Methoden sozialwissenschaftlicher Datenkonstruktion

G. Rohwer
U. Pötter

Oktober 2001

Fakultät für Sozialwissenschaft
Ruhr-Universität Bochum, GB 1
44780 Bochum

goetz.rohwer@ruhr-uni-bochum.de
ulrich.poetter@ruhr-uni-bochum.de

bmet.tex Oktober 2001

Vorwort

Der vorliegende Text beschäftigt sich mit Fragestellungen und Methoden sozialwissenschaftlicher Datenkonstruktion. Dabei orientieren wir uns an Varianten der Sozialforschung, die Daten als Ausgangspunkt für nachfolgende statistische Darstellungen und Analysen verwenden; wir sprechen summarisch von *statistischer Sozialforschung*. In diesem Kontext sind Daten als Werte statistischer Variablen definiert und Fragestellungen beziehen sich darauf, wie sie zustande kommen und verwendet werden.

Ein leitender Gesichtspunkt der Überlegungen besteht darin, daß Daten, insbesondere wenn sie als Werte statistischer Variablen aufgefaßt werden, nicht einfach vorhanden sind, sondern in gesellschaftlichen Verhältnissen – nicht nur, aber auch von Sozialwissenschaftlern – erzeugt werden. Infolgedessen kann man von Methoden der *Datenkonstruktion* sprechen.

Insofern in dem hier diskutierten Kontext Daten für nachfolgende statistische Darstellungen und Analysen verwendet werden sollen, gibt es einen Zusammenhang zwischen Methoden der Datenkonstruktion und statistischen Methoden. Es liegt nahe, sich diesen Zusammenhang folgendermaßen vorzustellen: Zuerst werden Daten gewonnen, dann werden sie mit Hilfe statistischer Methoden dargestellt und analysiert. Bei näherem Hinsehen zeigt sich jedoch, daß es eine ausgedehnte Grauzone gibt, in der statistische Methoden nicht zur Darstellung und Analyse vorgängig gewonnener Daten verwendet werden, sondern als Instrumente, um (dem Anspruch nach) neue Daten zu erzeugen. Ein damit verbundenes Problem besteht offenbar darin, daß unklar wird, ob und in welcher Weise überhaupt noch Aussagen mit empirischen Ansprüchen gemacht werden können.

Mit dem Buch verfolgen wir zwei Ziele. Das primäre Ziel besteht darin, eine Reihe von Methoden der Datenkonstruktion so darzustellen, daß man verstehen kann, wie sie funktionieren, und daß man sie, wenn man möchte, auch praktisch anwenden kann. Bei allen Methoden, mit denen wir uns beschäftigen, werden deshalb nicht nur ihre Grundgedanken, sondern (soweit möglich) auch ihre technischen Details besprochen.

Daneben verfolgen wir eine kritische Absicht. Wir möchten zeigen, daß es über die Frage, ob und ggf. wie *sinnvolle* Daten erzeugt werden können, nicht nur unterschiedliche, sondern auch – aus unserer Sicht – sehr fragwürdige Auffassungen gibt. Deshalb nimmt die Auseinandersetzung mit Auffassungen anderer Autoren einen vergleichsweise breiten Raum ein, der in Lehrbüchern normalerweise nicht üblich ist. Damit möchten wir jedoch nicht nur für ein größeres Problembewußtsein plädieren. Vor allem geht es uns darum, daß in der sozialwissenschaftlichen Methodenlehre überhaupt diskutiert – also auch kontrovers diskutiert – wird und dadurch sichtbar wird, daß sich hinter sogenannten Methodenfragen oft merkwürdige und fragwürdige Ansichten und Sinnvoraussetzungen verstecken.

Statistische Methoden im engeren Sinn des Wortes werden nicht behandelt; eine ausführliche Darstellung findet man in unseren „Grundzügen der sozialwissenschaftlichen Statistik“. Nur einige Grundbegriffe werden kurz erläutert, um Anschlußstellen sichtbar zu machen und Methoden der Datenkonstruktion von statistischen Methoden, mit denen Daten dargestellt und analysiert, nicht jedoch neue Daten erfunden werden sollen, abzugrenzen. Zum Verständnis der meisten Überlegungen sind auch weder besondere statistische noch mathematische Kenntnisse erforderlich. Einige Methoden der Datenkonstruktion wie etwa Varianten der Faktorenanalyse sind jedoch etwas komplizierter und erfordern zu ihrem Verständnis Grundbegriffe der Matrizenrechnung. Soweit erforderlich haben wir sie in einem Anhang dargestellt.

Für hilfreiche Diskussionen und Anmerkungen danken wir Stefan Bender, Klaus Klenovits, Karin Rinne und Bernhard Schimpl-Neimanns.

Bochum, Oktober 2001

G. Rohwer, U. Pötter

Inhalt

Kapitel 1

Einleitung

Empirische Sozialforschung kann durch die Absicht charakterisiert werden, daß Wissensbildung über gesellschaftliche Verhältnisse auf einen systematischen Prozeß der Datengewinnung gegründet werden soll, so daß über die jeweils verwendeten Daten, ihr Zustandekommen und ihre Bedeutung Rechenschaft abgelegt werden kann. Damit wird zugleich gesagt, daß Daten nicht als selbstverständliche und unproblematische Grundlage zur Verfügung stehen. Daten müssen erzeugt werden, und infolgedessen drücken sie stets auch aus, wie und für welche Zwecke sie erzeugt worden sind.[1] Das ist das Thema des vorliegenden Textes. Das Ziel besteht darin, ein Verständnis für einige der in der Literatur vorgeschlagenen und in der Forschungspraxis verwendeten Methoden sozialwissenschaftlicher Datenkonstruktion zu gewinnen.

1.1 Was heißt 'empirisch'?

1. Was ist mit dem Wort 'empirisch' gemeint, wenn von *empirischer* Sozialforschung gesprochen wird?[2] In einer Einführung in „Empirische Methoden“ (1993, S. 9) schreibt W. Laatz:

„*Empirie* heißt: Erfahrung, unmittelbar gegebene Wahrnehmung. Erfahrung – nicht nur im Sinne passiven Registrierens, sondern vielmehr als aktives Beobachten und Experimentieren – ist die zentrale Basis aller empirischen Wissenschaften.“

Hier wird 'Empirie' durch 'Erfahrung', diese dann durch 'Wahrnehmung' erläutert. Halten wir uns zunächst an den Begriff 'Erfahrung'. Bei H. J. Schneider (1993, S. 13) findet man folgende Erläuterungen:

„Ein Blick auf die Wortgeschichte zeigt, daß die Aussage, ein Mensch sei in einer Sache erfahren, ihm ursprünglich ein ungeschiedenes Ganzes von Können, Kennen, und Wissen zuspricht. Dabei ist dieses Sich-Auskennen etwas, das er, wie wir sagen, 'am eigenen Leibe' erworben hat, d.h. nicht durch Berichte, die er nur gehört oder gelesen hat: 'Erfahrung' steht dem bloßen 'Hörensagen' gegenüber. Mit dieser Leibbezogenheit gehört zum ursprünglichen Wortsinn auch die charakteristische Verflechtung von Aktivität und Passivität in der Handlung:

[1] Das wird in Einführungen in Methoden der Sozialforschung selten betont; als eine Ausnahme vgl. man jedoch Bateson 1984.

[2] Wir verwenden einfache Anführungszeichen, um auf sprachliche Ausdrücke zu verweisen; doppelte Anführungszeichen werden verwendet, um Zitate kenntlich zu machen oder um anzudeuten, daß ein Ausdruck metaphorisch und/oder unklar ist.

so wie jemand heute noch sagen kann, er habe die Fränkische Schweiz 'erwandert', konnte man früher sagen, man habe sie 'erfahren'; man setzt sich den dabei auftretenden Widerfahrnissen aus und verarbeitet sie handelnd."

Folgt man diesen Erläuterungen, kann 'Erfahrung' nicht umstandslos mit 'Wahrnehmung' gleichgesetzt werden, insbesondere nicht mit einer „unmittelbar gegebenen Wahrnehmung". Erfahrung muß vielmehr durch Tätigkeiten erworben werden. Noch eine weitere Schlußfolgerung wird nahegelegt: Erfahrungen sind zunächst an die *lokalen* Kontexte gebunden, in denen man sie „gemacht" hat.

2. Hält man sich an den so erläuterten Begriff der Erfahrung, kann man zwar daraus auch ein Verständnis von „Empirie" gewinnen; zugleich wird aber fragwürdig, ob die in der Sozialforschung gebräuchlichen Methoden der Datengewinnung „empirisch" genannt werden können. Die wohl wichtigste Methode bilden Umfragen (Interviews); daneben werden Daten aus der amtlichen Statistik verwendet, und nur ausnahmsweise werden Daten aus Beobachtungen gewonnen. Zum Beispiel ist gegenwärtig eine der wichtigsten Datenquellen das *Sozio-ökonomische Panel* (SOEP), in dem mehrere tausend Haushalte einmal jährlich nach einer Vielzahl ihrer Lebensumstände befragt werden. Es ist üblich, dies eine *empirische* Erfassung von Sachverhalten zu nennen.[3] Aber wer dann diese Daten zur weiteren Verwendung bekommt, lernt durch sie nicht die Menschen und ihre Lebensumstände kennen, die sich hinter den befragten Haushalten verbergen. Im Gegenteil, solches Kennenlernen wird durch strenge Anonymisierungsverfahren bewußt ausgeschlossen. Der Sozialforscher erhält vielmehr Daten von Menschen, die er nicht kennt, und über Situationen, die er nicht kennt und auch nie kennenlernen wird. — Die Frage, ob solche Verfahren der Datengewinnung „empirisch" *genannt* werden können, ist natürlich müßig. Aber eine wichtige Schlußfolgerung kann gleichwohl gezogen werden: Die so gewonnenen Daten können nicht dem Zweck dienen, *Erfahrungen* mit gesellschaftlichen Verhältnissen zu machen; und wenn sie einen Zweck haben, muß dieser auf andere Weise bestimmt werden.

3. Um das damit angesprochene Problem zu verdeutlichen, sei daran erinnert, daß in der historischen Entwicklung der Sozialwissenschaften der Gedanke eine wichtige Rolle gespielt hat, daß wissenschaftliche Theoriebildung nicht vom Alltagswissen der gesellschaftlich lebenden Menschen ausgehen sollte (man vgl. hierzu Matthes 1993). Besonders nachdrücklich wurde diese Auffassung von Emile Durkheim vertreten:

„Es ist notwendig, alle Vorbegriffe systematisch auszuschalten. [...] Der Soziologe muß also, sowohl bei der Bestimmung des Gegenstandes seiner Forschung als auch im Verlaufe seiner Beweisführung, sich des Gebrauchs von Begriffen entschieden entschlagen, die außerhalb der Wissenschaft und für die Bedürfnisse,

[3]Ausführliche Informationen über das SOEP findet man unter der folgenden Internet-Adresse: `www.diw.de/soep/`.

die nichts Wissenschaftliches an sich haben, gebildet werden. Er muß sich von den falschen Selbstverständlichkeiten befreien, die den Verstand des naiven Menschen beherrschen, und ein für allemal das Joch dieser empirischen Kategorien abschütteln, die durch lange Gewöhnung schließlich zu einer tyrannischen Macht werden." (Durkheim 1895/1984, S. 128f)

Wenn man diese Bemerkung positiv lesen will,[4] kann man sie so verstehen, daß ein Verständnis gesellschaftlicher Verhältnisse nicht ohne weiteres aus Erfahrungen mit sozialen Situationen gewonnen werden kann, sondern explizit konstruiert werden muß. Aber es stellt sich natürlich nicht nur die Frage, wie eine solche Konstruktion durchgeführt und begründet werden kann; es ist zunächst auch gar nicht einsehbar, welchen potentiell wissenswerten Sinn die Ergebnisse einer Sozialforschung haben könnten, die sich methodisch von Erfahrungen abkoppelt, um sich stattdessen auf konstruierte Daten zu stützen.

4. Um zu einem Verständnis zu gelangen, muß zunächst das Argument präzisiert werden, daß man sich gesellschaftliche Verhältnisse nicht durch Erfahrungen zugänglich machen kann. Damit soll natürlich nicht geleugnet werden, daß sich der größte Teil unseres Lebens *in sozialen Situationen* abspielt und daß man diese sozialen Situationen kennenlernt, indem man in ihnen aufwächst und an ihnen teilnimmt. Somit ist auch klar, daß man *diese* Situationen durch Erfahrung kennenlernt. Aber es sollte ein Unterschied gemacht werden: zwischen einer Bezugnahme auf soziale Situationen, die man durch Erfahrung kennengelernt hat bzw. kennenlernen kann, und einer Bezugnahme auf gesellschaftliche Verhältnisse. Wie immer man zu *Vorstellungen* über gesellschaftliche Verhältnisse gelangen kann, sie müssen sich jedenfalls fast ausschließlich auf Menschen und Situationen beziehen, die man nicht durch Erfahrung kennenlernen kann. Und wenn man sich dennoch Vorstellungen machen möchte, kann man sich dafür nicht ohne weiteres auf Erfahrungen – also Erfahrungen, die man in den jeweils eigenen lokalen Kontexten gemacht hat – berufen. Und wenn man nicht blindlings die Vorstellungen übernehmen möchte, die allerorten angeboten werden, muß man ihre Bildung als einen Konstruktionsprozeß reflektieren und hinterfragen.

[4]Eine andere Lesart ergibt sich aus folgender Bemerkung Durkheims: „Diese Idee, daß das gesellschaftliche Leben nicht mit Hilfe der Begriffe derjenigen erklärt werden sollte, die an ihm teilnehmen, sondern aus tiefer liegenden Ursachen, die dem Bewußtsein entgehen, halte ich für außerordentlich fruchtbar, und ich glaube auch, daß diese Ursachen hauptsächlich in der Art und Weise zu suchen sind, in welcher die assoziierten Individuen sich gesellschaftlich gruppieren. Nur so kann allem Anschein nach die Historie eine Wissenschaft werden und die Soziologie selbst existieren." (Zitiert nach dem Abdruck in Winch 1966, S. 35.) Mit dieser Idee zielt Durkheim offenbar auf eine Theoriebildung, die gesellschaftliche Verhältnisse nicht durch ihre Akteure, sondern die umgekehrt zunächst die gesellschaftlichen Verhältnisse und dann auch ihre Akteure durch „tieferliegende Ursachen" erklärbar machen soll, die sich einem bewußten Verständnis der Akteure entziehen.

5. Wichtig erscheint uns die Einsicht, daß der Konstruktionsprozeß bereits mit der Erzeugung von Daten beginnt. Die empirische Realität ist uns nicht als ein Feld empirischer Daten gegeben, sondern – im Bereich unserer Erfahrung – als ein Handlungsfeld, in dem man sich mehr oder weniger gut orientieren und bewegen kann. Ein erstes und wohl auch das wichtigste Problem besteht schon darin, wie man von hier aus zu einem Begriff gesellschaftlicher Verhältnisse gelangen kann, der nicht nur in metaphorischen Verallgemeinerungen der jeweils eigenen Lebenswelt besteht. Die Verfolgung dieser Frage würde uns allerdings in Fragen sozialwissenschaftlicher Theoriebildung verwickeln, denen im Rahmen dieses Textes nicht näher nachgegangen werden kann.[5] Zumindest als Problemstellung sollte man sich jedoch klarmachen, daß es keine selbstverständliche Antwort auf die Frage gibt, wie sich Sozialforschung empirisch auf gesellschaftliche Verhältnisse beziehen sollte.

6. Tatsächlich ist diese Frage auch innerhalb der Sozialforschung umstritten. Die meisten Lehrbücher über Methoden der Sozialforschung verbreiten die Auffassung, daß der primäre Realitätsbezug durch statistische Variablen hergestellt werden kann.[6] Gesellschaftliche Verhältnisse werden als eine Menge von Objekten, insbesondere Menschen, konzipiert, deren Merkmale und Beziehungen durch statistische Variablen erfaßt werden können. Man kann dementsprechend von einer *statistischen Sozialforschung* sprechen. Sie bildet auch den Hintergrund für den vorliegenden Text. Wir beschäftigen uns mit Methoden der Datenkonstruktion, die in der statistischen Sozialforschung verwendet werden. Daraus ergibt sich auch eine Abgrenzung von statistischen Methoden der Datenanalyse, die nicht der Konstruktion von Daten dienen, sondern darauf abzielen, gegebene Daten darzustellen und zu analysieren. Wir unterscheiden also zwischen *statistischen Methoden zur Datenanalyse* und *Methoden, die der Konstruktion von Daten dienen* sollen. Nur letztere bilden den Gegenstand dieses Textes.

1.2 Variablen und Daten

1. Statistische Sozialforschung verschafft sich ihren primären Realitätsbezug durch „Variablen“. Die zentrale Rolle, die diesem Begriff gegeben wird, kann man sich anhand der folgenden programmatischen Äußerung von E. K. Scheuch und D. Rüschemeyer (1956, S. 275) verdeutlichen:

„Abgesehen von einer dauernden Bindung an die empirische Kontrolle stellt sich die von der modernen Wissenschaftslehre ausgehende Soziologie zudem die Aufgabe, über die Beschreibung hinaus allgemeine Gesetze (das heißt Hypothesen über konstante Beziehungen zwischen Variablen) aufzustellen.“

[5] Als Einführung eignet sich eine Textsammlung von F. Fürstenberg 1995, in der versucht wird, ausgehend von einem Begriff „sozialer Handlungsfelder“ zu theoretischen Vorstellungen über gesellschaftliche Verhältnisse zu gelangen.

[6] Gegenwärtig verbreitete Einführungsbücher sind u.a.: Atteslander 1975, Bortz 1984, Friedrichs 1990, Laatz 1993, Kromrey 1994, Diekmann 1995, Bortz und Döring 1995, Schnell, Hill und Esser 1999, Schumann 1999.

Etwas ausführlicher heißt es bei F. N. Kerlinger (1964, S. 20):

„A *hypothesis* is a conjectual statement of the relation between two or more variables. Hypotheses are always in declarative sentence form, and they relate, either generally or specifically, variables to variables. There are two criteria for ‘good’ hypotheses and hypothesis statements. They are the same as two of those for problems and problem statements. One, hypotheses are statements about the relations between variables. Two, hypotheses carry clear implications for testing the stated relations.“

Also schon die Fragestellungen, für die sich die statistische Sozialforschung interessiert, sollen sich auf Variablen beziehen. Dementsprechend schreiben zum Beispiel J. Bortz und N. Döring (1995, S. 10):

„*Wissenschaftliche* oder auch *inhaltliche* Hypothesen, die aus gut begründeten Vorannahmen oder aus Theorien abgeleitet sind, stellen verbale Behauptungen über kausale oder nicht-kausale Beziehungen zwischen Variablen dar (z.B. ‘Das Leben in der Stadt ist psychisch belastender als das Leben auf dem Land.’).“

2. Aber was sind die „Variablen“, für deren Zusammenhänge man sich interessiert? Ein Autor gibt z.B. folgende Erläuterung:

„The concepts or ideas that we call variables are the phenomena of particular interest in our social science disciplines. They are called variables because they vary in amount or attribute for each individual (or group, or society, or state, or culture – whatever we happen to be observing).“ (Sirkin 1995, S. 8)

Aber beobachten kann man nur jeweils spezifische Situationen oder Objekte, ihre Merkmale und Verhaltensweisen, keine Variablen. Was aber sind Variablen? Sich gleichzeitig sowohl auf „concepts or ideas“ als auch auf „phenomena“ zu beziehen, ist offenbar konfus.[7] Bortz und Döring (1995, S. 5) geben folgende Erläuterung:

„Um Merkmalsunterschiede bei einer Gruppe von Objekten genau beschreiben zu können, wurde der Begriff *Variable* eingeführt. Eine Variable ist ein Symbol, das durch jedes Element einer spezifizierten Menge von Merkmalsausprägungen ersetzt werden kann.“

Aber beide Sätze passen nicht gut zusammen. Den ersten Satz kann man so verstehen, daß Variablen dazu dienen sollen, Merkmale von beobachtbaren Objekten zu repräsentieren. Im zweiten Satz werden jedoch Variablen mit Merkmalsräumen identifiziert, d.h. mit Mengen möglicher Merkmalswerte, die eine Variable, je nachdem auf welches Objekt sie sich bezieht,

[7] Auch die Vorstellung, daß Begriffe in irgendeinem „amount“ oder „attribute“ variieren können, erscheint merkwürdig. Sie wird aber nicht nur von Sirkin verwendet. Hier noch eine Formulierung von Dixon et al. 1987, S. 52: „What is a variable? A variable is a concept that varies in amount or kind. A variable is a concept of which it is possible to have more or less, or different kinds.“

annehmen kann.[8] Merkmalsräume und Variablen sollten allerdings unterschieden werden; und man kann eine Variable natürlich nicht durch jedes Element ihres Merkmalsraums ersetzen, sonst könnte man sich ja Daten beliebig selbst erzeugen.

3. Halten wir uns deshalb an den ersten Satz in dem zuletzt angeführten Zitat von Bortz und Döring. Ausgangspunkt ist die Vorstellung, daß man sich auf eine Menge von Objekten (Dinge, Menschen, Situationen, ...) beziehen kann, denen sich sinnvoll Merkmale zurechnen lassen. Wir werden im folgenden eine solche Menge von Objekten symbolisch durch

$$\Omega := \{\omega_1, \omega_2, \ldots, \omega_n\}$$

repräsentieren.[9] Hier verweist das Symbol Ω auf die Menge, und die Symbole $\omega_1, \omega_2, \ldots$ verweisen auf ihre Mitglieder, die einzelnen Objekte.

4. Um was für Objekte handelt es sich? In der historischen Entwicklung der Sozialstatistik hat man sich in erster Linie auf Menschen bezogen, und dies ist auch heute noch der Fall, wie z.B. ein Blick in ein statistisches Jahrbuch zeigen kann. Die Bezugnahme auf Menschen kommt auch darin zum Ausdruck, daß Statistiker oft das Wort 'Population' verwenden, um sich auf eine Gesamtheit von Objekten zu beziehen. Die Begriffsbildung ist jedoch so allgemein, daß man sich auf beliebige Objekte beziehen kann. In der Literatur zur Methodenlehre wird gelegentlich nicht einmal verlangt, daß es sich um empirisch wahrnehmbare Objekte handeln sollte. Z.B. schreibt J. Friedrichs (1990, S. 87f):

„Um einen Objektbereich, ein 'universe of discourse', zu strukturieren, wird man versuchen, Ordnung in die beobachteten Erscheinungen oder abstrakten Elemente zu bringen. Dabei kann es sich um Personen, um Städte, Normen, Formen sexuellen Verhaltens etc. handeln. Einer Menge von Elementen wird eine Menge von Merkmalen zugeordnet; man nimmt eine Klassifikation vor."

Diese im Prinzip grenzenlose Ausweitung eines Begriffs von Objekten, auf die sich statistische Variablen beziehen können, ist jedoch problematisch. Denn es besteht dann leicht die Gefahr, eine wesentliche Unterscheidung

[8] Diese Gleichsetzung findet auch bei folgender Definition statt, die sich bei Schnell, Hill und Esser 1999, S. 124, findet: „Variablen können als zusammenfassender Begriff für verschiedene Ausprägungen einer Eigenschaft (den 'Variablenwerten') angesehen werden; z.B. die Variable 'Ampelfarbe' kann die Variablenwerte 'rot, gelb, grün' annehmen."

[9] Wir unterscheiden in diesem Text die Zeichen '$=$' und '$:=$'. Ein Gleichheitszeichen mit vorangestelltem Doppelpunkt wird verwendet, um anzudeuten, daß eine definitorische Gleichsetzung vorgenommen wird, d.h. der Ausdruck auf der linken Seite wird durch den Ausdruck auf der rechten Seite definiert. Dagegen setzt ein einfaches Gleichheitszeichen voraus, daß beide Seiten schon definiert sind. Hier sei auch darauf hingewiesen, daß die Bedeutung von Symbolen kontextabhängig ist. Es muß stets auf den Kontext geachtet werden, in dem ein Symbol eingeführt worden ist, um seine Bedeutung zu verstehen.

zwischen Objekten und ihnen individuell oder relational zurechenbaren Eigenschaften und Verhaltensweisen zu übersehen. Betrachtet man statistische Begriffsbildungen immanent, kann man zwar von beliebigen Objekten ausgehen und ist nicht gezwungen, ihren ontologischen Status genauer zu klären. Sobald man sich jedoch mit den Begriffen der Statistik auf einen realen Gegenstandsbereich beziehen möchte, sollte man zunächst explizit die Objekte fixieren, die man *in diesem Gegenstandsbereich* zu statistisch charakterisierbaren Gesamtheiten zusammenfassen möchte. Dann erst hat man einen Ausgangspunkt, um von Eigenschaften – nämlich Eigenschaften dieser Objekte oder von Situationen, in denen sie sich befinden – auf nachvollziehbare Weise sprechen zu können.

5. Um diesen Punkt zu betonen, soll noch ein Beispiel zeigen, daß es erforderlich ist, genau darauf zu achten, wie sinnvolle Grundgesamtheiten konzipiert werden können.

„Als Grundgesamtheit wird somit die Menge aller statistischen Einheiten bezeichnet, über die man Aussagen gewinnen will. Diese muß klar umgrenzt sein und hängt natürlich von der interessierenden Fragestellung ab. Beim Münchner Mietspiegel setzt sich die Grundgesamtheit aus allen nach dem Mietgesetz relevanten Wohnungen Münchens zusammen. Im Rahmen politischer Umfragen etwa dazu, welche Partei man wählen würde, ist die Grundgesamtheit die zu dem jeweiligen Zeitpunkt wahlberechtigte Bevölkerung eines Landes. In diesen Beispielen ist die Population jeweils *endlich*. Es ist aber zu beachten, daß eine Grundgesamtheit nicht immer eine endliche Menge an statistischen Einheiten sein muß, deren Elemente konkret vorliegen. Sie kann auch *unendlich* groß oder sogar *hypothetisch* sein. Im Zusammenhang mit der Kreditwürdigkeitsüberprüfung bilden alle potentiellen Kunden die Grundgesamtheit, die somit hypothetisch ist. Als ein Beispiel für eine unendliche Grundgesamtheit kann die Menge aller möglichen Wartezeiten während der Betriebszeit einer U-Bahn angesehen werden." (Fahrmeir, Künstler, Pigeot, Tutz 1999, S. 13f)

Der Vorstellung, daß man sich auf unendliche Gesamtheiten beziehen kann, kann vielleicht in der Mathematik ein Sinn gegeben werden, aber nicht, wenn man sich auf reale Objekte oder Situationen beziehen möchte. Zum Beispiel die Behauptung, daß eine U-Bahn eine unendliche Menge von Wartezeiten erzeugen kann, klingt nicht nur merkwürdig, sondern ist sicherlich falsch. Noch problematischer ist der Gedanke, man könne Bezugnahmen auf reale Objekte und auf nur potentiell existierende Objekte umstandslos in einer dem Anspruch nach wohldefinierten Menge zusammenfassen. Im ersten Fall kann man sich auf Objekte in gewisser Weise als auf Gegebenheiten beziehen; man kann versuchen herauszufinden, wie sie beschaffen sind und sich verhalten. Bei nur potentiell existierenden Objekten ist dagegen bereits fragwürdig, wie man sinnvoll über sie sprechen kann.

6. Gelegentlich wird diskutiert, in welcher Weise davon gesprochen werden kann, daß Gesamtheiten überhaupt existieren. Z.B. schreibt ein Autor, der sich mit den Grundlagen der Statistik beschäftigt hat:

„Alle Gesamtheiten, mit denen man in der Statistik zu tun hat, sind Produkte des menschlichen Denkens. Daß die mathematischen Gesamtheiten, insbesondere die Gesamtheiten der Mengenlehre, der Maßtheorie und der Wahrscheinlichkeitstheorie, gedachte Gesamtheiten sind, versteht sich von selbst. Daß aber die empirischen Gesamtheiten an sich nicht existieren, bedarf einer Erklärung. Wohl sind die Einheiten, die Einzeldinge, aus denen die empirischen Gesamtheiten bestehen, wahrnehmbare Objekte; also existieren sie unabhängig vom menschlichen Geist. In diesem Sinne sind die empirischen Gesamtheiten 'real'. Die Einzeldinge werden aber vom Menschen durch Denken in Gesamtheiten zusammengefaßt.“ (Sagoroff 1973, S. 21)

In welcher Weise man von der Existenz von Einzeldingen und Gesamtheiten sprechen kann, sei dahingestellt. Wichtig ist jedoch die Feststellung, daß statistische Gesamtheiten durch Sozialwissenschaftler gebildet werden, um sich ein Bild von einer im einzelnen nicht überschaubaren Menge von Objekten zu machen. Ihre Zusammenfassung zu einer Gesamtheit ist in jedem Fall das Produkt einer gedanklichen Operation, ganz unabhängig davon, ob und in welcher Weise die Objekte existieren, die zu einer Gesamtheit zusammengefaßt werden.

