildet und die einzelnen Ausprägungen den verschiedenen Kategorien zugeordnet werden.

Eine *Häufigkeitsverteilung* einer univariaten Datenanalyse beschreibt, wie sich die untersuchten Personen auf die Merkmalsausprägungen eines Merkmals verteilen. Die Häufigkeitsverteilung der Ausprägungen eines Merkmals kann mittels Tabelle dargestellt werden, indem zu den einzelnen Ausprägungen die absoluten, relativen oder kumulierten (absoluten bzw. relativen) Häufigkeiten angegeben werden.

Tabelle 13-3: Angabe von absoluten und relativen Häufigkeiten für die Bearbeitungsdauer des Lernschritts von Studierenden

Dauer für die Bearbeitung des Lernschritts in Std.	Absolute Häufigkeit	Relative Häufigkeit (in %)
0,5	1	0,8
1,0	6	5,1
1,5	29	24,6
2,0	32	27,1
2,5	35	29,7
3,0	11	9,3
3,5	4	3,4
<i>Total</i>	<i>118</i>	<i>100,0</i>

Angabe von absoluten oder relativen Häufigkeiten. In einer Häufigkeitstabelle können die absoluten und/oder die relativen Häufigkeiten angegeben werden. Im Normalfall wird in Tabellen nur eine der beiden Häufigkeiten aufgeführt. Um die Unterschiede zwischen absoluten und relativen Häufigkeiten besser erfassen zu können, finden sich in Tabelle 13-3 jedoch beide Häufigkeitsangaben. Werden zu jedem Merkmal die Häufigkeiten seiner Ausprägungen ausgezählt, spricht man von *absoluten Häufigkeiten*. Wie aus der Tabelle ersichtlich ist, gaben beispielsweise 35 Studierende an, 2,5 Stunden für die Bearbeitung aufgewendet zu haben.

Angabe von relativen Häufigkeiten. Oft ist die Berechnung der prozentualen Anteile von Interesse, d. h. es werden die *relativen Häufigkeiten* angegeben. Dabei wird die Gesamtzahl der Personen, welche zu einem Merkmal gültige Antworten abgegeben haben, als 100% gesetzt und davon ausgehend die jeweiligen Anteile der einzelnen Ausprägungen ermittelt. Personen, die eine Frage nicht beantwortet haben, werden in der Regel nicht mitgezählt. Im Beispiel aus Tabelle 13-3 ist ersichtlich, dass 35 Studierende für die Bearbeitung rund 2,5 Stunden aufgewendet haben. Damit diese Information besser interpretiert werden kann, ist es nötig, dass man dies mit der Anzahl der Studierenden, die auf die anderen Ausprägungen fallen, oder mit dem Total der Studierenden vergleichen kann. Es soll deshalb die relative Häufigkeit berechnet werden. Diese berechnet sich wie folgt:

$$\text{relative Häufigkeit} = \frac{\text{absolute Häufigkeit}}{\text{Stichprobengröße } n}$$

Die relative Häufigkeit der Ausprägung „2,5 Stunden“ beträgt 0.297:

$$0.297 = \frac{35}{118}$$

Dies entspricht 29,7%. Das bedeutet, dass 29,7% der Studierenden für die Bearbeitung des Lernschritts rund 2,5 h aufgewendet haben. Die Summe aller relativen Häufigkeiten in einer Häufigkeitstabelle ist immer 1 bzw. 100%.

Die Angabe prozentualer Anteile hat gegenüber der Angabe absoluter Häufigkeiten den Vorteil, dass die Ausprägungen untereinander und über Merkmale hinweg – vor allem auch bei ungleich grossen Gruppen – besser miteinander verglichen werden können.

Bei der Angabe von relativen Häufigkeiten ist es notwendig, dass die Stichprobengröße n mitgeteilt wird, um anzugeben, auf wie viele Befragte sich die Prozentangaben beziehen. Dies geschieht z. B. dadurch, dass im Tabellentitel die Stichprobengröße n in Klammern angegeben wird. Im Beispiel könnte der Titel wie folgt lauten: „Tabelle 13-3: Angabe von absoluten und relativen Häufigkeiten für die Bearbeitungsdauer des Lernschritts von Studierenden ($n = 118$)“.

Angabe von kumulierten absoluten Häufigkeiten. Häufigkeiten können kumuliert werden. Es wird dabei zwischen *kumulierten relativen Häufigkeiten* und *ku-*

mulierten absoluten Häufigkeiten unterschieden. Die *kumulierte absolute Häufigkeit* gibt an, wie häufig ein bestimmter Ausprägungsgrad und alle niedrigeren Ausprägungsgrade eines Merkmals beobachtet wurden.

Beispiel zur Tabelle 13-3: Wie viele Studierende haben für die Bearbeitung des Lernschritts 1,5 Stunden oder weniger aufgewendet? Antwort: 36 Studierende (1+6+29 Studierende) haben 1,5 Stunden oder weniger zur Bearbeitung aufgewendet. Häufiger wird jedoch die kumulierte relative Häufigkeit angegeben.

Angabe von kumulierten relativen Häufigkeiten. Die kumulierte relative Häufigkeit gibt in Prozenten an, wie häufig eine bestimmte Ausprägung und alle niedrigeren Ausprägungen eines Merkmals beobachtet wurden. Die kumulierten relativen Häufigkeiten werden auch als *Prozentränge* bezeichnet.

Berechnung von kumulierten relativen Häufigkeiten. Der kumulierte Prozentsatz summiert zeilenweise die prozentuale Häufigkeit der Antworten auf. Im Beispiel aus Tabelle 13-4 bestimmt sich die kumulierte relative Häufigkeit von 2 Stunden als Summe der relativen Häufigkeiten der Ausprägungsgrade 0,5h, 1,0h, 1,5h und 2h. Es gaben rund 57,6% der Studierenden an, dass Sie für die Bearbeitung zwei Stunden oder weniger aufgewendet haben (0,8% + 5,1% + 24,6% + 27,1%).

Tabelle 13-4: Beispiel der Angabe von absoluten, relativen und kumulierten relativen Häufigkeiten für die Bearbeitungsdauer des Lernschritts (n = 118)

Dauer für die Bearbeitung des Lernschritts in Std.	Absolute Häufigkeit	Relative Häufigkeit (in %)	Kumulierte relative Häufigkeit (in %)
0,5	1	0,8	0,8
1,0	6	5,1	5,9
1,5	29	24,6	30,5
2,0	32	27,1	57,6
2,5	35	29,7	87,3
3,0	11	9,3	96,6
3,5	4	3,4	100,0
<i>Total</i>	<i>118</i>	<i>100,0</i>	

Wozu relative kumulierte Häufigkeiten einsetzen? Relative kumulierte Häufigkeiten können zur Beantwortung von drei Fragetypen benutzt werden:

- 1) *Fragetyp „Höchstens“:* Wie gross ist der Anteil der Studierenden, die höchstens 1,5 Stunden zur Bearbeitung aufgewendet haben? Antwort: 30,5% der Studierenden.
- 2) *Fragetyp „Grösser als“:* Wie hoch ist der Anteil der Studierenden, die mehr als 2,5 Stunden für die Bearbeitung einsetzten? Antwort: 12,7% der Studierenden = 100%-87,3%.

3) *Fragetyp „Zwischen“*: Wie hoch ist der Anteil der Studierenden, die zwischen 2 und 2,5 Stunden für die Bearbeitung aufgewendet haben? Antwort: 56,8% der Studierenden = 87,3%-30,5%.

Kategorienbildung bei Problemen der Anschaulichkeit aufgrund vieler Merkmalsausprägungen. Die bisher besprochene Darstellung einer Häufigkeitstabelle, in denen alle Merkmalsausprägungen aufgeführt sind, ist nur sinnvoll, wenn die Merkmale im Datensatz eine eng begrenzte Anzahl von Merkmalsausprägungen besitzen. Bei mehr als 10 bis 15 verschiedenen Merkmalsausprägungen sind die bisher vorgestellten Häufigkeitstabellen nicht mehr anschaulich. In Tabelle 13-5, in welcher die Altersausprägungen von Studierenden dargestellt sind, ist ersichtlich, dass es zu diesem Merkmal Ausprägungen von „19 Jahre“ bis „53 Jahre“ geben kann und somit 35 verschiedene Ausprägungen zum Merkmal „Alter“ vorkommen können.

Bildung von Ausprägungskategorien. Um eine bessere Überschaubarkeit zu erzielen, kann die Anzahl der Ausprägungen reduziert werden. Dies geschieht durch eine Zusammenfassung von Ausprägungen zu Gruppen, d.h. es werden *Kategorien* gebildet und die einzelnen Ausprägungen den verschiedenen Kategorien zugeordnet.

Im Beispiel wurden zum Merkmal „Alter“ Ausprägungskategorien gebildet. Die *Kategorienbreite* wurde auf fünf Jahre festgesetzt. Dadurch werden die 35 verschiedenen möglichen Ausprägungen auf 8 Ausprägungskategorien reduziert.

Tabelle 13-5: Beispiel zweier tabellarischer Häufigkeitsverteilungen – einmal ohne und einmal mit Kategorienbildung für das Merkmal „Alter“

Alter	Absolute Häufigkeit	Alterskategorie (Kategoriebreite = 5)	Absolute Häufigkeit
19	3	16–20	17
20	14	21–25	52
21	23	26–30	28
22	12	31–35	18
23	5	36–40	6
24	7	41–45	2
25	5	46–50	1
26	6	51–55	3
27	7		
...	...		
53	1		

Vorteile und Nachteile von Kategorienbildung. Bei der Festlegung von Ausprägungskategorien wird durch die Verdichtung der Merkmalsausprägungen zu Kategorien einerseits Übersichtlichkeit erzielt, andererseits gehen aber immer auch Informationen verloren: Je breiter die Kategorien dabei gewählt werden, desto grösser ist der Informationsverlust. Würde man im obigen Beispiel etwa nur zwei Kategorien festlegen, z. B. von „Personen bis und mit 40 Jahren“ und „Personen ab 41 Jahren“, so wären in der ersten Kategorie 121 Personen zu finden und in der zweiten Kategorie sechs Personen. Über die effektive Altersverteilung in der Stichprobe liessen sich kaum mehr differenzierte Aussagen bilden, ausser dass die Mehrheit der befragten Personen unter 40 Jahre alt ist. Bei der Kategorienbildung ist deshalb jeweils zwischen dem Verlust an Informationen und dem Gewinn an Übersichtlichkeit abzuwägen.

Beachten bei der Kategorienbildung. Eine generelle Regel für die Wahl einer sinnvollen Kategorienbreite gibt es nicht: Die Festlegung von Ausprägungskategorien kann nur inhaltlich, d. h. in Bezug auf die konkrete Zielsetzung, begründet werden.

Bei der Kategorienbildung ist darauf zu achten, dass eine Ausprägung eindeutig einer Kategorie zugeordnet werden kann. Überlappungen benachbarter Kategorien dürfen nicht bestehen, d. h. Kategorienbreiten von „20–25 Jahre“ und von „25–30 Jahre“ wären nicht zulässig, weil dann nicht eindeutig zu entscheiden wäre, welcher Kategorie eine 25-jährige Person zugeordnet werden soll. Man spricht davon, dass Kategorien *disjunkiv* (also einander ausschliessend) sein müssen.

13.3.2 Die grafische Darstellung

Es ist manchmal schwierig, die Verteilung von Ausprägungen eines Merkmals in einer Häufigkeitstabelle rasch zu überblicken. Deshalb werden Verteilungen mittels Diagrammen oftmals grafisch dargestellt.

Die grafische Darstellung einer Häufigkeitsverteilung weist gegenüber einer tabellarischen Darstellung folgende *Vorteile* auf:

- grössere Übersichtlichkeit,
- grössere Einprägsamkeit,
- grössere Attraktivität (die grafische Darstellung weckt Interesse und erhöht damit die Lesebereitschaft; sie lockert einen Text im Sinne einer Abwechslung auf).

Wenn aber exakte Zahlen interessieren, sind Tabellen einem Diagramm vorzuziehen. So wird in wissenschaftlichen Arbeiten auf die Wiedergabe von Häufigkeitstabellen oft nicht verzichtet.

Richtlinien zur Gestaltung von Diagrammen. Es ist zu bestimmen, ob bei grafischen Darstellungen die absoluten Häufigkeiten, die relativen Häufigkeiten oder auch beide angegeben werden sollen. Diese Entscheidung kann nur vor dem Hintergrund der inhaltlichen Fragestellung beantwortet werden. Relative Häufigkei-

ten werden gewählt, wenn die Daten bezüglich des gesamten Datensatzes bewertet werden sollen.

Für alle Typen von Diagrammen gilt, dass sie *eindeutig* und *vollständig* beschriftet werden müssen. Beispielsweise müssen bei einem Sektoridiagramm alle Sektoren eindeutig benannt werden. Bei einem Säulendiagramm sind alle Ausprägungen des Merkmals aufzuführen und x- und y-Achse zu beschriften. Der Koordinatenachse, welche die Häufigkeiten nennt, muss entnommen werden können, ob absolute oder relative Häufigkeiten dargestellt sind.

Grafik-Programme bieten sehr viele Möglichkeiten zur grafischen Gestaltung, von denen aber einige von den relevanten Informationen ablenken oder sie gar verschleiern. Dazu gehören:

- dominierende Hintergründe,
- Schraffuren, die verzerren,
- drei-dimensional (perspektivisch) dargestellte Abbildungen.

Verzerrende Darstellungen. Manche Darstellungsweisen können zudem ein irreführendes Bild der Daten geben. Beispielsweise lässt sich durch das Strecken der y-Achse der Eindruck grosser Unterschiede, beim Stauchen der y-Achse hingegen der Eindruck geringer Unterschiede erzeugen. Durch das Weglassen eines Teils des Achsenbereiches kann der subjektive Eindruck ebenfalls beeinflusst werden.

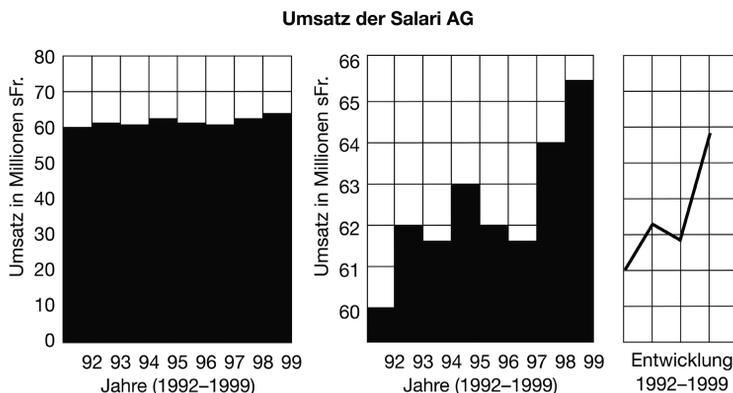


Abbildung 13-3: Grafische Darstellung des gleichen Datensatzes auf drei verschiedene Arten (in Anlehnung an Krämer, 2006b, S. 38–40)

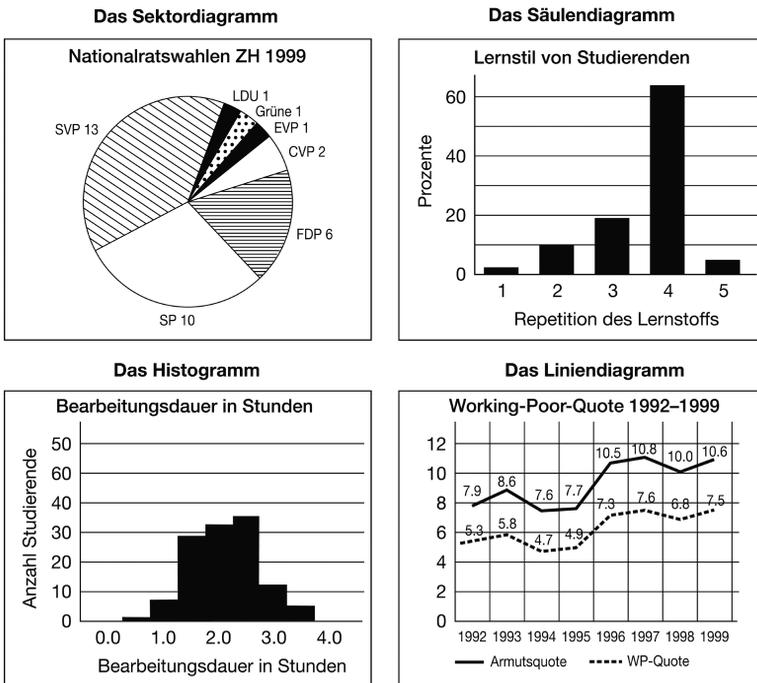
Anmerkungen: *Linkes Diagramm:* Korrekte, ehrliche Darstellung der Fakten. *Mittleres Diagramm:* Die y-Achse wurde gestreckt und zugleich abgeschnitten, ohne dass darauf hingewiesen wird. Die gleichen Umsatzzahlen präsentieren sich jetzt auf den ersten Blick viel positiver. *Rechtes Diagramm:* Um die Dynamik zu betonen, wird der Umsatz als Kurvendiagramm dargestellt. Damit die Stagnation bzw. der leichte Rückgang in gewissen Jahren nicht so auffällt, werden Kategorien gebildet und nur der durchschnittliche Umsatz von 2 Jahren dargestellt. Zudem wird die Bezeichnung der Achsen weggelassen.

Das negative Beispiel in Abbildung 13-3 zeigt, wie durch solche unterschiedlichen Darstellungsarten der subjektive Eindruck beim oberflächlichen Betrachten manipuliert werden kann.

Beim Umgang mit Diagrammen ist deshalb Folgendes zu beachten:

- Bei der grafischen Darstellung von Daten sind wir zu wissenschaftlicher Ehrlichkeit verpflichtet.
- Grafische Darstellungen in Publikationen müssen kritisch gelesen werden.

Die vier wichtigsten Typen von Diagrammen. Je nach *Skalenniveau eines Merkmals* und *Zielsetzung* wird ein bestimmter Darstellungstyp eines Diagramms gewählt. Im Folgenden werden die vier wichtigsten Typen von Diagrammen vorgestellt: Das *Sektordiagramm*, das *Säulendiagramm*, das *Histogramm* und das *Liniendiagramm*. Diese vier Typen sind in Abbildung 13-4 dargestellt.



© Statistisches Amt des Kantons Zürich, Quelle SAKE

Abbildung 13-4: Die vier wichtigsten Typen von Diagrammen

Das Sektordiagramm. Andere Bezeichnungen für das Sektordiagramm (*pie chart*) sind: Kreisdiagramm, Tortendiagramm, Kuchendiagramm.

- *Charakterisierung:* Ein Kreis wird so in Kreissektoren unterteilt, dass die Flächen der Kreissektoren zu den beobachteten Häufigkeiten der einzelnen Ausprägungen proportional sind.
- *Eignung:* Sektordiagramme eignen sich vor allem für die Darstellung von nominal skalierten Merkmalen. Das Sektordiagramm verdeutlicht die Verhältnisse zwischen den Häufigkeiten, mit denen eine Merkmalsausprägung beobachtet wurde, in der Regel besser als ein Säulendiagramm. Hingegen ist die Verwendung von Sektordiagrammen mit mehr als sieben Ausprägungskategorien (Segmenten) nicht mehr angemessen, da die Diagramme durch die vielen Segmente unübersichtlich werden.

Das Säulendiagramm. Das Säulendiagramm (*bar chart*) wird auch bezeichnet als Linien- oder Stabdiagramm (die Säulen sind zu Linien bzw. Stäben verschmälert), Balkendiagramm (die Säulen sind waagrecht angeordnet) oder Rechteckdiagramm.

- *Charakterisierung:* Die Häufigkeiten der Merkmalsausprägungen werden durch Säulen dargestellt. Die Höhe der Säulen spiegelt die Anzahl von Beobachtungen (absolute Häufigkeiten) oder den prozentualen Anteil der Beobachtungen (relative Häufigkeiten) wider.
- *Eignung:* Säulendiagramme können sowohl für nominal wie ordinalskalierte Merkmale verwendet werden.

Das Histogramm. Eine andere Bezeichnung für das Histogramm (*histogram*) lautet Flächendiagramm.

- *Charakterisierung:* Die Ausprägungsgrade des Merkmals sind intervall- oder verhältnisskaliert und beschreiben ein Ausprägungskontinuum. Damit schliessen die Ausprägungskategorien nahtlos aneinander an. Neben der Höhe der Säulen (Häufigkeit) ist auch deren Breite (Kategorienbreite) von Bedeutung.
- *Eignung:* Bei intervall- oder verhältnisskalierten Daten wird anstelle des Säulendiagramms das Histogramm gewählt. Häufig sind bei intervall- und verhältnisskalierten Daten zusammenfassende Kategorien nötig. Es ist zu beachten, dass dabei Informationen nicht nur verloren gehen, sondern die Daten auch – je nach Wahl der Kategorienbreite – anders wahrgenommen werden können.

Das Liniendiagramm. Liniendiagramme werden auch Häufigkeitspolygon, Kurvendiagramm oder Polygonzug genannt.