7. Eine wichtigere Frage ist, nach welchen Gesichtspunkten Objekte oder Situationen zu statistischen Gesamtheiten zusammengefaßt werden. In der Sozialforschung wird meistens von Unterscheidungen ausgegangen, die in der gesellschaftlichen Praxis üblich sind. Man unterscheidet zum Beispiel verschiedene Arten von Objekten, etwa Menschen, Tiere, sonstige Dinge, etwa Güter, die in bestimmten Kontexten erzeugt werden. Oder man bezieht sich auf soziale Gebilde, deren Abgrenzung umgangssprachlich oder/und für verwaltungstechnische Zwecke üblich ist, wie etwa Familien, Haushalte, Unternehmen. Auch regionale Abgrenzungen knüpfen zumeist an verwaltungstechnische Unterscheidungen an, indem man sich zum Beispiel auf die Studenten einer Universität oder die Bevölkerung einer Stadt oder Region bezieht.

8. Statistische Variablen sollen dazu dienen, die Mitglieder einer Menge von Objekten durch Merkmalswerte zu charakterisieren. Dafür benötigt man eine Vorstellung möglicher Merkmalswerte. Wir verwenden dafür den Begriff eines *Merkmalsraums*. Damit gemeint ist eine Menge von sich wechselseitig ausschließenden Merkmalen, durch die die Objekte, auf die man sich beziehen möchte, charakterisiert werden können. Zur Bezeichnung von statistischen Variablen verwenden wir Großbuchstaben, meistens $X, Y, Z, \ldots$, ggf. mit Indizes versehen. Die korrespondierenden Merkmalsräume werden durch Tilden kenntlich gemacht; z. B. bezeichnet $\tilde{\mathcal{X}}$ den Merkmalsraum der statistischen Variablen X. Einzelne *Merkmalswerte*, die wir auch Attribute, Merkmale oder Merkmalsausprägungen nennen, werden durch entsprechende Kleinbuchstaben angesprochen. Hat man z.B. die Merkmalswerte $\tilde{x}_1, \tilde{x}_2, \ldots$ festgelegt und schließen sich diese Merkmalswerte wechselseitig aus, kann daraus ein Merkmalsraum

$$\tilde{\mathcal{X}} := \{\, \tilde{x}_1, \tilde{x}_2, \ldots \}$$

gebildet werden. Weiterhin definieren wir: ein Merkmalsraum $\tilde{\mathcal{X}}$ heißt *vollständig bzgl. der Gesamtheit* Ω, wenn jedem Objekt in Ω genau ein Merkmalswert aus $\tilde{\mathcal{X}}$ zugeordnet werden kann. Eine *statistische Variable* kann dann als eine Funktion bzw. Abbildung

$$X : \Omega \longrightarrow \tilde{\mathcal{X}} \tag{1.2.1}$$

definiert werden, die jedem Objekt in Ω genau eines der Merkmale aus dem Merkmalsraum $\tilde{\mathcal{X}}$ zuordnet. Wenn also ω ein Objekt aus der Gesamtheit Ω ist, dann ist $X(\omega)$ der Merkmalswert, der dem Objekt ω durch die statistische Variable X zugeordnet wird. Diese bei den Objekten realisierten Werte einer statistischen Variablen werden oft mit korrespondierenden Kleinbuchstaben bezeichnet; man verwendet dann x_1 als Abkürzung für $X(\omega_1)$, x_2 als Abkürzung für $X(\omega_2)$ usw.

9. Es sollte beachtet werden, daß es einen begrifflichen Unterschied zwischen Merkmalswerten $(\tilde{x}_1, \tilde{x}_2, \ldots)$ und bei den Objekten einer gewissen Gesamtheit realisierten Merkmalswerten $(x_1, x_2, \ldots)$ gibt. Die Merkmalswerte sind durch den Merkmalsraum einer statistischen Variablen definitorisch vorgegeben; die Definition stellt zugleich sicher, daß sich alle Merkmalswerte voneinander unterscheiden. Eine Kenntnis dieser Merkmalswerte ist zwar erforderlich, um eine statistische Variable definieren zu können, vermittelt aber noch keinerlei Information über die Beschaffenheit der Realität. Dagegen zeigen die realisierten Merkmalswerte einer Variablen, welche der möglichen Merkmalswerte bei den Mitgliedern einer gewissen Gesamtheit feststellbar sind. Im Unterschied zu den Merkmalen in einem Merkmalsraum brauchen sich die realisierten Merkmalswerte nicht zu unterscheiden. Mehrere oder sogar alle Objekte können den gleichen Merkmalswert aufweisen; und andererseits ist es auch möglich, daß einige Merkmalswerte bei den Objekten einer Gesamtheit überhaupt nicht vorkommen.

10. Man kann sich gleichzeitig auf mehrere Merkmalsräume beziehen, um die Objekte einer Gesamtheit zu charakterisieren. Wenn es sich um Menschen handelt, kann man sie z.B. durch ihr Alter, ihr Geschlecht und ihren Erwerbsstatus charakterisieren; im Prinzip durch beliebig viele Merkmale. Dem dient der Begriff einer *mehrdimensionalen* statistischen Variablen. Die Anzahl ihrer Dimensionen entspricht der Anzahl der Merkmalsräume, mit denen die Definition beginnt.[10] Hat man zum Beispiel m

[10]Daß dieser kombinierte Merkmalsraum mehrere *Dimensionen* hat, ist allerdings zunächst nur eine rein formale Sprechweise, mit der zum Ausdruck gebracht werden soll, daß er sich aus mehreren einzelnen Merkmalsräumen zusammensetzt. Dementspre-

Merkmalsräume $\tilde{\mathcal{X}}_1, \ldots, \tilde{\mathcal{X}}_m$ festgelegt, kann daraus zunächst ein kombinierter m-dimensionaler Merkmalsraum $\tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$ gebildet werden.[11] Dann kann eine m-dimensionale statistische Variable

$$(X_1, \ldots, X_m) : \Omega \longrightarrow \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m \tag{1.2.2}$$

definiert werden, die jedem Objekt $\omega \in \Omega$ gleichzeitig Merkmalswerte $X_1(\omega), \ldots, X_m(\omega)$ zuordnet. Bei sozialstatistischen Erhebungen werden fast immer mehrere Merkmalsräume gleichzeitig betrachtet, so daß eine symbolische Repräsentation mit mehrdimensionalen Variablen beginnt.

11. Statt von 'statistischen Variablen' wird häufig abkürzend nur von 'Variablen' gesprochen. Meistens macht der Kontext klar, was gemeint ist. Wichtig ist jedoch, den Unterschied zwischen einfachen mathematischen Variablen, wie sie z.B. in der elementaren Algebra verwendet werden, und statistischen Variablen im Auge zu behalten. Statistische Variablen sind keine selbständigen Objekte, sondern Abbildungen einer Objektmenge in einen Merkmalsraum. Von Objekten kann man sagen, daß es sie in der Realität gibt; Variablen haben dagegen keine selbständige Existenz, sondern sind Teil einer Sprache, mit der über Objekte geredet werden soll.[12]

12. Denken wir noch einmal an die eingangs angeführte Erläuterung des Variablenbegriffs durch Bortz und Döring. Die Unklarheit kann jetzt genauer so formuliert werden: Sie besteht in einer Verwechslung von statistischen Variablen mit ihren Merkmalsräumen. Eine solche Verwechslung ist in der Methodenliteratur sehr verbreitet. Wir kritisieren sie nicht nur aus Gründen begrifflicher Klarheit, sondern vor allem, weil der Glaube, daß es sich bei Variablen um Merkmalsräume handelt, leicht der falschen Vorstellung Vorschub leistet, daß Variablen „selbständig existierende Gebilde" sind, deren Zusammenhänge unsere empirische Realität beherrschen. Aber um es noch einmal zu betonen: In der empirischen Realität gibt es keine Variablen, sondern zunächst Objekte der verschiedensten Art, Situationen und Ereignisse. Variablen sind demgegenüber „begriffliche Konstrukte", die dazu dienen sollen, über Dinge, Situationen und Ereignisse in einer durchaus spezifischen (statistischen) Weise reden zu können.

chend heißt es z.B. bei Halmos 1950, S. 152: „It is customary to refer to a product space $X = \times_{i=1}^{n} X_i$ as *n-dimensional*. This terminology is not meant to define dimension, nor to assert that n-dimensionality is an intrinsic structural property of a space; it serves merely to remind us of the way in which X was built from the components X_i."

[11] Das Symbol '$\times$' bezeichnet das kartesische Produkt von Mengen. Weitere Erläuterungen zu den Notationen findet man bei Rohwer und Pötter 2001, S. 21ff.

[12] Diese wichtige Unterscheidung wird nicht immer klar genug betont. Zum Beispiel heißt es bei Mueller et al. (1970, S. 9): „Any object or event which can vary in successive observations either in quantity or in quality may be termed a 'variable'." Solche Verwechslung von Objekten, Eigenschaften und Variablen kann leicht Konfusionen hervorrufen, indem sie den Eindruck entstehen läßt, man könne unmittelbar über Variablen reden, ohne ihren Realitätsbezug zu explizieren. Wir werden deshalb stets explizit unterscheiden: einerseits Objekte und andererseits Variablen, mit deren Hilfe eine Gesamtheit von Objekten charakterisiert werden kann.

13. Dies muß auch deshalb betont werden, weil statistische Variablen den grundlegenden begrifflichen Rahmen bilden, in dem man sich in der statistischen Sozialforschung auf gesellschaftliche Verhältnisse und ihre Akteure bezieht. Daraus gewinnt auch das Wort 'Daten' seine in dieser Variante der Sozialforschung übliche Bedeutung: *Daten sind Werte statistischer Variablen.* Dementsprechend läßt sich verstehen, was in der statistischen Sozialforschung mit *Datengewinnung* gemeint ist. Man beginnt mit der Definition einer statistischen Variablen in der Form (1.2.1) oder (1.2.2) und versucht dann, für die Mitglieder von Ω Werte der Variablen zu ermitteln.[13] Am Schluß dieses Prozesses der Datengewinnung hat man dann eine Datenmatrix, die folgendermaßen aussieht:

ω	X_1	$\cdots$	X_m
ω_1	x_{11}	$\cdots$	x_{1m}
$\vdots$	$\vdots$		$\vdots$
ω_n	x_{n1}	$\cdots$	x_{nm}

Jede Zeile entspricht einem Element von Ω und enthält die für dieses Element erfaßten Merkmalswerte.

14. Unsere empirische Realität besteht allerdings nicht aus Daten, sondern Daten als Werte statistischer Variablen sind Ergebnis eines Konstruktionsprozesses. Eine erste Unterscheidung ergibt sich daraus, von wem und zu welchen Zwecken Daten erzeugt werden. Dies geschieht zunächst – auch historisch gesehen – nicht durch Sozialwissenschaftler, sondern durch soziale Akteure, die daran interessiert sind, über Mengen von Objekten, insbesondere auch Menschen, Buch zu führen. Zunächst waren dies hauptsächlich staatliche und kirchliche Einrichtungen, dann auch zunehmend Unternehmen und Verbände. Daten dieser Art werden oft *prozeß-produzierte Daten* genannt. P. J. Müller (1977, S. 1) hat den Begriff so erläutert:

> „Unter prozeß-produzierten Daten werden all diejenigen Daten verstanden, die als Aufzeichnungen öffentlicher und privater Organisationen im Rahmen ihrer Tätigkeit und nicht nur zum Zwecke wissenschaftlicher bzw. statistischer Auswertung gesammelt werden bzw. wurden."

Exemplarisch kann man an die von den Sozialversicherungsträgern erzeugte Beschäftigtenstatistik denken, die seit einiger Zeit teilweise auch für sozialwissenschaftliche Forschungszwecke zur Verfügung steht.[14]

[13] Praktisch bezieht man sich meistens nur auf eine Teilmenge der Gesamtheit Ω und spricht dann von einer *Stichprobe*. Hier setzt die Stichprobentheorie an, in der überlegt wird, ob und ggf. wie mit Hilfe von Stichprobendaten Hypothesen über Merkmalsverteilungen in der zugrundeliegenden Gesamtheit reflektierbar gemacht werden können; man vgl. dazu Rohwer und Pötter 2001, Teil IV.

[14] Nähere Informationen über diese Datenquelle finden sich bei Bender et al. 1996.

15. Prozeß-produzierte Daten verdanken sich unmittelbar jeweils bestimmten Verwendungszwecken. Einen Gegensatz bildet die *amtliche Statistik*, die Daten produziert, ohne dabei von jeweils bestimmten Verwendungszwecken auszugehen. In Deutschland ist ihr zentraler Träger das Statistische Bundesamt, dessen Aufgaben in einem *Gesetz über die Statistik für Bundeszwecke* festgelegt sind. Über die von der amtlichen Statistik produzierten Daten und ihre thematischen Schwerpunkte kann man sich anhand der Statistischen Jahrbücher informieren.[15] Zentrale Themen sind vor allem die Bevölkerungs- und Wirtschaftsstatistik (man vgl. z.B. Rinne 1996). Insofern gibt es viele Berührungspunkte mit den Datenkonstruktionsmethoden der statistischen Sozialforschung. Dies betrifft auch teilweise die Methoden der Datengewinnung (Krug, Nourney und Schmidt 1999).

16. Schließlich werden Daten auch durch Sozialwissenschaftler erzeugt, sei es im Rahmen von Forschungsprojekten oder in der „angewandten" Sozialforschung, wie sie von Instituten, Verbänden, staatlichen Einrichtungen und Unternehmen betrieben wird.[16] Dafür gibt es eine Vielzahl unterschiedlicher Methoden. Neuere Methodenlehrbücher schlagen oft vor, drei Arten der Datengewinnung zu unterscheiden: (a) Umfragen bzw. Interviews, (b) Beobachtung und (c) Text- bzw. Inhaltsanalysen.[17] In den folgenden Kapiteln werden wir uns jedoch nur mit der Erzeugung von Daten durch Umfragen bzw. Interviews genauer beschäftigen und untersuchen, wie solche Daten zur Konstruktion neuer Daten verwendet werden.[18]

[15] Zahlreiche Informationen erhält man auch auf der Homepage des Statistischen Bundesamtes: `www.statistik-bund.de`.

[16] Ob und ggf. wie sich „angewandte" von „theoretisch orientierter" Sozialforschung unterscheiden läßt, ist eine offene Frage. Habermehl 1992, S. 9, sagt: „Von *angewandter Sozialforschung* wollen wir sprechen, wenn ein erfahrungswissenschaftliches Untersuchungsvorhaben unmittelbar auf die Lösung praktischer Probleme hin angelegt ist: Politikberatung, Marketing, Gemeindearbeit der Kirchen und so weiter."

[17] Schnell, Hill und Esser 1999, S. 297; Laatz 1993, Kap. 3–5; Kromrey 1994, Kap. 7.

[18] Die meisten in der empirischen Sozialforschung verwendeten Daten werden auf diese Weise gewonnen. Diekmann 1995, S. 372f, berichtet von einer Untersuchung, in der Artikel deutschsprachiger soziologischer Zeitschriften im Zeitraum 1989 – 1993 daraufhin untersucht wurden, auf welche Arten von Daten sie sich beziehen. Etwa 70 % verwenden Umfragedaten, 13 % prozeß-produzierte Daten, 11 % beziehen sich auf Texte (Text- bzw. Inhaltsanalyse) und nur 6 % gewinnen Daten aus Beobachtungen. Man vgl. auch die Angaben bei Scheuch 1999, S. 14.

Kapitel 2

Interviews und Tests

Die in der Sozialforschung verwendeten Daten stammen überwiegend aus Interviews bzw. Umfragen. In diesem Kapitel soll diese Form der Datengewinnung etwas genauer diskutiert und eine Unterscheidung zwischen Interviews und Tests besprochen werden.

2.1 Zum Verständnis von Interviews

1. In einem weit gefaßten Sinn sind Interviews Gespräche, in denen ein Gesprächsführer (der Interviewer) versucht, mit der Hilfe seiner Gesprächspartner (der Interviewten) Antworten auf vorab oder während des Gesprächs konzipierte Fragen zu gewinnen. Die Definition ist offensichtlich sehr allgemein, und dementsprechend vielfältig sind die sozialen Kontexte, in denen Interviews durchgeführt werden, ihre Formen und die mit ihrer Hilfe verfolgten Zwecke. Zum Beispiel diagnostische Interviews in der medizinischen Praxis, Prüfungsgespräche bei Examen, Einstellungsgespräche zur Auswahl von Bewerbern, Interviews mit berühmten Persönlichkeiten. In diesem Text werden wir uns jedoch mit der Vielfalt unterschiedlicher Zwecke und Formen von Interviews nicht näher beschäftigen, sondern Interviews nur unter dem Gesichtspunkt thematisieren, daß mit ihrer Hilfe auch Daten für die statistische Sozialforschung gewonnen werden können. Wir werden uns also auch nicht mit den verschiedenen Interviewformen beschäftigen, die im gesamten Spektrum der Sozialforschung mehr oder weniger häufig verwendet werden.

2. Im Kontext statistischer Sozialforschung sollen Interviews dem Zweck dienen, statistische Daten, d.h. Werte für statistische Variablen, zu gewinnen. Sie beziehen sich infolgedessen stets auf eine Gesamtheit von Befragungspersonen und setzen definierte Merkmalsräume voraus. Zur Datenerhebung werden deshalb meistens *standardisierte* Interviews durchgeführt, d.h. es werden vorab fixierte Fragestellungen verwendet, die dann allen Befragungspersonen in der gleichen Weise gestellt werden. Die äußere Form kann durchaus unterschiedlich sein. Zum Beispiel kann sich ein Interviewer persönlich an die zu befragenden Personen wenden, oder man kann ihnen einen Fragebogen zuschicken und hoffen, daß sie ihn ausfüllen und zurückschicken.

3. Unabhängig davon, in welcher Form ein Interview durchgeführt wird, besteht es darin, daß sich ein Interviewer – persönlich oder mittelbar – in einem kommunikativen Medium an einen oder mehrere andere Menschen wendet, um Antworten auf Fragen zu bekommen. Es wird zum Beispiel gefragt: Wie alt sind Sie? Wo sind Sie geboren? Waren Sie schon einmal verheiratet? Sind Sie gegenwärtig erwerbstätig? Usw. Es gibt drei elementare Voraussetzungen:

a) Die Fragen müssen für die Befragten verständlich sein, d.h. sie müssen in einer für den Interviewer und die Interviewten gleichermaßen verfügbaren Sprache formuliert sein.

b) Es muß sich um sinnvolle Fragen handeln, die deutlich machen, daß und wie eine zutreffende Antwort gegeben werden kann.

c) Es muß unterstellt werden können, daß die Interviewten in der Lage sind, die Fragen kompetent zu beantworten.

Bei diesen Anforderungen an die Fragestellungen eines Interviews handelt es sich um Sinnvoraussetzungen. Es gibt keine formal handhabbaren Kriterien, mit denen sie geprüft werden können. Das heißt auch: Es ist durchaus möglich, Menschen in Interviews mit unsinnigen Fragen zu konfrontieren, auf die infolgedessen keine sinnvollen Antworten gegeben werden können. Dennoch wird man auch dann oft genug Antworten erhalten, sei es, weil der Interviewer um eine Antwort bittet und seine Interviewpartner höflich genug sind, der Bitte zu entsprechen, oder sei es, daß sie die Sinnlosigkeit einer Frage nicht durchschauen und ihr einen beliebig erfundenen Sinn unterstellen. Infolgedessen können auch mit sinnlosen Fragestellungen Daten produziert werden. Es ist jedoch evident, daß man dann nur *Pseudo-Daten* gewinnt, denen keine Sachverhalte in der sozialen Realität entsprechen, sondern an und für sich sinnlose Daten, die bestenfalls in dieser Eigenschaft zu einem Thema gemacht werden können.

4. Für genuine Interviewsituationen ist es konstitutiv, daß man sich kommunikativ auf *gemeinsame* Fragestellungen beziehen kann und daß Fragen in der Annahme gestellt werden, daß der Interviewte sie nicht nur kompetent, sondern auch besser beantworten kann als der Fragesteller. Dies unterscheidet genuine Interviews zum Beispiel von Prüfungssituationen, in denen Fragen gestellt werden, um zu ermitteln, über welches Wissen und welche Fähigkeiten die zu prüfende Person verfügt. — Diese Unterscheidung kann verallgemeinert werden.

a) Einerseits kann man anderen Menschen Fragen stellen, weil man sich für deren Antworten interessiert. Das Interesse richtet sich dann auf diejenigen Sachverhalte, auf die sich die Fragen beziehen.

b) Andererseits kann man andere Menschen zum Gegenstand von Prüfungen machen. Um auf diese Intention zu verweisen, sprechen wir von *Tests*. Derjenige, der einen Test durchführt, möchte Einsichten in Eigenschaften und Fähigkeiten desjenigen gewinnen, der dem Test unterzogen wird.

Ihrer äußeren Form nach können Interviews beiden Zwecken dienen; zum Beispiel gibt es Prüfungs- und Einstellungsgespräche, bei denen es sich ihrer Funktion nach um Tests handelt. Die Unterscheidung der beiden Zwecksetzungen ist jedoch wichtig, und sie sollte deshalb auch terminologisch fixiert werden. Wir sprechen von *genuinen Interviews*, wenn man sich kommunikativ an eine Befragungsperson wendet, um mit ihrer Hilfe Antworten auf Fragen zu gewinnen, und dabei ihre Kompetenz zur Beantwortung der Fragen voraussetzen kann. Dagegen sprechen wir von *Tests*, wenn es sich um Verfahren (nicht unbedingt kommunikativer Art) handelt, mit deren Hilfe man Einsichten in Eigenschaften und Fähigkeiten von Personen (oder allgemein: beliebiger Objekte) gewinnen möchte, die dem Verfahren ausgesetzt werden.

5. Es ist bemerkenswert, daß in der Methodenliteratur zur Sozialforschung zwar in größtmöglicher Ausführlichkeit und Umständlichkeit unterschiedliche Varianten von Befragungstechniken diskutiert werden, die elementare Unterscheidung zwischen Interviews und Tests jedoch fast nie thematisiert wird. Gelegentlich wird die Unterscheidung sogar systematisch verschleiert; z.B. in der folgenden Definition von E. K. Scheuch (1973, S. 70f):

> „Unter Interview als Forschungsinstrument sei hier verstanden ein planmäßiges Vorgehen mit wissenschaftlicher Zielsetzung, bei dem die Versuchsperson durch eine Reihe gezielter Fragen oder mitgeteilter Stimuli zu verbalen Informationen veranlaßt werden soll.“

Folgt man dieser Definition, soll ein Interview offenbar zeigen, wie „Versuchspersonen“ – gelegentlich auch „Probanden“ genannt – mit verbalen Äußerungen auf „Stimuli“ reagieren.[1] Interviews werden also als Tests konzipiert, etwa so wie man mit geeigneten „Stimuli“ versuchen könnte herauszufinden, zu welchen „sprachlichen Reaktionen“ ein Papagei fähig ist.

6. Um die Unterscheidung zwischen Interviews und Tests besser zu verstehen, sollte darauf geachtet werden, daß man den Informationsgehalt „sprachlicher Äußerungen“ auf zwei unterschiedliche Weisen wahrnehmen

[1] Ähnliche Formulierungen sind tatsächlich weit verbreitet; hier einige Beispiele. McGaw und Watson 1976, S. 343: „... the survey is a reactive research technique, that is, the respondent is being presented with a stimulus and his reactions are observed.“ Költringer 1993, S. 7: „Das Interview als Instrument der Umfrageforschung kann unter Hinweis auf seine Zielsetzung definiert werden: es ist eine planmäßige Datensammlungsprozedur, bei der generalisierbare Stichproben von Versuchspersonen durch eine Reihe gezielter Fragen *zu vergleichbaren Reaktionen* veranlaßt werden, mit dem Zweck, eine wissenschaftliche Fragestellung zu beantworten.“ Denz 1989, S. 13: „Die Grundannahme, die hinter jeder Befragung, besonders hinter einer standardisierten, steht, ist eine experimentelle Logik, die an den Naturwissenschaften orientiert ist: Wenn jedem Befragten dieselbe Frage gestellt wird, kann eine unterschiedliche Antwort als unterschiedliche Einstellung, Meinung, Wissen usw. des Befragten interpretiert werden. Der einzige Faktor, der unterschiedliche Reaktionen auf ein und dieselbe Frage hervorruft, ist der Unterschied zwischen den Befragten.“

bzw. interpretieren kann. Angenommen, man stellt einer Person eine Frage. Dann kann man einerseits die Antwort in ihrer inhaltlichen Bedeutung wahrzunehmen versuchen. Andererseits kann man in der Antwort die „sprachliche Reaktion“ einer „Versuchsperson“ sehen, so wie dies von Scheuchs Definition eines Interviews nahegelegt wird. Die Antwort liefert dann eine Information über die Person, der die Frage gestellt worden ist. Man fragt zum Beispiel: Wie heißt der gegenwärtige Bundeskanzler,[2] will aber gar nicht wissen, wie er heißt, sondern will wissen, ob der Befragte es weiß und sein Wissen in geeigneten Worten mitteilen kann. In diesem Fall kann man vielleicht von einem „Test“ sprechen, aber sicherlich nicht von einem Interview im Sinne eines kommunikativen Verfahrens zur Informationsgewinnung.[3]

2.2 Tests und Dispositionen

1. Folgende Definition kann als Leitfaden dienen: Tests sind Verfahren, um Einsichten in Beschaffenheiten und Funktionsweisen von Objekten zu gewinnen. Das Hauptanwendungsgebiet von Tests im Sinne dieser allgemeinen Definition sind nicht-menschliche Materialien und Organismen und vor

[2]Diese Frage wird von Borg und Staufenbiel 1993, S. 28, als Beispiel eines „Intelligenz-Items“ angeführt.

[3]Auch bei den sog. „indirekten Befragungsmethoden“ handelt es sich nicht um genuine Interviews, denn bei diesen Methoden werden die Befragungspersonen absichtlich im unklaren darüber gelassen, was man über sie ermitteln möchte. Zur Illustration sei auf eine „Befragungsstrategie“ verwiesen, die von J. Wolf (1988, S. 44) für die Marktforschung empfohlen wurde: „Die Befragungsstrategie ist in den meisten Fällen darauf gerichtet, die befragten Personen solange wie möglich über den Befragungszweck im unklaren zu lassen. Die Befragten tendieren nämlich sonst unbewußt dazu, Fragen, deren Sinn sie zu durchschauen glauben, so zu beantworten, wie sie annehmen, daß es von ihnen erwartet wird, weil dies ihrem Selbstwertgefühl am besten entspricht.“ Es sei erwähnt, daß sich ähnliche Empfehlungen oder Überlegungen gelegentlich auch bei Autoren finden, die sich – vorgeblich – mit Fragen sozialwissenschaftlicher Datengewinnung beschäftigen. Zum Beispiel bei Holm 1974, S. 322: „Bei einer ‘geschlossenen’ Fragebatterie erkennt der Befragte bald die Zieldimension. Er durchschaut die ihm präsentierten Fragen. Das kann zwei Folgen haben: a) Der Befragte gibt bei den nachfolgenden Fragen die sozial erwünschte Antwort, oder b) er versucht, in seinen Antworten konsistent zu bleiben. [...] Das Durchschauen mit seinen beiden möglichen Folgen wird normalerweise als nachteiliger Effekt gewertet.“ Diese Sichtweise ist insbesondere bei psychologischen Tests verbreitet. Herkner 1991, S. 20, schreibt: „Um solchen störenden Überlegungen der Vpn [„Versuchspersonen“] vorzubeugen oder wenigstens deren Wirkungen abzuschwächen, wird den Vpn vielfach zu Beginn des Experiments eine plausible, aber falsche Erläuterung des Versuchs gegeben.“ Witte 1989, S. 37, schreibt: „Betrachtet man die Studien aus dem Bereich der Sozialpsychologie hinsichtlich der Verwendung von *Täuschungsmanövern*, so variiert der Prozentsatz von 80 % bei Studien zur Konformität über 44 % bei Studien zur Einstellungsänderung bis zu 33 % bei Studien zu Stress und ‘arousal’. Bei dieser großen Zahl von Experimenten mit Täuschung ist zu erwarten, daß kaum noch ein Experiment mit naiven ungetäuschten College-Versuchspersonen durchgeführt werden kann. Diese Probanden aber suchen hinter den vorgegebenen Instruktionen die Täuschungsmanöver.“

allem von Menschen hergestellte Artefakte. Hier soll es jedoch nur darum gehen, daß Tests in zunehmendem Maße auch verwendet werden, um Menschen zu prüfen, zu diskriminieren und zu selektieren. Eine Vorreiterrolle hat die sog. Testpsychologie gespielt;[4] aber die Vorstellung, daß es nützlich sein könnte, Menschen durch Tests besser einschätzbar und kalkulierbar zu machen, ist seither auch in Teilbereiche der Sozialforschung eingedrungen. Zwar werden in der Sozialforschung in der Regel keine Tests konstruiert und verwendet;[5] eine entsprechende Rhetorik – sowohl in der Verwendung des Wortes ‘Test’ als auch in einem korrespondierenden Reden vom „Messen“ (insbesondere von Meinungen und Einstellungen) – hat sich jedoch auch in Teilen der Sozialforschung verbreitet.

2. Obwohl in der Methodenliteratur zur Sozialforschung oft von Tests die Rede ist, werden nur selten Erläuterungen gegeben. Eine der Ausnahmen ist ein Methodenbuch von Bortz und Döring, das sich an der Testpsychologie orientiert und folgende Definition von Lienert übernimmt:[6]

> „Ein Test ist ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung.“ (Bortz und Döring 1995, S. 175)

Dem entspricht, daß diese Autoren das Ziel der Datenerhebung darin sehen, „die untersuchten Objekte bzgl. bestimmter Merkmale zu quantifizieren“ (ebda., S. 128).[7]

3. Am häufigsten wird in der Lehrbuchliteratur von Tests im Kontext einer Diskussion von Meßverfahren gesprochen. Da meistens auf eine explizite Definition des Wortes ‘Test’ verzichtet wird, gewinnt man den Eindruck, daß beide Worte – ‘Test’ und ‘Meßverfahren’ – weitgehend synonym ver-

[4]Man vgl. zur Einführung z.B. Lienert und Raatz 1998, Heidenreich 1984.

[5]Allerdings in Teilbereichen der „angewandten“ Sozialforschung, insbesondere in der Marktforschung; man vgl. z.B. Weis und Steinmetz 1998.

[6]Zur ursprünglichen Definition vgl. man Lienert und Raatz 1998, S. 1f.

[7]Eine etwas allgemeinere Definition des Testbegriffs hat Kerlinger in einem Buch über „Foundations of Behavioral Research“ (1964, S. 480) vorgeschlagen: „A *test* is a systematic procedure in which the individual tested is presented with a set of constructed stimuli to which he responds, the responses enabling the tester to assign the testee a numeral or set of numerals from which inferences can be made about the testee’s possession of whatever the test is supposed to measure.“ Zur Erläuterung wird ausgeführt: „Hundreds, perhaps thousands, of objective inferential tests and scales have been constructed and used. Watson estimates that at least 500 personality inventories have appeared commercially. [Das war 1959.] There are probably as many ability, aptitude, intelligence, and achievement tests. [...] Most tests and scales can be divided into the following classes: intelligence and aptitude tests, achievement tests, personality measures, attitude and value scales, and miscellaneous objective measures.“ (S. 481) Inzwischen gibt es auch kritische Stimmen, man vgl. z.B. Sacks 1999.

wendet werden.[8] Dem entspricht auch die Definition von Bortz und Döring: ein Test soll „möglichst quantitative Aussagen über den relativen Grad der individuellen Merkmalsausprägung“ liefern, und dieselbe Vorstellung verbindet man meistens mit einem Meßverfahren. Allerdings sollte an dieser Stelle an die Unterscheidung zwischen Interviews und Tests erinnert werden, denn der Sprachgebrauch geht vollends durcheinander, wenn auch Interviews als Meßverfahren bezeichnet werden; zum Beispiel:

„Die Qualität eines Meßverfahrens ist von mehreren Komponenten abhängig. Zunächst von der Qualität des Meßinstruments: Bei der Befragung ist das der Fragebogen oder ein Interviewerleitfaden. Dann aber auch von dessen Anwendung ...“ (Laatz 1993, S. 103)

In ähnlicher Weise haben auch andere Autoren Interviews als Meßverfahren bezeichnet.[9] Man kann sich jedoch überlegen, daß es unsinnig ist, ein Interview – oder einen Fragebogen für ein Interview – als ein Meßinstrument aufzufassen. Denn im Unterschied zu einem Test soll ein Interview dazu dienen, von einer Befragungsperson Aussagen zu gewinnen, für deren Inhalt man sich interessiert. Dabei ist es für die Charakterisierung des Interviews als Verfahren gleichgültig, ob es sich um Aussagen qualitativer oder quantitativer Art handelt. Aber auch wenn es sich um Aussagen handelt, die sich auf quantifizierbare Sachverhalte beziehen, wie z.B. bei Angaben zum Alter oder Einkommen, besteht das Interview nur darin, solche Aussagen zu fixieren. Eine Quantifizierung kommt hier weder durch das Interview noch durch die Aussage der Befragungsperson zustande; vielmehr wird – in diesen Fällen – für das Interview vorausgesetzt, daß es einen quantitativ bestimmten Sachverhalt gibt, auf den sich das Interview mit der Intention, diesen Sachverhalt festzustellen, beziehen kann.

4. Es gibt einen weiteren bemerkenswerten Unterschied. Bei Interviews geht es darum, Aussagen zu gewinnen, die sich auf empirisch zugängliche Sachverhalte beziehen oder Präferenzen zum Ausdruck bringen oder in der

[8] Fischer 1974, S. 20: „Testen bedeutet doch nichts anderes, als bestimmte Eigenschaften von Individuen zu messen.“ Bei Mehrens und Lehmann 1975, S. 5, heißt es: „The terms *test*, *measurement*, and *evaluation* are sometimes used interchangeably, but some users make distinctions among them. The term *test* is usually considered the narrowest of the three items. *Test*, in the narrowest sense, connotes the presentation of a standard set of questions to be answered. As a result of a person's answers to such a series of questions, we obtain a measure (that is, a numerical value) of a characteristic of that person. *Measurement* often connotes a broader concept: We can measure characteristics in ways other than by giving tests. Using observations, rating scales, or any other device that allows us to obtain information in a quantitative form is a measurement.“ Schumann 1999, S. 17, schreibt: „Meßinstrumente werden im Rahmen von Tests, die Auskunft über bestimmte Merkmalsausprägungen liefern sollen, eingesetzt.“ Im weiteren (S. 28) bezieht sich der Autor mit dem Wort ‘Meßinstrument’ auf die Fragen, die bei einem Interview oder Test gestellt werden.

[9] Zum Beispiel Kerlinger 1964, S. 468; Atteslander und Kopp 1984; Fowler 1988; Költringer 1993; Diekmann 1995, S. 371ff; Bortz und Döring 1995, S. 226; Schumann 1999, S. 52; Jacob und Eirmbter 2000, S. 36.

Form normativer Urteile Bewertungen ausdrücken. Ein Interview kann als ein Diskurs verstanden werden, in dem solche Aussagen entwickelt und sprachlich fixiert werden.[10] Tests sind demgegenüber nicht nur keine diskursiven Verfahren, sondern bezwecken auch einen anderen Typ von Aussagen, nämlich Aussagen über *dispositionale* Eigenschaften von Objekten. Dabei handelt es sich um Eigenschaften, die nur durch einen Kontext, in dem sich die dispositionale Eigenschaft in einer empirisch zugänglichen Verhaltensweise des Objekts zeigen kann, explizierbar sind. Infolgedessen wird allerdings die Unterscheidung zwischen dispositionalen Eigenschaften und Eigenschaften des Kontextes, in dem sich dispositionale Eigenschaften zeigen können, problematisch. Kurz gesagt: wie sich Dinge verhalten, hängt sowohl von ihren dispositionalen Eigenschaften als auch vom jeweiligen Kontext ihres Verhaltens ab. Daraus ergibt sich eine wichtige Implikation für das Verständnis von Tests, denn sie definieren den Kontext, in dem sich dispositionale Eigenschaften zeigen sollen. Was durch einen Test erfaßt werden kann, hängt infolgedessen *wesentlich* vom Test selbst ab.

5. Dieser Sachverhalt wird besonders deutlich, wenn psychologische Tests durch ein Reiz-Reaktions-Schema expliziert werden. Folgende Darstellung von R. B. Cattell (1973, S. 29) kann zur Illustration dienen.

„Persönlichkeit kann als das Verhalten eines Menschen in einer bestimmten Situation definiert werden. Das läßt sich folgendermaßen formulieren: $R = f(S, P)$, wobei R die Art und das Ausmaß der Verhaltensreaktion einer Person bedeutet, d.h. was sie sagt, denkt oder tut. Diese ist eine bestimmte Funktion der Reizsituation S, in welcher sich die Person befindet, und der Struktur P ihrer Persönlichkeit.“

Einige Autoren konzipieren dementsprechend die zentrale Fragestellung (sozial-)psychologischer Untersuchungen. Zum Beispiel heißt es in einem Lehrbuch der Sozialpsychologie von Herkner (1991, S. 22):

„Die zentrale psychologische Frage lautet: Welche Vorgänge in der (beobachtbaren) Umgebung eines Lebewesens verursachen welches Verhalten bzw. welche Verhaltensänderungen? Oder anders formuliert: Welche Vorgänge auf der Reizseite verursachen welche Vorgänge auf der Verhaltensseite?“

Einer solchen Betrachtungsweise verdankt sich auch die Vorstellung, daß man Interviews als Tests konzipieren könne; insbesondere eine Variante, die von Atteslander und Kopp (1984, S. 159) so erläutert wird:

„Vertreter, die im Interview eine Vereinheitlichung der Reize anstreben, arbeiten mit einem einfachen psychologischen Verhaltensmodell: dem Reiz-Reaktions-Schema. Das Interview wird dabei als Verhaltenssituation aufgefaßt, in der der Interviewer den Befragten mit einer Reihe von Fragen (die die zu untersuchenden Variablen enthalten) konfrontiert. Der Befragte sieht sich bestimmten

[10] Diese Betrachtungsweise macht auch deutlich, daß standardisierte Interviews, die mit vorab fixierten Fragen und Antwortmöglichkeiten operieren, einen Grenzfall darstellen, bei dem das Element der Diskursivität auf ein Minimum reduziert wird.

Reizen ausgesetzt, auf die er in Form seiner Antworten reagiert. Die Antworten können dann als die individuellen Reaktionen auf denselben Reiz angesehen werden, so daß sie zu Häufigkeiten zusammengefaßt werden und der entsprechenden Variablen zugeordnet werden können.“ (Atteslander und Koop 1984, S. 159)

6. Es sollte auch überlegt werden, welchen spezifischen Zwecken Tests dienen sollen. Cattell (1973, S. 287) gibt folgende Hinweise:

„Persönlichkeits-, Begabungs- und Motivationstests können in verschiedenen Bereichen nutzbringend angewendet werden. Man kann sie für die glücklichere Verwendung menschlicher Arbeitskräfte in den Berufen, für informative Werbung mit Rücksicht auf eine nützliche Warenverteilung, bei der Auslese politischer Führer, bei der Verhütung von Verbrechen und allgemein bei der Beeinflussung antisozialen Verhaltens, bei der Diagnose von Geisteskrankheit und der Beurteilung des therapeutischen Fortschritts wie bei der Beratung in Ehe- und Familienproblemen und der Bewertung sozialer Strömungen verwenden.“

Die Ausführungen von Lienert und Raatz in ihrem Lehrbuch über „Testaufbau und Testanalyse“ (1998, S. 6f) sind demgegenüber abstrakter und allgemeiner. Ebenfalls sprechen sie jedoch davon, daß ein Test als ein „Meßinstrument für psychologische Merkmale“ aufgefaßt werden kann; und sie formulieren dann folgende Anforderungen: „Ein guter Test soll als *Hauptgütekriterien* folgende drei Forderungen erfüllen: er soll objektiv, er soll reliabel, er soll valide sein.“ In der Methodenliteratur werden diese „Hauptgütekriterien“ für psychologische Tests meistens bei der Behandlung „sozialwissenschaftlicher Meßverfahren“ behandelt. Wir besprechen sie in Kapitel 8.

2.3 Varianten der Sozialforschung

1. Zwischen genuinen Interviews und Tests zu unterscheiden, ist nicht nur aus Gründen begrifflicher Genauigkeit wichtig. In gewisser Weise entspricht der Unterscheidung, daß man Sozialforschung auf zwei unterschiedliche Weisen konzipieren kann. Einerseits kann man sich von der Vorstellung leiten lassen, daß Sozialforschung dazu dienen soll, Einsichten in gesellschaftliche Verhältnisse zu gewinnen: wie sie beschaffen sind, wie sie sich historisch entwickelt haben und wie sie funktionieren. Gegenstand der Sozialforschung sind dann nicht die individuellen Gesellschaftsmitglieder, sondern gesellschaftliche Verhältnisse als Lebens- und Handlungsbedingungen von Akteuren. Insbesondere in der Methodenliteratur wird jedoch oft eine andere Auffassung vorgeschlagen: daß es Aufgabe der Sozialforschung sei, die individuellen Gesellschaftsmitglieder zu erforschen und ihre Verhaltensweisen voraussagbar zu machen. Man könnte von einer verhaltenstheoretisch konzipierten (im Unterschied zu einer soziologisch konzipierten) Sozialforschung sprechen;[11] im angelsächsischen Sprachraum gibt es

[11]Hinweise auf diese Unterscheidung findet man z.B. bei J. Coleman 1990, S. 1f.

für diese Variante der Sozialforschung den Ausdruck ‘behavioral science’.[12]

2. Je nachdem, welche der beiden Konzeptionen man verfolgt, stellen sich auch unterschiedliche Anforderungen an die Datengewinnung. Im ersten Fall benötigt man Daten über gesellschaftliche Verhältnisse als Lebens- und Handlungsbedingungen ihrer Akteure. Interviews dienen dann dem Zweck, Einsichten in die Lebens- und Handlungsbedingungen der jeweils befragten Personen zu gewinnen, um mit Hilfe der so gewonnenen Informationen Bilder (Modelle) gesellschaftlicher Verhältnisse zu konstruieren, anhand derer sich reflektieren läßt, wie die Verhältnisse beschaffen sind und funktionieren. Interviews sind dann nicht „Meßinstrumente“, um Eigenschaften der jeweils befragten Personen zu ermitteln, sondern Verfahren zur kommunikativen Informationsgewinnung über Sachverhalte, die durch Interviewer und Interviewte in einer gemeinsamen Perspektive objektiviert werden können.

3. Folgt man dagegen dem Ansatz der verhaltenstheoretisch konzipierten Sozialforschung, benötigt man in erster Linie Daten über die individuellen Akteure: ihre Eigenschaften, Meinungen, Einstellungen, Fähigkeiten. Dann wird es auch plausibel, von Tests zu sprechen: Man möchte herausfinden, wie die Akteure funktionieren, welche „Reaktionen“ sie zeigen, wenn sie unterschiedlichen „Stimuli“ ausgesetzt werden. Es ist verständlich, daß sich diese Auffassung von Interviews vor allem in Bereichen der angewandten Sozialforschung findet, wo es tatsächlich darum geht herauszufinden, wie sich Menschen durch geeignete „Stimuli“, z.B. kommerzielle, politische oder kirchliche Werbung, beeinflussen lassen.

[12]Dazu schrieb S. M. Lipset 1958, S. 477: „Die Übernahme des Begriffs ‘Verhaltenswissenschaften’ (behavioral sciences) im Gegensatz zu dem alten Begriff der Sozialwissenschaften während der letzten fünf Jahre hat die alte Diskussion über das gegenseitige Verhältnis der einzelnen Sozialwissenschaften neu belebt. Wie ich den Begriff der ‘Verhaltenswissenschaften’ verstehe, umfaßt er den größten Teil der Soziologie, Psychologie und Sozialanthropologie, ein Stück Politische Wissenschaft, etwas Nationalökonomie, und so gut wie nichts von der Geschichtswissenschaft. Im wesentlichen sollen ‘die Verhaltenswissenschaften’ jene Diziplinen umfassen, die sich mit dem wissenschaftlichen Studium des menschlichen Verhaltens beschäftigen.“

Kapitel 3

Reflexion von Erkenntnisinteressen

In der Methodenliteratur steht meisten die Diskussion unterschiedlicher Formen und Instrumente der Datengewinnung im Mittelpunkt. Von primärer Bedeutung ist demgegenüber, *was* durch Interviews oder Tests ermittelt werden soll. Dazu sollen in diesem Kapitel einige Überlegungen zur Diskussion gestellt werden.

3.1 Was möchte man wissen?

1. Die Frage, *was* mit Hilfe von Interviews oder Tests erfaßt werden soll, hängt natürlich im einzelnen von der jeweils verfolgten Fragestellung ab. Generalisierend kann man sich an der in Abschnitt 2.3 diskutierten Unterscheidung orientieren: ob man an Einsichten in gesellschaftliche Verhältnisse interessiert ist oder ob Eigenschaften und Reaktionsweisen der sozialen Akteure erfaßt werden sollen. Die Methodenliteratur orientiert sich überwiegend an der zweiten Alternative und definiert dementsprechend die Zwecke sozialwissenschaftlicher Interviews; z.B. Bortz und Döring in ihrem Kapitel über „Quantitative Methoden der Datenerhebung“:

„Der Datenerhebungsbegriff wird in diesem Kapitel sehr weit gefaßt. Die hier behandelte Datenerhebung dient generell dazu, die untersuchten Objekte bzgl. bestimmer Merkmale zu quantifizieren.“ (Bortz und Döring 1995, S. 128)

Umstandslos werden dann sowohl Tests als auch Befragungen (Interviews) dieser Zielsetzung subsummiert.

2. Gewissermaßen parallel zu der Vorstellung, daß die Hauptaufgabe darin bestehe, Eigenschaften – insbesondere auch dispositionale Eigenschaften – von Individuen zu ermitteln, betonen die meisten Lehrbücher die Erfassung von „Meinungen“ und „Einstellungen“; zum Beispiel:

„Questionnaires for quantitative research in the social sciences are usually designed with the intention of being operational definitions of concepts, instruments that reflect strength of attitudes, perceptions, views and opinions. This involves trying to measure and quantify how intensely people *feel* about issues, as opposed to what they know or can do.“ (Black 1999, S. 215)

Ähnlich heißt es bei Schnell, Hill und Esser (1999, S. 303f):

„Eine erste Entscheidung, die den Entwurf von Fragen betrifft, bezieht sich auf die Problematik, welche Informationen durch eine Frage von Befragten gewonnen werden sollen. Dabei wird im allgemeinen unterschieden zwischen: Fragen nach Einstellungen oder Meinungen von Befragten, Fragen nach Überzeugungen der

Befragten, Fragen nach Verhalten der Befragten, Fragen nach Eigenschaften der Befragten.

‘Einstellungsfragen’ oder ‘Meinungsfragen’ beziehen sich auf den Aspekt der Wünschbarkeit oder der negativen bzw. positiven Beurteilung, den Befragte mit bestimmten Statements verbinden. [...]

‘Überzeugungsfragen’ fragen danach, was Befragte für wahr oder falsch halten. Dabei können sich solche Fragen sowohl auf das ‘Wissen’ der Befragten in bezug auf bestimmte Sachverhalte beziehen, wie auch auf Problembereiche, zu denen niemand eine ‘richtige’ Antwort weiß. In jedem Fall zielen Überzeugungsfragen auf die Wahrnehmung und Einschätzung vergangener, gegenwärtiger oder zukünftiger Realität. Charakteristische Wendungen in Überzeugungsfragen sind ‘falsch/richtig’, ‘wahr/falsch’ usw. [...].

‘Verhaltensfragen’ in standardisierten Interviews beziehen sich prinzipiell auf Handlungen und Verhalten von Befragten bzw. eigentlich auf *Überzeugungen* der Befragten bezüglich ihres eigenen Verhaltens. [...]

Der Fragentyp, der sich auf Eigenschaften von Personen bezieht (‘Fragen nach Befragteneigenschaften’), umfaßt im allgemeinen Fragen nach personalen und demographischen Eigenschaften von Befragten wie Alter, Geschlecht, Ausbildung, Beruf, Einkommen, Familienstand, ethnische Zugehörigkeit, Parteizugehörigkeit, Konfession usw. Demographische Variablen werden in den meisten Fragebögen ‘routinemäßig’ erhoben; im allgemeinen und im einfachsten Fall, um (statistische) Zusammenhänge zwischen demographischen Eigenschaften von Personen und ihren Einstellungen, Überzeugungen und Verhaltensweisen zu ermitteln.“

3. Orientiert man sich an diesen und ähnlichen Ausführungen der Lehrbuchliteratur, scheint es so, als ginge es hauptsächlich darum, Meinungen und Einstellungen zu ermitteln. Das steht jedoch in einem eigentümlichen Mißverhältnis zu den Erkenntnisinteressen der Sozialforschung.[1] Die Intentionen der Lehrbuchliteratur werden zwar verständlich, wenn man daran denkt, daß es durchaus Überschneidungen der Sozialforschung mit gesellschaftlichen Praktiken gibt, die sich – mehr oder weniger verwissenschaftlicht – als Meinungs-, Markt- und Wahlforschung etabliert haben. Das Mißverhältnis wird jedoch deutlich, wenn man sich daran orientiert, daß Sozialforschung in erster Linie zur Aufklärung gesellschaftlicher Verhältnisse beitragen sollte. Denn dann geht es zunächst nicht um Meinungen und Einstellungen, sondern um die Beschaffenheit der gesellschaftlichen Verhältnisse als Lebens- und Handlungsbedingungen ihrer Akteure.

4. Tatsächlich spielen Meinungs- und Einstellungsfragen auch in den meisten für die Sozialforschung relevanten Datenquellen keine oder nur eine marginale Rolle. Zum Beispiel geht es bei allen wichtigen prozeß-produzierten Daten nicht um Meinungen und Einstellungen, sondern um soziale

[1]Da Erkenntnisinteressen in der Methodenliteratur selten erörtert werden, kommt auch dieses Mißverhältnis selten zur Sprache. Diskutiert wird es aber gelegentlich in kritischen Kommentaren; man vgl. z.B. Adorno 1957.

Sachverhalte, etwa in der Sozialhilfestatistik oder der Beschäftigtenstatistik der Sozialversicherungsträger. Ganz entsprechend verhält es sich mit der amtlichen Statistik. Ihr geht es zum Beispiel nicht darum, ob Menschen glauben, daß Preise steigen, sondern sie bemüht sich, durch Erhebung von Preisdaten methodisch abgesicherte Informationen über die Preisentwicklung unterschiedlicher Arten von Gütern zu ermitteln. Schließlich kann auch auf einige wichtige sozialwissenschaftliche Datensätze verwiesen werden. Zum Beispiel ist gegenwärtig das Sozio-ökonomische Panel (SOEP) eine der wichtigsten Datenquellen für die Sozialforschung. Es handelt sich um eine 1984 begonnene Panelerhebung, in der mehrere tausend Haushalte einmal jährlich nach einer Vielzahl ihrer Lebensumstände befragt werden. Dabei geht es fast ausschließlich um Lebens- und Handlungsbedingungen der in diesen Haushalten lebenden Menschen; nur wenige Fragen beziehen sich auf Meinungen und Einstellungen.[2]

5. Nicht nur die für die Lehrbuchliteratur typische Fokussierung auf Meinungen und Einstellungen sollte kritisiert werden, auch über die üblichen Einteilungen der durch Interviews zu erfassenden Sachverhalte sollte nachgedacht werden. Anhand der von Schnell, Hill und Esser vorgeschlagenen Gliederung, die bereits zitiert worden ist, sollen einige Unzulänglichkeiten deutlich gemacht werden.

a) Zunächst sprechen die Autoren von „Meinungen“ und „Einstellungen“, bei denen es darum geht, ob sich Befragte irgendetwas wünschen oder irgendetwas eher gut oder eher schlecht finden. Es gibt jedoch, wie schon erwähnt, keinerlei Hinweis darauf, ob Fragen dieser Art für die Erkenntnisinteressen der Sozialforschung überhaupt relevant sind und ob und ggf. wie in diesem Fall sinnvolle von sinnlosen Fragen unterschieden werden können.

b) Als nächstes sprechen die Autoren von „Überzeugungsfragen“, bei denen es darum geht, „was Befragte für wahr oder falsch halten“. Sie geben folgendes Beispiel an (S. 305):

Ist die folgende Aussage richtig oder falsch?

> „Im letzten Jahr wurden in der Bundesrepublik Deutschland
> mehr ausländische als deutsche Kinder geboren.“
>
> Richtig ( )
> Falsch ( )

[2]Andererseits gibt es natürlich auch zahlreiche sozialwissenschaftliche Datenerhebungen, die sich im wesentlichen Interessen der Meinungs- und Einstellungsforschung verdanken. Exemplarisch sei auf den ALLBUS hingewiesen, der seit 1980 in zweijährigem Abstand erstellt wird. Über die Geschichte des ALLBUS informieren Braun und Mohler 1991; aktuelle Informationen findet man im Internet: `www.gesis.org/ZA/` (Zentralarchiv für empirische Sozialforschung, Köln) und `www.gesis.org/ZUMA/` (Zentrum für Umfragen, Methoden und Analysen, Mannheim).

Aber warum sollte man in einem Interview eine Frage dieser Art stellen? Denn wäre man wirklich an der Frage interessiert, würde man sicherlich nicht Menschen fragen, die sehr wahrscheinlich keine zutreffende Antwort geben können; sondern man würde sich in den einschlägigen amtlichen Statistiken informieren. Es handelt sich auch offensichtlich nicht um eine Testfrage, mit der man herausbekommen möchte, ob Befragungspersonen eine zutreffende Antwort geben können, denn die Antwortmöglichkeit „Ich weiß es nicht“ wird gar nicht zugelassen. Die Frage ist tatsächlich irreführend: Sie soll Menschen dazu bringen, beliebige Vermutungen zu äußern. Sie verletzt also eine wesentliche Bedingung für sinnvolle Interviewfragen: daß man unterstellen kann, daß Befragte eine zutreffende Antwort geben können bzw., wenn dies nicht sicher ist, daß ihnen die Möglichkeit gegeben wird, zum Ausdruck zu bringen, daß sie die Frage nicht beantworten können. Diese Kritik richtet sich indessen nicht nur gegen die hier als Beispiel diskutierte Frage, sondern betrifft generell die „Überzeugungsfragen“, wie sie von Schnell, Hill und Esser konzipiert werden. Denn die Autoren sagen ausdrücklich, daß sich solche Fragen auch auf Problembereiche beziehen können, „zu denen niemand eine ‘richtige’ Antwort weiß.“ Es ist evident, daß als Resultat nur Pseudo-Daten entstehen können, wenn man Menschen Fragen stellt, die aufgrund sinnloser Formulierungen oder mangelhaften Wissens nicht vernünftig beantwortet werden können.

c) Als dritte Kategorie nennen die Autoren „Verhaltensfragen“, die sich auf das „eigene Verhalten“ der Befragten beziehen. Hierzu werden zwei Beispiele gegeben (S. 306):

A. Werden Sie sich noch im Laufe dieses Jahres einen Neuwagen mit einem Katalysatorsystem kaufen?

> Keinesfalls ( )
> Wahrscheinlich nicht ( )
> Vielleicht ( )
> Ziemlich wahrscheinlich ( )
> Ganz sicher ( )

B. Benutzen Sie empfängnisverhütende Mittel?

> Ja ( )
> Nein ( )

Offensichtlich unterscheiden sich beide Beispiele. Nur im zweiten Beispiel geht es tatsächlich um ein „Verhalten“, im ersten Beispiel bezieht sich die Frage dagegen auf eine möglicherweise vorhandene Kaufabsicht. Es ist jedoch nicht nur irreführend, beide Beispiele unter der Überschrift „Verhaltensfragen“ zusammenzufassen.[3] Gravierender ist, daß in der Art und

[3]Diese Konfusion ist durchaus verbreitet; z.B. schreibt Lewis-Beck 1995, S. 1: „Social science research begins with a question about human behavior. The questions are different, depending on the discipline. Here are some examples. Are women more likely to

Weise, wie die Autoren ihre Verhaltensfragen konzipieren, die eigentliche Problematik sozialwissenschaftlicher Bezugnahmen auf Tätigkeiten sozialer Akteure gar nicht sichtbar wird: daß es sich nämlich in den meisten sozialwissenschaftlich relevanten Fällen nicht um „eigenes Verhalten“, sondern um soziale Interaktionsprozesse handelt. Z.B. könnten Sozialwissenschaftler untersuchen, wie es dazu kommt, daß Menschen arbeitslos werden. Sie könnten dann versuchen, von arbeitslos gewordenen Menschen Informationen über den Prozeß zu gewinnen, durch den sie arbeitslos geworden sind. Aber die Fragen, die man dann stellen müßte, passen tatsächlich in keine der von Schnell, Hill und Esser angebotenen Kategorien.

d) Schließlich unterscheiden die Autoren noch Fragen, die sich auf „Eigenschaften von Personen“ beziehen, „im allgemeinen Fragen nach personalen und demographischen Eigenschaften von Befragten wie Alter, Geschlecht, Ausbildung, Beruf, Einkommen, Familienstand, ethnische Zugehörigkeit, Parteizugehörigkeit, Konfession usw.“ Solche Fragen sind sicherlich wichtig, wenn man Einsichten in gesellschaftliche Verhältnisse – im Sinne von Lebens- und Handlungsbedingungen von Akteuren – gewinnen will. Es ist jedoch irreführend, umstandslos von „Eigenschaften von Personen“ zu sprechen. Z.B. ist das Einkommen, das eine Person bezieht, keine Eigenschaft dieser Person, sondern charakterisiert eine Situation, in der sich die Person befindet. Hier wird noch einmal die Unterscheidung wichtig, ob man mit Sozialforschung zur Aufklärung gesellschaftlicher Verhältnisse oder zur kommerziellen oder politischen Verfügbarmachung ihrer individuellen Mitglieder beitragen möchte. Denn bei den „demographischen Eigenschaften“ handelt es sich nicht um naturgegebene Klassifikationen, sondern sie reflektieren – sowohl in ihrer Bedeutung als auch in ihren praktischen Verwendungszwecken – gesellschaftliche Institutionen. Natürlich werden sie in den gesellschaftlichen Handlungszusammenhängen oft als praktisch gültige Kategorien zur Klassifikation und Diskriminierung der Gesellschaftsmitglieder verwendet. Sozialforschung sollte demgegenüber Problembewußtsein schaffen und zeigen, wie die Unterschiede zustande kommen und sich verändern und welche Folgen sich aus ihrer Institutionalisierung ergeben.