- *Charakterisierung:* Die einzelnen Ausprägungen werden durch eine Gerade miteinander verbunden.
- *Eignung:* Liniendiagramme sind vor allem zur Darstellung von zeitlichen Verläufen eines untersuchten Merkmals geeignet. So zeigt die Abbildung 13-4 wie

sich die Armuts-Quote von 1992–1999 verändert. Da die Verbindungslinien zwischen benachbarten Kategorien einem Betrachter suggerieren, dass die Häufigkeiten „kontinuierlich“ zu- bzw. abnehmen, ist dieser Diagrammtyp zur grafischen Darstellung von Häufigkeitsverteilungen nicht zu empfehlen.

13.3.3 Verdichtung der Daten („Mittelwerte“ und Streuung)

Häufig will man Merkmale miteinander vergleichen. Es interessiert beispielsweise, ob männliche oder weibliche Studierende für die Bearbeitung eines Lernschritts durchschnittlich mehr Zeit aufwenden. Wie in Abbildung 13-5 ersichtlich ist, sind aus tabellarisch oder grafisch dargestellten Häufigkeitsverteilungen eindeutige Aussagen zu derartigen Fragen nur schwierig oder kaum ablesbar.

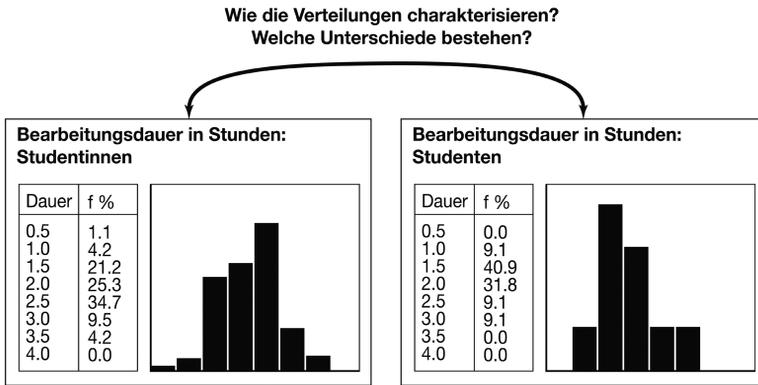


Abbildung 13-5: Über die Schwierigkeit, tabellarische und grafische Verteilungen miteinander zu vergleichen
Anmerkungen: Angabe der *Dauer* für die Bearbeitung in Stunden (von 0,5 bis 4,0 Stunden) und f % (= *frequency percent*) als Angabe der relativen Häufigkeit in Prozent.

Das Bestreben der deskriptiven Statistik, Beobachtungsdaten knapp und präzise zu charakterisieren, hat zur Entwicklung einer Anzahl von Kennwerten geführt. Diese sollen die Daten möglichst gut repräsentieren und zur Beschreibung der Verteilung verwendet werden können. Dabei werden viele Einzelinformationen zu wenigen, aber aussagekräftige Größen verdichtet. Zwei Gruppen von Kennwerten können zur Beschreibung von Häufigkeitsverteilungen unterschieden werden (vgl. Abbildung 13-6):

- 1) *Masse der zentralen Tendenz* („Mittelwerte“): Diese ermitteln den „Schwerpunkt“ einer Verteilung (siehe Kap. 13.3.4 „Masse der zentralen Tendenz ...“).
- 2) *Masszahlen zur Beschreibung der Streuung*: Sie beschreiben, wie dicht die einzelnen Daten beieinander liegen bzw. wie stark sie streuen (siehe Kap. 13.3.5 „Streuungsmaße“).

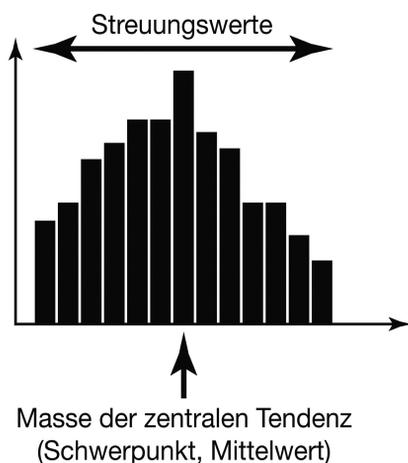


Abbildung 13-6: Charakterisierung von Häufigkeitsverteilungen durch Angabe von Streuungswerten und Massen der zentralen Tendenz

13.3.4 Masse der zentralen Tendenz („Mittelwerte“)

Der „Schwerpunkt“ einer Verteilung kann durch Angaben von „Mittelwerten“ – man spricht auch von Lagemassen oder Massen der zentralen Tendenz – charakterisiert werden.

Masszahlen zur Beschreibung der zentralen Tendenz: Zur Beschreibung der zentralen Tendenz lassen sich verschiedene Kennwerte unterscheiden. Die wichtigsten sind: (1) der arithmetische Mittelwert, (2) der Modus, (3) der Median. Diese Masse der zentralen Tendenz erfassen den durchschnittlichen, den typischen oder den zentralen Ausprägungsgrad einer Verteilung.

Der arithmetische Mittelwert M . Der arithmetische Mittelwert (*mean*) wird umgangssprachlich auch als Mittelwert oder Durchschnitt bezeichnet. Er wird in wissenschaftlichen Artikeln üblicherweise mit M abgekürzt.

- **Charakterisierung:** Wenn die Daten aus einer Stichprobenerhebung stammen, wird der arithmetische Mittelwert mit \bar{x} (lies „x-quer“) bezeichnet. Die Summe aller Ausprägungswerte wird dabei durch die Anzahl aller Beobachtungen, d. h. die Anzahl aller Messwerte, dividiert. Wenn beispielsweise eine Häufigkeitsverteilung mit den Werten 2, 4, 6 und 6 vorliegt, so beträgt der arithmetische Mittelwert $M = 4.5$ ($2 + 4 + 6 + 6 = 18$; $18/4 = 4.5$). Die mathematische Formel dazu lautet:

$$\bar{x} = \frac{\sum x}{n}$$

wobei: \bar{x} = der arithmetischer Mittelwert, $\sum x$ = Summe aller Werte (d. h. $x_1 + x_2 + \dots + x_n$) und n = Anzahl Beobachtungen.

- **Voraussetzungen:** Zur Berechnung des arithmetischen Mittelwertes sollte das Merkmal mindestens intervallskaliert sein. In vielen publizierten wissenschaftlichen Artikeln werden arithmetische Mittelwerte jedoch auch berechnet, wenn genau genommen nur ordinalskalierte Merkmale vorliegen. Beispielsweise werden bei Rating-Skalen (aus dem Englischen: *Rating* = Einschätzung) arithmetische Mittelwerte gebildet, auch wenn eine Kontroverse darüber besteht, ob diese als intervallskaliert interpretiert werden können. So wird eine Rating-Skala mit den Ausprägungen „1 = stimmt gar nicht“, „2 = stimmt wenig“, „3 = stimmt teils-teils“, „4 = stimmt ziemlich“ und „5 = stimmt völlig“ als intervallskaliert angesehen und der Mittelwert berechnet, weil angenommen wird, dass die Abstände zwischen zwei aufeinanderfolgenden Ausprägungen immer gleich gross sind (weiterführende Informationen vgl. Bortz & Döring, 2007, S. 181f).

- **Eignung:** Der arithmetische Mittelwert ist die am häufigsten verwendete Kennzahl. Vor allem für den Vergleich von Gruppen stellt der arithmetische Mittelwert eine wichtige Kennzahl zur Interpretation der Daten dar.

Da der arithmetische Mittelwert im Sinne eines Durchschnittswerts von allen Ausprägungswerten abhängig ist, kann bei kleinen Stichproben, bei schiefen Verteilungen (die Verteilung der Datenwerte weicht von einer symmetrischen Verteilung ab und ist z. B. rechts-schief bzw. links-steil) oder durch einzelne Extremwerte, die für die Verteilung nicht repräsentativ sind, ein verzerrter Eindruck entstehen. Dies ist in folgendem Beispiel der Fall: Es kommen die Ausprägungswerte „5“, „5“, „6“ und „20“ vor. Der arithmetische Mittelwert wäre somit $M = 9$, also deutlich höher als der Wertebereich, in welchem die Mehrheit der Beobachtungen angesiedelt sind.

Der Modalwert *Mo*. Andere Bezeichnungen für den Modalwert *Mo* (*mode*) sind: Modus, häufigster Wert oder dichtester Wert.

- **Charakterisierung:** Der Modalwert *Mo* ist diejenige Merkmalsausprägung, die am häufigsten vorkommt. Wenn beispielsweise die Häufigkeitsverteilung 3, 4, 4, 4, 4, 5, 6 gegeben ist, so beträgt der Modus $Mo = 4$.
- **Voraussetzungen:** Der Modalwert *Mo* kann prinzipiell bei jedem Skalenniveau berechnet werden.
- **Eignung:** Die Angabe des Modus *Mo* ist dann sinnvoll, wenn eine Häufigkeit die anderen Häufigkeiten dominiert oder die Verteilung in der „Umgebung“ des Modus *Mo* eine erkennbare Konzentration aufweist. Der Modalwert wird in Untersuchungen jedoch eher selten angegeben, da durch die Angabe des häu-

figsten Wertes die Daten nur grob charakterisiert werden können. Nützlich ist er vor allem bei nominalskalierten Daten, wenn bezeichnet werden soll, welche Ausprägungs-Kategorie am häufigsten vorgekommen ist. Von einer eigentlichen zentralen Tendenz kann aber kaum gesprochen werden.

Der Median Md . Der Median Md (*median*) wird auch als Zentralwert oder zentraler Wert bezeichnet.

- **Charakterisierung:** Nachdem die Ausprägungsgrade in eine Rangreihe gebracht wurden, d. h. die Ausprägungen vom niedrigsten zum höchsten Wert aufgelistet sind, kann der Median bestimmt werden: Liegen über einem Wert genau so viele Ausprägungen wie unter diesem Wert, so wird dieser Wert als Median bezeichnet. Dies bedeutet, dass 50 Prozent der Ausprägungen unter dem Median liegen und 50 Prozent der Ausprägungen über dem Median.
- **Medianbestimmung bei ungerader Anzahl Beobachtungen:** Die Berechnung des Medians ist einfach, wenn ein kleiner Datensatz mit einzelnen Ausprägungen eine ungerade Anzahl von Beobachtungen aufweist. Beispiel: Es seien die Ausprägungswerte 5, 8, 12, 15, 100 gegeben. Der Median Md ist 12, da er in der „Mitte“ der Beobachtungen liegt.
- **Medianbestimmung bei gerader Anzahl Beobachtungen:** Bei einer geraden Zahl von wenigen Beobachtungen – sofern sie intervallskaliert sind – berechnet sich der Median als arithmetischer Mittelwert der beiden zentralen Werte der geordneten Datenreihe. Beispiel: Werte: 1, 5, 8, 12, 15, 100; Median $Md = (8 + 12)/2 = 10$. Bei ordinalskalierten Daten mit einer geraden Anzahl von Beobachtungen ist es nicht möglich, die Werte der beiden mittleren Fälle zu halbieren wie bei intervallskalierten Daten. Man beschränkt sich deshalb darauf anzugeben, zwischen welche zwei Beobachtungen der Median zu liegen kommt.
- Wenn bei grösseren Datensätzen und intervallskalierten Merkmalen jedoch Ausprägungen mehrfach vorkommen, so wird der Median Md mit einer linearen Interpolation entweder manuell oder mit einer Statistiksoftware näherungsweise bestimmt und entspricht damit keinem tatsächlich beobachteten Ausprägungsgrad, sondern einem Wert zwischen zwei Ausprägungen.
- **Voraussetzungen:** Da zur Bestimmung des Medians die Ausprägungswerte in eine Rangordnung gebracht werden müssen, kann der Median Md nur bestimmt werden, wenn das Merkmal mindestens ordinalskaliert ist.
- **Eignung:** Im Gegensatz zum arithmetischen Mittel wirken sich einzelne extrem hohe oder extrem niedrige Ausprägungsgrade (Ausreisser oder Extremwerte) nicht auf den Median aus. Wenn beispielsweise die Werte 5, 8, 12, 15 und 100 vorliegen, dürfte die Angabe des Medians Md geeigneter sein, d. h. die Werte besser charakterisieren, als die Angabe des arithmetischen Mittelwerts M ($Md = 12$; $M = 26$). Auf den Median hat es somit keinen Einfluss, ob einzelne Werte weit oberhalb oder unterhalb des Medians Md liegen. Dies bedeutet, dass der Median

Md vor allem dann besonders nützlich ist, wenn stark asymmetrische (schiefe) Verteilungen oder Verteilungen mit einigen wenigen Extremwerten beschrieben werden sollen.

Vergleich der Kennwerte zur Beschreibung der zentralen Tendenz. Die drei Kennwerte zur zentralen Tendenz (arithmetischer Mittelwert, Modus und Median) des Datensatzes von Abbildung 13-5 sind in Tabelle 13-6 dargestellt.

Tabelle 13-6: Vergleich der Kennwerte zur Beschreibung der zentralen Tendenz zu „Bearbeitungsdauer in Stunden“ von Abbildung 13-5

Kennwerte	Studentinnen	Studenten
Arithmetischer Mittelwert M	2.17	1.84
Modus Mo	2.50	1.50
Median Md	2.00	1.75

Aufgrund der obigen Erläuterungen ist es nicht verwunderlich, dass die „Mittelwerte“ unterschiedlich ausfallen können: Es wird ersichtlich, dass die Studentinnen einen höheren arithmetischen Mittelwert M , einen höheren Median Md und einen höheren Modus Mo aufweisen als die Studenten.

13.3.5 Streuungsmasse

Mit der Beschreibung der zentralen Tendenz (arithmetischer Mittelwert, Median oder Modus) kann eine Verteilung hinsichtlich ihres „*Schwerpunktes*“ genauer charakterisiert werden. Zur weiteren Beschreibung einer Verteilung ist jedoch ein zusätzlicher Aspekt erforderlich: Es braucht Angaben zur *Streuung der Ausprägungen* bzw. zur *Homogenität/Heterogenität der Messwerte*.

Mit den *Masszahlen zur Beschreibung der Streuung (Dispersionsmasse)* sollen die in einer Untersuchung festgestellten Unterschiede angemessen beschrieben und quantifiziert werden. Folgende Masszahlen informieren u. a. über die Unterschiedlichkeit von Werten: Die Variabilität V , die Standardabweichung s und die Varianz s^2 .

Werden Studentinnen und Studenten nach der Bearbeitungsdauer für einen bestimmten Lernschritt befragt, so zeigt die Auswertung, dass sich die beiden Gruppen hinsichtlich der zentralen Tendenz unterscheiden (siehe Tabelle 13-6). Es stellt sich nun die Frage, ob sich die Merkmalsverteilungen innerhalb der beiden Gruppen auch hinsichtlich der Streuung unterscheiden. In Abbildung 13-7 ist dies angedeutet.

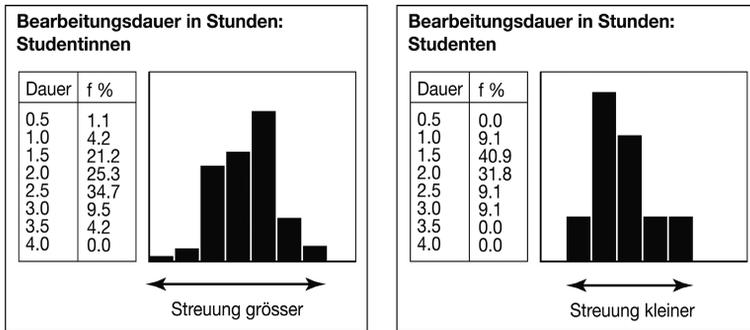


Abbildung 13-7: Vergleich zweier Häufigkeitsverteilungen zur Bearbeitungsdauer eines Online-Lernschritts zwischen Studentinnen und Studenten

Anmerkungen: Angabe der Dauer für die Bearbeitung in Stunden (von 0,5 bis 4,0 Stunden) und f % (= frequency percent): Angabe der relativen Häufigkeit in Prozent.

Die Variabilität V . Die Variabilität V (*range*) wird auch Spannweite oder Variationsbreite genannt.

- **Charakterisierung und Voraussetzungen:** Bei intervallskalierten Merkmalen entspricht die Variabilität der Differenz zwischen dem grössten und dem kleinsten beobachteten Merkmalswert: Im Beispiel aus Abbildung 13-7 weisen die Studentinnen als niedrigsten Wert 0,5 Stunden Bearbeitungsdauer auf und als höchsten Wert zeigt die Verteilung 3,5 Stunden an. Die Variabilität ist somit $V = 3,0$ Stunden ($3,5 - 0,5 = 3,0$). Bei ordinalskalierten Merkmalen kann die Variabilität ermittelt werden, indem nicht die Differenz, sondern die Spannweite durch die Nennung des grössten und kleinsten Merkmalswertes angegeben wird (im obigen Beispiel: Die Testwerte streuen bei Studentinnen zwischen 0,5 Stunden und 3,5 Stunden).
- **Eignung:** Die Spannweite kann einfach und rasch berechnet bzw. benannt werden. Sie kann jedoch durch das Vorhandensein von einzelnen Extremwerten, die für die Verteilung nicht repräsentativ sind, einen „verzerrten“ Eindruck hinterlassen. Beispielsweise seien folgende Werte einer Verteilung gegeben: 23, 66, 67, 72, 74, 76, 77, 77, 79, 80. Die Spannweite liegt somit zwischen 23 und 80. Einzelne Extremwerte an den beiden Enden einer Datenreihe (im Beispiel: 23) können die Spannweite erheblich vergrössern und mehr Variabilität suggerieren als tatsächlich vorhanden ist (90% der Werte liegen in Wahrheit zwischen den Werten 66 bis 80).

Die Variabilität gibt somit lediglich die Breite des Streubereichs an und beschreibt nicht, wie sich die einzelnen Beobachtungen in diesem Bereich verteilen.

Die Standardabweichung s (und Varianz s^2). Die Standardabweichung s wird in empirischen wissenschaftlichen Artikeln mit *SD* (*standard deviation*) abgekürzt.

Üblicherweise wird zum arithmetischen Mittelwert M immer die Standardabweichung SD angegeben.

Charakterisierung: Mit der Standardabweichung soll die „mittlere Abweichung“ der einzelnen Merkmalsausprägungen vom arithmetischen Mittelwert aller Merkmalsausprägungen ermittelt werden. Dies bedeutet, dass von sämtlichen beobachteten Ausprägungen die Differenz zum Mittelwert berechnet wird. Dies drückt sich in der Formel zur Berechnung der Standardabweichung aus:

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$$

wobei:

- $(x - \bar{x})$ die Abweichungen der Ausprägungen vom arithmetischen Mittelwert,
- $\sum(x - \bar{x})^2$ die Summe der quadrierten Abweichungen der Ausprägungen vom arithmetischen Mittelwert,
- n die Anzahl Beobachtungen bezeichnen.

Beziehung zwischen der Standardabweichung s und der Varianz s^2 . Ein anderes Mass hängt mit der Standardabweichung eng zusammen: Die Varianz s^2 . Wird bei der oben stehenden Formel zur Berechnung der Standardabweichung die Quadratwurzel nicht berechnet, so wird diese Masszahl Varianz s^2 genannt. Die Formel zur Berechnung lautet somit:

$$s^2 = \frac{\sum(x - \bar{x})^2}{n}$$

Die Varianz s^2 entspricht somit der Summe der quadrierten Abweichungen der Merkmalswerte vom arithmetischen Mittelwert, dividiert durch die Anzahl der Beobachtungen.

Die Varianz wird in empirischen wissenschaftlichen Artikeln kaum eingesetzt, um Aussagen zur Streuung zu machen. In der schliessenden Statistik (prüf- und entscheidungsstatistische Verfahren) haben jedoch die Varianz s^2 bzw. die Standardabweichung s eine herausragende Bedeutung (siehe Kap. 13.5 „Prüfstatistik“).

- *Voraussetzungen:* Das Merkmal muss mindestens intervallskaliert sein, da die Abstände zwischen den Merkmalsausprägungen und dem arithmetischen Mittelwert berechnet werden müssen.
- *Eignung:* Die Standardabweichung s wird als sinnvolles Mass zur Beschreibung der Variabilität angesehen, wenn die Daten näherungsweise symmetrisch und eingipflig sind, d. h. in der Verteilung nicht mehrere „Wertemaxima“ auftreten.

In der beschreibenden Statistik haben die Standardabweichung s und die Varianz s^2 wegen ihrer geringen Anschaulichkeit und der Schwierigkeit, diese Masszahlen zu interpretieren, keine so grosse Bedeutung. Es sei denn, man möchte – wie im Folgenden dargestellt – die Häufigkeitsverteilungen zweier oder mehrerer Stichpro-

ben miteinander vergleichen (siehe Abbildung 13-7): Bei der Bearbeitungsdauer weisen die Studentinnen eine grössere Standardabweichung ($SD=0.61$) als die Studenten ($SD=0.54$) auf. Die Studentinnen bilden somit eine heterogenere Gruppe als die Studenten, da die „Abweichung“ der Ausprägungen vom arithmetischen Mittelwert bei den Studentinnen höher ist. Ohne die Angabe der Streuung (Standardabweichung) wäre die Beschreibung der Verteilungen der beiden Gruppen nur unvollständig.

Aussagen zur Standardabweichung. Bei *normalverteilten Daten* liegen zwischen den Grenzen $\bar{x} - s$ und $\bar{x} + s$ 68,3% aller Beobachtungen; zwischen den Grenzen $\bar{x} - 2s$ und $\bar{x} + 2s$ 95,5% aller Beobachtungen. Dies ist in Abbildung 13-8 veranschaulicht.