6. Nicht nur die von Schnell, Hill und Esser vorgeschlagene Einteilung von Interviewfragen in unterschiedliche Kategorien ist fragwürdig; es läßt sich generell bezweifeln, ob eine sinnvolle Einteilung gefunden werden kann. Denn schließlich hängt es von der jeweils verfolgten Fragestellung ab, wie Sachverhalte konzipiert werden sollten, welcher Informationsbedarf sich daraus ergibt und ob und ggf. wie einige der erforderlichen Informationen durch Interviews gewonnen werden können. Wir werden deshalb nicht versuchen, ein besseres Einteilungsschema vorzuschlagen, sondern orientieren

vote than men? What makes for satisfaction on the job? Does higher unemployment lead to an increase in crime? To improve student performance, should the mode of instruction be discussion rather than lecture? If a community accepts a chemical waste treatment facility, are its citizens at higher risk for cancer?“

uns an einer heuristischen Unterscheidung zwischen sozialen Sachverhalten, Meinungsäußerungen und Einstellungen.

3.2 Soziale Sachverhalte

1. Gehen wir zunächst davon aus, daß das Erkenntnisinteresse empirischer Sozialforschung darauf zielt, Einsichten in Beschaffenheiten und Funktionsweisen gesellschaftlicher Verhältnisse zu gewinnen. Dann stellt sich die Frage, wie – mit welchen Begriffen und Vorstellungen – eine sinnvolle Bezugnahme auf gesellschaftliche Verhältnisse erfolgen kann. Aus der Perspektive der statistischen Sozialforschung liegt es nahe, gesellschaftliche Verhältnisse als Mengen von Objekten, insbesondere Menschen, zu konzipieren, die durch Eigenschaften und Beziehungen charakterisierbar sind. Diese Konzeption entspräche direkt dem Basiskonzept statistischer Variablen. Daten über gesellschaftliche Verhältnisse könnten dann als Informationen über Eigenschaften von Objekten und ihre Beziehungen verstanden werden. Diese Auffassung wird jedoch problematisch, sobald man sich näher überlegt, wie man sich, sowohl im Alltagsleben als auch in der Sozialforschung, sprachlich auf gesellschaftliche Verhältnisse bezieht.

2. Zur Erläuterung soll noch einmal eine bereits im vorangegangenen Abschnitt angedeutete Kritik aufgegriffen werden. Zum Beispiel beschäftigt sich Sozialforschung damit, daß Menschen erwerbstätig sind und dadurch ein Einkommen erhalten. Aber wenn eine Person ein Einkommen bezieht, ist das dann eine Eigenschaft dieser Person? Jedenfalls nicht in der gleichen Weise, wie man z.B. von der Körpergröße als Eigenschaft eines Menschen sprechen kann. Das Konzept statistischer Variablen ist zwar so allgemein, daß man Variablen auch verwenden kann, um jedem Mitglied einer Menge von Erwerbspersonen ein Einkommen zuzurechnen. Aber das ist eine rein formale Zurechnung, die nichts darüber aussagt, wie auf sinnvolle Weise davon gesprochen werden kann, daß eine Person ein Einkommen bezieht. Man könnte vielleicht sagen: ein bestimmtes Einkommen zu beziehen, ist eine Eigenschaft einer Situation, in der sich ein Mensch befindet. Viele Merkmale, durch die in der Sozialforschung Menschen charakterisiert werden, lassen sich besser verstehen, wenn man sie auf Situationen bezieht, in denen sich Menschen mehr oder weniger lange befinden (man vgl. auch Friedrichs 1974). Diese Sichtweise erscheint nicht nur aus Gründen begrifflicher Klarheit sinnvoll, sondern erlaubt auch einen Zugang zu weiterführenden Fragen: nicht nur festzustellen, in welchen Situationen sich Menschen befinden, sondern auch Überlegungen darüber anzustellen, wie es zu diesen Situationen gekommen ist und wie sie sich verändern.

3. Einige Soziologen haben demgegenüber vorgeschlagen, den primären Realitätsbezug durch einen Begriff „sozialen Handelns“ zu fassen. Ein berühmter Vertreter dieses Vorschlags ist Max Weber:

„Soziologie (im hier verstandenen Sinn dieses sehr vieldeutig gebrauchten Wortes) soll heißen: eine Wissenschaft, welche soziales Handeln deutend verstehen und dadurch in seinem Ablauf und seinen Wirkungen ursächlich erklären will. ‘Handeln’ soll dabei ein menschliches Verhalten (einerlei ob äußeres oder innerliches Tun, Unterlassen oder Dulden) heißen, wenn und insofern als der oder die Handelnden mit ihm einen subjektiven *Sinn* verbinden. ‘Soziales’ Handeln aber soll ein solches Handeln heißen, welches seinem von dem oder den Handelnden gemeinten Sinn nach auf das Verhalten *anderer* bezogen wird und daran in seinem Ablauf orientiert ist.“ (Weber 1976, S. 1)

Ob man soziales Handeln verstehen kann, soll an dieser Stelle nicht diskutiert werden. Die Frage ist, ob die Sachverhalte, mit denen sich die Sozialforschung beschäftigt, zunächst durch einen Begriff sozialen Handelns vergegenwärtigt werden können. Von Ausnahmen abgesehen muß diese Frage wohl verneint werden. Jedenfalls die statistische Sozialforschung beschäftigt sich fast immer nicht mit sozialem Handeln, sondern mit Sachverhalten; z.B. damit, daß Sozialhilfeempfänger Sozialhilfe beziehen, aber „Sozialhilfe beziehen“ ist kein soziales Handeln; oder mit Autounfällen, aber das ist ebenfalls kein soziales Handeln. Es erscheint sinnvoller, in diesen Fällen von *sozialen Sachverhalten* zu sprechen.

4. Aber es scheint doch evident zu sein, daß soziale Sachverhalte durch Aktivitäten sozialer Akteure hervorgebracht werden? Ist es dann nicht sinnvoll, beim sozialen Handeln zu beginnen und soziale Sachverhalte als dessen „Wirkungen“ aufzufassen, so wie dies Weber in dem angeführten Zitat nahelegt? Aber dann müßte man zunächst erklären, wie von „Wirkungen“ sozialen Handelns sinnvoll gesprochen werden kann. Wenn ich in den Supermarkt gehe, um einige Einkäufe zu machen, kann man das in Webers Sinn als eine „soziale Handlung“ bezeichnen; aber was sind die Wirkungen dieser Handlung? H. Esser, der in seiner Einführung in die Soziologie an die oben zitierte Definition Webers anschließt, sagt:

„Die Definition geht weiter davon aus, daß es die kollektiven Wirkungen – gesellschaftliche Institutionen, Strukturen, Kulturen und soziale Prozesse wie soziale Ordnung, soziale Ungleichheit, soziale Differenzierung oder sozialer Wandel u.a. – sind, die im Zentrum des Interesses der Untersuchungen und Analysen der Soziologie stehen – und *nicht* die individuellen Akteure und auch nicht deren individuelles Handeln.“ (Esser 1993, S. 4)

Das kann man der Definition zwar nicht entnehmen, verweist aber sicherlich auf Stichworte, mit denen sich Weber und andere Soziologen beschäftigt haben. Das Problem ist jedoch nicht nur, in welcher Weise man von „Institutionen“, „Strukturen“ und „Kulturen“ sinnvoll sprechen kann und ob und ggf. wie man sich darunter Sachverhalte vorstellen kann. Es ist vor allem überhaupt nicht selbstverständlich, ob und ggf. wie man das, was man sich vielleicht unter diesen Stichworten vorstellen kann, als Wirkungen sozialen Handelns auffassen kann. — Um einen Zugang zum Verständnis der statistischen Sozialforschung zu gewinnen, erscheint es deshalb sinnvoller, von sozialen Sachverhalten auszugehen und die Frage, ob und wie ihr Zustandekommen durch eine Bezugnahme auf Aktivitäten sozialer Akteure verstanden werden kann, zunächst offen zu lassen.

5. Was mit sozialen Sachverhalten gemeint ist, kann durch Beispiele erläutert werden, etwa: eine Person bezieht ein gewisses Einkommen; ein Student hat eine Klausur mit einer guten Note abgeschlossen; zwei Personen haben sich begrüßt. Unsere Sprache erlaubt es, in diesen Fällen von Sachverhalten zu sprechen, die man feststellen kann. Stets gibt es auch einen Bezug zu bestimmten Personen, von denen man sagen kann, daß sie an dem Sachverhalt *beteiligt* sind. Das zuletzt angeführte Beispiel zeigt auch, daß man sich soziale Sachverhalte oft als Ereignisse vorstellen kann, als „Vorkommnisse“, die in der sozialen Realität stattfinden. Nicht nur Begrüßungen, auch zum Beispiel, daß ein Autounfall stattgefunden hat, daß zwei Menschen geheiratet haben, daß ein neuer Supermarkt eröffnet worden ist usw. Stets sind Menschen beteiligt, und deshalb spricht man von *sozialen* Sachverhalten. Aber Sachverhalte können nicht als Eigenschaften der beteiligten Personen beschrieben werden. Man kann zwar Menschen u.a. dadurch charakterisieren, daß und wie sie an sozialen Sachverhalten beteiligt sind, durch sie betroffen sind, an ihrem Zustandekommen mitgewirkt haben; aber dadurch werden die Sachverhalte nicht zu Eigenschaften von Menschen.

6. Ergänzend sei darauf hingewiesen: Daß Menschen an einem sozialen Sachverhalt *beteiligt* sind, impliziert nicht, daß der Sachverhalt von Menschen *hervorgebracht* worden ist. Es sollten deshalb zwei Arten von Fragen unterschieden werden: Wie soziale Sachverhalte *zustandegekommen* sind, und ob und ggf. wie davon gesprochen werden kann, daß Akteure an ihrem Zustandekommen *mitgewirkt* haben. Sicherlich interessiert man sich oft vor allem für die zweite Art von Fragen, denn dann kann man auch daran denken, daß soziale Sachverhalte verändert werden können. Denkt man zum Beispiel an einen Autounfall, kann man sich nicht nur fragen, wie es dazu gekommen ist, sondern auch, welche Rolle die beteiligten Akteure gespielt haben. Und bei einer Person, die ein Einkommen bezieht, kann man sich überlegen, wie es dazu gekommen ist. Vielleicht kann festgestellt werden, daß dieser Sachverhalt die Folge eines Vertrags ist, der zuvor zwischen der Person und einem Arbeitgeber abgeschlossen worden ist. Aber wie ist es, wenn die Person eine Rente bezieht, welches soziale Handeln hat dann diesen Sachverhalt hervorgebracht und insbesondere über die Höhe der Rente entschieden?[4]

[4]Die Frage, in welcher Weise Akteure am Zustandekommen von sozialen Sachverhalten mitwirken, setzt voraus, daß man bereit ist, einige Objekte in der sozialen Realität als Akteure zu betrachten, andere dagegen nicht. Die Unterscheidung zwischen Objekten, die Akteure sind, und Objekten, die keine Akteure sind, ist jedoch nicht selbstverständlich. Die Frage, wie ein sozialer Sachverhalt zustandegekommen ist, bedarf deshalb

7. Die Konzeption sozialer Sachverhalte liefert *einen* möglichen begrifflichen Zugang zum Reden über gesellschaftliche Verhältnisse. Für die Sozialforschung erscheint dieser Ansatz vor allem deshalb sinnvoll, weil er sich mit der Vorstellung verbinden läßt, empirische Daten über gesellschaftliche Verhältnisse gewinnen zu können (natürlich auch deshalb, weil er sich an das umgangssprachliche Reden über gesellschaftliche Verhältnisse anschließt). Jeder hat eigene Erfahrungen mit *einigen* sozialen Sachverhalten, und es gibt eine gemeinsame Sprache, in der über soziale Sachverhalte und darüber, wie Akteure in ihr Zustandekommen involviert und von ihren Folgen betroffen sind, gesprochen werden kann. Dadurch können Erfahrungen ausgetauscht und kann die lokal beschränkte Perspektive jedes einzelnen Akteurs ausgeweitet werden. Insbesondere können Sozialwissenschaftler versuchen, sich Zugang zu den Erfahrungen zu verschaffen, die soziale Akteure mit jeweils ihren sozialen Kontexten gemacht haben, um daraus ein Bild gesellschaftlicher Verhältnisse zu gewinnen.

8. Schließlich kann man bei diesem Ansatz einen sinnvollen Zweck angeben, dem Interviews in der Sozialforschung dienen können: sich Zugang zu den Erfahrungen zu verschaffen, die die Gesellschaftsmitglieder mit jeweils ihren sozialen Kontexten gemacht haben. Insoweit Interviewer und Interviewte vergleichbare oder zumindest kommunizierbare Erfahrungen mit sozialen Sachverhalten gemacht haben, können sie sich verständigen. Es ist auch klar, daß es bei so konzipierten Interviews nicht um Tests geht, insbesondere nicht darum, die interviewten Personen zu testen. Sie werden nicht mit „Stimuli“ konfrontiert, um daraufhin ihre „Reaktionen“ zu ermitteln. Sondern Interviewer und Interviewte beziehen sich in einer erkenntnistheoretisch vergleichbaren kommunikativen Perspektive auf soziale Sachverhalte, wobei die das Interview begründende Annahme darin besteht, daß die interviewten Personen über die zu ermittelnden Sachverhalte besser informiert sind als der Interviewer.

9. Dieser Ansatz zum Verständnis von Interviews in der Sozialforschung kann in verschiedenen Varianten verfolgt werden. Restriktionen ergeben sich, wenn Daten für statistische Sozialforschung gewonnen werden sollen. Wie schon in Abschnitt 2.1 bemerkt worden ist, müssen dann die durch Interviews gewinnbaren Informationen in die Form von Werten statistischer Variablen gebracht werden. Die Informationen selbst können zwar

stets einer Erläuterung; insbesondere dann, wenn nach „Ursachen“ gefragt wird. Denn am Zustandekommen von Sachverhalten sind stets viele Objekte beteiligt, und es ist schließlich eine Frage weltanschaulicher und moralischer Überzeugungen, welche Objekte man als Akteure, als Verursacher und als verantwortlich für das Zustandekommen eines Sachverhalts ansehen will. Auch aus diesem Grund erscheint es nicht sinnvoll, soziale Sachverhalte von vornherein als „Produkte“ menschlicher Akteure zu definieren. Die Fragestellung, ob und ggf. wie soziale Sachverhalte als durch Menschen hervorgebracht betrachtet werden können, soll damit natürlich nicht aus der Sozialforschung ausgeschieden werden. Im Gegenteil, sie bildet für die meisten Bezugsprobleme ein wichtiges Motiv.

auf durchaus vielfältige Weisen gewonnen werden, schließlich müssen sie jedoch als Werte in einem Merkmalsraum fixiert werden. Soweit solche Daten aus Interviews gewonnen werden, werden deshalb in der Regel standardisierte Interviews verwendet, bei denen die Fragen vorgängig konzipierten Merkmalsräumen entsprechen.[5] Das schränkt natürlich die Möglichkeiten ein, die Erfahrungen der jeweils interviewten Personen in ihren Besonderheiten angemessen wahrnehmen zu können. Für die Datenerhebung kommen deshalb in erster Linie nur Sachverhalte infrage, deren Klassifizierung und Kennzeichnung auf einheitliche Weise institutionalisiert ist.[6] Insofern kann einerseits der Ansatz der statistischen Sozialforschung wegen seiner unvermeidlichen Beschränktheiten – im wesentlichen wegen der dann erforderlichen Definition von Daten durch statistische Variablen – kritisiert werden.[7] Andererseits läßt sich jedoch dieser Ansatz auch verteidigen. Denn zumindest eine wichtige Aufgabe der Sozialforschung besteht darin, Einsichten in Beschaffenheiten und Funktionsweisen von Institutionen zu gewinnen, also das gesellschaftliche Leben im Hinblick darauf zu untersuchen, daß und wie es institutionalisiert ist und welche Folgen sich daraus ergeben.

3.3 Meinungsäußerungen

1. Im vorangegangenen Abschnitt haben wir uns auf die Aufgabe bezogen, Informationen über soziale Sachverhalte zu gewinnen. Wie ebenfalls schon bemerkt wurde, steht dazu die in der Methodenliteratur übliche Betonung von Interviewfragen, die sich auf Meinungen und Einstellungen richten, in einem eigentümlichen Mißverhältnis. In diesem und im folgenden Abschnitt sollen deshalb einige ergänzende Überlegungen zur Diskussion gestellt werden.

2. Jeder kennt natürlich das Wort ‘Meinung’ und kann es in zahlreichen Kontexten sinnvoll verwenden. Wenn man sprachliche Äußerungen als Meinungen bezeichnet, stellt man einen Bezug zu den Subjekten her, die diese Meinungen äußern. Die Verwendung des Worts impliziert in der Regel auch

[5] Sicherlich ist es auch möglich, Informationen aus nicht-standardisierten Befragungen *nachträglich* in statistische Daten zu verwandeln.

[6] Diese Sinnvoraussetzung der statistischen Sozialforschung hat Adorno 1957, S. 515, etwas überspitzt folgendermaßen charakterisiert: „Eine zugleich atomistische und von Atomen zu Allgemeinheiten klassifikatorisch aufsteigende Sozialwissenschaft ist der Medusenspiegel einer zugleich atomisierten und nach abstrakten Klassifikationsbegriffen, denen der Verwaltung, eingerichteten Gesellschaft.“

[7] Etwa mit folgendem Argument von H. Blumer 1956, S. 392: „Meiner Beurteilung nach setzt der Interpretations- oder Definitionsprozeß, der in menschlichen Gruppen abläuft, der erfolgreichen Anwendung der Variablenanalyse auf menschliches Sozialleben die entscheidende Grenze. Dieser Prozeß, den ich für den Kern menschlichen Handelns halte, verleiht dem menschlichen Sozialleben jenen Charakter, der den logischen Prämissen der Variablenanalyse zu widersprechen scheint.“

die Unterstellung, daß es andere Meinungen gibt oder geben könnte. In dieser allgemeinen Bedeutung des Wortes besteht jedes Gespräch darin, daß Meinungen geäußert und zur Diskussion gestellt werden. Und auch jedes Interview kann so verstanden werden, daß Befragungspersonen Meinungen äußern. Der Interviewer fragt zum Beispiel: Wie alt sind Sie? Und der Interviewte äußert dann eine Meinung, zum Beispiel: Ich bin 30 Jahre alt. Und da sich die Frage auf einen Sachverhalt bezieht, könnte es sich um eine falsche Meinung handeln.

3. Was kann in der Form von Meinungen mitgeteilt werden? In erster Näherung erscheint es sinnvoll, jedenfalls drei Arten von Meinungsäußerungen zu unterscheiden.

a) Meinungen können sich auf *Sachverhalte* beziehen; durch Meinungsäußerungen werden dann – der Intention nach – Aussagen über Sachverhalte gemacht. Ganz unabhängig von der Art der Sachverhalte gibt es in diesem Fall eine wichtige Kommunikationsvoraussetzung: daß Meinungsäußerungen, wenn sie sich auf Sachverhalte beziehen, ein Wahrheitsanspruch unterstellt werden kann. Fragt man jemanden nach dem Weg zum Bahnhof, möchte man nicht eine beliebig ausgedachte Antwort, sondern eine Auskunft über einen Weg, der tatsächlich zum Bahnhof führt.

b) In der Form von Meinungsäußerungen können Menschen *Präferenzen* zum Ausdruck bringen. Das geschieht z.B. im Restaurant, wenn der Kellner fragt, was man bestellen möchte. Der Kellner erwartet dann nicht, daß man über irgendeinen Sachverhalt berichtet, sondern er möchte einfach wissen, was der Gast essen oder trinken möchte. Das kann ihm der Gast in der Form einer sprachlichen Äußerung mitteilen. Eine solche Äußerung bezieht sich jedoch nicht auf einen Sachverhalt; sie bezieht sich insbesondere nicht auf einen „mentalen Zustand“, den der Gast zunächst in seinem Kopf feststellt, um dann in der Form einer Tatsachenaussage darüber zu berichten.[8] Insofern ist schon die Formulierung, daß sich in diesem Fall die Meinungsäußerungen auf Präferenzen beziehen, metaphorisch. Man kann zwar Präferenzen *zum Ausdruck bringen* – in durchaus vielfältigen, nicht nur sprachlichen Formen; aber Präferenzen sind keine Sachverhalte, die man beschreiben oder in irgendeiner Weise feststellen könnte.[9]

c) Schließlich können in der Form von Meinungsäußerungen auch *Bewertungen* mitgeteilt werden. Zum Beispiel kann ein Lehrer Schulaufsätze bewerten, ein Theaterkritiker kann die Aufführung eines Schauspiels bewerten, Studenten können die Qualität des Mensaessens bewerten usw. Auch

[8] Lesenswerte Überlegungen zu dieser Frage findet man z.B. bei Schneider 1997.

[9] Gleichwohl ist es natürlich möglich, Menschen unter anderem durch von ihnen zum Ausdruck gebrachte Präferenzen zu charakterisieren. Aber dann werden nicht Präferenzen beschrieben, sondern man charakterisiert einen Menschen, indem man ihm Präferenzen zurechnet.

in diesen Fällen geht es ersichtlich nicht darum, über Sachverhalte zu berichten. Genauer: Eigenschaften der zu bewertenden Gegenstände spielen durchaus eine Rolle; bei einer ernstzunehmenden Bewertung erwartet man, daß sie nicht beliebig ausgedacht wird, sondern sich auf wahrnehmbare Eigenschaften der zu bewertenden Gegenstände oder Situationen bezieht. Gleichwohl besteht eine Bewertung nicht darin, über solche wahrnehmbaren Eigenschaften zu berichten. Hinweise auf solche Eigenschaften dienen vielmehr zur Begründung der schließlich geäußerten Bewertung. Ein solcher Begründungsanspruch unterscheidet Bewertungen andererseits von der Äußerung von Präferenzen. Einfach zu sagen, A gefällt mir besser als B, ist bestenfalls die Äußerung einer Präferenz, liefert jedoch keine Bewertung. Eine ernstzunehmende Bewertung muß sich nicht nur auf Eigenschaften der zu bewertenden Gegenstände beziehen, sondern auch in der Lage sein, die Kriterien und Gesichtspunkte, nach denen die Bewertung vorgenommen wird, zu explizieren.

4. Interviewfragen können sich auf Sachverhalte, Präferenzen oder Bewertungen beziehen. Konzipiert man korrespondierend zu solchen Fragen statistische Variablen für eine Gesamtheit von Befragungspersonen, liefern sie gleichermaßen Daten (im Sinne unserer Definition dieses Worts). Dennoch gibt es einen wichtigen Unterschied. Nur im ersten Fall entsprechen den Daten Sachverhalte. Wenn sich dagegen die Fragen auf Präferenzen oder Bewertungen beziehen, entsprechen den Daten keine Sachverhalte. Man kann hinzufügen, daß der Unterschied nicht darin besteht, daß es sich im ersten Fall um „objektive“, in den anderen Fällen um „subjektive“ Daten handelt, denn Meinungsäußerungen sind immer „subjektiv“.

5. Hier muß man allerdings etwas aufpassen. Unsere Überlegung hat sich bisher an der Frage orientiert, welche *Bedeutungen* Meinungsäußerungen haben können. Orientiert man sich an dieser Frage, gibt es klar erkennbare Unterschiede zwischen Aussagen, mit denen – der Intention nach – Sachverhalte festgestellt werden, und andererseits Aussagen, durch die Präferenzen oder Bewertungen zum Ausdruck gebracht werden. Das Argument, wie es bisher verwendet worden ist, besteht darin, daß mit Aussagen, durch die Präferenzen oder Bewertungen zum Ausdruck gebracht werden, keine Sachverhalte festgestellt werden. Dennoch ist es möglich, sich auch in diesen Fällen auf Sachverhalte zu beziehen. Dafür gibt es zwei unterschiedliche Möglichkeiten. Zunächst kann man sich direkt auf Meinungsäußerungen als „Sprechakte“ beziehen, die in Raum und Zeit stattfinden; z.B.: es ist doch eine Tatsache, daß Du gestern folgendes … gesagt hast. Diese Betrachtungsweise entspricht in gewisser Weise dem Vorschlag, Interviews im Sinne eines Reiz-Reaktions-Schemas aufzufassen. Folgte man dieser Auffassung, müßte man sagen, daß Interviews Daten über „Sprechakte“ liefern; und man müßte konsequenterweise auch sozialwissenschaftliche Aussagen, die auf der Grundlage solcher Daten entwickelt werden, als Aussagen über „Sprechakte“ von Menschen, die durch beliebig ausgedachte

„Stimuli“ provoziert worden sind, auffassen. Soweit ersichtlich ist, wird jedoch dieses Verständnis sozialwissenschaftlicher Aussagen auch von den Anhängern des Reiz-Reaktions-Schemas nicht vertreten.[10] Vielmehr folgen sie einer zweiten Variante, um aus Meinungsäußerungen Sachverhalte zu machen: Meinungsäußerungen werden als Hinweise auf „Einstellungen“ interpretiert, die dann den Subjekten der Meinungsäußerungen als sie charakterisierende „psychische Sachverhalte“ zugeschrieben werden.

6. Mit der Konstruktion von „Einstellungen“ werden wir uns im nächsten Abschnitt beschäftigen. Hier halten wir zunächst an der Voraussetzung fest, daß mit Interviews Meinungen erfragt werden können, *weil* man sich für ihren Inhalt interessiert. Dann ist es offenbar wichtig, sich mit der Bedeutung von Meinungsäußerungen zu beschäftigen und zu unterscheiden, ob sie sich auf Sachverhalte beziehen oder ob sie Präferenzen oder Bewertungen zum Ausdruck bringen. Soweit Sozialforschung an Einsichten in gesellschaftliche Verhältnisse interessiert ist, benötigt sie Daten über soziale Sachverhalte. Interviews sind dann, wie in Abschnitt 3.2 ausgeführt worden ist, eine potentiell wichtige Datenquelle, weil man durch sie Informationen über soziale Sachverhalte gewinnen kann. Andererseits sind Interviewfragen, die sich auf Präferenzen oder Bewertungen beziehen, für diesen Zweck nicht relevant.

7. Das heißt natürlich nicht, daß solche Fragen nicht gestellt werden sollten. Es gibt oft gute Gründe dafür, Menschen nach Präferenzen oder Bewertungen zu fragen. Jeder kennt genügend Beispiele aus dem praktischen Leben. Man kann sich auch viele Situationen vorstellen, in denen eine Ermittlung von Präferenzen oder Bewertungen bei einer Gesamtheit von Menschen sinnvoll sein kann. Meistens geht es dann um politische Fragen (in einem weiten Sinn des Wortes), also darum, wie irgendwelche Aspekte des gesellschaftlichen Lebens eingerichtet werden sollten. Es ist sicherlich nicht sinnlos, zur besseren Diskussion und zur Vorbereitung von Entscheidungen die Meinungen von Menschen in Erfahrung zu bringen, die an der Fragestellung interessiert sind oder interessiert sein könnten. So verstandene Meinungsforschung kann man als Bestandteil einer Demokratie verstehen, in der den Meinungen von Menschen Gehör verschafft werden sollte. Aber es handelt sich offenbar um unterschiedliche Erkenntnisinteressen. In der Sozialforschung geht es darum, empirisch ausweisbare und

[10] Obwohl es gelegentlich Formulierungen gibt, die eine solche Lesart nahelegen; zum Beispiel die folgende von M. Lodge 1982, S. 177: „Most of what we as social scientists claim to know about opinions and attitudes is founded on the analysis of responses to semantic stimuli. From a reading of the literature, it is clear that much of the empirical research on social, psychological, and political behavior is based on the analysis of one type of behavioral response – some variant of verbal behavior – what people say or, more typically, what people check on a labelled category scale, to one type of stimulus, words, typically words and phrases embedded in questionnaires administered to survey respondents.“

begründbare Einsichten in Beschaffenheiten und Funktionsweisen gesellschaftlicher Verhältnisse zu gewinnen. Meinungsforschung – in unserem Verständnis dieses etwas schillernden Wortes – strebt dagegen an, Meinungen von Menschen Gehör zu verschaffen.[11]

3.4 Einstellungskonstruktionen

1. Wenn man Menschen nach ihren Meinungen fragt, tut man dies normalerweise deshalb, weil man sich dafür interessiert, *was* sie zu sagen haben. Als Sozialforscher interessiert man sich für ihre Erfahrungen, um ein Bild gesellschaftlicher Verhältnisse zu gewinnen; als Meinungsforscher möchte man ihren Präferenzen und Bewertungen Gehör verschaffen. In einem deutlichen Kontrast dazu steht die statistische Einstellungsforschung, an der sich auch die sozialwissenschaftliche Methodenliteratur großenteils orientiert. Meinungen von Befragungspersonen werden dann nicht ermittelt, um sie in einem argumentativen Kontext zu verwenden. Meinungen interessieren dann nicht wegen ihres Inhalts, sondern als Indikatoren für Einstellungen.[12] Zugrunde liegt die Vorstellung, daß man den Befragungspersonen Einstellungen zurechnen kann, die sich durch eine Abfrage von Meinungen erfassen lassen. Wenn sich also überhaupt ein Erkenntnisinteresse unterstellen läßt, geht es in diesem Fall um die „psychische Verfassung“ der Befragungspersonen bzw. einer Gesamtheit von Menschen, aus der die Befragungspersonen ausgewählt worden sind.

2. Menschen durch Einstellungen zu charakterisieren, ist aus alltagspraktischen Handlungs- und Kommunikationszusammenhängen durchaus vertraut. Sozialpsychologische Verfahren der Einstellungskonstruktion bewegen sich jedoch in einem anderen Kontext. Den historischen Hintergrund bildet die Entwicklung der Psychophysik im 19. Jahrhundert.[13] Die sta-

[11] Dies entspricht nicht unbedingt dem Selbstverständnis von Meinungsforschern. Z.B. findet man in einer „Einführung in die Methoden der Demoskopie“ von E. Noelle-Neumann und T. Petersen (2000, S. 96f) den Hinweis, daß „bei vielen Untersuchungen nach Sachverhalten, Verhaltensweisen oder Einstellungen geforscht wird, die den Befragten unbewußt sind; sie können also auch bei gutem Willen, direkt befragt, keine Auskunft geben.“ Dementsprechend konzipieren die Autoren ihre Befragungsmethoden nicht als genuine Interviews, sondern als Tests (S. 59ff). Zum Beispiel: „Im demoskopischen Interview müssen Testfragen, Beobachtungsformen und Testaufgaben gefunden werden, die den Befragten veranlassen, seine Haltung zu offenbaren.“ (S. 83) „Man muß sich immer wieder aufs neue vergegenwärtigen: Das demoskopische Interview ist als Reaktions-Experiment zu verstehen, nicht als kluges Gespräch.“ (S. 85) Bei dieser Variante der Demoskopie handelt es sich also nicht um Meinungsforschung in der von uns vorgeschlagenen Definition.

[12] Z.B. sagen C. und D. Nachmias 1981, S. 209: „The study of opinion is of interest only in so far as it is a symbol of an attitude.“ Noch allgemeiner heißt es bei Fowler 1988, S. 74: „In surveys, answers are of interest not intrinsically but because of their relationship to something they are supposed to measure.“

[13] Eine lesenswerte historische Darstellung findet man bei Danziger 1990.

tistische Einstellungsforschung läßt sich als ein Abkömmling verstehen. Louis Thurstone, einer der Wegbereiter der Vorstellung, daß Einstellungen „gemessen" werden könnten, hat den Zusammenhang in einem Aufsatz über „The Measurement of Social Attitudes" so dargestellt:

„Several years ago when I was teaching conventional psychophysics, it seemed to me that psychophysics was really a very dull subject in spite of the fact that it did offer the satisfaction of clean and quantitative logic. [...] One way in which to retain the satisfactions that can be found in the logic of this subject is to change its content. We have tried this and it has seemed to some of us that psychophysics thereby takes on an entirely new aspect. Instead of asking a person, 'Which of these two little cylinders is the heavier?' (apologies to Mr. Boring for the stimulus error) – instead of asking a person which of two cylinders is the heavier, we might as well ask him something interesting, such as, 'Which of these two nationalities do you in general prefer to associate with?' or, 'Which of these two offenses do you consider to be in general the more serious?' or, 'Which of these two pictures or colored designs do you like better?' Questions of this sort of discrimination might be multiplied indefinitely and if they could be handled with some sort of psychophysical logic it is clear that we should have here the possibilities of objective description of more psychological significance than the sensory limen." (Thurstone 1932, S. 249f)

3. Die sich bis heute fortsetzenden Versuche, Einstellungen in einer Art von Analogie zu psychophysischen Experimenten „meßbar" zu machen, beruhen allerdings auf einer fragwürdigen Verwechslung.[14] Die Psychophysik, wie sie im 19. Jahrhundert von Gustav Fechner, Wilhelm Wundt u.a. entwickelt worden ist, interessiert sich für Wahrnehmungen bzw. Empfindungen; zum Beispiel für räumliche Wahrnehmungen oder Empfindungen von Lautstärken oder Tonhöhen. Das Erkenntnisinteresse zielt darauf, Einsichten in Zusammenhänge zwischen physikalisch charakterisierbaren Reizen und ihren subjektiven Wahrnehmungen zu gewinnen.[15] Bei Meinungen und Einstellungen handelt es sich jedoch nicht um subjektive Empfindungen. Wenn man zu irgendeiner Frage eine Meinung äußert, geschieht dies nicht so, daß man zunächst beobachtet, welche Meinung man hat, und dann das Ergebnis der Beobachtung mitteilt. Wenn man einen Ton hört, kann man mehr oder weniger gut beschreiben, wie man ihn hört. In der Alltagspsychologie könnte man davon sprechen, daß man dann eine Empfindung (des Tons) beschreibt.[16] Aber Meinungen und Einstellungen hat man nicht in

[14]Weitere Argumente, auf die wir hier nicht näher eingehen, richten sich dagegen, daß schon der Begriff 'Einstellung' nicht auf sinnvoll definierbare Sachverhalte verweist; vgl. Johnson 1936, S. 348ff.

[15]Guilford 1954, S. 20, hat das folgendermaßen ausgedrückt: „The classical psychophysics of Fechner, G. E. Müller, and Wundt was concerned primarily with the determination of sensory thresholds, or limens. Fechner's original interest was in the philosophical question of the relation between body and mind, between the physical and the psychical aspects of the word."

[16]Auf dieser Voraussetzung beruhen Untersuchungen zur Psychophysik der Wahrneh-

der Weise, wie man Empfindungen hat. Von Meinungen und Einstellungen so zu sprechen, als ob es sich um mit Empfindungen vergleichbare „psychische Tatsachen" handelt, ist irreführend.[17]

4. Viele Unklarheiten im Sprachgebrauch sind eine Folge dieser irreführenden Analogie. Man erinnere sich an die zu Beginn des Kapitels zitierte Äußerung von Black, daß es bei Meinungen und Einstellungen darum gehe, „how intensely people *feel* about issues, as opposed to what they know or can do". In seinem oben erwähnten Aufsatz sagte Thurstone: „We are measuring the degree of affect for or against the social objects listed." (S. 252) In beiden Formulierungen wird suggeriert, daß man sich das „Haben" von Meinungen und Einstellungen so vorstellen könne, wie man Empfindungen oder Gefühle hat. Die Analogie ist jedoch irreführend, denn es ist ein wichtiger Unterschied, ob man Menschen nach ihren Meinungen oder nach ihren Empfindungen fragt.[18] Fragt man z.B. jemanden nach

mung, wie sie z.B. von Campenhausen (1981) dargestellt werden. Dort heißt es (S. 6): „Wenn aber ein Reiz zu einer bewußten Wahrnehmung führt, dann kann man seine physikalische Natur auch mit der Methode der Psychophysik untersuchen." Auch bei der Konzeption psychophysischer Experimente gibt es einen bemerkenswerten Kontrast zur „Messung von Einstellungen". Campenhausen schreibt (S. 7f): „Drei Arbeitsregeln für psychophysische Experimente zur Natur des Reizes müssen darum immer beachtet werden: Die Information a), die übertragen werden soll, muß vor dem Experiment vereinbart sein. Der Versuchsleiter (Vl.) und die Versuchsperson (Vp.) müssen sich auf diejenige Wahrnehmung, auf die es im Experiment ankommen soll, vorher einigen. Sie können beispielsweise festlegen, daß bei einer akustischen Wahrnehmung die Tonhöhe unabhängig von der Lautheit, der Klangfarbe und von zusätzlichen Geräuschen beurteilt werden soll. Der Vl. wird weiterhin b) darauf achten müssen, daß sich die Vp. aktiv an dem Experiment beteiligt, daß ihr also nichts wegen mangelnder Aufmerksamkeit entgeht. [...] Schließlich c) muß man damit rechnen, daß sich die Vp. manchmal irrt und beispielsweise behauptet, etwas wahrzunehmen, wenn gar kein Reiz geboten wurde."

[17]Hier sei noch ein neueres Zitat von Micko 1970, S. 15, angeführt, in dem die Verwechslungen gut zum Ausdruck kommen: „Wir können die Psychophysik als Studium des Zusammenhangs von Reizstärke und Empfindungsstärke definieren, oder als Zusammenhang anderer, einander entsprechender quantitativer Variabler von Reiz und Empfindung. In diesem Fall läßt sich der Begriff zweifelslos nicht so erweitern, daß wir von einer Psychophysik der Reaktionsvariablen sprechen können. Das ist nur möglich, wenn wir uns an die beinahe ebenso alte, neutralere Formulierung halten, derzufolge wir in der Psychophysik den Zusammenhang von psychologischen und physikalischen Maßen einer jeweils betrachteten Reizvariablen untersuchen. Die psychologische Messung unterscheidet sich von der physikalischen dadurch, daß anstelle von Metermaß, Waage, Uhr etc. menschliche Beobachter als 'Meßinstrument' verwendet werden. Entweder registrieren wir Urteile eines Beobachters wie die Anzeigen eines Meßinstruments, oder wir gewinnen die Meßwerte mit Hilfe eines Meßmodells aus Merkmalen seiner verbalen oder nicht-verbalen Reaktionen auf die dargebotenen Reize."

[18]Es ist bemerkenswert, daß die Problematik einer solchen Analogie in der Literatur zur „Einstellungsmessung" nicht diskutiert wird. Gelegentlich, z.B. von Anderson 1982, S. 145, wird zwar darauf hingewiesen, daß es Unterschiede zwischen physikalisch beschreibbaren Reizen der klassischen Psychophysik und „sozialen Einstellungsobjekten" gibt; aber die wesentliche Unterscheidung – zwischen Empfindungen und Meinungsäußerungen – gerät gar nicht ins Blickfeld. Infolgedessen wird auch nicht auf genaue sprachliche Formulierungen geachtet. Zum Beispiel heißt es bei Sjöberg 1982, S. 199: „People ex-

dem Alter oder der gegenwärtigen Erwerbstätigkeit, möchte man normalerweise nicht wissen, was er bei dieser Frage empfindet (obwohl auch das vielleicht interessant sein mag). Natürlich werden Meinungsäußerungen oft von Affekten begleitet und lassen sich Gefühls- und Meinungsäußerungen oft nur schwer unterscheiden;[19] die Unterscheidung ist jedoch für menschliche Kommunikation konstitutiv.

5. Der irreführende Versuch, die Ermittlung von Meinungen und Einstellungen konzeptionell an Vorstellungen der Psychophysik anzuschließen, hat sicherlich zur Verbreitung der Vorstellung beigetragen, daß Interviews im Sinne eines Reiz-Reaktions-Schemas aufgefaßt werden können. Ein solches Schema ist ja bei psychophysischen Fragestellungen nicht unbedingt sinnlos. Menschen werden mit gewissen Reizen konfrontiert, z.B. mit Tönen unterschiedlicher Lautstärke, und sollen dann über korrespondierende „Reaktionen" berichten. Aber die Formulierung deutet schon an, daß in diesem Zusammenhang mit „Reaktionen" nicht „sprachliche Reaktionen" gemeint sind, sondern den Reizen korrespondierende Empfindungen, über die dann berichtet werden soll. Das Erkenntnisinteresse zielt darauf, Einsichten in Zusammenhänge zwischen Reizen und Empfindungen zu gewinnen. Dabei wird unterstellt, daß solche Empfindungen als „psychische Tatsachen" konzipiert werden können, die sich beschreiben lassen. Das Reiz-Reaktions-Schema wird jedoch irreführend, wenn man es aus dem Kontext psychophysischer Fragestellungen löst und auf Interviews zu übertragen versucht, also die Interviewfragen als Reize und die dann erfolgenden Meinungsäußerungen als Reaktionen darstellt. Daraus resultiert nicht nur ein falsches Verständnis von Interviews; die falsche theoretische Konzeption macht auch die produzierten Einstellungskonstrukte sinnlos. Denn an die Stelle einer kommunikativen Ermittlung von Meinungen tritt eine nur noch verfahrenstechnisch rekonstruierbare Generierung „sprachlicher Reaktionen". So kann man sich bestenfalls noch darüber streiten, was eine größere Beliebigkeit aufweist: der Prozeß, durch den „sprachliche Reaktionen" erzeugt werden, oder der Prozeß, der sie in Einstellungskonstrukte verwandelt.

6. An Beispielen kann man sich verdeutlichen, wie „sprachliche Reaktionen" generiert werden, um daraus Einstellungen zu konstruieren. Den Befragungspersonen wird – da man einen fiktiven Sachverhalt „messen" möchte – typischerweise gar keine Gelegenheit gegeben, eine Meinung zu entwickeln, sondern sie werden mit vorformulierten „Statements" konfrontiert, zu denen sie dann eine „Meinung" äußern sollen. Wie man sich erinnert, ist diese Vorgehensweise bei Schnell, Hill und Esser schon in ihrer Definition von Meinungs- und Einstellungsfragen enthalten, denn sie sagen ausdrücklich, daß es in diesen Fällen um eine Stellungnahme zu „Statements" geht (vgl. Abschnitt 3.1). Bortz und Döring begründen diese Vorgehensweise folgendermaßen:

> „Die Art der Formulierung des Fragebogenitems – als *Frage* oder als *Behauptung (Statement)* – richtet sich nach den untersuchten Inhalten. (Beispiel: 'Sind Sie der Ansicht, daß der Gesetzgeber für Autobahnen ein Tempolimit vorschreiben sollte?' Oder: 'Der Gesetzgeber sollte für Autobahnen ein Tempolimit vorschreiben!'). Zur Erkundung von Positionen, Meinungen und Einstellungen sind Behauptungen, deren Zutreffen der Befragte einzustufen hat, besser geeignet als Fragen. Mit ihrer Hilfe läßt sich die interessierende Position oder Meinung prononcierter und differenzierter erfassen als mit Fragen, die zum gleichen Inhalt gestellt werden. Die Frage ist üblicherweise allgemeiner formuliert und hält das angesprochene Problem prinzipiell offen. Realistische, tatsächlich alltäglich zu hörende Behauptungen sind demgegenüber direkter und veranlassen durch geschickte, ggf. gar provozierende Wortwahl auch zweifelnde, unsichere Befragungspersonen zu eindeutigen Stellungnahmen." (Bortz und Döring 1995, S. 233)

Man sollte sich klarmachen, daß es bei der Provokation „sprachlicher Reaktionen" für die Konstruktion von Einstellungen nicht um Meinungsäußerungen von Befragungspersonen geht, bei denen man sich für ihren Inhalt und ggf. Wahrheitsanspruch interessiert. Bei den „sprachlichen Reaktionen", die man für Einstellungskonstruktionen gewinnen möchte, geht es ausdrücklich nicht darum, „was Befragte für wahr oder falsch halten". Bortz und Döring sagen zum Beispiel:

> „Für die Ermittlung von Einstellungen sind Itemformulierungen ungeeignet, mit denen wahre Sachverhalte dargestellt werden (Beispiel: 'Eine schlechte berufliche Qualifikation erhöht das Risiko für Arbeitslosigkeit.' Eine Zustimmung zu diesem Item würde keine Meinung, sondern allenfalls Fachkenntnisse über die Zusammenhänge von beruflicher Qualifikation und Arbeitslosigkeit signalisieren. Für eine Einstellungsmessung besser geeignet wäre z.B. die Formulierung: 'Eine schlechte berufliche Qualifikation sollte das Risiko für Arbeitslosigkeit erhöhen')." (Bortz und Döring 1995, S. 233)

7. Ein weiteres Beispiel soll den Gegensatz illustrieren. Es handelt sich um die Skala D04 aus dem ZUMA-Handbuch sozialwissenschaftlicher Skalen (Allmendinger, Schmidt und Wegener 1983). In diesem Beispiel geht es um einen „Test", mit dem festgestellt werden soll, wie Menschen Attribut- und insbesondere Ursachenzuschreibungen vornehmen. Die Details der Fragestellung brauchen hier nicht diskutiert zu werden, bemerkenswert ist jedoch, wie die „Testanweisung" aussieht, mit der „sprachliche Reaktionen"

---

perience and express evaluations, likes and dislikes, for social objects and concepts. Such likes and dislikes are called attitudes, and they may be measured by means of a number of methods available for scaling subjective experience." Herkner 1991, S. 17, schreibt: Die Sozialpsychologie „versucht zu erklären, wie das Erleben (Meinungen, Gefühle, Absichten usw.) und Verhalten von Personen durch andere Personen beeinflußt wird." Wer eine Meinung äußert, liefert also dadurch die Beschreibung eines „psychischen Erlebnisses"? Wenn man, wie zum Beispiel Feger und Graumann 1983, S. 76, der Auffassung ist, daß „Erleben und Verhalten die zentralen Begriffe sind, die den Gegenstand der Psychologie umschreiben", dann ist es jedenfalls fragwürdig, ob Meinungsäußerungen dem psychologischen „Gegenstandsbereich" subsummiert werden können.

[19] Einen lesenswerten Einstieg in die Diskussion von Zusammenhängen zwischen „Denken" und „Fühlen" findet man bei Zajonc 1980.

ermittelt werden sollen:

„*Testanweisung:* Hier finden Sie eine Reihe bunt zusammengewürfelter Feststellungen, die für jeden von uns wichtig sind. Immer zwei Feststellungen gehören zusammen (A und B). Bitte immer erst beide lesen, dann eine davon ankreuzen. Und zwar die, die Ihrer Meinung nach eher auf Sie zutrifft. Da uns Ihre persönliche Meinung interessiert, gibt es keine richtigen und falschen Antworten wie in einer Prüfung. Wenn mal beide Feststellungen nicht richtig hinhauen oder auf Sie nicht ganz passen sollten, versuchen Sie sich reinzudenken. Dann kreuzen Sie einfach das an, was noch am ehesten auf Sie zutrifft. Nur wenn es gar nicht anders geht, in Notfällen sozusagen, können Sie Ihr Kreuz in das kleine Kästchen C darunter setzen ('keine Meinung'). Wenn jemand dort sein Kreuzchen macht, sind wir natürlich so schlau wie vorher. Wir wissen dann ja nicht, welche Meinung Sie zu den Feststellungen haben, so daß die ganze Mühe umsonst sein kann. Bitte also, wenn möglich, A oder B ankreuzen."

Es geht also ausdrücklich nicht darum, eine möglichst *begründete* Meinung zu entwickeln und im Zweifelsfall zu sagen, daß man keine vernünftige Meinung äußern kann. Im Gegenteil, die „Testpersonen" werden dazu aufgefordert, *beliebige* Meinungen zu produzieren.[20]

8. Das meistens verwendete Mittel, um „differenzierte" Stellungnahmen zu provozieren, sind sog. Rating-Skalen. Zur Illustration sei ein weiteres Beispiel aus dem ZUMA-Handbuch sozialwissenschaftlicher Skalen angeführt. Es handelt sich um eine von K. U. Mayer und W. Müller entwickelte Skala bzw. in der Terminologie des Handbuchs um einen „Test" (A03) zur „Wahrnehmung sozialer Ungleichheit". Es gibt insgesamt acht „Statements", eines heißt zum Beispiel so:

„Die Bundesrepublik ist eine offene Gesellschaft. Was man im Leben erreicht, hängt nicht mehr vom Elternhaus ab, aus dem man kommt, sondern von den Fähigkeiten, die man hat, und der Bildung, die man erwirbt."

Die Befragungspersonen haben dann die Möglichkeit, eine von vier Angaben zu machen:

Stimme voll zu	( )
Stimme eher zu	( )
Stimme eher nicht zu	( )
Stimme überhaupt nicht zu	( )

Die Problematik ist offensichtlich. Es wird ein in allen Aspekten seiner Formulierung unpräzises und fragwürdiges „Statement" vorgegeben, gleichwohl sollen die Befragungspersonen eine differenzierte Antwort geben. In

[20] Dies ist übrigens kein abwegiges Beispiel, sondern entspricht Empfehlungen in der Lehrbuchliteratur, etwa von Bailey 1982, S. 120f: „Opinion questions are especially difficult. The respondent often does not have an opinion because he or she has never thought about the topic. He or she is concerned about appearing stupid, and must be reassured that there is no right or wrong answer."

einem Diskurs, der sich mit dem sachlichen Thema beschäftigte, das sich hinter dem „Statement" andeutet, würde vermutlich niemand eine solche Vorgehensweise akzeptieren, sondern man würde Meinungen zum Thema entwickeln und wechselseitig hinterfragen. Bei diesem Test wird jedoch den Befragungspersonen gar keine Möglichkeit gegeben, sich zum Thema zu äußern, sondern sie sollen sich zu einem vorgegebenen und schon in seiner Formulierung fragwürdigen „Statement" äußern. — Aber so scheint es nur. Tatsächlich wird von den Befragungspersonen gar keine Äußerung zu dem sachlichen Thema verlangt, um das es möglicherweise gehen könnte. Diese Möglichkeit wird schon durch die Form der Fragestellung ausgeschlossen. Denn zu einem fragwürdigen „Statement" kann man sich nicht dadurch sinnvoll äußern, daß man einen „Grad der Zustimmung" angibt. Erwartet wird tatsächlich etwas anderes: daß die Befragungspersonen eine quantifizierende Aussage über einen „psychischen Sachverhalt" machen, der sich in ihrem Kopf verbirgt.[21] Das entspricht zwar äußerlich der Konzeption der Psychophysik, wie sie oben angedeutet worden ist, macht in diesem Fall jedoch keinen Sinn. Oder möchte man herausfinden, was Befragungspersonen bei der Vorlage eines „Statements" *empfinden*? Aber angenommen, daß es tatsächlich darum gehen sollte, Empfindungen zu ermitteln. Wie könnte oder sollte sich eine Befragungsperson verhalten, wenn sie versuchen würde, eine solche Aufforderung ernst zu nehmen?

9. Wie in der Methodenliteratur ausgeführt wird, müssen zur „Einstellungsmessung" nicht unbedingt „Statements" verwendet werden. Es kann sich um beliebige „Stimuli" handeln, mit denen Befragungspersonen konfrontiert werden, um „Reaktionen" zu provozieren, die Rückschlüsse auf ihre Einstellungen erlauben sollen. Einige Autoren sprechen deshalb ganz pauschal von „Einstellungsobjekten"; zum Beispiel Dunn-Rankin in einem Buch über Skalierungsmethoden:

„Scaling consists of ordering things in some meaningful way such as labeling the units on a thermometer. [...] While the objects in the previous case are mercury heights that vary under temperature changes, scaling in the social sciences deals with *psychological* objects. *Psychological objects* can be tangible, such as chairs and postcards, but they can also be almost anything that is perceived by the senses and that results in some cognitive affect. Psychological objects can be colors, words, tones, and sentences as well as houses, gold stars, and movie stars. Psychological objects are most often sentences or statements, such as 'There will always be wars' or 'I hate war.' With young children the objects are often pictures. In marketing analysis, psychological objects are the products of industry: cars, soap, TV's, and toothpaste." (Dunn-Rankin 1983, S. 3f)

Fassen wir kurz zusammen: Es sollte deutlich geworden sein, daß es bei der Konstruktion von Einstellungen gar nicht um Meinungsäußerungen von

[21] In einer Formulierung von H. Feger 1982, S. 220: „All procedures of direct attitude measurement can be understood as attempts to project structural assumptions into self-descriptive statements of persons whose attitudes are about to be measured."

Menschen geht, vielmehr darum (in einer Formulierung von Hamblin 1974, S. 61), „how people perceive and react to various aspects of their physical and social environments.“ Die Befragungspersonen werden also getäuscht. Das Interview, mit dem sie scheinbar um eine Meinungsäußerung gebeten werden, soll gar nicht dazu dienen, ihre Meinung zu ermitteln; vielmehr soll es dazu dienen, Aussagen über ihre „psychische Verfassung“ zu konstruieren. Aber welche Erkenntnisinteressen könnten eine solche Tätigkeit rechtfertigen und warum sollten sich Menschen dafür als „Versuchspersonen“ zur Verfügung stellen?

Kapitel 4

Varianten der Quantifizierung

Statistische Sozialforschung wird oft als „quantitative Sozialforschung“ bezeichnet. Das ist zwar nicht falsch, sollte aber differenziert werden. In diesem Kapitel beginnen wir mit einer Diskussion der Frage, in welchen verschiedenen Weisen von ‘Quantifizierung’ gesprochen werden kann. Mit einer spezifischen Variante, sog. Skalierungsverfahren, werden wir uns in späteren Kapiteln noch ausführlicher beschäftigen.

4.1 Numerische Repräsentationen

1. Es ist üblich, für Merkmalsräume statistischer Variablen numerische Repräsentationen zu verwenden. Man denke an eine statistische Variable

$$X : \Omega \longrightarrow \tilde{\mathcal{X}}$$

die jedem Mitglied von Ω einen Merkmalswert in einem Merkmalsraum $\tilde{\mathcal{X}}$ zuordnet. Den Begriff ‘Merkmalsraum’ haben wir bisher (in Abschnitt 1.2) einfach als eine Menge von Attributen (Merkmalsausprägungen) eingeführt, die verwendet werden können, um die Mitglieder von Ω bzw. Aspekte von Situationen, in denen sie sich befinden, zu charakterisieren. Es sind also zunächst Attribute, keine Zahlen. Wenn zum Beispiel Ω eine Menge von Menschen ist, könnte man einen Merkmalsraum

$$\tilde{\mathcal{X}} := \{\text{‘männlich’}, \text{‘weiblich’}\}$$

verwenden, um mit Hilfe einer Variablen X das Geschlecht der Mitglieder von Ω zu erfassen. Von einer *numerischen Repräsentation* für den Merkmalsraum einer statistischen Variablen wird gesprochen, wenn man jedes durch ihn definierte Attribut durch eine jeweils spezifische Zahl repräsentiert. Zum Beispiel könnte man das Attribut ‘männlich’ durch die Zahl 0, das Attribut ‘weiblich’ durch die Zahl 1 repräsentieren. Dann gelangt man zu einem *numerischen Merkmalsraum*, der aus den beiden Zahlen 0 und 1 besteht, für die jeweils eine bestimmte Bedeutung vereinbart worden ist.

2. In der Statistik wird normalerweise vorausgesetzt, daß für Merkmalsräume statistischer Variablen eine numerische Repräsentation bereits gegeben ist.[1] In den folgenden Überlegungen unterscheiden wir jedoch zunächst

[1] Die Statistik hat es insoweit nur mit Zahlen zu tun. Es ist jedoch klar, daß man auf die Bedeutung dieser Zahlen Bezug nehmen muß, wenn man schließlich zu inhaltlich interpretierbaren Aussagen kommen will. In unserem Beispiel möchte man nicht wissen, wieviele Nullen und wieviele Einsen es in der Gesamtheit Ω gibt, sondern wieviele Männer und wieviele Frauen.

zwischen Merkmalsräumen und ihren numerischen Repräsentationen und verwenden die Schreibweise

$$\phi : \tilde{\mathcal{X}} \longrightarrow \mathbf{R}$$

um explizit einen Zusammenhang herzustellen. Hierbei bezeichnet $\mathbf{R}$ die Menge der reellen Zahlen und $\tilde{\mathcal{X}}$ ist der Merkmalsraum einer statistischen Variablen, bei dessen Elementen es sich um Merkmalswerte (Attribute) handelt. Die Abbildung ϕ ordnet jedem Merkmalswert $\tilde{x} \in \tilde{\mathcal{X}}$ eine bestimmte Zahl $\phi(\tilde{x}) \in \mathbf{R}$ zu, die den Merkmalswert numerisch repräsentiert. Wenn ϕ außerdem eine injektive Funktion ist, so daß gilt

$$\tilde{x} \neq \tilde{x}' \implies \phi(\tilde{x}) \neq \phi(\tilde{x}')$$

wird ϕ eine *numerische Repräsentation* für den Merkmalsraum $\tilde{\mathcal{X}}$ genannt. Offenbar kann man eine numerische Repräsentation ϕ verwenden, um aus einem Merkmalsraum $\tilde{\mathcal{X}}$ einen numerischen Merkmalsraum

$$\phi(\tilde{\mathcal{X}}) = \{\phi(\tilde{x}) \,|\, \tilde{x} \in \tilde{\mathcal{X}}\} \subset \mathbf{R}$$

zu bilden.

3. Wichtig ist, daß eine Verwendung numerischer Repräsentationen für Merkmalsräume noch keinerlei Quantifizierung impliziert. Man denke noch einmal an das Beispiel. Dadurch, daß das Attribut 'männlich' durch die Zahl 0, das Attribut 'weiblich' durch die Zahl 1 repräsentiert wird, hat man natürlich das Merkmal 'Geschlecht' nicht quantifiziert. Die Frage der Quantifizierung stellt sich zunächst an einer anderen Stelle: um welche Art von Attributen es sich handelt, die ein Merkmalsraum zusammenfassen soll. Hier gibt es eine wichtige Unterscheidung:

a) Man kann einen Merkmalsraum verwenden, um qualitative Unterschiede zwischen Objekten zu erfassen, und

b) man kann einen Merkmalsraum verwenden, um Objekte durch quantifizierbare Merkmale zu charakterisieren.

4. Quantifizierbare Merkmale werden auch als *Größen* bezeichnet. Wir orientieren uns an folgender Definition, die von Hermann von Helmholtz (1887, S. 96) gegeben wurde:

„Objecte oder Attribute von Objecten, die mit ähnlichen verglichen den Unterschied des grösser, gleich oder kleiner zulassen, nennen wir *Grössen*. Können wir sie durch eine benannte Zahl ausdrücken, so nennen wir diese den *Werth* der Grösse, das Verfahren, wodurch wir die benannte Zahl finden, *Messung*."