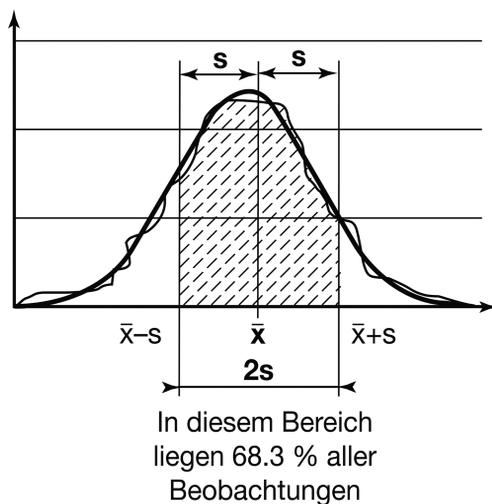


Abbildung 13-8: Darstellung des Zusammenhangs zwischen dem arithmetischen Mittelwert \bar{x} und $\bar{x} - s$ und $\bar{x} + s$

Dies soll mit einem Beispiel verdeutlicht werden: Angenommen wird, dass in Abbildung 13-8 der arithmetische Mittelwert $M=50$ und die Standardabweichung $SD=15$ sei und dass näherungsweise normalverteilte Ausprägungen vorliegen, was in Abbildung 13-8 der Fall ist. Wird vom Mittelwert die Standardabweichung subtrahiert ($50 - 15 = 35$) bzw. addiert ($50 + 15 = 65$), so bedeutet dies, dass bei normalverteilten Daten 68,3% aller Werte zwischen 35 und 65 liegen; 95,5% aller Werte liegen zwischen 20 und 80 ($50 - 30 = 20$, $50 + 30 = 80$).

Vergleich von Kennwerten zur Beschreibung der Streuung. Zum in Abbildung 13-7 dargestellten Datensatz sind die Kennwerte zur Beschreibung der Streuung in nachfolgender Tabelle 13-7 aufgeführt.

Tabelle 13-7: Vergleich der Kennwerte zur Beschreibung der Streuung zum Fallbeispiel von Abbildung 13-7

Kennwerte	Studentinnen	Studenten
Variabilität V (Spannweite)	$V = 3.00$	$V = 2.00$
Standardabweichung s	$s = 0.61$	$s = 0.54$
Varianz s^2	$s^2 = 0.37$	$s^2 = 0.30$

Der Eindruck, dass die Streuung unter den Studentinnen grösser ist als unter den Studenten (siehe Abbildung 13-7) zeigt sich in der Variabilität V wie in der Standardabweichung s . Durch die Berechnung von $\bar{x} - s$ und $\bar{x} + s$ kann Folgendes ausgesagt werden: Rund zwei Drittel der Studentinnen (68,3%) brauchten für die Bearbeitung des Lernschritts zwischen 1,56 Stunden und 2,78 Stunden; rund zwei Drittel der Studenten (68,3%) brauchten für die Bearbeitung des Lernschritts zwischen 1,30 Stunden und 2,38 Stunden.

13.4 Deskriptive Statistik II: Beziehungen zwischen zwei Merkmalen

Fragestellungen zu Zusammenhängen. Verfahren zur Analyse von Zusammenhängen zwischen zwei Merkmalen (Variablen) spielen in den Bildungs- und Sozialwissenschaften eine grosse Rolle. Fragestellungen, die eine solche *bivariate* Betrachtung (gleichzeitige Betrachtung von zwei Variablen) erfordern, sind z. B.:

- Hängt der Erziehungsstil von Eltern mit Persönlichkeitsmerkmalen ihrer Kinder zusammen?
- Besteht zwischen der Punktzahl im Aufnahmetest und dem späteren Studienerfolg (Gesamtpunktzahl in allen Fächern nach dem ersten Ausbildungsjahr) ein Zusammenhang?
- Besteht zwischen der Art des Studierens (Vor- und Nachbereitung von Vorlesungen, Mitglied einer Lerngruppe, Einsatz von Lernstrategien etc.) und dem Prüfungserfolg ein Zusammenhang?

Deterministische vs. stochastische Zusammenhänge. Zusammenhänge gibt es auch in den Naturwissenschaften und der Mathematik. Beispielsweise lassen sich Gesetze der klassischen Physik anhand mathematischer Funktionen beschreiben. Zusammenhänge, die anhand mathematischer Funktionen exakt beschrieben werden können, nennt man *deterministische* Zusammenhänge. Solche Zusammenhänge gibt es in den Bildungs- und Sozialwissenschaften praktisch nie. Man spricht in diesem Fall von *stochastischen* (zufallsabhängigen) Zusammenhängen. Beispielsweise kann man aufgrund der Art des Studierens nicht exakt und fehlerfrei den Prü-

fungserfolg eines Studenten/einer Studentin voraussagen, auch wenn zwischen der „Art des Studierens“ und dem „Prüfungserfolg“ ein gewisser Zusammenhang bestehen dürfte. Viele andere Faktoren wie beispielsweise die „Müdigkeit“, die „Motivation“ oder die „Art der Prüfung“ beeinflussen ebenfalls den „Prüfungserfolg“.

Während man im Alltag von Beziehung, Zusammenhang, Übereinstimmung oder Verbindung spricht, gebraucht die Statistik die Begriffe *Korrelation*, *Assoziation* oder *Kontingenz*.

Gewisse Autoren und Autorinnen sprechen bei der Analyse der Beziehung zweier Variablen ...

- von *Korrelationen*, wenn sie metrisch, d. h. intervall- oder verhältnisskaliert sind,
- von *Assoziationen*, wenn sie ordinalskaliert sind,
- von *Kontingenzen*, wenn die beiden Variablen nominalskaliert sind.

In der Forschungspraxis wird diese Terminologie jedoch nicht immer eingehalten. Die Stärke der Beziehung kann in Abhängigkeit vom Skalenniveau der beteiligten Variablen mit einer Masszahl ausgedrückt werden: Dem Korrelationskoeffizienten bzw. dem Assoziations- oder Kontingenzmass.

Die verschiedenen Möglichkeiten zur Darstellung der Beziehung zwischen zwei Variablen. Beziehungen zwischen zwei Variablen können *tabellarisch* und *grafisch* dargestellt werden. Darauf wird in den folgenden beiden Kapiteln eingegangen (siehe Kap. 13.4.1 und Kap. 13.4.2). Es sei vorweggenommen, dass damit zwar ein Eindruck von der Beziehung zwischen zwei Variablen gewonnen werden kann, aber ...

- sobald eine Tabelle über viele Zellen verfügt, ist es schwierig, überhaupt eine Beziehung zwischen zwei Variablen zu erkennen,
- ein aus einer grafischen Darstellung entnommener Zusammenhang gründet auf einer subjektiven, nicht standardisierten Betrachtung und ist – wenn man die Abbildung weglässt – nur schwer beschreibbar.

Das Bestreben der deskriptiven Statistik, Beobachtungen verdichtet bzw. zusammenfassend zu charakterisieren, hat zur Entwicklung einer Anzahl von Masszahlen und statistischer Verfahren geführt, mit der eine Beziehung zwischen zwei Merkmalen gekennzeichnet werden kann:

- Die *Regressionsanalyse* hat die Aufgabe, die „Art“ eines statistischen Zusammenhangs zusammenfassend zu beschreiben.
- Bei der *Korrelation* wird die „Stärke“ des Zusammenhangs mit einem einzigen Zahlwert ausgedrückt.

In Kapitel 13.4.3 bis Kapitel 13.4.6 werden Möglichkeiten aufgezeigt, wie die Beziehung zwischen zwei Merkmalen numerisch charakterisiert und interpretiert werden kann. Das folgende Fallbeispiel dient der Illustration der weiteren Ausführungen.

► **Fallbeispiel einer Untersuchung zur Fragestellung, ob zwischen autoritativem Erziehungsstil von Eltern und Persönlichkeitsmerkmalen ihrer Kinder ein Zusammenhang besteht**

Problemstellung: Ein autoritativer Erziehungsstil ist u. a. gekennzeichnet durch „Anregung zu Selbständigkeit und Eigenaktivität“, häufiger „Äußerung von Sympathie und Lob“ und „Unterstützung und Hilfe bei Problemen“. Eine Forscherin will nun der Frage nachgehen, ob ein solcher autoritativer Erziehungsstil von Eltern mit Persönlichkeitsmerkmalen ihrer Kinder zusammenhängt. Es wird erwartet, dass zwischen dem Erziehungsstil der Eltern und gewissen Persönlichkeitsmerkmalen ihrer Kinder ein Zusammenhang besteht.

Fragestellung: Besteht zwischen einem autoritativen Erziehungsstil von Eltern und Persönlichkeitsmerkmalen ihrer Kinder ein Zusammenhang?

Vorgehen: 50 bis 100 Eltern sollen einen Fragebogen ausfüllen, welcher Fragen zur autoritativen Erziehung und zu Persönlichkeitsmerkmalen ihrer Kinder enthält.

Einschätzung des autoritativen Erziehungsstils der Eltern: Die Eltern müssen anhand von 15 auf sie bezogenen Aussagen einschätzen, inwieweit sie einen autoritativen Erziehungsstil pflegen. Dabei werden den Eltern jeweils vier Antwortmöglichkeiten geboten: Von „trifft gar nicht zu“ bis „trifft sehr zu“ (für die Auswertung wurden diesen Antwortausprägungen die Werte „1“ bis „4“ zugewiesen). Der Summenwert aller 15 Aussagen sagt aus, ob die Eltern einen autoritativen Erziehungsstil pflegen (geringst möglicher Ausprägungsgrad = 15, höchstmöglicher Ausprägungsgrad = 60).

Einschätzung der Eltern zu verschiedenen Persönlichkeitsmerkmalen ihrer Kinder: Die Persönlichkeitsmerkmale der Kinder werden erfasst, indem Eltern die Persönlichkeit ihrer Kinder anhand einer Kurzversion des Fragebogens NEO-FFI (Costa & McCrae, 1992), welcher aus 30 zu beurteilenden Aussagen besteht, einschätzen. Dieser Fragebogen erfasst fünf Dimensionen der Persönlichkeit:

- *Neurotizismus* bezeichnet die emotionale Stabilität vs. Labilität von Personen.
- *Extraversion* erfasst, wie stark Personen u. a. als gesellig, selbstsicher und aktiv beschrieben werden können.
- *Offenheit für Erfahrungen* erfasst, wie stark Personen ein Interesse an neuen Erfahrungen, Erlebnissen und Eindrücken besitzen.
- *Gewissenhaftigkeit* erfasst, ob Personen eine hohe Impuls- sowie Selbstkontrolle aufweisen.
- *Verträglichkeit* misst die altruistische (im Gegensatz zum Egoismus stehende Rücksichtnahme auf andere) Verhaltenstendenz sowie das Harmoniebedürfnis von Personen.

Für jede der fünf Dimensionen wurde ein Summenwert gebildet, in dem die Werte der jeweiligen 6 Aussagen addiert wurden (Summenwert einer Dimension im Minimum: 6mal die Ausprägung 0 = 0; Summenwert einer Dimension im Maximum: 6mal die Ausprägung 4 = 24).

13.4.1 Die tabellarische Darstellung mittels Kreuztabelle

Kreuztabellen werden auch als Mehrfeldertafeln, Kontingenztabellen (*contingency table*) oder Kontingenztafeln bezeichnet. Unter *Kontingenz* wird die Häufigkeit des gemeinsamen Auftretens zweier nominalskalierter (oder kategorialer) Variablen verstanden.

In *Kreuztabellen* (Kontingenztabellen, Kontingenztafeln) werden die gemeinsamen Häufigkeitsverteilungen zweier nominal- oder ordinalskalierteter Merkmale dargestellt. Auch metrische Merkmale werden manchmal zu Überblickszwecken mittels Kreuztabellen dargestellt.

Charakterisierung. Kreuztabellen dienen dazu, die Häufigkeitsverteilungen zweier Variablen (oder bei multivariaten Auswertungen mehrerer Variablen) tabellarisch darzustellen.

In Tabelle 13-8 ist eine der möglichen Kreuztabellen zum Fallbeispiel aufgeführt: Der Zusammenhang zwischen dem Merkmal *autoritativer Erziehungsstil der Eltern* und dem Persönlichkeitsmerkmal *Verträglichkeit ihrer Kinder*.

Tabelle 13-8: Beispiel einer Kreuztabelle zu *autoritativem Erziehungsstil der Eltern* und dem Persönlichkeitsmerkmal *Verträglichkeit ihrer Kinder* ($n = 54$)

Autoritativer Erziehungsstil der Eltern	Verträglichkeit (relative Häufigkeit in %)						Zeilensumme (absolute Häufigkeiten)
	≤ 4	5–8	9–12	13–16	17–20	21–24	
≤ 24	0	0	0	0	0	0	0
25–33	0	1,9	3,7	13,0	5,6	0	13
34–42	0	0	3,7	20,4	44,4	0	37
43–51	0	0	0	3,7	3,7	0	4
52–60	0	0	0	0	0	0	0

Anmerkung: Dimension *Verträglichkeit*: geringst möglicher Ausprägungsgrad = 0, höchst möglicher Ausprägungsgrad = 24. Merkmal *autoritativer Erziehungsstil*: geringst möglicher Ausprägungsgrad = 15, höchst möglicher Ausprägungsgrad = 60.

Voraussetzungen und Eignung. Kreuztabellen können grundsätzlich bei Merkmalen aller Skalenniveaus angewendet werden. Am ehesten werden sie jedoch bei nominal- und allenfalls ordinalskalierten Merkmalen eingesetzt. Bei intervall- (aber auch ordinal-) skalierten Daten kann eine Kreuztabelle aufgrund der zumeist vielen Ausprägungen schnell unübersichtlich werden.

Kategorisierung von Daten. Bei Darstellung von intervall- und ordinalskalierten Daten in einer Kreuztabelle wird deshalb eine *Kategorisierung der Daten* häufig unumgänglich sein, da sonst zu viele Zellen entstehen würden und damit allfällige Zusammenhänge zwischen den Ausprägungen zweier Merkmale kaum mehr ersicht- lich wären. Auch im Fallbeispiel wurden deshalb die Ausprägungen kategorisiert (siehe Tabelle 13-8): Die Kategorienbreite wurde bei autoritativem Erziehungsstil mit „9“ gewählt, bei Verträglichkeit mit „4“ (zur Kategorisierung von Ausprägungen siehe Kap. 13.3.1). Obwohl die Übersichtlichkeit durch die Kategorisierung

erhöht wurde, sind aus der Kreuztabelle im Fallbeispiel die Zusammenhänge nur schwer erkennbar. Es fällt jedoch auf, dass niedrige Werte in *autoritativem Erziehungsstil der Eltern* eher mit niedrigen Werten in *Verträglichkeit* einherzugehen scheinen, höhere Werte in *autoritativem Erziehungsstil der Eltern* eher mit höheren Werten in *Verträglichkeit*.

13.4.2 Die Grafische Darstellung mittels Streudiagramm

Charakterisierung. Beim Streudiagramm (*scatterplot*) werden die für die einzelnen Personen registrierten Kombinationen von zwei Merkmalsausprägungen (Wertepaare) in ein kartesisches Koordinatensystem eingetragen.

In Abbildung 13-9 sind die in Tabelle 13-8 aufgeführten Daten der Kreuztabelle mittels Streudiagramm grafisch dargestellt.

Die Zusammenhänge zwischen den beiden Merkmalen sind bei der grafischen Darstellung mittels Streudiagramm auf einen Blick erkennbar: Niedrige Werte in autoritativem Erziehungsstil der Eltern gehen eher mit niedrigen Werten in Verträglichkeit ihrer Kinder einher, höhere Werte in autoritativem Erziehungsstil der Eltern mit eher höheren Werten in Verträglichkeit ihrer Kinder. Es ist jedoch schwierig, über die Stärke des Zusammenhangs weitergehende Aussagen zu machen.

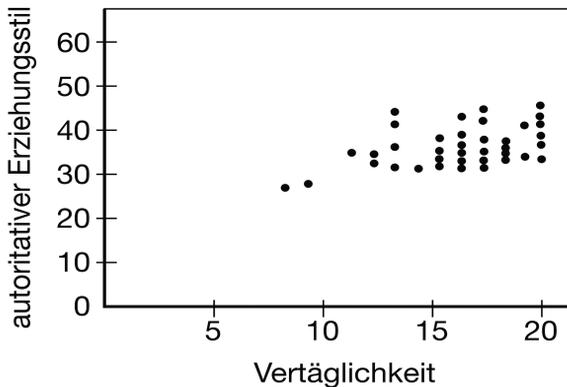


Abbildung 13-9: Beispiel eines Streudiagramms zu *autoritativem Erziehungsstil der Eltern* und dem Persönlichkeitsmerkmal *Verträglichkeit ihrer Kinder* ($n = 54$)

Anmerkungen: Dimension „Verträglichkeit“: geringst möglicher Ausprägungsgrad = 0, höchst möglicher Ausprägungsgrad = 24; Dimension „autoritativer Erziehungsstil“: geringst möglicher Ausprägungsgrad = 15, höchst möglicher Ausprägungsgrad = 60.

Wenn Wertepaare mehrfach vorkommen. Wenn in einem Streudiagramm gewisse Wertepaare mehrfach vorkommen, überlagern sich die Beobachtungen. Solche Überlagerungen sollen – wenn dies die Grafik-Software ermöglicht – in der Abbildung zum Ausdruck gebracht werden: Wenn Wertepaare mehrfach vorliegen, wird eine Symbolik gewählt, die sichtbar macht, wie häufig ein Wertepaar vorkommt (z. B. durch Angabe der Farbintensität oder Grösse der Datenpunkte). Im Streudiagramm aus Abbildung 13-9 sind Überlagerungen *nicht* ersichtlich. Von insgesamt 54 vorhandenen Fragebögen resp. Fällen werden im Streudiagramm lediglich 38 Punkte angezeigt, d. h. es ist nicht sichtbar, ob sich bei einem Datenpunkt mehrere Wertepaare überlagern.

Voraussetzungen und Eignung. Die Darstellung von Wertepaaren in einem Streudiagramm ist sinnvoll, wenn intervall- oder ordinalskalierte Merkmale vorliegen. Streudiagramme sind nützlich, um Extremwerte (*Ausreisser*) zu identifizieren. Dies wäre beispielsweise der Fall, wenn in Abbildung 13-9 ein Wertepaar mit den Ausprägungen in *Verträglichkeit*=3 und in *autoritativem Erziehungsstil*=45 vorkommen würde.

Zudem können Streudiagramme erste Hinweise dazu liefern, ob ein Zusammenhang linear oder nicht-linear ist.

Prüfung, ob eine lineare oder nicht-lineare Beziehung vorliegt. Wenn zwischen zwei Merkmalen ein Zusammenhang besteht, kann die Beziehung linear oder nicht-linear (nonlinear) sein.

Eine *Beziehung* ist *linear*, wenn auf die Veränderung eines Parameters (x-Wert) stets eine dazu proportionale Änderung eines anderen Parameters (y-Wert) erfolgt.

Verschiedene nicht-lineare Beziehungen können unterschieden werden: z. B. *exponentielle*, *parabolische*, *kubische* oder *logarithmische Beziehungen*. In Abbildung 13-10 sind verschiedene Beziehungen dargestellt.

Sollen lineare oder nicht-lineare Zusammenhänge numerisch charakterisiert werden, so bieten sich je nach Art der Beziehung *unterschiedliche Verfahren der Berechnung* an. Im Folgenden wird *nur auf lineare Modelle* näher eingegangen, da lineare Beziehungen einerseits den einfachsten und andererseits den am häufigsten auftretenden Anwendungsfall darstellen.

Es soll betont werden, dass Verfahren zur Berechnung von linearen Beziehungen nicht angewendet werden sollen, ohne vorher zu *prüfen, ob die Annahme eines Modells der linearen Beziehung überhaupt angebracht ist*. Die visuelle Darstellung der Wertepaare mittels eines Streudiagramms kann erste Hinweise liefern, ob eine Beziehung linear bzw. non-linear ist. Wird das falsche Modell gewählt, so können zwischen zwei Merkmalen enttäuschend niedrige Zusammenhänge resultieren (anstatt stärkere Zusammenhänge bei Wahl eines anderen Modells).

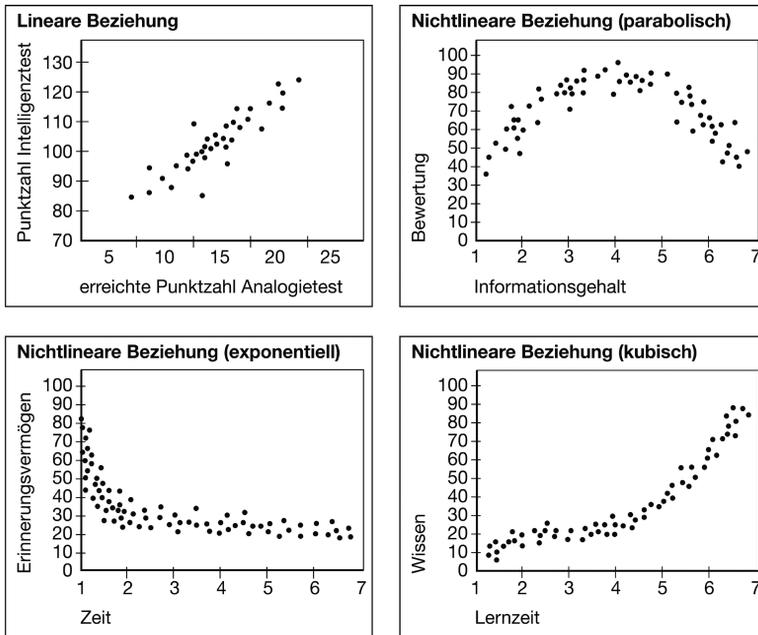


Abbildung 13-10: Darstellung linearer und nicht-linearer Beziehungen zwischen zwei Merkmalen mittels Streudiagramm

13.4.3 Merkmalsvoraussagen: Die Regressionsanalyse

Eine Vorbemerkung: Im Folgenden wird nur auf die *lineare* Regressionsanalyse näher eingegangen. Für die Ermittlung *nicht-linearer Funktionskurven* konsultiere man entsprechende Fachliteratur (z. B. Bortz & Schuster, 2016).