Die Unterscheidung verläuft zwischen Merkmalsräumen, deren Merkmalswerte in eine lineare Ordnung gebracht werden können, und Merkmalsräumen, bei denen das nicht der Fall ist. Im ersten Fall sprechen wir von

quantitativen Merkmalsräumen, die sich auf *Größenbegriffe* beziehen, und nennen die möglichen Merkmalswerte die *Werte einer Größe*; im zweiten Fall sprechen wir von *qualitativen Merkmalsräumen*.[2]

5. Auch wenn $\tilde{\mathcal{X}}$ ein quantitativer Merkmalsraum ist, handelt es sich bei seinen Elementen zunächst um Merkmalswerte und nicht um Zahlen. Die Annahme, daß es für $\tilde{\mathcal{X}}$ eine *lineare Ordnung* gibt, besagt vielmehr, daß es eine Relation $\prec$ gibt, so daß folgende Bedingungen erfüllt sind:

a) Für alle $\tilde{x}, \tilde{x}' \in \tilde{\mathcal{X}}$ gilt: $\tilde{x} = \tilde{x}'$ oder $\tilde{x} \prec \tilde{x}'$ ($\tilde{x}'$ ist größer als $\tilde{x}$) oder $\tilde{x}' \prec \tilde{x}$ ($\tilde{x}$ ist größer als $\tilde{x}'$).

b) Die Relation ist transitiv, d.h. für alle $\tilde{x}, \tilde{x}', \tilde{x}'' \in \tilde{\mathcal{X}}$ gilt: $\tilde{x} \prec \tilde{x}'$ und $\tilde{x}' \prec \tilde{x}'' \implies \tilde{x} \prec \tilde{x}''$.

Es ist jedoch klar, daß man sich oft eine numerische Repräsentation ϕ für den Merkmalsraum $\tilde{\mathcal{X}}$ verschaffen kann, so daß gilt:

$$\tilde{x} \prec \tilde{x}' \iff \phi(\tilde{x}) < \phi(\tilde{x}') \qquad (4.1.1)$$

Wenn von der numerischen Repräsentation eines quantitativen Merkmalsraums gesprochen wird, soll damit stets auch gemeint sein, daß sie die Bedingung (4.1.1) erfüllt.

6. Sowohl für qualitative als auch für quantitative Merkmalsräume kann man im allgemeinen beliebig viele unterschiedliche numerische Repräsentationen bilden; auch dann, wenn man die Bedingung (4.1.1) berücksichtigt, was natürlich nur bei quantitativen Merkmalsräumen sinnvoll ist. Dies ist wichtig, weil statistische Aussagen in vielen Fällen davon abhängen, welche numerische Repräsentation für die Merkmalsräume statistischer Variablen verwendet werden. Das heißt indessen nur, daß die jeweils verwendete numerische Repräsentation explizit angegeben werden muß. Wie bereits bemerkt wurde, geschieht dies in der Statistik üblicherweise dadurch, daß man von vornherein eine jeweils bestimmte numerische Repräsentation als gegeben annimmt, also von numerischen Merkmalsräumen ausgeht. Ist $\tilde{\mathcal{X}}$ der Merkmalsraum einer Variablen und ϕ eine numerische Repräsentation, werden also $\tilde{\mathcal{X}}$ und $\phi(\tilde{\mathcal{X}})$ formal identifiziert, so daß man bereits $\tilde{\mathcal{X}}$ als eine Menge von Zahlen auffassen kann, für die separat eine Bedeutung vereinbart ist, die es erlaubt, die Zahlen als Merkmalswerte zu interpretieren. Wir werden im weiteren dieser Konvention folgen und annehmen, daß Merkmalsräume stets als Teilmengen der reellen Zahlen gegeben sind.

[2] Es sei hier schon erwähnt, daß viele Autoren im Anschluß an Stevens (1946) bei quantitativen Merkmalsräumen unterschiedliche Arten von „Skalenniveaus" unterscheiden. Diese Unterscheidungen sind jedoch zunächst nicht erforderlich, und wir werden uns mit ihnen erst in Kapitel 9 bei der Diskussion von Metriken beschäftigen.

7. Von *Quantifizierung* kann man zunächst dann sprechen, wenn zur Charakterisierung von Objekten bzw. von Situationen, in denen sie sich befinden, quantitative Merkmale verwendet werden. Aber wie kommen diese Quantifizierungen zustande? Es sind natürlich Sozialwissenschaftler, die die Daten für ihre Variablen erzeugen und in diesem Zusammenhang sowohl qualitative als auch quantitative Merkmale erfassen. In beiden Fällen wird jedoch zunächst an Kenntnisse angeknüpft, die in der gesellschaftlichen Praxis üblich sind. Dies gilt jedenfalls dann, wenn Daten durch Interviews gewonnen werden; denn dann ist die Datengewinnung darauf angewiesen, daß die befragten Personen verstehen, welche Merkmale erfaßt werden sollen. Insbesondere wenn ein quantitatives Merkmal – zum Beispiel Alter oder Höhe des Einkommens – erfaßt werden soll, müssen die Befragten mit der entsprechenden Quantifizierung vertraut sein. Insoweit kann man also nicht davon sprechen, daß von Sozialwissenschaftlern eine Quantifizierung von Merkmalen oder Sachverhalten vorgenommen wird, die in der gesellschaftlichen Praxis, auf die sie sich beziehen, nicht schon üblich ist.

8. Soweit es sich um qualitative Merkmale handelt, werden zur numerischen Repräsentation natürliche Zahlen verwendet, wobei entweder mit der Zahl 0 oder mit der Zahl 1 begonnen wird. Oft ist es praktisch, mit 0 zu beginnen und einen Merkmalsraum, der m Attribute zusammenfaßt, durch die Zahlenmenge $\{0, 1, \ldots, m-1\}$ zu repräsentieren. Diese Konvention wurde bereits in dem oben angeführten Beispiel zur Kodierung des Merkmalsraums „Geschlecht“ verwendet.

9. Bei quantitativen Merkmalen unterscheiden wir zunächst: Zählgrößen, Meßgrößen und monetäre Größen.

a) Am einfachsten zu verstehen sind *Zählgrößen*, die dadurch zustande kommen, daß bei einer Menge zusammengehöriger Dinge gezählt wird, wieviele es gibt. Zum Beispiel: Anzahl der Personen in einem Haushalt, Anzahl der Beschäftigten in einem Unternehmen, Anzahl der Arztbesuche während einer gewissen Zeitspanne. Zur numerischen Repräsentation von Zählgrößen ist es üblich, die natürlichen Zahlen zu verwenden, denn sie verdanken sich ja zunächst diesem Zweck, die Anzahl einer Menge von Dingen anzugeben. Dann ist auch unmittelbar verständlich, welche Bedeutung die für die numerische Repräsentation verwendeten Zahlen haben.

b) Außer Zählgrößen werden auch quantitative Merkmale verwendet, die durch Meßverfahren definiert sind. Wir sprechen dann von *Meßgrößen*. Beispiele sind Gewichte, Längen und Zeitdauern; insbesondere im Bereich technischer Geräte und bei medizinischen Untersuchungen gibt es noch zahlreiche weitere Meßgrößen. Wiederum kann man eine Unterscheidung treffen zwischen Meßgrößen, die in der gesellschaftlichen Praxis allgemein üblich und verständlich sind, und Meßgrößen, die von Wissenschaftlern, Technikern und Medizinern zunächst für spezifische Zwecke erfunden werden. Die Grenzen sind allerdings fließend und unterliegen einem historischen Wandel. Sicherlich kann man jedoch sagen, daß Gewichte, Längen und Zeitdauern allgemein übliche Meßgrößen sind; und das heißt auch, daß nicht nur die Bedeutung dieser Meßgrößen bekannt ist, sondern daß entsprechende Meßgeräte allgemein verfügbar sind und verwendet werden.

c) Schließlich gibt es noch eine dritte Art quantitativer Größen: *monetäre Größen*, insbesondere Preisgrößen und verschiedene Arten von Einkommensgrößen. Sie müssen von Meßgrößen unterschieden werden, denn sie kommen nicht durch Meßverfahren zustande, sondern werden festgesetzt oder ausgehandelt. Ökonomen sprechen von Preisbildungsprozessen, die natürlich im einzelnen sehr unterschiedlich beschaffen sein können.

10. Außer diesen elementaren Arten von Größen gibt es eine Vielzahl weiterer. Meistens entstehen sie dadurch, daß ausgehend von elementaren Größen neue Größenbegriffe gebildet werden; z.B. Geschwindigkeit, Anteil der Mietausgaben am Haushaltseinkommen, Anzahl der Arztbesuche pro Jahr. Auch mit Methoden sozialwissenschaftlicher Datenkonstruktion wird oft das Ziel verfolgt, neue Größenbegriffe zu begründen. Ob man zu sinnvollen Größenbegriffen gelangt, kann sicherlich nur in jedem Einzelfall geprüft werden. Ein generelles Problem besteht jedoch darin, daß bereits die Unterscheidung zwischen qualitativen und quantitativen Merkmalsräumen nicht objektiv vorgegeben ist. Die in einem Merkmalsraum zusammengefaßten Merkmalswerte können immer *irgendwie* in eine lineare Ordnung gebracht werden, so daß es für die Quantifizierung von Merkmalsräumen keine natürlichen Grenzen gibt. Infolgedessen muß in jedem Einzelfall geprüft werden, ob man zu einem sinnvollen Größenbegriff gelangt.

4.2 Statistische Quantifizierungen

1. Das Wort ‘Quantifizierung’ bezieht sich zunächst darauf, ob Objekte oder Situationen, in denen sie sich befinden, durch quantifizierbare Eigenschaften charakterisiert werden. Je nachdem erhält man Daten als Werte in qualitativen oder quantitativen Merkmalsräumen. Im Anschluß daran kann man allerdings fragen, was mit solchen Daten weiter geschieht und in welcher Weise dabei neue Quantifizierungen vorgenommen werden. Dafür erscheint es uns wichtig, vor allem eine Unterscheidung zwischen statistischen Quantifizierungen und Skalierungsverfahren zu bedenken. In diesem Abschnitt soll besprochen werden, wie bei statistischen Verwendungen von Daten von einer Quantifizierung gesprochen werden kann.

2. In Abschnitt 4.1 wurde besprochen, daß zur Repräsentation von Merkmalsräumen statistischer Variablen stets Zahlen verwendet werden. Aber es wurde noch nicht überlegt, warum man das macht. Ein Grund läßt sich jedoch leicht angeben: Man möchte mit diesen Variablen rechnen können. Betrachten wir als Beispiel eine Variable

$$X : \Omega \longrightarrow \tilde{\mathcal{X}} := \{1, 2, 3, \ldots\}$$

die die Haushaltsgröße (Anzahl der Mitglieder) der in Ω zusammengefaßten Haushalte angibt. In diesem Fall liefert bereits die Definition des Merkmalsraums eine sinnvolle numerische Repräsentation, und man kann sie verwenden, um zum Beispiel die durchschnittliche Haushaltsgröße zu berechnen, also

$$\mathrm{M}(X) = \frac{1}{n} \sum_{\omega \in \Omega} X(\omega)$$

wobei angenommen wird, daß es n Haushalte in Ω gibt.

3. Sobald man anfängt, mit Variablen zu rechnen, stellt sich allerdings die Frage, ob das Rechnen einen Sinn ergibt. Aber dafür gibt es – wie bei allen Sinnfragen – keine harten Kriterien. Vielmehr ist es die Aufgabe desjenigen, der ein Rechenverfahren vorschlägt oder anwendet, zu zeigen, daß es zu sinnvollen (interessanten, wissenswerten, ...) Rechenergebnissen führen kann, und ggf. Bedingungen anzugeben, die dafür erforderlich sind. Zum Beispiel kann man sich überlegen, daß einfache Durchschnittsbildungen zwar immer errechnet werden können, aber nicht immer sinnvoll sind. Oder angenommen, um noch ein weiteres Beispiel anzuführen, daß man sich auf zwei Variablen beziehen kann: Miete und Einkommen. Dann erscheint es durchaus sinnvoll, den Anteil der Miete am Einkommen zu berechnen. Aber die einfache Rechenoperation, Miete und Einkommen zu addieren, würde kein sinnvoll interpretierbares Resultat liefern.

4. Beim Nachdenken über solche Sinnfragen ist es nützlich, sich zu vergegenwärtigen, daß statistische Variablen *Abbildungen* einer Objektmenge in einen Merkmalsraum sind. Man kann infolgedessen drei Gesichtspunkte unterscheiden.

a) Man kann sich auf den Merkmalsraum einer statistischen Variablen beziehen, also mit den numerisch repräsentierten Merkmalswerten rechnen, ohne dabei an eine bestimmte Menge von Objekten zu denken, denen solche Merkmalswerte zukommen könnten.

b) Man kann sich auf die Werte einer statistischen Variablen beziehen und sie als eine „Menge von Daten“ auffassen.[3] Einer statistischen Variablen entspricht dann ein Datenvektor.

[3] Spricht man hier von einer „Menge von Daten“, sollte man daran denken, daß es sich nicht um eine Menge im gewöhnlichen Sinn des Wortes handelt; denn bei verschiedenen Objekten kann durchaus der gleiche Merkmalswert auftreten.

c) Man kann sich darauf beziehen, daß eine statistische Variable eine Abbildung ist, also begrifflich etwas anderes als ein Datenvektor oder eine „Menge von Daten“. Diese Betrachtungsweise ist grundlegend, wenn neue Variablen – also zunächst ein neuer Merkmalsraum und dann eine neue Abbildung – konstruiert werden sollen.

5. Wenn mit Variablen im Rahmen *statistischer* Überlegungen gerechnet wird, kommt noch ein weiterer Gesichtspunkt hinzu. Der Grundgedanke ist, statistische Variablen durch ihre Verteilungen zu charakterisieren. Das Ziel besteht darin, diese Verteilungen zu ermitteln, darzustellen und Zusammenhänge zwischen Verteilungen zu untersuchen. In einer kurzen Formulierung: Statistische Aussagen sind stets Aussagen über Verteilungen statistischer Variablen, also Aussagen über Gesamtheiten von Objekten, nicht über ihre einzelnen Mitglieder. Sich dies zu vergegenwärtigen ist wichtig, um zu verstehen, wie in der Statistik mit Variablen gerechnet wird. Es wird dann in erster Linie mit den Verteilungen von Variablen gerechnet. Denken wir noch einmal an das oben angeführte Beispiel, in dem der Mittelwert von Haushaltsgrößen berechnet wurde. Angenommen, man hat die Zahl 2.6 gefunden. Wenn man dann versucht, den Sinn der Aussage „Die durchschnittliche Haushaltsgröße beträgt 2.6“ zu erklären, muß man sich auf eine Verteilung von Haushaltsgrößen in einer Gesamtheit von Haushalten beziehen. Die Aussage bezieht sich nicht unmittelbar auf irgendwelche bestimmten Objekte (Haushalte), auch nicht auf irgendein neu erfundenes Objekt (einen „durchschnittlichen Haushalt“), sondern sie bezieht sich auf eine statistische Verteilung, in diesem Fall auf eine Verteilung von Haushaltsgrößen bei einer Gesamtheit von Haushalten.

6. Ein möglichst genaues Verständnis statistischer Variablen und ihrer Verwendung für statistische Aussagen ist wichtig, um zu verstehen, wie sich solche Aussagen auf empirische Realität beziehen. Wie schon betont wurde, sind statistische Variablen keine Objekte, die es irgendwo in der empirischen Realität gibt. Sondern es sind sprachliche Konstruktionen, um über Objekte bzw. Situationen, in denen sie sich befinden, zu reden. Also stellt sich auch die Frage, wie es zu den Sachverhalten kommt, die durch statistische Aussagen dargestellt werden, und wie sie sich verändern. Wie kommt es z.B. zu einer bestimmten durchschnittlichen Haushaltsgröße und wie kommt es zu Veränderungen? Offenbar ist es erforderlich, zunächst an die Verteilung der Haushaltsgrößen zu denken, die durch ihren Mittelwert nur sehr abstrakt charakterisiert wird. Um dann aber weiterzukommen, muß man an die Prozesse denken, durch die sich Haushalte bilden, verändern und auflösen; und es stellt sich die Frage, ob und ggf. wie mit statistischen Methoden Einsichten in solche Prozesse gewonnen werden können. Auf diese Frage soll hier indessen nur hingewiesen werden, da sie nicht mehr zu unserem gegenwärtigen Thema, den Methoden sozialwissenschaftlicher Datenkonstruktion, gehört.

7. Die genaue Klärung des Begriffs einer statistischen Verteilung erfolgt in Einführungen in die Statistik. Hier genügt es deshalb, an einige grundlegende Aspekte zu erinnern. Um möglichst einfache Begriffserläuterungen geben zu können, beziehen wir uns auf eine Variable

$$X : \Omega \longrightarrow \tilde{\mathcal{X}} := \{ \tilde{x}_1, \ldots, \tilde{x}_m \}$$

bei der es m unterschiedliche Merkmalswerte geben kann. Hat man die Werte dieser Variablen bei den Mitgliedern von Ω festgestellt, kann man für jede Teilmenge des Merkmalsraums ausrechnen, wieviele Mitglieder von Ω einen Merkmalswert aufweisen, der in diese Teilmenge fällt. Wenn man dann von diesen absoluten zu relativen Häufigkeiten übergeht, kommt man zum Begriff einer statistischen Verteilung. Eine Definition kann folgendermaßen gegeben werden. Es sei zunächst $\tilde{X}$ eine beliebige Teilmenge von $\tilde{\mathcal{X}}$ (wir nennen solche Teilmengen *Merkmalsmengen*), und es sei n die Anzahl der Mitglieder von Ω. Dann ist die relative Häufigkeit von $\tilde{X}$ in Ω durch

$$\mathrm{P}[X](\tilde{X}) := \frac{1}{n} \left| \{ \omega \in \Omega \,|\, X(\omega) \in \tilde{X} \} \right|$$

definiert.[4] Betrachtet man nun $\mathrm{P}[X]$ als eine Funktion

$$\tilde{X} \longrightarrow \mathrm{P}[X](\tilde{X})$$

die jeder Teilmenge $\tilde{X} \subseteq \tilde{\mathcal{X}}$ ihre relative Häufigkeit in Ω zuordnet, wird diese Funktion eine *statistische Verteilung* oder auch eine *Häufigkeitsfunktion* (der Variablen X) genannt.

8. Man kann sich leicht klarmachen, daß Häufigkeitsfunktionen additiv sind, d.h. wenn $\tilde{X}_1$ und $\tilde{X}_2$ zwei disjunkte Teilmengen von $\tilde{\mathcal{X}}$ sind,[5] gilt:

$$\mathrm{P}[X](\tilde{X}_1 \cup \tilde{X}_2) = \mathrm{P}[X](\tilde{X}_1) + \mathrm{P}[X](\tilde{X}_2)$$

Um die Verteilung einer statistischen Variablen darzustellen, genügt es deshalb, die relative Häufigkeit für jeden möglichen Merkmalswert zu ermitteln. Zur Illustration betrachten wir eine Gesamtheit, die aus 10 Personen besteht, und eine Variable X, die deren Geschlecht erfaßt. Der Merkmalsraum ist $\{0, 1\}$, 0 steht für Männer, 1 für Frauen; und man hat z.B. folgende Werte gefunden: $0, 0, 1, 0, 1, 1, 0, 0, 0, 1$. Die Verteilung der Variablen ist dann durch

$$\mathrm{P}[X](\{0\}) = 6/10 \quad \text{und} \quad \mathrm{P}[X](\{1\}) = 4/10$$

gegeben. Das bedeutet: 60 % der Mitglieder dieser Gesamtheit sind Männer, 40 % sind Frauen.

[4] Wenn M eine endliche Menge ist, bedeutet $|M|$ die Anzahl ihrer Elemente.

[5] Zwei Mengen heißen *disjunkt*, wenn sie keine gemeinsamen Elemente haben.

9. Der Begriff einer statistischen Verteilung bildet den Ausgangspunkt für die meisten statistischen Verfahren und Modellbildungen, mit deren Hilfe Verteilungen dargestellt und charakterisiert werden können. Die Vielzahl der statistischen Verfahren verdankt sich vor allem der Tatsache, daß der Verteilungsbegriff für mehrdimensionale Variablen verallgemeinert werden kann. Davon ausgehend kann man nämlich Zusammenhänge zwischen Variablen durch Eigenschaften ihrer gemeinsamen Verteilung charakterisieren. Hier geht es jedoch nicht um eine Einführung in diese Verfahren. Auf den Begriff einer statistischen Verteilung wird hier hingewiesen, um deutlich zu machen, in welcher Weise in diesem Zusammenhang von 'Quantifizierung' gesprochen werden kann; und um später eine wichtige Unterscheidung zwischen statistischen Verfahren, die am Begriff der Verteilung ansetzen, und sog. Skalierungsverfahren deutlich machen zu können.

10. Es ist evident, daß mit der Bildung statistischer Verteilungen eine Quantifizierung vorgenommen wird. Man sollte sich jedoch klarmachen, wie und was hierbei quantifiziert wird. Die Quantifizierung bezieht sich auf das Vorkommen von Merkmalen in einer Gesamtheit; Eigenschaften der Gesamtheit werden quantifiziert, nicht Eigenschaften ihrer einzelnen Mitglieder. Verteilungsaussagen sind stets Aussagen über Gesamtheiten, keine Aussagen über die individuellen Mitglieder der Gesamtheit. — Zum Verständnis dieser Feststellung sollte beachtet werden, daß unsere Sprache zweideutig ist, wenn im Plural über Individuen oder ganz allgemein über Objekte gesprochen wird. Eine Aussage der Art „Für die Individuen aus Ω gilt ...“ kann bedeuten:

(a) Für jedes Individuum aus Ω gilt ...; oder

(b) Für die Gesamtheit der Individuen aus Ω, also für Ω gilt ...

Aussagen über statistische Verteilungen sind stets Aussagen des Typs (b), nicht Aussagen des Typs (a). Will man empirisches Wissen über statistische Verteilungen bilden, benötigt man zwar zunächst Informationen über die individuellen Objekte bzw. über die jeweils spezifischen Situationen, in denen sie sich befinden; in diesem Sinn ist statistische Datenerhebung individuenbezogen. Sobald man aber Merkmals*verteilungen* betrachtet, findet ein Perspektivenwechsel statt. Die Aufmerksamkeit richtet sich dann auf die Gesamtheit der Objekte bzw. Situationen, nicht mehr auf die individuellen Objekte, anhand derer die Informationen gewonnen worden sind.

11. Diese der statistischen Methode eigentümliche Abstraktion von individueller Identität wurde vom *International Statistical Institute* in einer „Declaration of Professional Ethics“ (1986, S. 238) so formuliert: „Statistical data are unconcerned with individual identities. They are collected to answer questions such as 'how many?' or 'what proportions?', not 'who?'. The identities and records of co-operating (or non-cooperating) subjects should therefore be kept confidential, whether or not confidentiality has been explicitly pledged.“ Die Unterscheidung verläuft zwischen Aussagen über Individuen und Aussagen über Gesamtheiten von Individuen. Statistische Aussagen beziehen sich, wie deutlich geworden sein sollte, stets auf Gesamtheiten.

4.3 Skalierungsverfahren

1. Allerdings gibt es bei dieser Unterscheidung gleichsam eine offene Flanke. Denn unsere Sprache erlaubt es, Individuen auch durch Merkmale von Gesamtheiten zu charakterisieren, deren Mitglieder sie sind. Insofern ist es möglich, statistische Verfahren auch zu verwenden, um Individuen zu charakterisieren.[6] Wir werden in späteren Kapiteln einige Beispiele kennenlernen. Es sollte jedoch betont werden, daß dann statistische Verfahren auf eine wesentlich unterschiedliche Weise verwendet werden. Um diesen Unterschied zu betonen, sprechen wir von statistischen Verfahren nur dann, wenn sich das Erkenntnisinteresse auf Verteilungsaussagen richtet. Dagegen sprechen wir von *Skalierungsverfahren*, wenn das Ziel darin liegt, die individuellen Mitglieder einer Gesamtheit zu charakterisieren.

2. Diese Unterscheidung ist insbesondere für unser Generalthema wichtig: Methoden der Datenkonstruktion. Sowohl statistische Methoden als auch Skalierungsverfahren beruhen auf einer vorgängigen Ermittlung individuell zurechenbarer Daten. Statistische Methoden nehmen diese Daten als Ausgangspunkt, um Merkmalsverteilungen und ihre Zusammenhänge zu untersuchen. Sie erzeugen keine neuen Daten, sondern liefern Bilder bzw. Modelle für vorgegebene Mengen von Daten bzw. die gesellschaftlichen Verhältnisse, auf die die Daten verweisen. Anders verhält es sich mit Skalierungsverfahren. Sie beginnen nicht nur mit individuellen Daten, sondern verlängern den Prozeß der Datenkonstruktion, indem sie darauf abzielen,

[6] Diese Möglichkeit spielte bereits bei der Erfindung und Entwicklung statistischer Verfahren eine wichtige Rolle. Exemplarisch sei auf Francis Galton 1889, S. 35ff, verwiesen. Galton betont zunächst, daß man nicht nur Durchschnittswerte, sondern statistische Verteilungen insgesamt betrachten sollte: „The knowledge of an average value is a meagre piece of information. [...] We require no more than a fairly just and comprehensive method of expressing the way in which each measurable quality is distributed among the members of any group, whether the group consists of brothers or of members of any particular social, local, or other body of persons, or whether it is co-extensive with an entire nation or race.“ Dann aber geht es so weiter: „A knowledge of the distribution of any quality enables us to ascertain the Rank that each man holds among his fellows, in respect to that quality. This is a valuable piece of knowledge in this struggling and competitive world, where success is to the foremost, and failure to the hindmost, irrespective of absolute efficiency. [...] When the distribution of any faculty has been ascertained, we can tell from the measurement, say of our child, how he ranks among other children in respect to that faculty, whether it be a physical gift, or one of health, or of intellect, or of morals. As the years go by, we may learn by the same means whether he is making his way towards the front, whether he just holds his place, or whether he is falling back towards the rear. Similarly as regards the position of our class, or of our nation, among other classes and other nations.“

neue Daten über Individuen zu konstruieren.[7]

3. Es sollte deshalb kritisiert werden, wenn in Methodenlehrbüchern „Datenerhebung“ von vornherein im Kontext von Skalierungsintentionen konzipiert wird, wie dies z.B. von Bortz (1984, S. 74) vorgeschlagen wird:

> „Der Datenerhebungsbegriff wird in diesem Kapitel sehr weit gefaßt. Die hier behandelte Datenerhebung dient generell dazu, die untersuchten Objekte bzgl. bestimmter Merkmale zu quantifizieren. Hierfür sind in den Sozialwissenschaften zuweilen recht komplizierte Prozeduren erforderlich, die aus den ursprünglich erhobenen ‘Rohdaten’ aggregierte Maßzahlen ableiten (z.B. eine gewichtete Summe oder eine Merkmalsskalierung nach der Paarvergleichsmethode). Am Ende dieser Prozeduren stehen Zahlen, die über Merkmalsausprägungen der untersuchten Objekte informieren.“

Diese Aussage sollte schon deshalb kritisiert werden, weil es sich in der Sozialforschung keineswegs immer darum handelt, quantifizierbare Sachverhalte zu erfassen. Vor allem sollte aber kritisiert werden, daß als Zweck der Datenerhebung die Skalierung von Individuen angegeben wird. Diese Intention wäre auch nur schwer mit der oben zitierten Aussage der „Declaration of Professional Ethics“ vereinbar.

4. Dennoch nimmt die Beschäftigung mit Skalierungsverfahren in der sozialwissenschaftlichen Methodenliteratur einen breiten Raum ein. Insbesondere gibt es einen engen Zusammenhang zwischen Skalierungsverfahren und Versuchen, spezifisch sozialwissenschaftliche Größenbegriffe zu begründen.[8] Wir werden darauf in späteren Kapiteln genauer eingehen. Vorläufig genügt eine grobe Unterscheidung von Fragestellungen, die im Zusammenhang mit dem Stichwort ‘Skalierungsverfahren’ verfolgt werden.[9]

[7] Sehr deutlich wird dies in einem Buch über Skalierungsmethoden von A. van der Ven (1977, S. 17f), in dem der Autor zwischen „Beobachtungen“, die man empirisch ermitteln kann, und „Daten“, die durch Skalierungsverfahren erzeugt werden, unterscheidet.

[8] Darauf wird besonders von Sozialwissenschaftlern Wert gelegt, in deren Verständnis eine „exakte“ Wissenschaft mehr oder weniger gleichbedeutend mit einer „quantifizierenden“ Wissenschaft ist. Z.B. schrieben E. K. Scheuch und D. Rüschemeyer in einem programmatischen Aufsatz über „Soziologie und Statistik“ 1958, S. 278f: „Die im Augenblick vielversprechendste Begegnung von Soziologie und theoretischer Statistik vollzieht sich bei der Konstruktion von Erhebungsinstrumenten, wogegen man bisher bei der Anwendung statistischer Verfahren vornehmlich an die Analyse des gesammelten Materials dachte. Dabei denken wir insbesondere an die Fortschritte auf dem Gebiet des ‘Messens’ sozialer Phänomene, nämlich beim Aufbau sogenannter ‘Skalen’. [...] Es scheint uns aber von grundsätzlicher Bedeutung zu sein, daß es hier – durch Verwendung statistischer Verfahren von nur mittlerem Schwierigkeitsgrad – gelungen ist, Daten zu quantifizieren, die man bisher ausschließlich qualitativ betrachtet hat. Man darf sogar noch weiter gehen und die Behauptung aufstellen, daß der Unterschied zwischen qualitativer und quantitativer Analyse überhaupt nur ein solcher der Betrachtung ist und sich nicht zwingend aus der Realität ergibt [...].“

[9] Es sei angemerkt, daß die Worte ‘Skalierung’ und ‘Skalierungsverfahren’ in der Literatur in vielen unterschiedlichen Bedeutungen verwendet werden; man vgl. z.B. Borg und Staufenbiel 1993, S. 212. Wir erläutern unseren Sprachgebrauch in Abschnitt 5.2.

a) Zunächst geht es um die Frage, wie man aus gegebenen Variablen neue Variablen (Indizes) konstruieren kann. Eine wichtige Teilfrage besteht darin, unter welchen Bedingungen solche Indizes als quantitative Variablen aufgefaßt werden können.
b) Eine weitere Frage bezieht sich darauf, wie man für qualitative Merkmalsräume numerische Repräsentationen finden kann, die (scheinbar) als eine Quantifizierung verstanden werden können.
c) Schließlich gibt es noch eine Vielzahl von Ansätzen, um empirische Variablen als Funktionen konstruierter („latenter") Variablen darstellbar zu machen. Oft spricht man dann von Skalierungsmodellen. Das wichtigste Beispiel ist die Faktorenanalyse.

Bei diesen Fragestellungen geht es stets – direkt oder indirekt – um Methoden der Datenkonstruktion. Darüberhinaus gibt es noch ein breites Spektrum an Methoden, die sich gewissermaßen in einem Grenzgebiet zwischen Statistik und Datenkonstruktion bewegen; exemplarisch sei auf die Korrespondenzanalyse und auf Methoden der multidimensionalen Skalierung und der Clusteranalyse hingewiesen. Darauf werden wir genauer in einem separaten Text eingehen, der sich mit Methoden zur Analyse und Darstellung relationaler Daten beschäftigt.

Kapitel 5

Konstruierte Variablen

Die meisten Methoden, die in diesem Text diskutiert werden, bestehen im Kern darin, daß aus gegebenen Variablen neue Variablen konstruiert werden. In diesem Kapitel beschäftigen wir uns mit dem begrifflichen Rahmen, mit dem Reden von Indizes und Skalen, mit einer Unterscheidung zwischen verteilungsabhängigen und -unabhängigen Konstruktionsverfahren und mit einigen Varianten der Standardisierung konstruierter Variablen.

5.1 Der begriffliche Rahmen

1. Als Ausgangspunkt kann folgende Unterscheidung dienen: Einerseits gibt es *empirische Variablen*, deren Werte sich direkt aus Beobachtungen, Interviews oder Meßverfahren ermitteln lassen; andererseits gibt es *konstruierte Variablen*, die als Funktionen empirischer (oder anderer bereits konstruierter) Variablen entstehen. Der Sinn der Unterscheidung ergibt sich durch die Frage, welche Bedeutung den Werten einer Variablen gegeben werden kann. Die Werte empirischer Variablen gewinnen ihre Bedeutung aus einem Prozeß der Datengewinnung und der dafür verwendeten Sprache. Die Bedeutung der Werte konstruierter Variablen muß demgegenüber durch das Konstruktionsverfahren begründet werden.

2. Ein etwas anderer Blick auf die Unterscheidung ergibt sich, wenn man darauf achtet, daß es gewissermaßen zwei Stufen in der Konstruktion von Variablen (und also auch von Daten) gibt. Die erste Stufe besteht darin, Variablen zu definieren, deren Werte man empirisch ermitteln kann. Die zweite Stufe besteht darin, aus den empirischen Variablen mit Hilfe von Rechenverfahren neue Variablen zu konstruieren. Die Bedeutung konstruierter Variablen kann also nicht unmittelbar im Hinblick auf eine empirisch erfaßbare Realität expliziert werden, sondern hängt auch von dem Rechenverfahren ab, das zur Konstruktion der neuen Variablen verwendet wird. Eine wichtige Frage ist deshalb, welchen Zwecken die Konstruktion neuer Variablen dienen soll. Hier wird wiederum die Unterscheidung zwischen statistischen Methoden (im engeren Sinn) und Skalierungsverfahren wichtig. Statistische Methoden zielen darauf ab, Verteilungen und Verteilungszusammenhänge zwischen Variablen zu ermitteln und darzustellen; konstruierte Variablen können gelegentlich ein Hilfsmittel bilden, um diesen Zweck zu erreichen. Bei Skalierungsverfahren geht es demgegenüber darum, die individuellen Mitglieder einer Gesamtheit zu charakterisieren und einzuordnen, und diesem Zweck dienen dann auch konstruierte Variablen: die individuellen Mitglieder einer Gesamtheit mit Hilfe der Werte konstruierter Variablen zu charakterisieren und einzuordnen. Auf eine kurze Formel gebracht, besteht der Unterschied darin: Statistische Methoden verwenden Daten, Skalierungsverfahren produzieren Daten.

3. Bevor wir unterschiedliche Verwendungsweisen genauer diskutieren, soll präzisiert werden, wie im folgenden in einer allgemeinen Form von konstruierten Variablen gesprochen werden soll. Den Ausgangspunkt bildet eine m-dimensionale statistische Variable

$$(X_1, \ldots, X_m) : \Omega \longrightarrow \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$$

Hiervon ausgehend können neue Variablen $X^* : \Omega \longrightarrow \tilde{\mathcal{X}}^*$ gebildet werden, deren Werte sich aus den Variablen $X_1, \ldots, X_m$ bestimmen lassen. Im einfachsten Fall verwendet man eine Funktion

$$g : \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m \longrightarrow \tilde{\mathcal{X}}^*$$

die einen neuen Merkmalsraum $\tilde{\mathcal{X}}^*$ erzeugt. Eine solche Funktion g ordnet jedem Element des Merkmalsraums $\tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$ einen Wert in einem neuen Merkmalsraum $\tilde{\mathcal{X}}^*$ zu. Hat man diese Funktion definiert, kann man die Werte der neuen Variablen X^* durch

$$X^*(\omega) := g(X_1(\omega), \ldots, X_m(\omega))$$

definieren. In diesem Fall hängt der Wert der neuen Variablen X^* für ein Argument ω, also $X^*(\omega)$, nur von den Werten der Variablen $X_1, \ldots, X_m$ für das gleiche Argument ω ab. Allgemeiner kann man bei der Bildung neuer Variablen auch Rechenvorschriften verwenden, die Informationen über die gesamten Funktionen $X_1, \ldots, X_m$ berücksichtigen. Man kann z.B. zunächst den Mittelwert einer Variablen X, also $\mathrm{M}(X)$, berechnen und die neue Variable X^* durch $X^*(\omega) := X(\omega) - \mathrm{M}(X)$ definieren. Wir schreiben deshalb die neue Variable X^* in der Form $X^* := g(X_1, \ldots, X_m)$, wobei der Buchstabe g die zur Definition der neuen Variablen verwendete Funktion bezeichnet. Bei dieser Schreibweise muß man allerdings etwas aufpassen, denn $X_1, \ldots, X_m$ sind keine Zahlen, sondern statistische Variablen, also selbst Abbildungen bzw. Funktionen. Folgende Schreibweise macht das explizit:

$$X^*(\omega) = \underbrace{g(X_1, \ldots, X_m)}_{X^*}(\omega)$$

4. Zur Illustration betrachten wir ein Beispiel für die Verwendung konstruierter Variablen im Rahmen der Statistik, nämlich bei Regressionsmodellen, die dazu dienen sollen, Einsichten in Zusammenhänge zwischen Variablen zu gewinnen. Hat man außer den Variablen $X_1, \ldots, X_m$ noch eine weitere quantitative Variable Y, kann man zum Beispiel ein einfaches lineares Regressionsmodell ausprobieren:

$$\mathrm{M}(Y|X_1 = x_1, \ldots, X_m = x_m) = \beta_0 + x_1\beta_1 + \cdots + x_m\beta_m$$

das eine lineare Abhängigkeit der bedingten Mittelwerte von Y von den Werten der Variablen $X_1, \ldots, X_m$ annimmt. Hat man bestimmte Parameter $\hat{\beta}_0, \ldots, \hat{\beta}_m$ ermittelt, kann man sie zur Definition einer neuen Variablen

$$X^* := \hat{\beta}_0 + X_1\hat{\beta}_1 + \cdots + X_m\hat{\beta}_m$$

verwenden, so daß $X^*(\omega)$ der durch die Werte $X_1(\omega), \ldots, X_m(\omega)$ bedingte Mittelwert von Y ist. Die Konstruktion macht deutlich, daß $X^*(\omega)$ keine dem Individuum ω zurechenbare Eigenschaft ist. Vielmehr dient das Verfahren einem anderen Zweck: Einsichten in einen möglicherweise vorhandenen Zusammenhang zwischen den empirischen Variablen $X_1, \ldots, X_m$ und Y zu gewinnen.

5.2 Indizes und Skalen

1. Die Beschäftigung mit konstruierten Variablen bzw. mit Verfahren zur Konstruktion neuer Variablen nimmt in der sozialwissenschaftlichen Methodenliteratur einen breiten Raum ein und findet typischerweise im Kontext einer Diskussion von Skalierungsverfahren statt. Zunächst ist bemerkenswert, daß sich in diesem Kontext auch ein spezifischer Sprachgebrauch verbreitet hat. Konstruierte Variablen werden als *Indizes* bezeichnet und die Variablen, aus denen ein Index konstruiert wird, werden *Indikatoren* genannt.[1] Z.B. heißt es bei Schnell, Hill und Esser (1999, S. 160): „Unter einem '*Index*' wird eine Zusammenfassung von mehreren Einzelindikatoren zu einer neuen Variablen verstanden.“ Diekmann (1995, S. 208) schreibt:

„Allgemein definieren wir einen Index als Variable, deren Werte sich aus einer Rechenoperation mehrerer anderer Variablen ergeben. Welche Rechenoperationen (Summierung, gewichtete Summe, Multiplikation usw.) und welche Variablen der Indexkonstruktion zugrunde gelegt werden, sollte natürlich 'inhaltlich' begründbar sein.“

Ein Index, so wie dies Wort in der Methodenliteratur verwendet wird, ist also zunächst allgemein gleichbedeutend mit unserem Begriff einer konstruierten Variablen.[2]

[1] Besozzi und Zehnpfennig 1976, S. 13, bemerken, „daß die Termini 'Indikator' und 'Index' in der Test- und Skalierungs-Theorie meist durch die Begriffe 'item' und 'Gesamt-Score' (total score) ersetzt werden.“

[2] Man beachte, daß das Wort 'Index' in vielen unterschiedlichen Bedeutungen verwendet wird, auch im Rahmen der Statistik. Z.B. wird von „Segregationsindizes“ gesprochen oder von Indizes, mit denen das Ausmaß der Einkommensungleichheit erfaßt werden soll. Man sollte sich jedoch klarmachen, daß es sich in diesen Fällen nicht in der Weise um Indizes handelt, wie dies Wort oben im Anschluß an seine Verwendung in der Methodenliteratur erläutert worden ist. Die in der Statistik konstruierten „Maßzahlen“ sind in diesem Sinn keine Indizes, denn mit ihnen wird eine ganz andere Absicht verfolgt: sie sollen dazu dienen, statistische Verteilungen zu charakterisieren; man denke z.B. an Begriffsbildungen wie 'Mittelwert' und 'Varianz' zur Charakterisierung von Verteilun-

2. Wichtig ist, welchem Zweck ein Index dienen soll. Bei den genannten Autoren steht die Indexbildung stets im Kontext von Verfahren, mit deren Hilfe Objekte durch „Meßwerte" charakterisiert werden sollen. Die Vorstellung ist, daß Indizes „Meßwerte" zur Charakterisierung von Individuen liefern können. Ganz explizit sagen z.B. Bortz und Döring (1995, S. 133): „Ein Index ist ein Meßwert für ein komplexes Merkmal, der aus den Meßwerten mehrerer Indikatorvariablen zusammengesetzt ist."[3] Zum Verständnis sollte beachtet werden, daß Bortz und Döring wie auch viele andere Autoren in der sozialwissenschaftlichen Methodenliteratur in einer vollständig unspezifischen Weise von ‘Meßverfahren’ sprechen.[4] Sieht man von dieser Rhetorik ab, mit der wir uns in Kapitel 8 genauer beschäftigen, geht es um die Frage, ob und ggf. wie durch Indizes Größenbegriffe gebildet werden können, d.h. ob und ggf. wie sich für den Merkmalsraum eines Index eine sinnvoll interpretierbare lineare Ordnung begründen läßt.

3. Leider ist der Sprachgebrauch in der Methodenliteratur unklar. Z.B. schreibt Laatz (1993, S. 261):

„Indices und Skalen sind nicht – wie häufig verbreitet – Erhebungstechniken, sondern *zusammenfassende Maßzahlen*. In ihnen werden verschiedene Messungen zu einem Ausdruck zusammengefaßt. In gewisser Weise stellen sie einen Schritt der Datenaufbereitung dar. Mit Hilfe von Indices und Skalen wird ein multidimensionaler *Eigenschaftsraum* auf eine Dimension *reduziert*. Skalen – im hier gebrauchten Wortsinne – sind Indices, die mehrere Messungen derselben Dimensionen zusammenfassen und mit Hilfe eines *Skalierungsverfahrens* entwickelt wurden. Skalierungsverfahren prüfen insbesondere die interne Konsistenz der Teilinstrumente und dienen der Auswahl geeigneter Teilmeßinstrumente."

Hier wird also zwischen Skalierungsverfahren und Meßverfahren unterschieden. Skalierungsverfahren sind hiernach nicht Meßverfahren, sondern Verfahren zur Konstruktion und Beurteilung von Verfahren, mit denen Eigenschaften quantitativ erfaßt werden sollen. Ähnlich heißt es auch bei Schnell, Hill und Esser (1999, S. 160): „Sowohl die Indexbildung als auch Skalierungsverfahren sind Auswertungsverfahren, keine Datenerhebungs- oder Meßverfahren." Das Wort ‘Auswertungsverfahren’ macht allerdings nicht deutlich, worauf es auch diesen Autoren ankommt; etwas später (S. 173) sagen sie nämlich:

gen. In diesen Fällen werden keine neuen statistischen Variablen konstruiert – also auch keine neuen Daten erzeugt –, sondern es werden Begriffe entwickelt, um ein- oder mehrdimensionale Verteilungen zu charakterisieren oder um zwei oder mehr Verteilungen zu vergleichen. Auch mit diesen Begriffsbildungen wird eine Quantifizierung vorgenommen; aber es werden keine Objekte, Individuen oder Situationen quantifiziert, sondern statistische Verteilungen.

[3] Man beachte in diesem Zitat auch die Verwechslung von Indizes im Sinne von Variablen mit ihren möglichen Werten.

[4] Zum Beispiel spricht Kromrey 1994, S. 158, davon, „was wir bei den Untersuchungseinheiten – z.B. Personen – messen, d.h. beobachten oder sonstwie feststellen können."

„‘Skalierungsverfahren’ sind Methoden zur Konstruktion von Meßinstrumenten. Das Resultat der Durchführung eines Skalierungsverfahrens ist eine ‘Skala’."

Diese Definition ist jedoch einerseits zu weit, andererseits etwas zu eng gefaßt. Sie ist zu weit gefaßt, weil sie keinen Hinweis darauf liefert, mit welchen Arten von „Meßinstrumenten" sich Skalierungsverfahren beschäftigen. Im Kontext psychologischer und sozialwissenschaftlicher Anwendungen sind es typischerweise Indizes (im Sinne konstruierter Variablen). Die Definition ist aber auch etwas zu eng, da sich Skalierungsverfahren nicht nur mit der Konstruktion, sondern auch mit der Begründung und Beurteilung von Indizes beschäftigen. Um für diesen Text eine Orientierung zu schaffen, werden wir folgende Definition verwenden: *Skalierungsverfahren* sind Verfahren zur Konstruktion, Begründung und Beurteilung quantifizierender Indizes, also von Merkmalsräumen, die als Größenbegriffe Verwendung finden sollen. Davon unterscheiden wir *Klassifikationsverfahren* als Verfahren zur Konstruktion, Begründung und Beurteilung von klassifizierenden Indizes.

4. Das Wort ‘Skala’ wird in der Literatur sowohl in einem weiten als auch in einem engeren Sinn verwendet. In einem weiten Sinn ist es gleichbedeutend mit einer numerischen Repräsentation für den Merkmalsraum einer Variablen. Den Kontext bildet also eine statistische Variable

$$X : \Omega \longrightarrow \tilde{\mathcal{X}}$$

durch die jedem Mitglied ω aus der Gesamtheit Ω ein Merkmalswert $X(\omega)$ in einem Merkmalsraum $\tilde{\mathcal{X}}$ zugeordnet wird. Der Merkmalsraum ist dabei zunächst eine Menge möglicher Merkmalswerte. Eine *Skala* entsteht dadurch, daß man explizit eine numerische Repräsentation

$$\phi : \tilde{\mathcal{X}} \longrightarrow \mathbf{R}$$

einführt, die jedem Merkmalswert $\tilde{x} \in \tilde{\mathcal{X}}$ eine bestimmte Zahl $\phi(\tilde{x}) \in \mathbf{R}$ zuordnet, so wie dies in Abschnitt 4.1 besprochen worden ist.[5] In einem engeren Sinn impliziert der Begriff ‘Skala’, daß es sich um die numerische Repräsentation eines quantitativen Merkmalsraums handelt. Wir werden im folgenden das Wort ‘Skala’ im weiten Sinn des Wortes verwenden, um die Frage offen zu halten, ob und ggf. wie für den Merkmalsraum eines Index eine sinnvoll interpretierbare lineare Ordnung begründet werden kann. Wie ebenfalls bereits in Abschnitt 4.1 besprochen worden ist, nehmen wir auch an, daß mit jeder Definition eines Merkmalsraums zugleich eine Skala festgelegt wird, so daß man den Merkmalsraum als eine Teilmenge der reellen Zahlen auffassen kann.

[5] Wir verwenden das Wort ‘Skala’ sowohl für die Abbildung ϕ als auch für den durch sie erzeugten numerischen Merkmalsraum $\phi(\tilde{\mathcal{X}})$.

5. Wie verhalten sich nun die Worte 'Skala' und 'Index' zueinander? In dem oben angeführten Zitat schlägt Laatz vor, die Worte zu unterscheiden. Nach seinem Vorschlag sind Skalen ein Spezialfall von Indizes, nämlich „Indices, die mehrere Messungen derselben Dimension zusammenfassen". Diese Auffassung paßt jedoch nicht besonders gut zu unserem allgemeinen Begriff einer Skala und auch nicht zum Begriff eines Index im Sinne einer konstruierten Variablen. Denken wir noch einmal an das allgemeine Verfahren zur Bildung von Indizes. Gegeben ist eine Reihe von Variablen

$$(X_1, \ldots, X_m) : \Omega \longrightarrow \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$$

und daraus wird eine neue Variable $X^* : \Omega \longrightarrow \tilde{\mathcal{X}}^*$ gebildet. Der Merkmalsraum $\tilde{\mathcal{X}}^*$ ist natürlich im Sinne unserer Definition wiederum eine Skala, diesmal eine Skala für die neue Variable X^*. Es handelt sich um eine abgeleitete Skala;[6] sie ergibt sich aus den Skalen für die Variablen (Indikatoren) $X_1, \ldots, X_m$ und der Funktion g, durch die der Index definiert wird. Eine damit konsistente Unterscheidung zwischen den Worten 'Skala' und 'Index' verläuft also folgendermaßen:

- Ein Index ist eine Variable, die aus anderen Variablen nach einer festgelegten Vorschrift gebildet werden kann.
- Eine Skala besteht in der numerischen Repräsentation eines Merkmalsraums, wobei es sich insbesondere um den Merkmalsraum eines Index handeln kann.

6. Beziehen wir uns jetzt noch einmal auf die Formulierung von Laatz: „Mit Hilfe von Indices und Skalen wird ein multidimensionaler Eigenschaftsraum auf eine Dimension reduziert." Nach unserem Sprachgebrauch kann man dies nur für konstruierte Variablen (Indizes) sagen. Ihr Ausgangspunkt ist eine m-dimensionale Variable und also auch ein multidimensionaler Eigenschaftsraum, d.h. ein kombinierter Merkmalsraum für die m Variablen, auf den wir durch die Symbolik $\tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$ verweisen. Die Konstruktion eines Index besteht dann darin, aus den m Variablen eine neue zu bilden; und das impliziert natürlich, daß dadurch aus dem m-dimensionalen Merkmalsraum für die zunächst gegebenen Variablen ein neuer eindimensionaler Merkmalsraum für den Index gebildet wird. Insofern kann man sagen, daß durch die Bildung eines Index ein m-dimensionaler Merkmalsraum auf einen eindimensionalen Merkmalsraum reduziert wird. Allerdings ist es möglich, aus einer gegebenen Menge von Variablen mehr als nur einen Index zu konstruieren und diese Indizes dann wiederum zu einer mehrdimensionalen Variablen zu vereinigen. Dies zeigt noch einmal, daß in diesem Kontext der Dimensionsbegriff eine rein formale Bedeutung hat.

[6] Im weiten Sinn des Wortes; denn ob und ggf. wie mit dem Index X^* ein neuer Größenbegriff begründet werden kann, ist zunächst eine vollständig offene Frage.

5.3 Verteilungsabhängigkeit

1. Um einschätzbar zu machen, welche Ansprüche mit einer Konstruktion von Indizes verbunden werden können, ist es wichtig, daß es zwei wesentlich unterschiedliche Arten von Konstruktionsverfahren gibt. Man kann sich das folgendermaßen verdeutlichen. Möchte man aus gegebenen Variablen $X_1, \ldots, X_m$ eine neue Variable X^* bilden, setzt dies voraus, daß zunächst ein neuer Merkmalsraum für die Variable X^* begründet wird. Unter Verwendung der in Abschnitt 5.1 eingeführten Notation kann dies durch die Definition einer Funktion g geschehen, durch die der Merkmalsraum $\tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$ in einen neuen Merkmalsraum $\tilde{\mathcal{X}}^*$ abgebildet wird. Wichtig ist, wie die Funktion g zustande kommt, denn dafür gibt es zwei wesentlich unterschiedliche Möglichkeiten:

a) Man kann diese Funktion ohne Rückgriff auf Informationen über die Mitglieder von Ω definieren. Wir sagen dann, daß die neue Variable X^* mit einem *verteilungsunabhängigen* Verfahren konstruiert wird.

b) Andererseits kann man Funktionen verwenden, deren Definition davon abhängt, welche Werte für die Variablen $X_1, \ldots, X_m$ bei den Mitgliedern von Ω gegeben sind. Wir sagen dann, daß die neue Variable X^* mit einem *verteilungsabhängigen* Verfahren konstruiert wird.

2. Betrachten wir zunächst ein Beispiel für den ersten Fall: X_1 erfaßt die Einkommen bei einer Menge von Haushalten und X_2 ihre Mietausgaben. Dann kann man einen Index $X^* := X_2/X_1$ bilden, um den Anteil der Mietausgaben am Einkommen zu erfassen. In diesem Beispiel ist es evident, daß der Wert von X^* bei einem Haushalt ω nur davon abhängt, welches Einkommen dieser Haushalt bezieht und welche Mietzahlungen er leisten muß. Der Wert von X^* bei einem Haushalt ω ist also – soweit dies von der Indexbildung abhängt – ganz unabhängig von den Werten von X^* bei anderen Haushalten.

3. Bei vielen Indexkonstruktionen ist dies jedoch nicht der Fall, sondern der Wert, den ein Index X^* bei einem jeweils bestimmten Mitglied $\omega \in \Omega$ annimmt, hängt von der statistischen Verteilung der Variablen $X_1, \ldots, X_m$ in der Gesamtheit Ω ab. Ein einfaches Beispiel liefert die Abweichung der individuellen von der durchschnittlichen Haushaltsgröße. Erfaßt X die Haushaltsgrößen bei einer Menge von Haushalten, wird in diesem Beispiel die neue Variable durch $X^* := X - \mathrm{M}(X)$ definiert. Für jeden Haushalt $\omega \in \Omega$ erhält man also den Wert der neuen Variablen durch $X^*(\omega) = X(\omega) - \mathrm{M}(X)$ Offenbar hängt dieser Wert nicht nur von der Haushaltsgröße des Haushalts ω ab, sondern auch von $\mathrm{M}(X)$, also den Haushaltsgrößen aller Haushalte in Ω. Oder anders ausgedrückt: in diesem Fall hängt X^* nicht nur vom Merkmalsraum $\tilde{\mathcal{X}}$, sondern auch von der Verteilung der statistischen Variablen X ab.

4. Bei statistischen Analysen ist Verteilungsabhängigkeit nicht nur unproblematisch, sondern macht gerade eine wesentliche Eigenschaft statistischer Aussagen deutlich: Es handelt sich – direkt oder indirekt – stets um Aussagen über statistische Verteilungen, und sie beziehen sich mithin stets auf eine Gesamtheit von Individuen, Objekten oder Situationen. Verteilungsabhängige Indizes sind jedoch problematisch, wenn das Ziel nicht darin besteht, statistische Aussagen zu formulieren, sondern wenn durch sie einzelne Individuen charakterisiert werden sollen. Denn was immer dann mit einem solchen Index erfaßt wird, es handelt sich jedenfalls nicht um eine einem Individuum individuell zurechenbare Eigenschaft.

5. Dies liefert zugleich einen wichtigen Grund dafür, daß verteilungsabhängige Indizes nicht als Meßverfahren verstanden werden können. Sie liefern keine Verfahren, die unabhängig von einer Menge kontingenter Daten, wie sie typischerweise in Form einer Stichprobe verfügbar sind, definiert werden könnten. Nicht nur die Werte solcher Indizes, sondern *auch ihre Definition* verändert sich mit den jeweils für die Konstruktion verwendeten Daten.[7] Es ist deshalb wichtig, bei der Besprechung von Konstruktionsverfahren für Indizes darauf zu achten, ob und ggf. wie eine Verteilungsabhängigkeit vorliegt. Wie sich zeigen wird, implizieren insbesondere die meisten Skalierungsverfahren verteilungsabhängige Variablenkonstruktionen (Indizes), und die entsprechenden Datenkonstruktionsverfahren sind schon deshalb nicht sinnvoll als Meßverfahren charakterisierbar.

5.4 Standardisierungen

1. In diesem Zusammenhang ist es auch nützlich, daran zu erinnern, daß Datenkonstruktionsverfahren oft eine Standardisierung statistischer Variablen beinhalten und daß schon dadurch eine Verteilungsabhängigkeit erzeugt wird. Zur Erläuterung beziehen wir uns auf eine Variable

$$X : \Omega \longrightarrow \tilde{\mathcal{X}}$$

Im allgemeinen besteht eine Standardisierung von X darin, daß man eine neue Variable $g(X)$ konstruiert, wobei die Funktion g die Transformation angibt. Standardisierte Variablen sind also konstruierte Variablen. Unterschiedliche Arten der Standardisierung kommen dadurch zustande, daß unterschiedliche Transformationen verwendet werden.

2. Eine einfache Standardisierung besteht in der *Zentrierung* einer Variablen: $g(X) := X - \mathrm{M}(X)$, wobei $\mathrm{M}(X)$ der Mittelwert von X ist; $g(X)$

[7]Die Problematik verteilungsabhängiger Indizes in der psychologischen Testtheorie wird z.B. von Fischer 1974, S. 133ff, diskutiert. Man vgl. auch eine schon ältere Arbeit von Gulliksen 1946, in der der Autor glaubte, dieses Problem lösen zu können.

hat dann den Mittelwert 0. Eine weitere Transformation, die oft zur Standardisierung einer Variablen verwendet wird, bezieht sich auf die Varianz. Wenn $\mathrm{V}(X)$ die Varianz von X bezeichnet, hat die neue Variable $g(X) := X/\sqrt{\mathrm{V}(X)}$ die Varianz 1. Man spricht dann auch von einer Normierung der Varianz einer Variablen. Beide Varianten können zusammengefaßt werden, üblicherweise in der folgenden Form:

$$g(X) := \frac{X - \mathrm{M}(X)}{\sqrt{\mathrm{V}(X)}} \tag{5.4.1}$$

Die neue Variable $g(X)$ hat dann den Mittelwert 0 und die Varianz 1. Wenn in der Literatur ohne nähere Erläuterungen von der Standardisierung einer Variablen gesprochen wird, ist meistens diese Form gemeint. Manchmal wird auch von einer *Z-Transformation* gesprochen.

3. Warum werden Standardisierungen durchgeführt? Oft dienen sie nur dem Zweck, statistische Rechnungen zu vereinfachen. Z.B. vereinfacht sich die Darstellung von Kovarianz- und Korrelationsmatrizen, wenn man zunächst einfache Standardisierungen (Z-Transformationen) vornimmt (vgl. Abschnitt A.2). Vor allem in der Literatur über psychologische Tests werden noch viele andere Standardisierungen verwendet.[8] Zur Begründung wird oft angeführt, daß durch Standardisierungen unterschiedliche Tests (oder Indizes) „vergleichbar" gemacht werden können.[9] Demgegenüber sollte jedoch betont werden, daß standardisierte Indizes stets verteilungsabhängig sind, so wie dies in Abschnitt 5.3 erläutert worden ist. Der Wert eines standardisierten Index für irgendein bestimmtes Objekt oder Individuum hängt stets auch von den in der Gesamtheit realisierten Daten ab, die zur Standardisierung verwendet werden.