Vorhersage eines Merkmals. Wenn ein Merkmal mit einem anderen zusammenhängt, kann der Ausprägungsgrad des einen Merkmals zur Vorhersage des anderen eingesetzt werden. Wie soll beispielsweise eine Institution Studierende auslesen, wenn sich für einen bestimmten Studiengang mehr Studierende anmelden als ausgewählt werden können? Eine Möglichkeit besteht darin, nur solchen Personen den Zugang zum Studium zu ermöglichen, welche das Studium (möglichst) erfolgreich absolvieren dürften. Ein Aufnahmetest könnte deshalb zukünftig erfolgreiche von weniger erfolgreichen Studierenden identifizieren helfen.

Mit *Zusammenhangsanalysen* (siehe Kap. 13.4.4) kann in einem ersten Schritt überprüft werden, ob zwischen dem *Aufnahmetest* und dem *Studienerfolg* (= Gesamtpunktzahl in allen Fächern nach dem ersten Studienjahr) überhaupt ein

Zusammenhang besteht und wie hoch dieser ausfällt. Wenn zwischen Aufnahmetest und Studienerfolg ein hoher Zusammenhang besteht, dann liegen aufgrund des Aufnahmetests Informationen vor, um geeignete Studierende auslesen zu können.

Mit der *Regressionsanalyse* kann durch die mathematische Verknüpfung von zwei Merkmalen der Ausprägungsgrad des einen Merkmals zur Vorhersage des anderen eingesetzt werden. Dabei wird üblicherweise unterschieden zwischen Prädiktorvariable (die zur Vorhersage eingesetzt wird) und Kriteriumsvariable (die vorhergesagt werden soll).

Aufgrund der „Punktzahl im Aufnahmetest“ lässt sich berechnen, wie viele „Punkte in allen Fächern nach dem ersten Studienjahr“ zu erwarten sind. Wenn beispielsweise eine Studentin im Aufnahmetest 100 Punkte erreicht, so könnte die Regressionsanalyse ergeben, dass bei dieser Studentin nach dem ersten Studienjahr eine Gesamtpunktzahl aus allen Fächern von 150 Punkten erwartet werden kann. Erreicht ein Student im Aufnahmetest 50 Punkte, so wird bei diesem Studenten aufgrund der Regressionsrechnung nach dem ersten Studienjahr nur eine Gesamtpunktzahl von 125 prognostiziert.

Andere Beispiele, bei welchen Regressionsanalysen eingesetzt werden können:

- Welches Körpergewicht ist bei einzelnen Personen aufgrund der Körpergröße zu erwarten?
- Welche Schlafdauer kann aufgrund einer bestimmten Dosis eines Schlafmittels vorhergesagt werden?

Einfachregression oder multiple Regression? In diesem Kapitel wird jeweils von einer unabhängigen Variablen (wie z. B. Aufnahmetest) ausgegangen, man spricht dabei auch von *Einfachregression*. Um genauere Aussagen zu ermöglichen, könnten nicht nur diese eine Variable *Aufnahmetest*, sondern auch noch andere Variablen wie *letzte Zeugnisnoten*, *Ergebnisse eines Interviews*, eine *Selbstbeurteilung der Studierenden* etc. berücksichtigt werden, um den Studienerfolg besser voraussagen zu können. Wenn *mehrere* unabhängige Variablen berücksichtigt werden, spricht man von *multipler Regression*, d. h. die Ausprägung eines bestimmten Merkmals wird durch *mehrere* Variablen zu erklären versucht.

Annahme einer deterministischen Beziehung zwischen zwei Variablen. Um die Regression besser verstehen zu können, soll zuerst vom einfacheren Fall einer *deterministischen Beziehung* ausgegangen werden. Eine deterministische lineare Beziehung zwischen zwei Variablen kann durch folgende Gleichung beschrieben werden (siehe auch Abbildung 13-11):

$$y = a + b \cdot x$$

Es bedeuten:

y = Werte auf der y -Achse (abhängige Variable),

x = Werte auf der x -Achse (unabhängige Variable),

a = die Höhenlage (d. h. der Schnittpunkt der Geraden mit der y -Achse),

b = Steigung der Geraden.

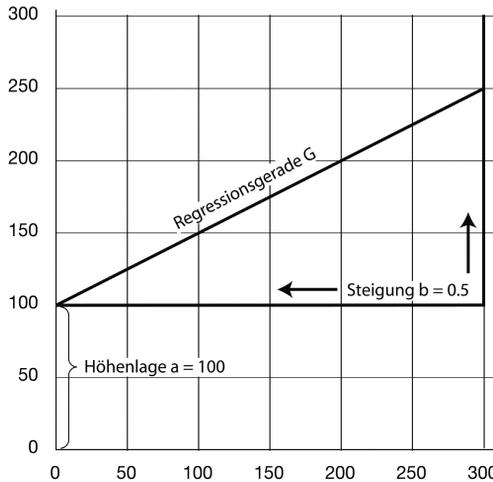


Abbildung 13-11: Eine deterministische Beziehung zwischen zwei metrischen Variablen

Man bezeichnet „ a “ und „ b “ als *Regressionskoeffizienten*. Aus Abbildung 13-11 ist ersichtlich, dass zwischen den beiden Merkmalen folgende lineare Beziehung besteht, d. h. alle Punkte liegen auf der Geraden:

$$y = 100 + 0,5 \cdot x$$

wobei Höhenlage $a = 100$, Steigung der Geraden $b = 0,5$.

Wenn x bekannt ist, kann y eindeutig vorausgesagt werden. Dies soll am eingangs erwähnten Beispiel nun erläutert werden:

Wenn im Aufnahmetest 100 Punkte: Welcher Studienerfolg (Gesamtpunktzahl) wird vorausgesagt? Der Einfachheit halber wird angenommen, dass zwischen „Aufnahmetest“ (x) und „Studienerfolg“ (y) eine deterministische Beziehung besteht, welche durch die Gleichung $y = 100 + 0,5 \cdot x$ ausgedrückt werden kann. Erreicht nun eine Person im Aufnahmetest beispielsweise $x = 100$ Punkte, so wird sie aufgrund der Berechnung als „Gesamtpunktzahl aller Fächer nach dem ersten Jahr“ $y = 150$ Punkte erreichen ($y = 100 + 0,5 \cdot 100$).

Natürlich gibt es keine solche deterministische Beziehung zwischen den zwei Variablen, deshalb wird im nächsten Abschnitt auf den „realistischen“ Fall einer stochastischen Beziehung eingegangen.

Die stochastische (zufallsabhängige) Beziehung zwischen zwei Variablen. Es wurde bereits darauf hingewiesen, dass in den Sozialwissenschaften zwischen zwei Merkmalen fast immer ein stochastischer Zusammenhang besteht und somit vom Ausprägungsgrad des einen Merkmals nicht exakt und fehlerfrei auf den Ausprägungsgrad des anderen Merkmals geschlossen werden kann. Die Punkte werden in diesem Fall nicht mehr alle auf der Geraden liegen, sondern sie werden mehr oder weniger von der Geraden abweichen, d. h. es entsteht ein sogenannter *Punkteschwarm*, wie in Abbildung 13-12 dargestellt.

Im Beispiel dürften zusätzlich viele andere Faktoren dafür mit verantwortlich sein, dass ausgehend von den erreichten Punktzahlen im Aufnahmetest der „Studienerfolg“ nicht exakt vorausgesagt werden kann. „Abweichungen“ sind z. B. zurückzuführen auf „Tagesform“, „Ermüdungseffekte“, „Prüfungsangst“, „Engagement im ersten Studienjahr“, „investierte Lernzeit“, „Lernstrategien“ usw.

Die Regressionsgerade, um den Gesamttrend aller Punkte wiederzugeben. Es stellt sich nun die Frage, wie der Gesamttrend aller dieser Punkte im Punkteschwarm am besten wiedergegeben werden kann. Dazu dient die Regressionsrechnung. Mit ihr soll diejenige Gerade ermittelt werden, die den Gesamttrend aller Punkte am besten wiedergibt. Dazu wird eine Gerade (die so genannte Regressionsgerade) so in den Punkteschwarm gelegt, dass die Abweichungen der Punkte in y-Richtung möglichst klein sind (siehe Abbildung 13-12).

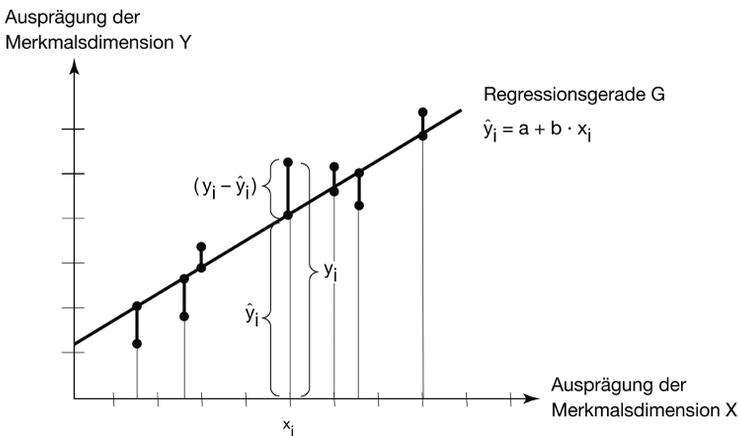


Abbildung 13-12: Punkteschwarm mit Regressionsgerade und Abweichungen der Punkte in y-Richtung

Als *Regressionsgerade* wird diejenige Gerade bezeichnet, welche die Summe der quadrierten Abweichungen der Punkte von der Geraden minimiert. Diese Methode wird Summe-Fehler-Quadrat-(SFQ-)Methode oder die Kleinste-Quadrat-Methode (*least square method*) genannt.

Berechnungsbeispiel. Durch die mathematische Beschreibung der Regressionsfunktion ist es möglich zu berechnen, welcher y -Wert auf Grund eines x -Wertes zu erwarten ist. Im erwähnten Beispiel von „Aufnahmetest“ (x) und „Studienerfolg“ (y) kann folgende Frage gestellt werden: Welche *Gesamtpunktzahl* kann für eine Person, die im Aufnahmetest 100 Punkte erreicht hat, prognostiziert werden? Durch die Berechnung wird – ausgehend vom Resultat im Aufnahmetest – für den „Studienerfolg“ eine Gesamtpunktzahl von 150 erwartet (wenn $a = 100$ und $b = 0.5$), wobei \hat{y} den vorhergesagten y -Wert darstellt.

Zur Genauigkeit von Prognosen. Es ist einsichtig, dass solche Prognosen die Realität nicht unbedingt treffend einschätzen und damit nur bedingt sinnvoll bzw. nützlich sein können. Eine Prognose wird umso besser und treffender sein, je mehr Datenpunkte im Punkteschwarm vorhanden sind und je näher sich diese beobachteten Datenpunkte an der Regressionsgeraden befinden (d. h. je höher die Korrelation ist; vgl. dazu das nächste Kapitel).

13.4.4 Merkmalszusammenhänge: Korrelation bei metrisch-skalierten Variablen

Während sich die Regressionsanalyse mit der Beschreibung der „Art“ des statistischen Zusammenhangs beschäftigt, beschäftigt sich die *Korrelationsanalyse* oder *Zusammenhangsanalyse* mit der Quantifizierung der *Stärke* eines Zusammenhangs zwischen zwei Merkmalen, ausgedrückt durch den *Korrelationskoeffizienten* r .

Verschiedene Korrelationskoeffizienten – je nach Skalenniveau. Die Wahl der Prozedur für die Berechnung des Korrelationskoeffizienten hängt vom Messniveau der Variablen ab. Je nach Skalenniveau sind unterschiedliche Prozeduren zur Berechnung des Korrelationskoeffizienten nötig. Die Produkt-Moment-Korrelation oder *Pearson-Bravais-Korrelation* (*Pearson correlation*) wird häufig einfach als „Korrelation“ bezeichnet und der Korrelationskoeffizient mit r abgekürzt.

Die Stärke des Zusammenhangs zwischen zwei Merkmalen wird durch den *Korrelationskoeffizienten* r ausgedrückt. Wenn zwei metrische Variablen vorliegen, wird die *Produkt-Moment-Korrelation* (auch als Pearson-Bravais-Korrelation, *Pearson correlation*, bezeichnet) berechnet.

Die Verwendung der *Produkt-Moment-Korrelation* setzt ein metrisches Skalenniveau in den beiden betrachteten Merkmalen voraus. Auf dieses Zusammenhangsmass

wird in diesem Kapitel eingegangen; auf Zusammenhangsmasse für nominal- und ordinalskalierte Daten wird in Kapitel 13.4.5 eingegangen.

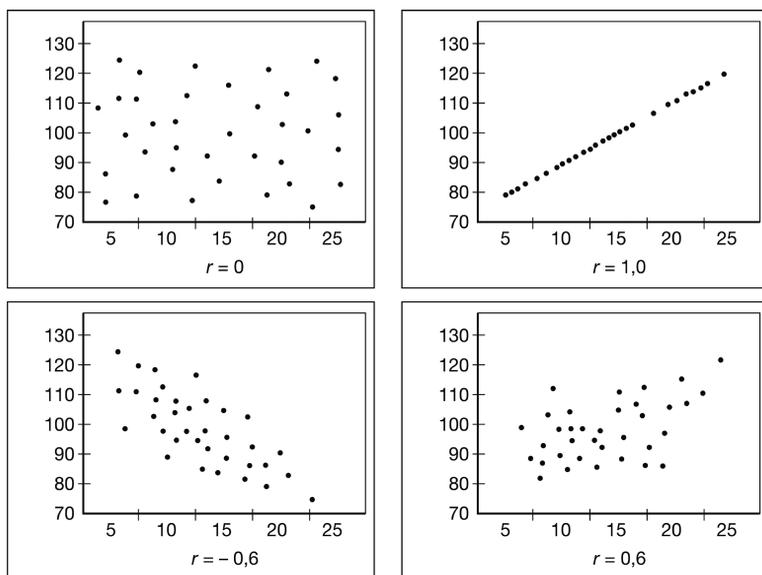


Abbildung 13-13: Typische Streudiagramme und zugehörige Korrelations-Koeffizienten

Der Wertebereich des Korrelationskoeffizienten r . Der Korrelationskoeffizient r kann Werte zwischen -1.00 und 1.00 annehmen. Wenn der Zusammenhang gleichsinnig ist, d. h. der Korrelationskoeffizient ein positives Vorzeichen aufweist (z. B. $r = .45$, $r = .72$, $r = .89$), dann bedeutet dies: Je grösser der Wert der einen Variablen wird, desto grösser wird der korrespondierende Wert in der anderen Variable sein. Ist der Korrelationskoeffizient hingegen negativ (z. B. $r = -.40$, $r = -.34$, $r = -.69$), so drückt dies einen *gegenläufigen* Zusammenhang aus: Je grösser der Wert der einen Variablen wird, desto kleiner wird der korrespondierende Wert in der anderen Variable ausfallen. Abbildung 13-13 zeigt beispielhaft die Korrelationskoeffizienten r zu vier verschiedenen Streudiagrammen.

Interpretation des Korrelationskoeffizienten r . Auch wenn der Korrelationskoeffizient r mit den Zuordnungen verbal beschrieben und die Stärke des Zusammenhangs eingeschätzt werden kann (siehe Tabelle 13-9), sollte bei der Interpretation der Stärke des Zusammenhangs berücksichtigt werden, welche Variablen untersucht wurden: So soll der Wert des Korrelationskoeffizienten hinsichtlich der Werte beurteilt werden, die im Rahmen ähnlich gelagerter Fragestellungen ermittelt wur-

den. Für „weich gemessene“ Merkmale wie Einstellungsskalen dürfte ein Wert von 0.5 bereits ein „Maximum“ darstellen und als relativ hoch (verglichen mit anderen Untersuchungen) beurteilt werden.

Tabelle 13-9: Verbale Beschreibungen des Korrelationskoeffizienten r

Wert des Korrelationskoeffizienten	Verbale Beschreibung
$ r \leq 0.2$	sehr geringe Korrelation
$0.2 < r \leq 0.5$	geringe Korrelation
$0.5 < r \leq 0.7$	mittlere Korrelation
$0.7 < r \leq 0.9$	hohe Korrelation
$0.9 < r \leq 1.0$	sehr hohe Korrelation

Zur Berechnung des Korrelationskoeffizienten. Die *Kovarianz* kennzeichnet den Grad des „Miteinander-Variierens“ zweier Merkmale. Zwei Merkmale *variieren miteinander*, wenn geringe Ausprägungen in einem Merkmal mit geringen Ausprägungen im anderen Merkmal korrespondieren. Treten hingegen hohe Ausprägungen im einen Merkmal auf, so werden auch im anderen Merkmal eher höhere Ausprägungen festgestellt. Dieses „Miteinander-Variieren“ wird als *Kovariieren* bezeichnet. Für dieses Kovariieren gepaarter Daten kann eine Masszahl angegeben werden, die als *Kovarianz cov* (x, y) [*covariance*] bezeichnet wird:

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

wobei:

\bar{x}, \bar{y} = arithmetische Mittelwerte der in Abszissen- resp. Ordinatenrichtung aufgetragenen Merkmalsausprägungen,

x_i, y_i = Koordinaten eines bestimmten Punktes i im Streudiagramm,

n = Anzahl Punkte im Streudiagramm.

Der Korrelationskoeffizient r berechnet sich nun wie folgt: Die Kovarianz $\text{cov}(x, y)$ der beiden Merkmalsausprägungen wird durch das Produkt der Standardabweichungen s_x und s_y dividiert.

Aufgrund der Durchschnittsbildung bei der Berechnung des Korrelations-Koeffizienten r wird klar, warum die *Produkt-Moment-Korrelation* nur für intervall- und verhältnisskalierte Merkmale zulässig ist: Da der Korrelationskoeffizient r stark auf Ausreisser in den Beobachtungen reagiert, sollten die vorliegenden Daten idealerweise normalverteilten Merkmalen entstammen.

Der Determinationskoeffizient r^2 . Eine weitere nützliche Masszahl ist der Determinationskoeffizient r^2 , der als Quadrat des Korrelationskoeffizienten r definiert ist.

Diese Masszahl wird auch als *Bestimmtheitsmass* bezeichnet. Der Determinationskoeffizient r^2 weist besondere Eigenschaften auf:

Der *Determinationskoeffizient* r^2 : Der Ausprägungsgrad von $r^2 \cdot 100\%$ gibt an, welcher prozentuale Anteil der Varianz des einen Merkmals aufgrund der Ausprägungen des anderen Merkmals erklärbar ist.

Beispielsweise existiert zwischen dem Merkmal *autoritativer Erziehungsstil* und dem Merkmal *Verträglichkeit der Kinder* eine Korrelation von $r = .47$. Der Determinationskoeffizient beträgt somit $r^2 = .22$. Dieser Determinationskoeffizient r^2 zeigt den Anteil der (statistisch) erklärten Varianz, d.h. 22% der Streuung (bzw. Heterogenität) in der Verträglichkeit der Kinder sind statistisch auf den autoritativen Erziehungsstil zurückzuführen (d.h. durch diesen determiniert). Der Rest der Streuung, d.h. rund 78%, bleibt statistisch unerklärt und ist somit auf andere, unbekannte Faktoren wie z. B. „andere soziale Erfahrungen“ (mit Lehrpersonen, Mitschülerinnen und Mitschüler etc.), „Charakter“, „momentaner Gesundheitszustand“ etc. zurückzuführen.

13.4.5 Merkmalszusammenhänge: Korrelation bei nicht-metrisch-skalierten Variablen

Im Folgenden wird darauf eingegangen, welche Möglichkeiten bestehen, um eine Korrelation zu berechnen, wenn keine intervallskalierten Variablen vorliegen.

Die Rangkorrelation nach Spearman: Beschreibung der „Stärke“ eines Zusammenhangs zwischen zwei ordinalskalierten Variablen. Für Variablen, die stark von der Normalverteilung abweichen, und für ordinalskalierte Variablen, eignet sich der *Rangkorrelationskoeffizient r_s nach Spearman* (*Spearman correlation*), um die Stärke der Beziehung zwischen zwei Variablen zu beschreiben. Ebenfalls wird der Rangkorrelationskoeffizient r_s berechnet, wenn gewisse Zweifel über das Intervall- bzw. Verhältnisskalenniveau bestehen.

Bei der Berechnung des Rangkorrelationskoeffizienten r_s werden die einzelnen Beobachtungen geordnet und jedem Wert eine Rangzahl zugewiesen. Es entstehen so n -Paare mit Rangzahlen. Aus diesen Rängen wird der Korrelationskoeffizient r_s berechnet, der ebenfalls, wie die Produkt-Moment-Korrelation, im Wertebereich zwischen -1 und $+1$ definiert ist.

Beschreibung der „Stärke“ eines Zusammenhangs zwischen zwei nominalskalierten Variablen. Zur Messung der Stärke des Zusammenhangs zwischen zwei nominalskalierten Merkmalen können ebenfalls geeignete Masszahlen (*Assoziationsmasse* bzw. *Kontingenzmasse*) ausgewählt und berechnet werden.

Fragen nach möglichen Beziehungen auf dem Niveau einer Nominalskala könnten beispielsweise wie folgt lauten:

- Besteht zwischen *Geschlechtszugehörigkeit* („männlich“, „weiblich“) und ausgewähltem *Studienggebiet* („Sprach- und Kulturwissenschaft“, „Naturwissenschaft“, „Kunst“, „Medizin“, „Rechtswissenschaft“, „Wirtschaftswissenschaft“ und „Sozialwissenschaft“ ein Zusammenhang?
- Besteht zwischen *Familienstatus* („ledig“, „verheiratet“, „geschieden“, „verwitwet“) und *Erziehungsstil* („autoritativ“, „autoritär“, „laissez-faire“) eine Beziehung?

Da die Kategorien nominalskalierter Merkmale beliebig angeordnet werden können, ist einsichtig, dass man nicht von einer positiven oder negativen Beziehung sprechen kann und deshalb die Masszahlen zur Charakterisierung einer Beziehung zwischen nominalskalierten Merkmalen vorzeichenlos sein müssen.

Solche Assoziations- bzw. Kontingenzmasse sind beispielsweise die *Prozentsatzdifferenz* ($d\%$), *Lambda* (nach Goodman & Kruskal) oder die *Masszahlen auf der Basis von Chi-Quadrat* (Phi-Koeffizient, Cramer's V, Kontingenzkoeffizient C). Für die Berechnung dieser für nominalskalierte Merkmale konzipierten Assoziationsmasse ist entsprechende Fachliteratur zu konsultieren (z. B. Bortz & Schuster, 2016; Benninghaus, 2005).

Beschreibung der „Stärke“ eines Zusammenhangs zwischen zwei Variablen unterschiedlichen Messniveaus. Es ist jederzeit möglich, Daten so zu behandeln, als seien sie auf einem niedrigeren Skalenniveau gemessen worden (aber nicht umgekehrt!). So ist es beispielsweise möglich, metrische Daten so zu behandeln, als seien sie ordinalskaliert. So können Zusammenhänge zwischen einer metrischen und einer ordinalskalierten Variable mit dem für ordinalskalierte Merkmale verwendete Rangkorrelationskoeffizienten r_s nach Spearman beschrieben werden.

Zum Fallbeispiel, welches eingangs des Kapitels 13.4 aufgeführt ist, wird deshalb die Rangkorrelation r_s nach Spearman berechnet.

► **Fallbeispiel einer Untersuchung mit der Fragestellung, ob zwischen autoritativem Erziehungsstil von Eltern und Persönlichkeitsmerkmalen ihrer Kinder ein Zusammenhang besteht (Fortsetzung)**

Angabe von Masszahlen zur Beschreibung der Stärke des Zusammenhangs. Mit der Variable *autoritativer Erziehungsstil der Eltern* liegt eine ordinalskalierte Variable vor. Die Persönlichkeitsfaktoren werden ebenfalls als ordinalskaliert angesehen. Beziehungen ordinalskalierter Merkmale werden mittels Rangkorrelation r_s nach Spearman bestimmt. Zwischen *autoritativem Erziehungsstil der Eltern* und den fünf Persönlichkeitsfaktoren *Neurotizismus*, *Extraversion*, *Offenheit für Erfahrungen*, *Gewissenhaftigkeit* und *Verträglichkeit* wurden folgende Korrelationen berechnet:

Rangkorrelationen r_s (nach Spearman) zwischen autoritativem Erziehungsstil der Eltern und fünf Persönlichkeitsfaktoren ihrer Kinder (n = 54).

Persönlichkeitsfaktor	Autoritativer Erziehungsstil
Neurotizismus	-.45

Persönlichkeitsfaktor	Autoritativer Erziehungsstil
Extraversion	.15
Offenheit für Erfahrungen	.12
Gewissenhaftigkeit	.46
Verträglichkeit	.47

Zwischen *autoritativem Erziehungsstil der Eltern* und den Persönlichkeitsfaktoren *Extraversion* und *Offenheit von Erfahrungen* bestehen sehr geringe Korrelationen ($r_s = .15$ bzw. $r_s = .12$). Zwischen den übrigen Persönlichkeitsfaktoren und dem autoritativen Erziehungsstil der Eltern bestehen knapp mittlere Korrelationen: Während zwischen *autoritativem Erziehungsstil der Eltern* und *Neurotizismus* eine mittlere, negative Korrelation festzustellen ist ($r_s = -.45$), konnten zwischen *autoritativem Erziehungsstil der Eltern* und *Gewissenhaftigkeit* bzw. *Verträglichkeit* mittlere, positive Korrelationen festgestellt werden ($r_s = .46$ bzw. $r_s = .47$).

13.4.6 Interpretation von Korrelationen

Welches ist Ursache, welches ist Wirkung? Häufig werden Korrelationen fälschlicherweise *kausal* interpretiert, d. h. es wird eine Beziehung zwischen Ursache und Wirkung zweier Ereignisse hergestellt und damit Fehlinterpretationen provoziert. Beispielsweise besteht zwischen dem Merkmal *Alter* und dem Merkmal *Blutdruck* ein Zusammenhang: Wird dieser Zusammenhang kausal interpretiert, so nimmt man an, dass das *Alter* den *Blutdruck* beeinflusst bzw. dass das Lebensalter die Ursache für die Höhe des Blutdrucks ist.

Ob dem so ist, kann nicht einfach so behauptet werden, da Korrelationen lediglich über die Grösse eines Zusammenhangs zwischen Variablen Auskunft geben, über eine mögliche Wirkung der einen Variable auf die andere Variable jedoch keine Aussage machen. Grundsätzlich sind – betrachtet man den Zusammenhang zweier Variablen – verschiedene kausale Zusammenhänge möglich:

Merkmal „X“ beeinflusst Merkmal „Y“ kausal. Beispiel: *Vorbereitung auf die Prüfung* (Merkmal X) und *Prüfungsergebnis* (Merkmal Y) korrelieren miteinander, wobei die Vorbereitung auf die Prüfung (zeitlich vorausgehend) das Prüfungsergebnis (zeitlich nachfolgend) beeinflusst.

Merkmal „Y“ beeinflusst Merkmal „X“ kausal – oder nicht? Beispiel: Krämer (2006b) berichtet von einer Schlagzeile in einer Zeitung, dass nach einer amerikanischen Studie 87 Prozent der Meditierenden „eine sehr niedrige Herzinfarkt-Rate“ gehabt hätten. Es wurde angenommen, dass sich *Meditieren* (Merkmal X) positiv auf eine *niedrigere Herzinfarkt-Rate* (Merkmal Y) auswirkt. Dies könnte nach Krämer (2006b, S. 174) zutreffen, der kausale Zusammenhang könnte aber gerade auch umgekehrt sein: „Auch hier kann die Ursache dort liegen, wo die Meldung sie vermutet, vielleicht aber auch anderswo: Menschen, für die Stress und Herzinfarkt [Merkmal Y] Fremdwörter sind, meditieren gern [Merkmal X]“.

Scheinkorrelation: Merkmal „X“ und Merkmal „Y“ werden von einer dritten bzw. mehreren anderen Variablen beeinflusst. Beispiel (nach Krämer, 2006a): Bei erwachsenen Männern besteht eine bemerkenswert negative Korrelation zwischen dem *Einkommen* (Merkmal X) und der *Anzahl Haare auf dem Kopf* (Merkmal Y), d. h. je höher das Einkommen, desto geringer die Anzahl Haare auf dem Kopf und umgekehrt. Aber weder sind die Haare für das Einkommen noch ist das Einkommen für die Haare verantwortlich zu machen. Vielmehr kommt diese negative Korrelation dadurch zustande, dass beide Merkmale von einem dritten Merkmal (dem *Lebensalter*) abhängen: Mit wachsendem Alter nimmt das Einkommen in der Regel zu und die Haare fallen häufiger aus. Man spricht hier von einer *Scheinkorrelation* (= hohe Korrelation zwischen zwei Merkmalen, die inhaltlich nicht gerechtfertigt ist, sondern nur einen scheinbaren Zusammenhang aufweisen). Solche übersehenen Hintergrundvariablen produzieren Nonsenskorrelationen zuhauf (Krämer, 2006b):

Angefangen bei den Klapperstörchen, deren Zahl hoch positiv mit den bundesdeutschen Geburten korreliert, über die Zahl der unverheirateten Tanten eines Menschen und den Kalziumgehalt seines Skelettes (negative Korrelation), Heuschnupfen und Weizenpreis (negative Korrelation), Schuhgröße und Lesbarkeit der Handschrift (positive Korrelation) bis zu Ausländeranteil und Kriminalität (positive Korrelation) spannt sich ein weiter Bogen eines falsch verstandenen bzw. absichtlich missbrauchten Korrelationsbegriffs. (Krämer, 2006a, S. 170)

In der Literatur wird zur Erklärung von Scheinkorrelation häufig das im Zitat erwähnte Beispiel der Klapperstörche verwendet: Obwohl zwischen dem Vorkommen von Störchen und der Geburtenrate verschiedener Länder ein deutlicher Zusammenhang besteht, wissen wir aufgrund von Erfahrungen, dass kein kausaler Zusammenhang zwischen Störchen und Geburten besteht. Offensichtlich muss eine dritte Variable (oder mehrere andere Variablen) mit diesen zusammenhängen. Es könnte sein, dass dies die Variable „wirtschaftliches Entwicklungsniveau eines Landes“ ist: Je höher der Entwicklungsstand eines Landes, desto weniger Störche gibt es und umso niedriger ist die Geburtenrate in solchen Ländern.

Merkmal „X“ und Merkmal „Y“ beeinflussen sich gegenseitig kausal. Beispiel: *Sympathie* (Merkmal X) und *Kontakt* (Merkmal Y) beeinflussen sich gegenseitig.

Schlussfolgerungen zur Interpretation einer Korrelation. Deshalb ist bei der Interpretation von Korrelationen unbedingt Folgendes zu beachten:

Ein Korrelationskoeffizient liefert *keine* Informationen, welche der oben aufgeführten vier Interpretationen richtig ist. Man sollte daher eine hohe Korrelation als einen Hinweis auf einen *möglichen* engen Zusammenhang zwischen zwei Merkmalen verstehen. Um auf einen kausalen Zusammenhang schliessen zu können, sind weitere sachlogische Überlegungen erforderlich. Wollte man Ursache und Wirkung beweisen, so wäre dies am ehesten mit Hilfe eines Experiments möglich.

13.5 Prüfstatistik

Auch wenn die deskriptive Statistik wichtige Informationen über eine Stichprobe bereitstellt, so ist sie zur Überprüfung wissenschaftlicher Hypothesen ungeeignet. Die deskriptive Statistik ermöglicht zwar, eine bestimmte Stichprobe über Häufigkeiten und statistische Kennwerte (z. B. Mittelwerte und Standardabweichungen) zu beschreiben. Sie kann aber im Unterschied zur Prüfstatistik keine Aussagen machen, die über die Stichprobe hinausgehen und damit keine Geltung für die Grundgesamtheit beanspruchen. Anstatt Prüfstatistik wird auch von *Inferenzstatistik*, *Entscheidungsstatistik*, *hypothesenprüfender Statistik* oder *schliessender Statistik* gesprochen.

► Beispiel: Untersuchung über Fernseh-Konsum und schulische Motivation bei Primarschulkindern

In einer Studie soll untersucht werden, ...

1. wie häufig Kinder fernsehen;
2. über welche schulische Motivation Kinder verfügen, die einen hohen oder niedrigen Fernsehkonsum aufweisen;
3. ob sich Kinder mit hohem Fernsehkonsum von solchen mit niedrigem Fernsehkonsum hinsichtlich schulischer Motivation unterscheiden.

Generalisierender Anspruch der Prüfstatistik. Die deskriptive Statistik kann die ersten beiden Fragen nach der Häufigkeit des Fernsehkonsums und nach der durchschnittlichen schulischen Motivation beantworten, nicht aber die dritte Frage nach dem Unterschied in schulischer Motivation zwischen Kindern mit hohem Fernsehkonsum und Kindern mit niedrigem Fernsehkonsum. Diese letzte Frage ist nämlich nicht mehr einfach beschreibend oder deskriptiv, sondern sie hat einen generalisierenden Anspruch. Der generalisierende Anspruch zeigt sich darin, dass man wissen möchte, ob ein allfälliger Unterschied in der schulischen Motivation von Kindern mit niedrigem und hohem Fernsehkonsum nicht nur für die untersuchte Stichprobe gilt, sondern ob er generelle Gültigkeit besitzt und damit für alle Kinder mit niedrigem oder hohem Fernsehkonsum gilt. Mit anderen Worten: Kann von der Stichprobe auf die Grundgesamtheit rückgeschlossen werden?

Mit Grundgesamtheit sind all diejenigen Personen gemeint, die durch die Stichprobe repräsentiert werden sollen.

In diesem Beispiel also alle existierenden Kinder mit niedrigem und hohem Fernsehkonsum. Anstatt von Grundgesamtheit wird in der statistischen Literatur auch von *Population* gesprochen (siehe Kap. 5.4 „Stichproben“). Aufgrund der gewählten Stichprobe sollen mit Hilfe prüfstatistischer Verfahren Aussagen für die Grundgesamtheit gemacht werden können.

Die Inferenz- oder Prüfstatistik trifft auf der Basis von Stichprobenresultaten Schlussfolgerungen über die Grundgesamtheit. Sie basiert auf der Wahrscheinlichkeitstheorie und ermöglicht unter der Annahme eines Restrisikos, Aussagen über die Bedeutsamkeit (Signifikanz) von Unterschieden oder Zusammenhängen zu machen.

Irrtumswahrscheinlichkeit. Der Schluss von der Stichprobe auf die Grundgesamtheit ist immer mit einem Restrisiko – man spricht dabei von Irrtumswahrscheinlichkeit – verbunden. Zum einen sind Erhebungsinstrumente immer mit einem Messfehler behaftet. Messungen eines Merkmals (z. B. mittels Fragen zur schulischen Motivation) sind nie perfekt. Es gibt immer in einem bestimmten Ausmass Messfehler. Dies zeigt sich beispielsweise darin, dass unterschiedliche Messinstrumente, welche aber das gleiche Merkmal messen sollen (z. B. verschiedene Fragen zur schulischen Motivation), mehr oder weniger unterschiedliche Ergebnisse hervorbringen. Erhebungsinstrumente messen also immer nur annähernd das interessierende Merkmal. Zum anderen stellt die Stichprobe ja gerade eine Auswahl von Personen der Grundgesamtheit dar und kann deshalb Ergebnisse hervorbringen, die die Verhältnisse in der Grundgesamtheit nicht angemessen wiedergeben. Mittels statistischer Tests kann nun ermittelt werden, mit welcher Wahrscheinlichkeit ein Unterschied oder ein Zusammenhang „tatsächlich“ (d. h. in der Grundgesamtheit) besteht oder „zufällig“ aufgrund der Stichprobenauswahl entstanden ist.

13.5.1 Grundprinzipien prüfstatistischer Verfahren

Das Grundprinzip und zentrale Begriffe der Prüfstatistik werden in diesem Kapitel am eingangs erwähnten Fallbeispiel kurz aufgezeigt. Für eine vertiefte Auseinandersetzung ist weiterführende Literatur zu konsultieren. Anschliessend werden einzelne Verfahren vorgestellt und schliesslich auf den Unterschied zwischen deskriptiven und prüfstatistischen Fragestellungen und die Interpretation deren Befunde eingegangen.

► Beispiel: Fragestellung als Ausgangspunkt

In der Untersuchung wurde festgestellt, dass in der Stadt Bern 10-jährige Primarschulkinder mit hohem Fernsehkonsum pro Tag in der schulischen Motivation einen Mittelwert von $M_{\text{Khoch}} = 1.79$, solche mit niedrigem Fernsehkonsum einen Mittelwert von $M_{\text{Kniedrig}} = 4.20$ aufweisen. An der Studie nahmen 120 Kinder teil. Es stellt sich nun die Frage, ob tatsächlich ein solcher Unterschied in der schulischen Motivation zwischen Kindern mit hohem und solchen mit niedrigem Fernsehkonsum besteht, oder ob dieser Befund nur zufällig entstanden ist.

Formulierung einer statistisch bearbeitbaren Fragestellung. Darf der festgestellte Unterschied im Fernsehkonsum generalisiert werden? Mit anderen Worten: Besteht ein solcher Unterschied zwischen Kindern mit hohem Fernsehkonsum und solchen mit niedrigem Fernsehkonsum nicht nur bei den befragten Kindern der Stichprobe, sondern bei allen Kindern der Stadt Bern?

Festlegen der „Grösse“, die prüfstatistisch untersucht werden soll (Prüfgrösse). Im Fallbeispiel ist die Prüfgrösse die Differenz zwischen dem Mittelwert der Kinder mit niedrigem Fernsehkonsum und dem Mittelwert der Kinder mit hohem Fernsehkonsum $\bar{x}_{\text{Kniedrig}} - \bar{x}_{\text{Khoch}}$. Der Ausprägungsgrad der Prüfgrösse in dieser Untersuchung beträgt also: $4.20 - 1.79 = 2.41$

Formulierung einer Nullhypothese (Arbeitshypothese) und einer Alternativhypothese (Forschungshypothese). In der Untersuchung soll herausgefunden werden, ob der Unterschied in der schulischen Motivation zwischen Kindern mit hohem und niedrigem Fernsehkonsum zufällig ist oder nicht. Diese Vermutung (Hypothese) gilt es nun zu überprüfen. Damit solche Vermutungen überprüft werden können, wird eine sogenannte *Nullhypothese* und eine *Alternativhypothese* formuliert:

- Die *Nullhypothese* besagt, dass es nicht auszuschliessen ist, dass der beobachtete Unterschied in den Stichprobemittelwerten zufällig zustande gekommen ist. Im Fallbeispiel: Der beobachtete Unterschied von 2.41 in schulischer Motivation zwischen Kindern mit hohem und solchen mit niedrigem Fernsehkonsum ist zufällig zustande gekommen.
- Die *Alternativhypothese* lautet: Der Unterschied in den Stichprobemittelwerten ist nicht zufällig entstanden; er darf damit generalisiert und interpretiert werden. Im Fallbeispiel: Es besteht ein Unterschied hinsichtlich schulischer Motivation zwischen Kindern mit hohem und solchen mit niedrigem Fernsehkonsum.

Prüfverfahren: Annahme oder Ablehnung der Nullhypothese anhand Prüfgrösse und Prüfverteilung. Anhand des konkreten Ausprägungsgrads der Prüfgrösse und der sogenannten Prüfverteilung kann über eine Hypothese entschieden werden. Die Prüfgrösse wurde schon weiter oben eingeführt: $\bar{x}_{\text{Kniedrig}} - \bar{x}_{\text{Khoch}}$. Unter Prüfverteilung ist die theoretische Verteilung der Prüfgrösse gemeint, wenn die Nullhypothese gültig ist. In einem Gedankenexperiment soll die Prüfverteilung für das Fallbeispiel illustriert werden (siehe Abbildung 13-14).

In der Stichprobe, die der Untersuchung des Fallbeispiels zugrunde liegt, beträgt die Differenz in der schulischen Motivation zwischen Kindern mit hohem Fernsehkonsum und Kindern mit niedrigem Fernsehkonsum 2.41 (siehe Abbildung 13-14). Würde man diese Untersuchung nochmals mit anderen Kindern der Stadt Bern durchführen, ist anzunehmen, dass man etwas andere Mittelwerte in schulischer Motivation erhalten würde. Bei einer zweiten Stichprobe könnte sich beispielsweise ein Unterschied von 2.33 zeigen. Würde man insgesamt fünfmal eine solche Un-

tersuchung durchführen, so würde diese Differenz in schulischer Motivation jedes Mal etwas anders ausfallen: 2.47, 2.33, 1.45, 1.56, 1.93. In Abbildung 13-14b ist dargestellt, welche Differenzen auftreten könnten, wenn 100 Stichproben gezogen würden. Welcher Mittelwertsunterschied ist nun der richtige? Würden 1.000 Stichproben gezogen, so würden die Häufigkeitsverteilung der Differenzen im Fallbeispiel eine Normalverteilung andeuten (siehe Abbildung 13-14c). Würden unendlich viele Stichproben gezogen (was faktisch unmöglich ist), so würde dies eine Normalverteilung ergeben (siehe Abbildung 13-14d). Diese ist nichts anderes als die *theoretische Verteilung der Prüfgröße*.

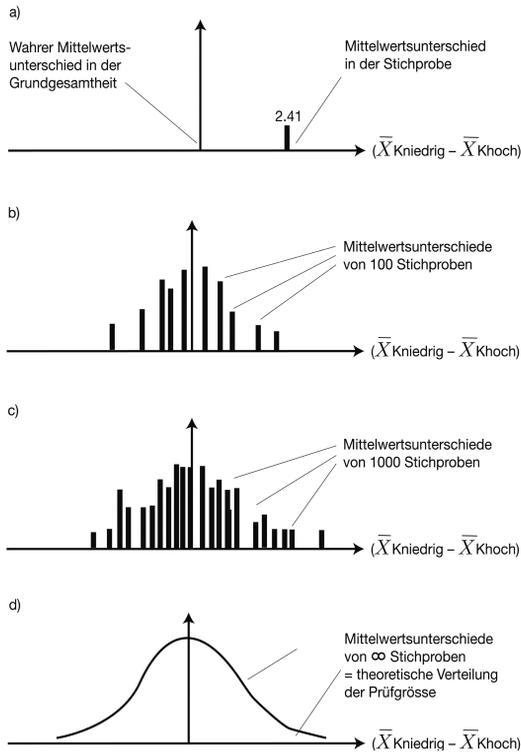


Abbildung 13-14: Mittelwertsunterschiede, wenn a) eine Stichprobe, b) 100 Stichproben, c) 1000 Stichproben, d) ∞ Stichproben gezogen werden

Die im Fallbeispiel aufgeführte Prüfgröße, die Prüfverteilung und die Art, wie ihre Verteilungsparameter geschätzt werden, unterscheiden sich nach Art des gewählten Prüfverfahrens. Einige solcher Verfahren werden in Kapitel 13.5.3 „Prüfstatistische Tests“ kurz vorgestellt.

Übertretenswahrscheinlichkeit und Bestimmung des Signifikanzniveaus. Es kann nun aufgrund dieser theoretischen Verteilung berechnet werden, mit welcher Wahrscheinlichkeit die im Fallbeispiel aufgeführte Differenz in der schulischen Motivation zwischen Kindern mit hohem und niedrigem Fernsehkonsum überschritten wird. In Abbildung 13-15 ist diese Differenz von 2.41 in die Prüfverteilung und die Überschreitungswahrscheinlichkeit von 2.41 (schraffiert) eingetragen worden.

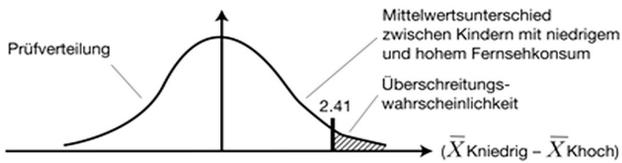


Abbildung 13-15: Überschreitungswahrscheinlichkeit

Die Überschreitungswahrscheinlichkeit gibt an, mit welcher Wahrscheinlichkeit der beobachtete Unterschied in den Stichprobenwerten zufällig zustande gekommen ist.

Es muss nun ein Grenzwert oder kritischer Wert bestimmt werden, damit entschieden werden kann, ob die Nullhypothese oder die Alternativhypothese richtig ist. Dieser Grenzwert soll so gewählt werden, dass mit „hinreichender Sicherheit“ gesagt werden kann, dass der Unterschied nicht durch Zufall zustande gekommen ist. In den Sozialwissenschaften wird häufig eine Wahrscheinlichkeit von mindestens 95% als ausreichend grosse Sicherheit angesehen. Entsprechend beträgt dann die Irrtumswahrscheinlichkeit, auch Restrisiko genannt, 5% oder weniger.

Das Restrisiko, dass wir bezüglich einer Hypothese falsch entscheiden, wird als Signifikanzniveau bezeichnet. Ein irrtümliches Ergebnis ist ein Ergebnis, das nur für die Stichprobe und nicht für die Grundgesamtheit besteht.

Irrtumswahrscheinlichkeiten. Es ist üblich, dass bei der Anwendung von prüfstatistischen Verfahren von Irrtumswahrscheinlichkeit gesprochen wird; das statistische Symbol dafür ist p . Die Irrtumswahrscheinlichkeit p wird als Dezimalzahl angegeben (siehe Tabelle 13-10). Eingebürgert haben sich drei Grenzwerte, man spricht hier vom sogenannten Signifikanzniveau von 5%, 1% oder 0,1%. Das 1%-Kriterium ist ein strengerer Massstab für Signifikanz als das 5%-Kriterium. Wenn beispielsweise ein Mittelwertsunterschied mit dem 1%-Kriterium signifikant ist, bedeutet dies, dass die Wahrscheinlichkeit für einen zufällig entstandenen Mittelwertsunterschied 1:100 beträgt.

Tabelle 13-10: Signifikanzniveaus für die Prüfung auf „Unterschiede“

Irrtumswahrscheinlichkeit	Angabe als Dezimalzahl	verbale Umschreibung	Abkürzungen für die Signifikanz
> 5%	$p > .05$	nicht signifikant	n. s.
≤ 5%	$p \leq .05$	signifikant	*
≤ 1%	$p \leq .01$	hoch signifikant	**
≤ 0,1%	$p \leq .001$	höchst signifikant	***

Ein- und zweiseitige Hypothesenprüfungen. In Abbildung 13-16 ist grafisch dargestellt, wie die Überschreitungswahrscheinlichkeit beurteilt wird. Der Grenzwert kann einseitig oder zweiseitig festgelegt werden (siehe Abbildung 13-16a und 13-16b), je nachdem, ob die Hypothese *einseitig* (gerichtet; *one-tailed hypothesis*) oder *zweiseitig* (ungerichtet; *two-tailed hypothesis*) geprüft werden soll. Eine einseitige Prüfung liegt dann vor, wenn nur eine Seite der Normalverteilung zur Verwerfung der Null-Hypothese verwendet wird. Der p -Wert von 5% liegt dann gesamthaft auf nur einer Seite der Prüfverteilung. Wird die Hypothese zweiseitig getestet, wird die Überschreitungswahrscheinlichkeit auf beide Seiten der Prüfverteilung aufgeteilt. In diesem Fall beträgt der p -Wert auf beiden Seiten 2,5%. Ein zweiseitiger Test wird verwendet, wenn man von einem Unterschied zwischen zwei Gruppen ausgeht, ohne zu wissen, in welche Richtung sich die Gruppen unterscheiden werden. Ein einseitiger Test wird dann berechnet, wenn eine konkrete Erwartung zur Richtung des Unterschiedes vorliegt (z. B. Kinder mit hohem Fernsehkonsum haben eine niedrigere schulische Motivation als Kinder mit niedrigem Fernsehkonsum).

Nicht signifikantes Ergebnis ($p > 5\%$). Abbildung 13-16c zeigt den Fall an, dass die Überschreitungswahrscheinlichkeit $p > 5\%$, d. h. nicht signifikant, ist: Die Wahrscheinlichkeit, dass der beobachtete Unterschied in den Stichprobenmittelwerten zufällig zustande kam, ist so gross, dass der Zufall bei der Interpretation der Ergebnisse nicht ausgeschlossen werden kann. Die Nullhypothese H_0 wird angenommen. Bei einem nicht signifikanten Befund kann deshalb nur die Aussage gemacht werden, dass der zwischen den Stichproben beobachtete Mittelwertsunterschied auch zufällig entstanden sein kann. In unserem Fallbeispiel (unter Annahme, dass ein Unterschied von 2.41 ein nicht signifikantes Resultat zeigt): Der Unterschied in „schulischer Motivation“ zwischen Kindern mit hohem und solchen mit niedrigem Fernsehkonsum kann auch zufällig entstanden sein.

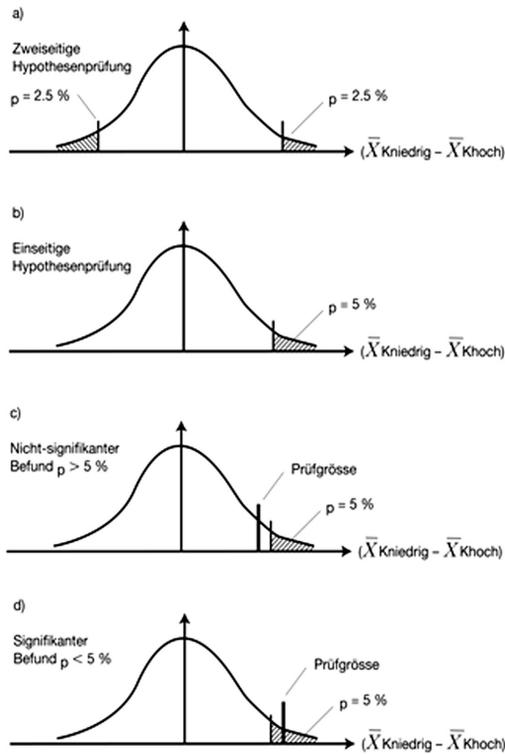


Abbildung 13-16: Beurteilung der Überschreitungswahrscheinlichkeit

Signifikantes Ergebnis. Was bedeutet nun ein Ausprägungsgrad der Prüfgröße $p \leq 5\%$? Die Nullhypothese H_0 wird hier zugunsten der Alternativhypothese H_1 abgelehnt. Es besteht ein signifikanter Unterschied. Dies bedeutet, dass bei der Interpretation des beobachteten Unterschieds in den Stichprobenmittelwerten der Zufall mit einer Irrtumswahrscheinlichkeit $p \leq 5\%$ ausgeschlossen werden kann. Erweist sich in unserem Fallbeispiel die Differenz in schulischer Motivation zwischen Kindern mit hohem und solchen mit niedrigem Fernsehkonsum als signifikant ($p \leq 5\%$), so ist diese um 2.41 bessere schulische Motivation bei Kindern mit geringem Fernsehkonsum nicht zufällig zustande gekommen. Damit darf dieser Befund auf die Population verallgemeinert werden: Es kann davon ausgegangen werden, dass in der Population Kinder mit geringem Fernsehkonsum eine höhere schulische Motivation aufweisen als Kinder mit hohem Fernsehkonsum. Diese Aussage kann mit einer Irrtumswahrscheinlichkeit $p \leq 5\%$ angenommen werden.

Irrtümer bei Hypothesentests. In der prüfstatistischen Literatur wird in diesem Zusammenhang von zwei verschiedenen Fehlern gesprochen, die man bei der Schlussfolgerung von der Stichprobe auf die Grundgesamtheit begehen kann (siehe Tabelle 13-11). Wenn man zum Beispiel den Mittelwertsunterschied in der schulischen Motivation zwischen Kindern mit hohem Fernsehkonsum und Kindern mit niedrigem Fernsehkonsum als bedeutsam (signifikant) beurteilt, dieser Unterschied aber in der Grundgesamtheit nicht besteht, begeht man den Fehler 1. Ordnung (oder *Alpha-Fehler*). Die eigentlich zutreffende Nullhypothese wird abgelehnt. Falls man aber eine Nullhypothese, die eigentlich falsch wäre (d. h. in der Grundgesamtheit nicht zutrifft), als richtig annimmt, begeht man den Fehler 2. Ordnung (oder *Beta-Fehler*).

In prüfstatistischen Verfahren wird lediglich das Risiko, einen Fehler 1. Ordnung zu begehen, ermittelt. Man möchte also wissen, wie hoch das Risiko ist, dass man fälschlicherweise von einem Unterschied ausgeht, obwohl dieser gar nicht besteht. Falls dieses Risiko sehr klein ist, kann man die Null-Hypothese verwerfen.

13.5.2 Effektstärke

Ob ein Ergebnis signifikant ausfällt, ist nicht allein von der Grösse des Unterschieds oder der Stärke eines Zusammenhangs abhängig. Beispielsweise könnte in zwei verschiedenen Studien der Unterschied in der schulischen Motivation zwischen Kindern mit niedrigem und hohem Fernsehkonsum beide Male genau gleich sein und trotzdem könnte der Unterschied im einen Fall signifikant (Irrtumswahrscheinlichkeit $p \leq .05$), im anderen Fall nicht signifikant ausfallen (Irrtumswahrscheinlichkeit $p > .05$). Dies wäre auf unterschiedliche Stichprobengrößen der beiden Studien zurückzuführen.

Tabelle 13-11: Beziehung zwischen Zustand in der Grundgesamtheit, Entscheidung und Fehler im Hypothesentesten

Entscheid	„Tatsächlicher Zustand“ der Grundgesamtheit	
	Nullhypothese ist wahr	Nullhypothese ist falsch
Nullhypothese wird abgelehnt	Fehler 1. Ordnung (Alpha-Fehler)	Richtige Entscheidung
Nullhypothese wird nicht abgelehnt	Richtige Entscheidung	Fehler 2. Ordnung (Beta-Fehler)

Wenn in der einen Studie 200 Kinder untersucht wurden, in der anderen hingegen nur 20 Kinder, dann ist es in der ersten Studie aufgrund der grösseren Stichprobe wahrscheinlicher, dass der Unterschied von einer halben Note nicht zufällig entstanden ist, sondern die „tatsächlichen“ Verhältnisse in der Grundgesamtheit

wiedergibt. Die Wahrscheinlichkeit statistisch signifikante Unterschiede oder Zusammenhänge in einer Studie zu finden, erhöht sich also mit zunehmender Stichprobengröße.

Wie geht man mit dieser Unsicherheit um, dass in Abhängigkeit der Stichprobengröße einmal ein Ergebnis signifikant ausfällt und das andere Mal nicht? Wie kann bestimmt werden, ob ein Ergebnis nicht einfach nur statistisch signifikant ist, sondern ob es auch bedeutsam und von praktischer Relevanz ist?

Um diesen Unsicherheiten entgegenzuwirken, werden in der sozialwissenschaftlichen Statistik neben der Signifikanz auch Masse für die *Effektstärke* angegeben.

Effektstärken sind statistische Masse, die die Stärke oder Grösse eines Unterschiedes oder eines Zusammenhanges angeben. Sie ermöglichen damit zu bestimmen, ob ein Resultat „nur“ statistisch signifikant oder zusätzlich auch praktisch bedeutsam ist.

Durch die Angabe von Effektstärken kann also verhindert werden, dass zu viel in ein statistisch signifikantes Ergebnis, das über eine sehr grosse Stichprobe gewonnen wurde, hinein interpretiert wird. Effektstärken um $.20$ gelten als *schwache Effekte*, solche um $.50$ gelten als *mittlere oder moderate Effekte*, während Effektstärken über $.80$ als *starke Effekte* gelten.

Wenn sich zum Beispiel in einer Untersuchung mit über 1.000 Kindern solche mit niedrigem und hohem Fernsehkonsum in ihrer schulischen Motivation signifikant unterscheiden, die Effektstärke aber einen sehr geringen Wert anzeigt ($< .20$), so ist bei der Interpretation der Bedeutung dieses Ergebnisses Vorsicht geboten. Wenn hingegen die Effektstärke $.50$ oder höher ist, so zeigt dies, dass der Unterschied auch in praktischer, d. h. in erzieherischer Hinsicht relevant sein dürfte.

Je nach statistischem Verfahren werden unterschiedliche Effektstärkenmasse berechnet, wie zum Beispiel der Determinationskoeffizient bei Korrelationen, R^2 bei Regressionsanalysen oder η^2 (Eta-Quadrat) bei Varianzanalysen.

13.5.3 Unterschieds- und Zusammenhangshypothesen

Innerhalb der Prüfstatistik gibt es verschiedene Tests zur Überprüfung von Hypothesen. Je nach Hypothese und Skalenniveau der interessierenden Variablen werden unterschiedliche statistische Tests verwendet. Fragestellungen und Hypothesen lassen sich danach unterscheiden, ob Mittelwerte verglichen und damit *Unterschiede* geprüft oder ob *Zusammenhänge* zwischen Variablen untersucht werden. So können im erwähnten Fallbeispiel einerseits Unterschiede in der schulischen Motivation zwischen Kindern mit hohem und solchen mit niedrigem Fernsehkonsum untersucht, andererseits kann aber auch der Zusammenhang zwischen Fernsehkonsum und schulischer Motivation überprüft werden. Im ersten Fall müssten wir die Signifikanz des Unterschiedes zwischen Kindern mit niedrigem und hohem Fern-

sehkonsument testen, im anderen Fall die Signifikanz des Zusammenhangs zwischen Fernsehkonsum und schulischer Motivation.

An dieser Stelle sollen häufig verwendete prüfstatistische Verfahren zu Unterschieds- und Zusammenhangshypothesen vorgestellt werden. Es kann nur auf die Grundideen dieser Tests und nicht auf statistische Details eingegangen werden. Die folgenden Ausführungen sollen vor allem ermöglichen, die Auswertungen in empirischen sozialwissenschaftlichen Artikeln besser verstehen zu können. Für eine ausführlichere Darstellung prüfstatistischer Grundlagen wird am Ende des Kapitels auf weiterführende Literatur verwiesen.

13.5.4 Tests für Unterschiedshypothesen

Die statistischen Tests zur Prüfung von Unterschiedshypothesen sind vom Skalenniveau (und anderen Voraussetzungen) abhängig, auf dem die interessierende Variable eingeschätzt wurde. Wurde ein Merkmal auf Intervallskalenniveau eingeschätzt (siehe Kap. 13.1 „Grundlage: Schätzen bzw. Messen und Skalenniveaus“), so basiert die Verteilung der Ausprägungsgrade häufig auf dem Mittelwert und der Varianz bzw. der Standardabweichung. Auf diese sogenannten *parametrischen Prüfverfahren* wird zuerst eingegangen, bevor am Schluss *nicht-parametrische (verteilungsfreie) Verfahren* aufgeführt werden.

t-Test. Dieser Test prüft, ob hinsichtlich eines Merkmals zwischen den Mittelwerten zweier verschiedener Gruppen signifikante Unterschiede bestehen. In Abbildung 13-17 ist das zugehörige Wirkungsmodell dargestellt.



Abbildung 13-17: Wirkungsmodell für den t-Test

Der t-Test setzt bei der untersuchten Merkmalsvariable (abhängige Variable) Intervallskalenniveau voraus. t-Tests werden unterschieden in

- (a) t-Tests für unabhängige Stichproben und
- (b) t-Tests für abhängige Stichproben.

(a) *T-Test für unabhängige Stichproben:* Dieser Test vergleicht die Mittelwerte zweier Gruppen, deren Mitglieder entweder der einen oder der anderen, aber *niemals beiden Gruppen* angehören. So gehören Kinder entweder zur Gruppe mit hohem oder dann aber zur Gruppe mit niedrigem Fernsehkonsum. Ein weiteres Beispiel

für unabhängige Gruppen ist Zugehörigkeit zu einer bestimmten Schule (Schule A, Schule B).

Was bedeutet in diesem Beispiel der Wert $t = 11.07$? Dieser Wert ist die Prüfgrösse (siehe Kap. 13.5.1 „Grundprinzip prüfstatistischer Verfahren“) und gibt an, wie gross der Unterschied zwischen den zu vergleichenden Mittelwerten ist. Je grösser dieser t-Wert, desto höher fällt die Signifikanz aus.

(b) *Der t-Test für abhängige Stichproben.* Dieser Test vergleicht die Mittelwerte zweier Gruppen, deren Mitglieder *beiden Gruppen angehören*. Dies ist z. B. immer dann der Fall, wenn Versuchspersonen zu zwei Messzeitpunkten an einer Untersuchung teilgenommen haben. Man spricht in diesem Fall von einer sogenannten *Messwiederholung*. Die Ausprägungen einer Person zum Zeitpunkt 1 und zum Zeitpunkt 2 werden einander dabei paarweise zugeordnet.

Wenn beispielsweise der Fernsehkonsum bei Kindern einmal mit 12 Jahren erhoben wird und dann bei den gleichen Kindern nochmals mit 14 Jahren, so kann mit dem t-Test für abhängige Stichproben überprüft werden, ob sich der Fernsehkonsum verändert hat.

► Beispiel: t-Test für unabhängige Stichproben

Fragestellung: Wie unterscheidet sich die schulische Motivation bei Kindern mit hohem Fernsehkonsum und solchen mit niedrigem Fernsehkonsum?

Methodisches Vorgehen: Die Variablen „schulische Motivation“ und „Fernsehkonsum“ wurden über einen Fragebogen erfasst. Es wurden 10 Fragen zur schulischen Motivation gestellt (z. B. „ich lerne, weil mich der Unterrichtsstoff interessiert“ oder „ich lerne, weil ich den Stoff verstehen möchte“). Die Fragen konnten auf einer fünf-stufigen Skala beantwortet werden (1 = „trifft überhaupt nicht zu“ bis 5 = „trifft völlig zu“). Die Variable „schulische Motivation“ wurde gebildet, indem der Mittelwert aus den 10 Fragen zur schulischen Motivation berechnet wurde. Die Variable „schulische Motivation“ konnte somit Ausprägungen von 1 bis 5 annehmen, wobei 1 eine niedrige schulische Motivation und 5 eine hohe schulische Motivation ausdrückt.

Die Variable „Fernsehkonsum“ wurde erfasst, indem die Kinder nach der Häufigkeit des täglichen Fernsehkonsums gefragt wurden („wie viele Stunden pro Tag schaust du fernsehen?“). Die Kinder wurden zwei Gruppen zugeteilt: In die Gruppe mit hohem und in die Gruppe mit niedrigem Fernsehkonsum. Von den 57 befragten Kindern wurden auf diese Weise 38 Kinder mit hohem Fernsehkonsum und 19 Kinder mit niedrigem Fernsehkonsum identifiziert.

Ergebnisse: Um zu überprüfen, ob sich Kinder mit niedrigem und hohem Fernsehkonsum unterscheiden, wurde ein t-Test berechnet mit Fernsehkonsum als unabhängiger Variable (hoher vs. niedriger Fernsehkonsum) und schulischer Motivation als abhängiger Variable. Der t-Test wurde zweiseitig getestet, da in der Fragestellung keine Annahme über die Richtung des Unterschiedes formuliert ist.

Der t-Test ergab, dass sich Kinder mit hohem Fernsehkonsum ($M = 1.79$, $SD = 0.63$) und Kinder mit niedrigem Fernsehkonsum ($M = 4.20$, $SD = 0.85$) höchst signifikant voneinander unterschieden, $t(55) = 11.07$, $p < .001$. Die Mittelwertsunterschiede zwischen Kindern mit

hohem und niedrigem Fernsehkonsum zeigten die Richtung des Unterschiedes: Kinder mit niedrigem Fernsehkonsum schätzten ihre schulische Motivation höher ein als Kinder mit hohem Fernsehkonsum. Da dieser Unterschied mit einer sehr niedrigen Irrtumswahrscheinlichkeit ($p < .001$) auch in der Grundgesamtheit besteht, kann von einem hoch signifikanten Ergebnis gesprochen werden.

Varianzanalyse. Dieses Testverfahren, häufig mit ANOVA abgekürzt, wird wie der t-Test zum Vergleich von Mittelwerten eingesetzt. Mit der Varianzanalyse werden aber *nicht nur zwei, sondern mehrere Gruppen* bezüglich ihren Mittelwerten in einem Merkmal miteinander verglichen (z. B. drei oder mehr Altersgruppen). Wiederrum setzt dieser Test Intervallskalenniveau bei der abhängigen Variable voraus. Die unabhängige Variable darf ein beliebiges Skalenniveau aufweisen. Es existieren für verschiedene Fragestellungen und Hypothesen verschiedene Formen der Varianzanalyse.

Folgende Aspekte sind für die Einordnung varianzanalytischer Verfahren entscheidend:

- (a) einfaktoriell vs. mehrfaktoriell: Gibt es eine oder mehrere unabhängige Variablen?
- (b) univariat vs. multivariat: Gibt es eine oder mehrere abhängige Variablen?
- (c) mit Messwiederholung vs. ohne Messwiederholung: Wurde die Messungen einmal oder mehrmals bei den gleichen Personen durchgeführt?

(a) *Einfaktorielle und multifaktorielle Varianzanalysen:* Wenn beispielsweise Lernende, die hohen, mittleren und niedrigen Fernsehkonsum aufweisen, in ihrer schulischen Motivation miteinander verglichen werden sollen, so wird eine *einfaktorielle Varianzanalyse* durchgeführt. Einfaktoriell bedeutet, dass nur eine unabhängige Variable (Fernsehkonsum) und nicht mehrere (z. B. Fernsehkonsum und Geschlecht) verwendet werden. Man spricht bei der unabhängigen Variablen auch von Gruppenvariablen, da mehrere Gruppen miteinander verglichen werden sollen (im Beispiel zu Abbildung 13-18 sind es Kinder mit hohem, mittlerem und niedrigem Fernsehkonsum).

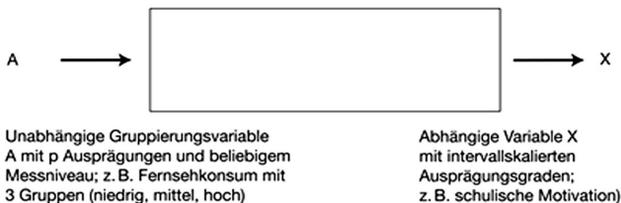


Abbildung 13-18: Wirkungsmodell für die einfaktorielle, univariate Varianzanalyse

► **Beispiel: Einfaktorielle, univariate Varianzanalyse ohne Messwiederholung**

Fragestellung: Bestehen zwischen Kindern mit hohem, mittlerem und niedrigem Fernsehkonsum Unterschiede hinsichtlich schulischer Motivation?

Methode: Die Forscher der oben berichteten Studie überlegten, dass der Unterschied zwischen Kindern mit niedrigem und hohem Fernsehkonsum zu groß, und deshalb eine Aufteilung auf drei Gruppen (niedriger, mittlerer und hoher Fernsehkonsum) angemessener sein könnte.

Ergebnis: In unten stehender Tabelle sind die Mittelwerte der drei Gruppen aufgeführt. Es wurde eine Varianzanalyse durchgeführt mit Fernsehkonsum (hoch, mittel, niedrig) als Gruppenvariable (man spricht dabei auch von Faktor) und schulischer Motivation als abhängiger Variable. Es liess sich hinsichtlich schulischer Motivation ein höchst signifikanter Unterschied des Fernsehkonsums feststellen, $F(2, 54) = 11.94, p < .001$. Die Effektstärke dieses signifikanten Unterschiedes ($\eta = .31$) ist als moderat einzuschätzen. Der Effekt ist somit nicht nur statistisch, sondern auch praktisch bedeutsam.

Tabelle 13-12: Mittelwerte, Standardabweichungen (in Klammern) von Kindern mit niedrigem, mittlerem und hohem Fernsehkonsum

	niedrig ($n = 15$)	mittel ($n = 23$)	hoch ($n = 19$)
Fernsehkonsum	4.53 (0.92)	2.47 (1.28)	2.47 (1.35)

Der F-Wert von 11.94 stellt die Prüfgrösse dar. Auch hier gilt: Je höher der F-Wert, desto höher fällt die Signifikanz aus. Im Unterschied zum t-Test, bei welchem nur zwei Gruppen miteinander verglichen werden, ist bei der Varianzanalyse mit mehreren Gruppenvergleichen nicht gleich klar, auf welche der Gruppenunterschiede das signifikante Ergebnis zurückzuführen ist. Unterscheiden sich in diesem Beispiel alle drei Gruppen oder nur zwei Gruppen? Tabelle 13-12 zeigt die arithmetischen Mittelwerte und Standardabweichungen der drei Gruppen von Kindern zum Fernsehkonsum. Falls letzteres der Fall ist, welche zwei Gruppen unterscheiden sich signifikant? Aus diesem Grund werden Folgeanalysen, sogenannte *Post-hoc Tests*, durchgeführt. Post-hoc Tests testen die Mittelwerte aller Gruppen der Gruppierungsvariable (oder des Faktors) gegeneinander. Umfasst die Gruppierungsvariable wie im Beispiel drei Gruppen so müssen drei Einzelvergleiche durchgeführt werden. Bei vier Gruppen müssen sechs Vergleiche berechnet werden, etc. Bekannte Post-hoc Tests heissen Bonferroni, Fisher's LSD, Duncan, Newman-Keuls, Scheffé und Tukey.

► **Beispiel: Durchführung von Post-hoc Tests**

Um den signifikanten Effekt lokalisieren zu können, wurden Post-hoc Tests nach Tukey durchgeführt.

Ergebnisse: Die Ergebnisse der Einzelvergleiche zeigten, dass sich die Kinder mit mittlerem und hohem Fernsehkonsum von den Kindern mit niedrigem Fernsehkonsum unterschieden

(beide $p < .01$). Die Kinder mit mittlerem Fernsehkonsum und solche mit hohem Fernsehkonsum unterschieden sich aber nicht ($p > .05$).

Angenommen, die Forschenden interessieren sich im Weiteren, ob sich diese Unterschiede für Kinder auch zeigen, wenn zwischen dem Konsum anspruchsloser oder anspruchsvoller Fernsehsendungen unterschieden wird. Bei dieser erweiterten Fragestellung wird gleichzeitig der Einfluss von zwei Faktoren oder Gruppierungsvariablen (Fernsehkonsum-Gruppen und Anspruchsniveau) auf die schulische Motivation varianzanalytisch untersucht. Da in diesem Beispiel nicht mehr nur eine, sondern zwei Gruppierungsvariablen in die Varianzanalyse eingehen, muss hier eine *mehrfaktorielle Varianzanalyse* gerechnet werden (siehe dazugehöriges varianzanalytisches Wirkungsmodell in Abbildung 13-19).

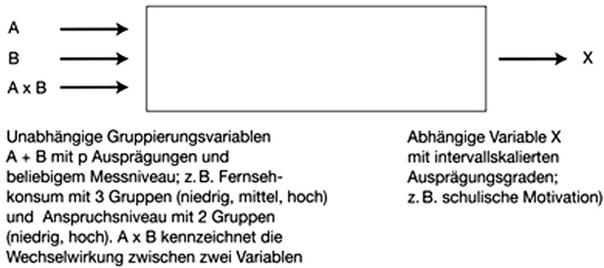


Abbildung 13-19: Wirkungsmodell der mehrfaktoriellen Varianzanalyse

Mit einer mehrfaktoriellen Varianzanalyse erhält man differenziertere Informationen, als wenn für die Gruppierungsvariablen separate Analysen durchgeführt werden. Denn sie ermöglicht zusätzlich etwas über die Wechselwirkung zwischen zwei oder mehreren Gruppierungsvariablen herauszufinden. Zum Beispiel wäre interessant zu erfahren, ob erhöhter Fernsehkonsum nur dann mit niedrigen Werten in der schulischen Motivation einhergeht, wenn das Anspruchsniveau der Fernsehsendungen niedrig ist. Umgekehrt sollten Lernende, welche häufig anspruchsvolle Sendungen schauen, keine niedrigen Werte in der schulischen Motivation aufweisen. Falls sich in mehrfaktoriellen Varianzanalysen solche Wechselwirkungen zwischen den Gruppierungsvariablen zeigen, wird von *Interaktionseffekten* gesprochen.

► Beispiel: Mehrfaktorielle univariate Varianzanalyse ohne Messwiederholung (ANOVA)

Fragestellung: Unterscheiden sich Kinder mit unterschiedlich häufigem Fernsehkonsum und solche, die unterschiedlich anspruchsvolle Sendungen konsumieren in der schulischen Motivation?

Methode: Im Fragebogen zum Fernsehkonsum wurde zusätzlich mit einer offenen Frage erhoben, was für Sendungen geschaut werden. Die von den Kindern genannten Sendungen wurden dann den Kategorien „hohes Anspruchsniveau“ und „niedriges Anspruchsniveau“ zugeordnet.

Ergebnisse: Die deskriptiven Statistiken sind in unten stehender Tabelle aufgeführt. Es wurde eine 2 (Anspruchsniveau: hoch vs. niedrig) \times 3 (Fernsehkonsum: niedrig, mittel, hoch) Varianzanalyse mit schulischer Motivation als abhängiger Variable gerechnet. Die Analyse ergab einen höchst signifikanten Haupteffekt des Fernsehkonsums, $F(2, 51) = 12.99$, $p < .001$. Post-hoc-Analysen zeigten, dass Kinder mit mittleren und hohem Fernsehkonsum eine niedrigere schulische Motivation aufwiesen als Kinder mit niedrigem Fernsehkonsum (beide $p < .01$). Im Weiteren zeigte sich ein signifikanter Haupteffekt des Anspruchsniveaus der konsumierten Sendungen, $F(1, 51) = 10.99$, $p < .01$. Die Mittelwerte (vgl. Tabelle 13-13 unten) zeigen, dass Kinder, die anspruchsvolle Sendungen konsumierten, eine höhere schulische Motivation aufwiesen als solche, die anspruchsvolle Sendungen schauten.

Schliesslich zeigte sich ein signifikanter Interaktionseffekt zwischen Anspruchsniveau und Fernsehkonsum, $F(2, 51) = 3.65$, $p < .05$. Dieser Interaktionseffekt ist in unten stehender Abbildung 13-20 veranschaulicht.

Es ist sichtbar, dass der Unterschied zwischen Kindern mit niedrigem Fernsehkonsum und Kindern mit mittlerem oder hohem Fernsehkonsum bei Kindern, die wenig anspruchsvolle Sendungen schauten, deutlicher ist. Dieser Eindruck konnte auch statistisch bestätigt werden, indem zwei separate einfaktorielle Varianzanalysen getrennt für die beiden Anspruchsniveau-Gruppen durchgeführt wurden. Diese zeigten, dass der Fernsehkonsum lediglich bei den Kindern, die anspruchsvolle Sendungen schauten, einen höchst signifikanten Effekt auf die schulische Motivation zeigte, $F(2, 28) = 15.59$, $p < .001$; bei den Kindern, die anspruchsvolle Sendungen schauten, war die schulische Motivation hoch, unabhängig davon, ob sie wenig, durchschnittlich oder viel Fernsehen schauten.

Interpretation: Dieser Befund zeigt, dass die Häufigkeit des Fernsehkonsums nicht bei allen Kindern mit geringer schulischer Motivation zusammenhängt. Entscheidend für diesen Zusammenhang ist das Anspruchsniveau der Sendungen, die konsumiert werden. Häufiges Fernsehen geht nur bei den Lernenden, die anspruchsvolle Sendungen schauen, mit niedriger schulischer Motivation einher. Lernende, die anspruchsvolle Sendungen schauen, weisen unabhängig von der Häufigkeit des Fernsehkonsums eine hohe schulische Motivation auf.

Tabelle 13-13: Mittelwerte und Standardabweichungen in schulischer Motivation (abhängige Variable) bei Kindern mit hohem, mittlerem und niedrigem Fernsehkonsum (unabhängige Variable)

Anspruchsniveau	Fernsehkonsum		
	niedrig ($n = 15$)	mittel ($n = 23$)	hoch ($n = 19$)
niedrig ($n = 31$)	4.63 (0.74)	2.55 (1.44)	1.92 (1.35)
hoch ($n = 26$)	4.43 (1.13)	4.2 (0.43)	4.2 (0.90)

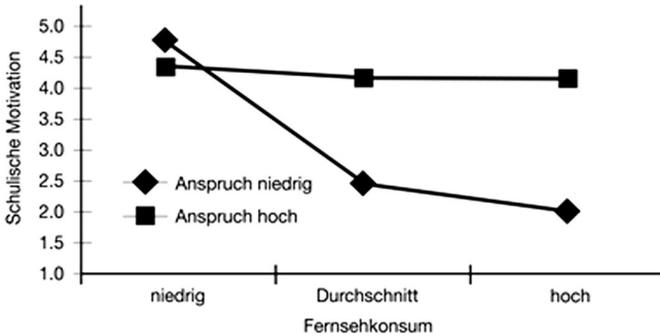


Abbildung 13-20: Interaktion zwischen Anspruchsniveau und Fernsehkonsum

(b) *Univariate und multivariate Varianzanalysen:* Die oben beschriebenen Varianzanalysen beschreiben univariate Varianzanalysen, bei welchen lediglich *eine* abhängige Variable einbezogen wird. Die *Multivariate Varianzanalyse*, auch *MANOVA* genannt, ermöglicht die gleichzeitige Überprüfung von Mittelwertsunterschieden in mehreren abhängigen Variablen (siehe Abbildung 13-21). Sie ist gegenüber separaten univariaten Varianzanalysen zu bevorzugen, wenn sich die verschiedenen abhängigen Variablen einem übergeordneten Konstrukt zuordnen lassen bzw. davon ausgegangen werden kann, dass es inhaltliche Bezüge resp. Abhängigkeiten zwischen den Variablen gibt. Die multivariate Varianzanalyse überprüft nun, ob sich gesamthaft in einem Konstrukt, das mehrere abhängige Variablen umfasst, Unterschiede zwischen Gruppen (z. B. Kinder mit niedrigem und hohem Fernsehkonsum) feststellen lassen. Warum können anstelle einer multivariaten Varianzanalyse nicht einzelne univariate Varianzanalysen durchgeführt werden?

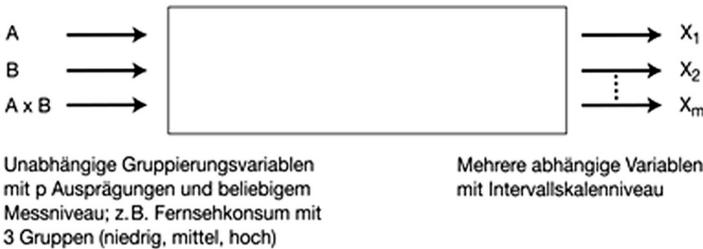


Abbildung 13-21: Wirkungsmodell für die multivariate Varianzanalyse

► **Beispiel: Einfaktorielle, multivariate Varianzanalyse (MANOVA)**

Angenommen eine Forscherin hat nicht die generelle schulische Motivation, sondern die fächerspezifische schulische Motivation erhoben (Motivation für Mathematik, Motivation für Deutsch, Motivation für Biologie). Die fächerspezifischen Motivationen lassen sich dem übergeordneten Konstrukt „schulische Motivation“ zuordnen. Möchte die Forscherin allgemein wissen, ob sich Kinder mit niedrigem und hohem Fernsehkonsum in der schulischen Motivation unterscheiden, muss sie eine multivariate Varianzanalyse rechnen, bei welcher gleichzeitig alle Motivationsskalen in *eine* Analyse eingegeben werden.

Ein Problem separater univariater Varianzanalysen besteht darin, dass sich der Fehler erster Ordnung (Alpha-Fehler) bei der Durchführung separater univariater Varianzanalysen kumuliert. Wenn für jeden einzelnen Test eine Irrtumswahrscheinlichkeit von .05 vorausgesetzt wird, so ist dies ein weniger strenges Kriterium als wenn bei einer einzelnen multivariaten Varianzanalyse, die alle abhängigen Variablen enthält, die gleiche Irrtumswahrscheinlichkeit vorausgesetzt wird. Werden im Beispiel oben drei separate univariate Varianzanalysen gerechnet, so ist die Wahrscheinlichkeit, in einer der drei Tests ein signifikantes Ergebnis zu finden, dreimal höher als wenn alle drei fächerspezifischen Motivationsvariablen in einer multivariaten Varianzanalyse auf Mittelwertsunterschiede getestet werden. Ist man an einer Aussage interessiert, die sich auf das übergeordnete Konstrukt bezieht, so ist dieses strengere Kriterium und damit eine multivariate Varianzanalyse gegenüber einzelnen univariaten Varianzanalysen zu bevorzugen.

Im Weiteren sind Variablen, welche das gleiche Konstrukt erfassen, häufig korreliert. So sollten die drei oben beschriebenen Motivationsvariablen untereinander korrelieren, wenn man von einem übergeordneten Konstrukt „schulische Motivation“ ausgeht. Falls generelle Aussagen zu Unterschieden in der schulischen Motivation von Interesse sind, müssen multivariate Varianzanalysen durchgeführt werden, um den Zusammenhängen zwischen den abhängigen Variablen Rechnung zu tragen.

Es gibt verschiedene Prüfgrößen, welche in multivariaten Varianzanalysen berechnet werden: Pillai-Spur, Wils-Lambda, Hotelling-Spur, Wurzel nach Roy.

(c) *Varianzanalysen mit Messwiederholung*: Gleich wie beim t-Test gibt es auch bei Varianzanalysen die Möglichkeit Daten auszuwerten, die zu mehreren Messzeitpunkten erhoben wurden.

Viele Fragestellungen lassen sich erst beantworten, wenn die gleichen Personen mit den gleichen Messinstrumenten mehr als einmal untersucht werden. Dies gilt insbesondere für experimentelle oder entwicklungspsychologische Fragestellungen (siehe Kap. 5.1 „Forschungszugänge“). Besonders wichtig ist, dass zu den verschiedenen Messzeitpunkten die gleichen Instrumente verwendet werden, da sich sonst bei allfälligen Unterschieden nicht sagen lässt, ob diese auf die unterschiedlichen Messzeitpunkte oder auf die unterschiedlichen Messinstrumente zurückgehen.

In Varianzanalysen mit Messwiederholungen können beliebig weitere Gruppierungs-Variablen (z. B. Geschlecht) einbezogen werden. Diejenigen Variablen, deren Veränderung über die Zeit untersucht wird, werden Messwiederholungs-Variablen oder Innersubjekt-Variablen genannt. Die Gruppierungsvariablen, die über die Zeit konstant sind (z. B. Geschlecht), werden Zwischensubjekt-Variablen genannt.

Nicht-parametrische (verteilungsfreie) Tests zur Überprüfung von Unterschiedshypothesen. In den Sozialwissenschaften wird den erfassten Variablen häufig Intervallskalenniveau zugeschrieben, obwohl sie meist „nur“ Ordinalskalenniveau aufweisen. Intervallskalenniveau setzt voraus, dass die Abstände zwischen den Merkmalsausprägungen einer Variable alle identisch sind (siehe Kap. 13.1 „Grundlage: Schätzen bzw. Messen und Skalenniveau“). Diese Annahme ist bei Antwortskalen, wie sie zum Beispiel in Fragebögen verwendet werden, nicht unproblematisch. Auch wenn Forschende davon ausgehen, dass die semantischen Abstände einer mehrstufigen Antwortskala in einem Fragebogen (z. B. *trifft gar nicht zu*, *trifft eher nicht zu*, *trifft eher zu*, *trifft völlig zu*) alle gleich sind, ist fraglich, ob dies für die Personen, die solche Einschätzungen vornehmen, gleichermassen gilt.

► **Beispiel: Einfaktorielle, univariate Varianzanalyse mit Messwiederholung**

Fragestellung: Unterscheiden sich Kinder mit hohem, mittlerem und niedrigem Fernsehkonsum hinsichtlich schulischer Motivation in der 3. und 5. Klasse?

Methode: Die Kinder wurden mit dem gleichen Fragebogen zum Fernsehkonsum und zur schulischen Motivation zunächst in der dritten und später in der fünften Klasse befragt. Somit liegen zu zwei Messzeitpunkten Informationen über die Häufigkeit des Fernsehkonsums (niedrig, durchschnittlich, hoch) und die schulische Motivation vor.

Ergebnis: Es wurde eine Varianzanalyse mit Messwiederholung mit schulischer Motivation als Innersubjekt-Variable und Fernsehkonsum als Zwischensubjekt-Variable (niedrig, mittel, hoch) durchgeführt. Die Analyse ergab einen höchst signifikanten Effekt der schulischen Motivation, $F(1, 38) = 11.75, p < .001$. Ein Vergleich der Mittelwerte zum ersten und zweiten Messzeitpunkt zeigte, dass die schulische Motivation bei den älteren Kindern ($M = 3.32, SD = 1.52$) niedriger war als bei den jüngeren Kindern ($M = 4.24, SD = 0.92$). Im Weiteren zeigte sich ein signifikanter Effekt des Fernsehkonsums, $F(2, 38) = 4.31, p < .05$. Ein signifikanter Interaktionseffekt zwischen Fernsehkonsum und schulischer Motivation, $F(2, 38) = 3.42, p < .05$, zeigte aber, dass dieser Effekt nicht unabhängig vom Messzeitpunkt war (siehe Abbildung 13-22 unten, welche den Interaktionseffekt veranschaulicht). Separate einfaktorielle Varianzanalysen für die beiden Altersgruppen ergaben lediglich zum zweiten Messzeitpunkt einen hoch signifikanten Effekt des Fernsehkonsums auf die schulische Motivation, $F(2, 40) = 5.66, p < .01$. Zum ersten Messzeitpunkt war die schulische Motivation der Kinder hoch, unabhängig von der Häufigkeit des Fernsehkonsums. Nachfolgeanalysen des signifikanten Effektes zum zweiten Messzeitpunkt ergaben, dass sich die Kinder mit hohem und mittlerem Fernsehkonsum signifikant von solchen mit niedrigem Fernsehkonsum unterschieden (beide $p < .05$).

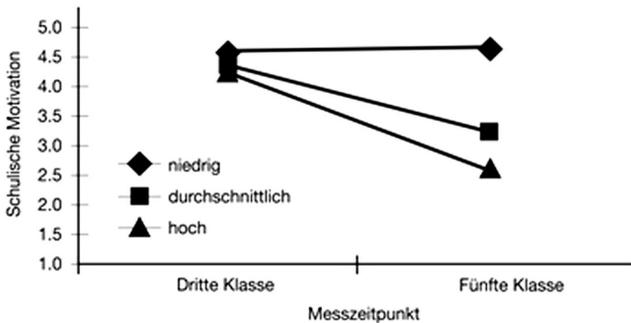


Abbildung 13-22: Schulische Motivation und Fernsehkonsum in der dritten und fünften Klasse

So ist beispielsweise anzunehmen, dass für eine Person der Unterschied zwischen „trifft eher nicht zu“ und „trifft eher zu“ psychologisch bedeutsamer ist als der Unterschied zwischen „trifft eher zu“ und „trifft völlig zu“, da es bei Ersterem um eine grundsätzlichere Entscheidung geht (*trifft zu* oder *trifft nicht zu*), bei letzterem hingegen lediglich um eine Präzisierung.

Trotzdem ist bei vielen Variablen die Anwendung parametrischer Verfahren (d. h. Verfahren, die Intervallskalenniveau voraussetzen) legitim. Wichtig ist nämlich in erster Linie, dass die erfassten Variablen nicht bedeutsam von der Annahme der Normalverteilung abweichen. Ist dies der Fall kann Intervallskalenniveau unterstellt und die dazu passenden parametrischen statistischen Verfahren durchgeführt werden. Weichen aber die Verteilungen der Variablen von der Normalverteilung ab, müssen alternative Tests, sogenannte *nicht-parametrische (verteilungsfreie) Tests*, durchgeführt werden. Für die meisten Fragestellungen gibt es zu parametrischen Tests analoge nicht-parametrische Verfahren. Einige der wichtigsten nichtparametrischen Tests sind in Tabelle 13-14 aufgeführt.

13.5.5 Tests für Zusammenhangshypothesen

Der Chi-Quadrat-Test. Dieser Test überprüft den Zusammenhang zwischen zwei nominal- oder höher skalierten Variablen, zum Beispiel zwischen Geschlecht (Knabe, Mädchen) und Fernsehkonsum (hoch, niedrig). Grundlage zur Berechnung der Prüfgröße Chi-Quadrat (χ^2) sind die *beobachteten und erwarteten Häufigkeiten* der verschiedenen Zellen in der Kreuztabelle (siehe Kap. 13.4.1 „Die tabellarische Darstellung mittels Kreuztabelle“). Die *beobachteten Werte* bezeichnen die Häufigkeiten der Personen, welche in den einzelnen Zellen tatsächlich ausgezählt wurden. Die *erwarteten Häufigkeiten* bezeichnen die Häufigkeiten, die zustande kämen, wenn die Personen zufällig auf die Zellen verteilt wären. Die erwarteten Häufigkeiten basieren somit auf der Annahme, dass kein Zusammenhang zwischen den beiden

nominalskalierten Variablen besteht. Die erwarteten Werte einer bestimmten Zelle (e_{ij}) lassen sich folgendermassen berechnen:

$$e_{ij} = \frac{\text{Zeilensumme}_i \cdot \text{Spaltensumme}_j}{\text{Gesamtsumme } n}$$

wobei:

- Zeilensumme_{*i*} = Zeilensumme, in welcher sich die Zelle e_{ij} befindet
- Spaltensumme_{*i*} = Spaltensumme, in welcher sich die Zelle e_{ij} befindet
- Gesamtsumme n = Gesamtzahl aller Beobachtungen.

Tabelle 13-14: Häufig eingesetzte nicht-parametrische Tests

Fragestellung	Test und Prüfgrösse (in Klammern)	Beispiele zu Ergebnissen
Wie unterscheiden sich zwei unabhängige Stichproben?	U-Test nach Mann-Whitney (U)	Es gab keinen signifikanten Unterschied zwischen Frauen und Männern im Alkoholkonsum ($U = 135.50$, $N1 = 20$, $N2 = 20$, $p = \text{ns}$).
Wie unterscheiden sich zwei abhängige Stichproben?	Wilcoxon-Test (z)	Die Lernenden zeigten nach den Sommerferien eine signifikant höhere schulische Motivation als vor den Sommerferien, $z = 2.53$, $p < .05$.
Wie unterscheiden sich mehr als zwei unabhängige Stichproben?	H-Test nach Kruskal und Wallis (H)	Es zeigte sich ein signifikanter Effekt des autoritären Erziehungsstils (hoch, mittel, niedrig) auf das aggressive Verhalten der Kinder, $H = 8.66$, $p < .05$. Nachfolgeanalysen mit dem U-Test nach Mann-Whitney zeigten, dass hoch autoritäre Eltern aggressivere Kinder hatten als wenig autoritäre Eltern ($z = 2.77$, $p < .05$). Die Kinder durchschnittlich autoritärer Eltern unterschieden sich nicht von den anderen beiden Gruppen.
Wie unterscheiden sich mehr als zwei abhängige Stichproben?	Friedman-Test (χ^2)	Das Gewicht der Teilnehmerinnen und Teilnehmer veränderte sich über die zweimonatige Diät nicht signifikant, $\chi^2(2) = .20$, ns.

Der Chi-Quadrat-Test überprüft nun, ob die beobachteten Häufigkeiten bedeutsam von den erwarteten Werten abweichen. Chi-Quadrat (χ^2) stellt die Prüfgrösse dar, welche die Stärke der Abweichung der beobachteten von den erwarteten Werten und somit auch die Stärke des Zusammenhangs zwischen den beiden Variablen angibt.

► Beispiel: Chi-Quadrat-Test

Fragestellung: Welcher Zusammenhang besteht zwischen Geschlecht und Fernsehkonsum bei Kindern?

Ergebnisse: Zur Untersuchung des Zusammenhangs zwischen Geschlecht und Fernsehkonsum wurde ein Chi-Quadrat-Test durchgeführt. In unten stehender Tabelle 13-15 sind die beobachteten und erwarteten Werte aufgeführt. Die Auswertung ergab einen hoch signifikanten Zusammenhang zwischen Geschlecht und Fernsehkonsum, $\chi^2(1, N=57) = 6.88, p < .01$. Ein Vergleich der beobachteten mit den erwarteten Werten zeigt, dass die Mädchen häufiger als erwartet in der Gruppe der Kinder mit hohem Fernsehkonsum waren, während die Jungen häufiger als erwartet bei den Kindern mit niedrigem Fernsehkonsum waren. Entsprechend zeigte sich bei den Kindern mit niedrigem Fernsehkonsum, dass die Mädchen weniger als erwartet in dieser Gruppe vorkommen, die Jungen hingegen häufiger als erwartet.

Tabelle 13-15: Geschlecht und Fernsehkonsum

Fernsehkonsum	Geschlecht		Total (Zeile)
	Mädchen	Jungen	
Niedrig			
<i>Beobachtet</i>	14	24	38
<i>Erwartet</i>	18.67	19.33	
Hoch			
<i>Beobachtet</i>	14	5	19
<i>Erwartet</i>	9.33	9.67	
Total (Spalte)	28	29	57

Der Chi-Quadrat-Test muss sich nicht auf die Untersuchung von zweistufigen nominalskalierten Variablen beschränken, wie dies im oben aufgeführten Beispiel der Fall ist. Der Test kann auch mit nominalskalierten Variablen mit mehr als zwei Stufen durchgeführt werden. Zum Beispiel könnte untersucht werden, ob der Fernsehkonsum (bestehend aus den drei Gruppen: hoch, mittel und niedrig) mit der sozialen Schicht (niedrig, mittel und hoch) zusammenhängt. Daraus würde sich eine Kreuztabelle mit 3x3 Zellen ergeben.

Korrelationen und Regressionsanalyse. Möchte man den Zusammenhang zwischen ordinal- oder intervallskalierten Variablen bestimmen, werden Korrelationen oder Regressionsanalysen berechnet (siehe Kap. 13.4 „Deskriptive Statistik II: Beziehungen zwischen zwei Merkmalen“). Um bestimmen zu können, ob die Korrelation zwischen zwei Variablen signifikant ist, muss auch hier ein Signifikanztest durchgeführt werden. Dabei werden der Korrelationskoeffizient r und die Irrtumswahrscheinlichkeit p angegeben. Soll nicht nur die Korrelation von zwei Variablen, sondern von mehreren Variablen angegeben werden, so ist es übersichtlicher, wenn

die Korrelationskoeffizienten und deren Signifikanz nicht alle im Text beschrieben, sondern tabellarisch dargestellt werden (siehe Beispiel unten).

Multiple Regressionsanalyse. Ähnlich wie bei der mehrfaktoriellen Varianzanalyse besteht auch bei der Regressionsanalyse die Möglichkeit, gleichzeitig den Einfluss mehrerer unabhängiger Variablen auf eine abhängige Variable zu untersuchen. Anstelle von unabhängigen Variablen wird im Kontext von Regressionsanalysen auch von den *Prädiktorvariablen* gesprochen, anstelle von abhängigen Variablen von *Kriteriumsvariablen*.

► Beispiel: Korrelationen

Fragestellung: Wie hängt die schulische Motivation von Kindern mit der Häufigkeit ihres Fernsehkonsums, ihrem zeitlichen Aufwand für Hausaufgaben und den schulischen Aspirationen ihrer Eltern zusammen?

Methode: Die schulischen Aspirationen der Eltern aus Sicht der Kinder wurden mit einem Fragebogen erhoben. Es wurden sechs Fragen zum elterlichen Interesse und Engagement für die schulischen Aktivitäten und Leistungen gestellt (z. B. *meinen Eltern ist es wichtig, dass ich gut in der Schule bin*). Die Skala „Aspirationen der Eltern“ wurde gebildet, in dem der Mittelwert der Antworten zu den sechs Fragen berechnet wurde.

Ergebnis: Die Analysen ergaben, dass schulische Motivation negativ mit der Häufigkeit des Fernsehkonsums ($r = -.37$) und positiv mit dem zeitlichen Aufwand für Hausaufgaben ($r = .51$) und den elterlichen Aspirationen aus Sicht der Kinder ($r = .63$) korreliert war (siehe Tabelle 13-16). Kinder, welche sich als schulisch motiviert einschätzten, berichteten, dass sie wenig Fernsehen konsumierten, viel Zeit in Hausaufgaben investierten und Eltern hatten, die an ihren schulischen Leistungen und Aktivitäten interessiert waren. Im Weiteren zeigte sich ein positiver Zusammenhang zwischen elterlichen Aspirationen und investierter Zeit in Hausaufgaben. Umso mehr die Eltern an den schulischen Leistungen und Aktivitäten der Kinder interessiert waren, desto mehr Zeit verbrachten die Lernenden mit Hausaufgaben.

Tabelle 13-16: Korrelationen zwischen schulischer Motivation, Fernsehkonsum, Hausaufgaben und Aspirationen der Eltern

	Schulische Motivation	Fernsehkonsum	Hausaufgaben	Aspirationen der Eltern
Schulische Motivation	–			
Fernsehkonsum	–.37**	–		
Hausaufgaben	.51***	–.02	–	
Eltern	.63***	–.06	.48***	–

Anmerkungen: $N = 57$. * $p < .05$. ** $p < .01$. *** $p < .001$

Zum einen zeigt die regressionsanalytische Auswertung, ob die verschiedenen Prädiktorvariablen, deren Einfluss auf eine Kriteriumsvariable untersucht werden soll,

alle zusammen signifikant zur Vorhersage der schulischen Motivation beitragen. Dabei gibt der Kennwert ΔR^2 an, wie viel Varianz in der abhängigen Variable durch die Prädiktorvariablen aufgeklärt werden kann – oder umgangssprachlich ausgedrückt: Wie gut die Prädiktorvariablen die Kriteriumsvariable vorhersagen können. ΔR^2 variiert zwischen 0 und 1. Ein Wert von 0 bedeutet, dass die Prädiktorvariablen keinen Beitrag (d. h. 0%) zur Vorhersage der Kriteriumsvariable leisten; Ein Wert von 1 bedeutet, dass die Kriteriumsvariable vollständig (d. h. 100%) durch die Prädiktorvariablen vorhergesagt werden kann. Im Weiteren ermöglicht die multiple Regressionsanalyse, den Einfluss der verschiedenen Prädiktorvariablen auf die Kriteriumsvariable sichtbar zu machen und miteinander zu vergleichen. Die Regressionskoeffizienten (siehe Kap. 13.4.3 „Merkmalsvorausagen: Die Regressionsanalyse“) sind dazu ungeeignet, da diese Werte abhängig sind vom Wertebereich der jeweiligen Prädiktorvariable. Die sogenannten *Beta-Gewichte* hingegen sind standardisierte Werte und eignen sich für den Vergleich der Bedeutung verschiedener Prädiktorvariablen in der Vorhersage der Kriteriumsvariablen.

Die Regressionsanalyse setzt sowohl für die Prädiktor- als auch für die Kriteriumsvariablen Intervallskalenniveau voraus. Allerdings ist die Regressionsanalyse robust gegenüber der Verletzung dieser Voraussetzung. Deshalb werden auch häufig ordinalskalierte oder gar nominalskalierte Variablen als Prädiktorvariablen verwendet.

► Beispiel: Multiple Regressionsanalyse

Fragestellung: Kann die schulische Motivation durch den Fernsehkonsum, die elterlichen schulischen Aspirationen und den zeitlichen Aufwand für Hausaufgaben signifikant vorhergesagt werden?

Ergebnisse: Es wurde eine Regressionsanalyse mit schulischer Motivation als abhängiger Variable und Fernsehkonsum, elterlichen Aspirationen und zeitlichem Aufwand für Hausaufgaben als unabhängige Variablen durchgeführt. Die Analyse zeigte, dass die schulische Motivation durch die drei unabhängigen Variablen höchst signifikant vorhergesagt werden konnte ($\Delta R^2 = .54, p < .001, n = 57$). Die Prädiktorvariablen erklärten somit 54% der Varianz in der schulischen Motivation. Die Beta-Gewichte der drei Prädiktorvariablen sind in unten stehender Tabelle aufgeführt. Diese zeigen, dass die Variable „elterliche schulische Aspirationen“ nach Kontrolle der anderen beiden Prädiktorvariablen am meisten zur Vorhersage der schulischen Motivation beiträgt.

Tabelle 13-17: Kennwerte der Regressionsanalyse mit schulischer Motivation als Kriteriumsvariable und Fernsehkonsum, elterlichen Aspirationen und zeitlichem Aufwand für Hausaufgaben als Prädiktorvariablen

	<i>B</i>	<i>SD</i>	β	<i>p</i>
Fernsehkonsum	-.24	.06	-.34	.001
Hausaufgaben	-.55	.19	-.29	.01
Eltern	.40	.09	.46	.001

13.5.6 Deskriptive und prüfstatistische Fragestellungen

Abschliessend wird auf die Bedeutung von deskriptiven und prüfstatistischen Fragestellungen eingegangen. Falls keine hypothesenprüfenden Verfahren eingesetzt werden, ist es wichtig, dass nur Fragestellungen formuliert werden, die mittels deskriptiver Statistik zu beantworten sind.

► Beispiel: Prüfstatistische und deskriptive ausgerichtete Fragestellungen

Prüfstatistische ausgerichtete Fragestellung:

Unterscheiden sich die Schulleistungen von Kindern mit hohem Fernsehkonsum von denen mit niedrigem Fernsehkonsum?

Kommentar: Die Frage ist ohne prüfstatistische Überprüfung nicht zu beantworten, da die Frage nach einer generalisierenden Antwort verlangt, welche über eine deskriptive Beschreibung der Stichprobe hinausgeht.

Deskriptive Fragestellung:

Welche Schulleistungen zeigen Kinder mit hohem Fernsehkonsum und solche mit niedrigem Fernsehkonsum?

Kommentar: Diese Frage lässt sich ohne prüfstatistische Kenntnisse mittels deskriptiver Statistik durch die tabellarische oder grafische Darstellung der Häufigkeitsverteilung oder durch die Berechnung von Mittelwerten und Standardabweichungen beantworten.

Ebenso ist darauf zu achten, dass Daten, die deskriptiv ausgewertet wurden, nicht prüfstatistisch interpretiert werden. Wenn beispielsweise auf deskriptiver Basis die schulische Motivation von Kindern mit hohem Fernsehkonsum und Kindern mit niedrigem Fernsehkonsum berechnet wurden, so dürfen keine verallgemeinernden Aussagen gemacht werden – auch dann nicht, wenn ein deutlicher Unterschied sichtbar ist. Dafür braucht es eindeutige Kriterien, welche prüfstatistische Tests bereitstellen können.

► Beispiel: Interpretationen deskriptiver und prüfstatistischer Befunde

(a) Die Kinder mit hohem Fernsehkonsum ($M=4.23$) zeigten um eine halbe Note schlechtere Leistungen als die Kinder mit niedrigem Fernsehkonsum ($M=4.73$).

Kommentar: Dieser Schluss ist zulässig. Diese Aussage ist rein beschreibend auf die Stichprobe bezogen und macht keine generalisierende Schlussfolgerung über die tatsächlichen Verhältnisse in der Grundgesamtheit.

(b) Kinder mit hohem Fernsehkonsum zeigen schlechtere Leistungen als Kinder mit niedrigem Fernsehkonsum.

Kommentar: Dies ist ein „unzulässiger Schluss“. Diese Aussage ist nicht mehr rein beschreibend, sondern formuliert den Unterschied als Tatsache und generalisiert damit – ohne prüfstatistische Überprüfung – auf die Grundgesamtheit.

(c) Kinder mit hohem Fernsehkonsum zeigen schlechtere Leistungen als Kinder mit niedrigem Fernsehkonsum, $t(186) = -4.23, p < .001$.

Kommentar: Dies ist ein „zulässiger Schluss“. Die Aussage in Beispiel 2 ist jetzt zulässig, da sie mittels eines prüfstatistischen Tests abgesichert wurde.

Weiterführende Literatur

— Deskriptive Statistik

Atteslander, P., Ulrich, G.-S. & Hadjar, A. (2023). *Methoden der empirischen Sozialforschung*. Erich Schmidt Verlag.

Bortz, J. & Schuster, C. (2016). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Springer.

Schäfer, T. (2016). *Methodenlehre und Statistik*. Springer.

— Prüfstatistik

Bortz, J. & Schuster, C. (2016). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Springer.

Bühl, A. (2018). *SPSS 25. Einführung in die moderne Datenanalyse ab SPSS 25* (16. Aufl.). Pearson.

Bühner, M. & Ziegler, M. (2017). *Statistik für Psychologen und Sozialwissenschaftler* (2. Aufl.). Pearson.

Dempster, M. & Hanna, D. (2019). *Statistik und Forschungsmethoden für Psychologen und Sozialwissenschaftler für Dummies*. John Wiley & Sons.

Teil IV: Analyse von Daten, 9783882561689, 2023
wurde mit IP-Adresse 134.100.177.044 aus dem Netz der SUB Hamburg am Mai 20, 2025 um 08:22:42 (UTC) heruntergeladen.
Das Weitergeben und Kopieren dieses Dokuments ist nicht zulässig.