[8]Eine einfache Methode besteht in der Bildung von Prozenträngen. Das Verfahren verwendet die Verteilungsfunktion der Variablen X, also

$$F[X](x) := \mathrm{P}[X](\{\omega \in \Omega \mid X(\omega) \leq x\})$$

Ist dann $\tilde{x} \in \tilde{\mathcal{X}}$ ein möglicher Wert der Variablen X, wird $100 \cdot F[X](\tilde{x})$ sein *Prozentrang* genannt.

[9]So heißt es z.B. bei Cattell 1967, S. 241: „While talking of scoring tests we might point out that the psychologist expresses most of his measures in standard scores because his raw scores, unlike these of the physical scientist, are not very meaningful as they stand. The metre, the gramme, and the degree of centigrade are all related, but every one of 900 tests has its own score units." Cattell bezieht sich hier auf die Standardisierung entsprechend (5.4.1). Aber auch Normalisierungen werden oft verwendet; vgl. z.B. Mehrens und Lehmann 1975, S. 146. Anastasi 1982, S. 79, gibt als Begründung an: „In order to achieve comparability of scores from dissimilarly shaped distributions, nonlinear transformations may be employed to fit the scores to any specified type of distribution curve."

Kapitel 6

Aspekte der Indexbildung

Bei den meisten Formen sozialwissenschaftlicher Datenkonstruktion handelt es sich um Varianten der Indexbildung. In diesem Kapitel besprechen wir anhand von Beispielen eine Reihe von Problemen, die bei der Indexbildung beachtet werden sollten. Dabei verwenden wir das Wort ‘Index’ stets so, wie es in Abschnitt 5.1 erläutert worden ist. Die Probleme sind nicht technischer Art, sondern betreffen den Sinn und die jeweils intendierte Verwendung von Indizes.

6.1 Einfache Beispiele

1. Wir beginnen mit einigen einfachen Beispielen, bei denen der Sinn der Indexbildung leicht durchschaubar ist. Ein besonders einfaches Beispiel wurde bereits erwähnt. Es gibt zwei Variablen: X_1 erfaßt das Einkommen von Haushalten, X_2 ihre Ausgaben für Miete. Dann kann man einen Index

$$X^* := X_2/X_1$$

bilden, um den Anteil der Mietkosten am Einkommen zu erfassen. Auf diese Weise entsteht ein unmittelbar verständlicher quantitativer Index, mit dessen Hilfe Haushalte unterschieden, geordnet und verglichen werden können. Zur Illustration kann Abb. 6.1.1 dienen. Die Daten haben wir der 50 %-Teilstichprobe des SOEP (vgl. Abschnitt 1.1) für das Jahr 1984 entnommen, die für Lehrzwecke verfügbar ist. In dieser Teilstichprobe gibt es 2798 Haushalte. 2645 dieser Haushalte haben eine gültige Angabe über ihr monatliches Haushaltseinkommen gemacht. Davon haben 1678 Haushalte in Mietwohnungen gewohnt und auch die Höhe ihrer Miete angegeben. Das durchschnittliche Einkommen beträgt 2420 DM, die durchschnittliche Miethöhe 420 DM. Nur 7 dieser Haushalte hatten ein Einkommen über 8000 DM. Wir beschränken uns deshalb auf die Haushalte, deren Einkommen unter 8000 DM lag; in unserem Datensatz gibt es also 1671 Haushalte. Abb. 6.1.1 zeigt mittels eines Streuungsdiagramms den Zusammenhang zwischen Haushaltseinkommen und Mietausgaben für diese 1671 Haushalte. Man kann sich zum Beispiel überlegen, in welchem Bereich des Diagramms sich Haushalte befinden, bei denen der Mietanteil am Einkommen zwischen 30 und 40 Prozent liegt.

2. Als zweites Beispiel sehen wir uns Preisindizes an, wie sie vom Statistischen Bundesamt konstruiert werden. Obwohl es in diesem Fall eigentlich nicht erforderlich ist, auf unseren Begriff statistischer Variablen zurückzugreifen, werden wir diesen Begriffsrahmen verwenden, um Preisindizes

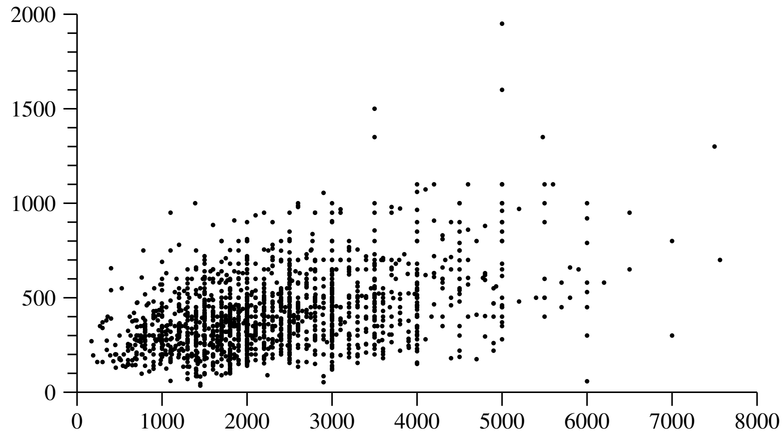


Abb. 6.1.1 Streuungsdiagramm für 1671 Beobachtungen des Haushaltseinkommens (X-Achse) und der Miethöhe (Y-Achse) im Jahr 1984.

als Beispiele für Indizes, wie sie in Abschnitt 5.1 definiert worden sind, verständlich zu machen. Als Gesamtheit Ω gehen wir von einer Menge von Regionen aus (statistische Variabilität bezieht sich dann auf regionale Preisunterschiede).[1] Es wird auf insgesamt m Arten von Gütern Bezug genommen. Die Variable X_j erfaßt den Preis einer Einheitsmenge des Gutes j. Dann kann für jede Region $\omega \in \Omega$ ein Preisindex

$$P_\omega(a_1, \ldots, a_m) := \sum_{j=1}^{m} a_j X_j(\omega)$$

definiert werden. Es handelt sich um einen gewichteten additiven Index, der natürlich davon abhängt, welche Gewichte $(a_1, \ldots, a_m)$ verwendet werden. Bei der Konstruktion von Preisindizes wird eine Menge von Gewichten oft als ein „Warenkorb“ bezeichnet: a_j gibt an, wieviel Einheiten der j-ten Gütersorte im Warenkorb berücksichtigt werden sollen. Die Definition von Warenkörben hängt vom Verwendungszweck des Index ab. Bei der Konstruktion von Preisindizes für die Lebenshaltung orientiert sich das Statistische Bundesamt an den Verbrauchsausgaben von Haushalten (unterschiedlicher Art), die in gelegentlich durchgeführten „Einkommens- und Verbrauchsstichproben“ ermittelt werden.[2]

3. Ein weiteres einfach durchschaubares Beispiel vermitteln Indizes für *Äquivalenzeinkommen*. Zugrunde liegt die Frage, wie Haushaltseinkommen sinnvoll verglichen werden können. Das Problem resultiert daraus,

[1] Bei der Statistik der Verbraucherpreise bezieht sich das Statistische Bundesamt auf etwa 120 Gemeinden.

[2] Einzelheiten werden ausführlich in der Literatur zur Wirtschafts- und Sozialstatistik diskutiert; man vgl. zum Beispiel Neubauer 1996.

daß Haushalte unterschiedlich viele Mitglieder haben und zum Beispiel ein Einkommen von 3000 DM/Monat bei einem 1-Personen-Haushalt und bei einem 4-Personen-Haushalt sicherlich auf unterschiedliche Einkommenssituationen verweist. Um Haushaltseinkommen dennoch vergleichbar zu machen, könnte man einen Index für das Pro-Kopf-Einkommen verwenden, also $Y^* := Y/H$, wobei Y das Haushaltseinkommen und H die Anzahl der Mitglieder des Haushalts erfaßt. Allerdings kann dieser einfache Index mit dem Argument kritisiert werden, daß sich durch das Zusammenleben in Haushalten Ersparnisse ergeben, die bei der Konstruktion vergleichbarer Einkommenspositionen berücksichtigt werden sollten, etwa daß einige Güter gemeinsam genutzt werden können und nicht für jedes Haushaltsmitglied gesondert angeschafft werden müssen. Die Frage ist dann, wie diese Überlegung bei der Indexkonstruktion berücksichtigt werden kann. Dazu gibt es eine Reihe unterschiedlicher Vorschläge. Ein einfaches Schema, in dessen Rahmen sich ein Spektrum unterschiedlicher Indizes konstruieren läßt, wurde von Buhmann u.a. (1988) vorgeschlagen:

$$Y^*_\delta := Y/H^\delta$$

wobei $0 \leq \delta \leq 1$. Extremfälle sind das unveränderte Haushaltseinkommen ($\delta = 0$) und das Pro-Kopf-Einkommen ($\delta = 1$). Natürlich stellt sich die Frage, wie ein geeigneter Wert für δ gefunden werden kann. Einige Überlegungen dazu finden sich bei den genannten Autoren. Es sei jedoch erwähnt, daß man sich nicht unbedingt für nur einen bestimmten Parameterwert entscheiden muß, sondern daß man eine Untersuchung, bei der Äquivalenzeinkommen eine Rolle spielen, auch so anlegen kann, daß die Abhängigkeit der schließlich erzielten Ergebnisse von unterschiedlichen Äquivalenzdefinitionen sichtbar wird.

4. Schließlich sei noch ein Beispiel anderer Art angeführt. Es handelt sich um die Bewertung von Leistungen in einer Statistik-Klausur, also um einen Test, so wie dies Wort in Kapitel 2 besprochen worden ist. Das allgemeine Schema kann folgendermaßen erläutert werden. Ω ist die Menge der Personen, die an dem Test teilnehmen. Es gibt m Aufgaben $A_1, \ldots, A_m$; die Variablen $X_1, \ldots, X_m$ erfassen, wieviele Punkte in den Aufgaben erzielt worden sind. Die Merkmalsräume sind durch $\tilde{\mathcal{X}}_j := \{0, 1, \ldots, n_j\}$ (für $j = 1, \ldots, m$) gegeben. In der Aufgabe A_j können also bis zu n_j Punkte erzielt werden. Zur Leistungsbewertung kann ein einfacher additiver Index verwendet werden, also

$$X^* := \sum_{j=1}^{m} X_j$$

wenn man davon ausgeht, daß jeder Punkt, unabhängig davon, in welcher Aufgabe er erzielt worden ist, mit gleichem Gewicht in die Gesamtbewertung eingehen soll. Wiederum handelt es sich um einen einfach durchschaubaren quantifizierenden Index, der in diesem Fall auf dem Abzählen von

Tabelle 6.1.1 Ergebnisse einer Statistik-Klausur (5 Aufgaben).

X_1	X_2	X_3	X_4	X_5	X^*
2	10	0	8	15	35
2	10	5	3	15	35
4	10	0	5	0	19
8	5	0	4	1	18
8	7	0	10	7	32
8	10	0	9	15	42
8	10	5	6	0	29
10	0	5	1	15	31
10	5	5	0	2	22
10	5	5	5	15	40
10	5	5	5	15	40
10	10	0	0	13	33
10	10	0	0	14	34
10	10	0	4	12	36
10	10	0	4	12	36
10	10	0	6	14	40
10	10	0	8	15	43
10	10	0	9	15	44
10	10	0	10	7	37
10	10	0	10	15	45
10	10	0	10	15	45
10	10	1	10	9	40
10	10	1	10	9	40
10	10	2	6	10	38
10	10	2	6	10	38
10	10	3	6	7	36
10	10	3	6	10	39
10	10	5	4	9	38
10	10	5	4	12	41
10	10	5	8	13	46
10	10	5	8	15	48
10	10	5	8	15	48
10	10	5	8	15	48
10	10	5	9	9	43
10	10	5	9	9	43
10	10	5	10	0	35
10	10	5	10	13	48
10	10	5	10	14	49
10	10	5	10	15	50
10	10	5	10	15	50
10	10	5	10	15	50
10	10	5	10	15	50
10	10	5	10	15	50
10	10	5	10	15	50
10	10	5	10	15	50
10	10	5	10	15	50

Punkten beruht. — Zur Illustration sei auf die Ergebnisse einer Statistik-Klausur hingewiesen, die in einer unserer Lehrveranstaltungen durchgeführt worden ist. Es gab fünf Aufgaben, bei denen die folgenden Punktzahlen erreicht werden konnten:

$$n_1 = 10, \ n_2 = 10, \ n_3 = 5, \ n_4 = 10, \ n_5 = 15$$

Tabelle 6.1.1 zeigt die Ergebnisse, die von 46 Klausurteilnehmern erzielt worden sind. In späteren Kapiteln werden wir diese Daten zur Illustration einiger Rechenverfahren verwenden.

6.2 Fragen der Begriffsbildung

1. Die im vorangegangenen Abschnitt diskutierten Beispiele zeigen, daß Indizes für unterschiedliche Zwecke konstruiert werden können. In diesem Abschnitt geht es um die Frage, welche begriffsbildenden Ansprüche mit der Konstruktion von Indizes verbunden werden können. — Betrachten wir unter diesem Gesichtspunkt die bisher angeführten Beispiele. In jedem dieser Beispiele wird ein neuer Begriff gebildet: 'Anteil der Mietausgaben am Einkommen', 'Preisindex', 'Äquivalenzeinkommen', 'Punktzahl bei einem Test'. Man versteht diese Begriffe, wenn man die Verfahren untersucht, mit denen der jeweilige Index konstruiert wird, oder anders formuliert: Die Begriffe gewinnen ihren Sinn aus dem Konstruktionsverfahren. Daß ein sinnvoller Begriff entsteht, folgt natürlich nicht schon daraus, daß das Rechenverfahren nachvollziehbar ist, sondern kann sich schließlich erst in den jeweils intendierten Verwendungszusammenhängen zeigen, also daran, wie mit einem Index-Begriff argumentiert wird und welche Aussagen und möglicherweise praktisch folgenreichen Konsequenzen sich aus den Begriffsverwendungen ergeben; z.B. wenn ein spezifischer Begriff des Äquivalenzeinkommens verwendet wird, um sozialpolitische Fragen zu erörtern. Man kann dann durch eine Bezugnahme auf den Verwendungszusammenhang den Sinn oder die Angemessenheit eines Begriffs in Frage stellen.

2. Es ist bemerkenswert, daß es in der sozialwissenschaftlichen Methodenliteratur noch ein wesentlich anderes Verständnis der begrifflichen Ansprüche von Indizes gibt. Es wird nämlich die Meinung vertreten, daß mit Hilfe von Indizes auch schon in Verwendung befindliche Begriffe präzisierend rekonstruiert werden können. Ein wichtiger Vorreiter dieser Auffassung war Paul Lazarsfeld, der treffend von einem „flow from concepts to measurement procedures" gesprochen hat (Lazarsfeld 1993, S. 239ff). In der neueren Methodenliteratur wird oft von „Operationalisierung" gesprochen, z.B. gesagt, daß Indizes zur Operationalisierung von schon in Verwendung befindlichen Begriffen dienen können; und meistens wird suggeriert, daß dadurch Begriffe in ihrem „empirischen Bedeutungsgehalt" erfaßt und präzisiert werden können. Z.B. heißt es bei Schnell, Hill und Esser (1999, S. 122):

„Zumindest zu Beginn eines Forschungsprojektes sind Theorien in den Sozialwissenschaften häufig weder explizit und eindeutig formuliert, noch sind die verwendeten Begriffe eindeutig definiert und von anderen Begriffen abgegrenzt. Die 'Theorien' für einen bestimmten Gegenstandsbereich bestehen gelegentlich nur aus vagen Vorstellungen und Ideen, die die Auffassung des Gegenstandsbereichs stukturieren. Beispiele für solche strukturierenden Begriffe oder 'Konzepte' sind z.B. 'Macht', 'Herrschaft', 'Integration' oder 'Identität'."

Hiervon ausgehend werden dann zwei Aufgaben formuliert: Die jeweils verwendeten Begriffe sollten zunächst einer sprachlichen Klärung unterzogen werden; dann sollte in einem zweiten Schritt für die geklärten Begriffe nach „Operationalisierungen" gesucht werden.

3. Um zu überlegen, ob man einen sinnvollen Zusammenhang zwischen „theoretischen Begriffen" und empirisch erfaßbarer Realität durch „Operationalisierungen" herstellen kann, muß man zunächst verstehen, was gemeint sein könnte. Schnell, Hill und Esser (1999, S. 123f) geben folgende Erläuterung:

„Die 'Operationalisierung' eines theoretischen Begriffes besteht aus der Angabe einer Anweisung, wie Objekten mit Eigenschaften (Merkmalen), die der theoretische Begriff bezeichnet, beobachtbare Sachverhalte zugeordnet werden können."

Ähnliche Erläuterungen finden sich bei anderen Autoren.[3] Es gibt jedoch mehrere Probleme. Ein erstes Problem betrifft die Frage, wie man sich mit Begriffen auf empirische Realität beziehen kann. Den angeführten Zitaten liegt die Vorstellung zugrunde, daß Begriffe Aspekte der empirischen Realität *bezeichnen*; bei Schnell, Hill und Esser bezeichnen sie Eigenschaften, bei Bortz und Döhrung bezeichnen sie Sachverhalte. Aber das kann man bestenfalls von Namen sagen. Die meisten Begriffe, die in Sätzen vorkommen, mit denen man sich auf empirische Realität bezieht, bezeichnen nicht Objekte oder Sachverhalte oder deren Eigenschaften, sondern bringen Unterscheidungen oder Beziehungen zum Ausdruck, auf die hingewiesen werden soll. Das hat zunächst Folgen für die Aufgabe, die Bedeutung von Begriffen zu klären. Es wäre irreführend, diese Aufgabe dadurch zu verfolgen, daß man nach Sachverhalten oder Eigenschaften sucht, die durch einen Begriff bezeichnet werden. Vielmehr muß man sich überlegen, ob und wie ein Begriff im Kontext von Sätzen verwendet werden kann, mit denen sich sinnvolle Aussagen formulieren lassen. Man denke z.B. an das Wort 'Macht'. Mit diesem Wort können ohne weiteres überhaupt keine sinnvollen Sätze gebildet werden. Aber natürlich kann man sich überlegen, was

[3] Z.B. bei Bortz und Döring 1995, S. 63: „Eine *operationale Definition* standardisiert einen Begriff durch die Angabe der Operationen, die zur Erfassung des durch den Begriff bezeichneten Sachverhaltes notwendig sind, oder durch Angabe der meßbaren Ereignisse, die das Vorliegen dieses Sachverhaltes anzeigen (Indikatoren)." Dixon, Bouma und Atkinson schreiben in ihrem „Handbook of Social Science Research" 1987, S. 50f: „Up to this point we have dealt with concepts; academic performance, study, nutrition, marital happiness. These are abstract ideas. If we are going to do research involving these concepts we will have to find ways of making these abstract concepts concrete and measurable. The activity of finding measurable variables for concepts is called 'operationalization'. An operational definition of a concept goes beyond a usual definition. It defines a concept in terms that can be measured, that is, in empirical terms."

es bedeutet bzw. bedeuten sollte, einen Akteur mächtig zu nennen, oder mächtiger als einen anderen Akteur. Dann wird man vielleicht feststellen, daß man sich auf Handlungsmöglichkeiten von Akteuren beziehen muß. Das aber sind weder Sachverhalte noch Eigenschaften von Objekten.

4. Ein weiteres Problem betrifft die implizite Unterscheidung zwischen „theoretischen Begriffen", die sich vermeintlich auf „nicht-beobachtbare Sachverhalte" beziehen, und „empirischen Begriffen", mit denen man sich vermeintlich unmittelbar auf empirische Realität beziehen kann.[4] Denn erst durch eine solche Unterscheidung könnte „Operationalisierung" als eine spezifische Aufgabe sinnvoll erscheinen. Aber schon die Unterscheidung ist fragwürdig. Eine Äußerung von Kromrey (1994, S. 114) – im Kontext von Überlegungen „über den Zusammenhang von sozialer Herkunft und Schul- sowie Berufserfolg" – kann das Problem verdeutlichen:

> „Die definierten Begriffe ‘sozialer Status des Elternhauses’, ‘Bildung’ und ‘Berufserfolg’ der Kinder haben alle einen empirischen Bezug: Es gibt in der sozialen Realität so etwas wie eine soziale Rangordnung in der Gesellschaft, wie Bildung einer Person oder wie Berufserfolg. Aber alle Begriffe sind zunächst praktisch nicht direkt beobachtbar (wahrnehmbar, erfahrbar), sondern nur indirekt auf dem Umweg über die Beobachtung von Sachverhalten, die etwas mit Statuseinschätzung, mit Bildung, mit Berufserfolg zu tun haben. Wir benötigen also *Indikatoren*, die uns das Vorliegen der mit den Begriffen bezeichneten Sachverhalte anzeigen, ‘indizieren’."

Der Gedankengang ist ziemlich merkwürdig. Zunächst wird von Begriffen wie etwa ‘Bildung’ und ‘Berufserfolg’ ausgegangen. Dann wird das Problem jedoch nicht darin gesehen, wie mit diesen oder verwandten Begriffen sinnvolle Sätze gebildet werden können, sondern darin, daß diese Begriffe „nicht direkt beobachtbar" sind. Aber schon die Formulierung ist irritierend, denn Begriffe können natürlich niemals beobachtet werden.[5] Unterstellen wir jedoch, daß Kromrey eigentlich etwas anderes meint: daß nicht ohne weiteres klar ist, wie man sich mit Begriffen wie ‘Bildung’, ‘Berufserfolg’ und ‘soziale Rangordnung’ sinnvoll und nachvollziehbar auf empirische Realität beziehen kann. Aber diesem Problem wird gar nicht nachgegangen. Stattdessen wird ohne jede nähere Überlegung einfach angenommen, daß es ein „irgendetwas", das durch diese Begriffe „bezeichnet" wird, in der Realität schon geben wird und daß das Problem vielmehr darin bestehe, daß man dieses „irgendetwas" nicht beobachten könne, z.B. „eine soziale Rangordnung in der Gesellschaft". Aber woher weiß Kromrey, daß es dies „irgendetwas" in der sozialen Realität gibt? Die Tatsache, daß es das *Wort* ‘soziale Rangordnung’ gibt und daß es durchaus möglich ist, Sätze

[4] Diese Unterscheidung wurde eine Zeitlang von dem Wissenschaftstheoretiker R. Carnap vertreten (vgl. z.B. Carnap 1956) und hat dann auch in der sozialwissenschaftlichen Methodenliteratur eine gewisse Verbreitung gefunden.

[5] Auch Schnell, Hill und Esser 1999, S. 125, verwenden diese merkwürdige Formulierung, daß „theoretische Begriffe nicht direkt beobachtbar" sind.

zu bilden, in denen dieses Wort vorkommt, liefert dafür sicherlich keine Begründung. Man muß sich – darin ist Kromrey natürlich zuzustimmen – mit der sozialen Realität beschäftigen, also untersuchen, wie die sozialen Beziehungen zwischen den Menschen beschaffen sind und ob zum Beispiel einige Verhaltensweisen durch die Annahme verstanden werden können, daß ihre Akteure von Rangunterschieden ausgehen. Vielleicht findet man als Ergebnis einer solchen empirischen Untersuchung auch ein Verständnis des Begriffs ‘soziale Rangordnung’. Aber ohne solche Überlegungen auch nur ansatzweise angestellt zu haben, kann gar nicht davon ausgegangen werden, daß der Begriff irgendeine nachvollziehbare Bedeutung hat; und es wäre natürlich erst recht unsinnig, ihn – vielleicht deshalb? – einen „theoretischen Begriff" zu nennen.

5. Die zunächst wichtige Aufgabe sollte also darin bestehen, die Begriffe zu *explizieren*, die in den weiteren Überlegungen – und ggf. zur Definition statistischer Variablen – verwendet werden sollen.[6] Wie meistens bei denjenigen Autoren, die Indizes zur „Operationalisierung" von „theoretischen Begriffen" vorschlagen, nimmt jedoch Kromreys Gedankengang eine andere Wendung. Sich auf soziale Realität zu beziehen, soll nicht dem Zweck dienen, vorgängige „theoretische Begriffe" in ihrem potentiellen Sinngehalt zu explizieren. Stattdessen möchte man „Indikatoren" für das „irgendetwas" finden, das vermeintlich durch die vorgängigen Begriffe „bezeichnet" wird. Es ist allerdings evident: Wenn diese Begriffe nichts bezeichnen, kann es dafür auch keine Indikatoren geben. Das schließt selbstverständlich nicht aus, Aspekte der sozialen Realität durch statistische Variablen zu erfassen und diese Variablen als Indikatoren zur Bildung von Indizes zu verwenden. Aber es sollte deutlich geworden sein, daß weder die als Indikatoren verwendeten Variablen noch die aus ihnen gebildeten Indizes ohne weiteres als Explikationen vorgängiger „theoretischer Begriffe" aufgefaßt werden können. Es verhält sich genau umgekehrt: Durch Indexbildung werden nicht vorgängige „theoretische Begriffe" operationalisiert, sondern *neue Begriffe gebildet*; Begriffe, die möglicherweise – abhängig von ihrer Explikation – zur Bildung sinnvoller Aussagen verwendet werden können. Das heißt aber auch: Wer einen Index konstruiert, übernimmt dadurch die Verantwortung für den Begriff, den er bildet, und muß zeigen, daß und wie

[6] Die hier erforderliche Explikationsaufgabe hat zunächst gar nichts damit zu tun, „einen vagen, vieldeutigen und unexakten Begriff in einen präzisen, eindeutigen und exakten Begriff" zu überführen, wie dies gelegentlich in der Methodenliteratur gefordert wird; vgl. z.B. Groeben und Westmeyer 1975, S. 57. Denn Begriffe kann man auf *beliebige* Weisen „präzise" machen. Die Explikation eines Begriffs sollte vielmehr zeigen, wie man sich mit ihm auf empirische Realität beziehen kann, d.h. wie man mit seiner Hilfe sinnvolle und verständliche Sätze bilden kann, die etwas über die Beschaffenheit unserer Erfahrungswelt aussagen. Dafür ist es im allgemeinen unzureichend, sich auf Daten zu beziehen (die ja auch erst als Werte definierter Variablen entstehen); sondern man muß auf „Empirie" im Sinne von Erfahrung zurückgreifen, wie dies in der Einleitung erläutert worden ist.

mit dem Begriff sinnvolle Aussagen gemacht werden können.

6. Natürlich ist die Konstruktion von Indizes keineswegs die einzige Möglichkeit, um (theoretische) Begriffe zu bilden. Dies sollte betont werden, da in der Methodenliteratur gelegentlich der Eindruck erweckt wird, als könne die gesamte sozialwissenschaftliche Begriffsbildung auf empirische und konstruierte Variablen reduziert werden. Zum Beispiel geht das oben angeführte Zitat von Schnell, Hill und Esser so weiter:

„Die meisten allgemeinen Konzepte sind viel zu unklar, als daß direkt Meßanweisungen gegeben werden können. Aus diesem Grund beginnt die Operationalisierung eines theoretischen Begriffs meist mit der Klärung, welche theoretischen Aspekte eines bestimmten Gegenstandsbereichs (‘Dimensionen’) durch den theoretischen Begriff bezeichnet werden, bzw. welche Dimensionen das Konzept anspricht. Dieser rein theoretische Klärungsprozeß wird in der deutschsprachigen Literatur etwas überschwenglich in Anlehnung an Zetterberg (1973) meist als ‘dimensionale Analyse’ bezeichnet. Da der Begriff ‘dimensionale Analyse’ in den Naturwissenschaften eine andere Bedeutung besitzt, nämlich die Bestimmung der ‘Dimension’ einer physikalischen Größe (z.B. besitzt die Fallbeschleunigung die Dimension m/s^2), verwenden wir den Begriff ‘Konzeptspezifikation’ in Anlehnung an die angelsächsischen Begriffe ‘conceptualization’ und ‘concept specification’.“ (S. 122)

Etwas später (S. 124) heißt es dann:

„Ein theoretischer Begriff behauptet die Existenz (mindestens) einer Dimension, auf der Objekte unterschieden werden können. Im einfachsten Fall können Objekte (‘Merkmalsträger’, z.B. Moleküle, Computer, Personen, Gesellschaften, Sonnensysteme etc.) lediglich danach unterschieden werden, ob ihnen ein bestimmtes Merkmal (oder eine bestimmte Eigenschaft) zukommt oder nicht. Welche Merkmale der theoretische Begriff bezeichnet, liegt als Ergebnis der Konzeptspezifikation vor. Man kann Merkmale danach klassifizieren, ob sie allen oder nur einigen Objekten des Gegenstandsbereichs zukommen. Als ‘Konstante’ bezeichnet man Eigenschaften, die allen jeweils betrachteten Objekten zukommen; z.B. die Eigenschaft ‘enthält mindestens ein Kohlenstoffatom’ ist für alle lebenden Organismen eine Konstante. Eigenschaften, die bei verschiedenen Merkmalsträgern in mehr als einer Ausprägung vorkommen, werden als ‘Variablen’ bezeichnet.“

Daß es sich bei der Klärung von Begriffen um einen „rein theoretischen Klärungsprozeß“ handelt, haben wir bereits im Zusammenhang mit Kromreys Gedankengang kritisiert (man vgl. auch Bulmer 1979). Es geht vielmehr darum, Begriffe zu finden, mit denen man sich sinnvoll auf Aspekte unserer Erfahrungswelt beziehen kann, und sie durch eine Darlegung ihrer *empirischen* Bezüge und Erläuterung von Sätzen, die man mit ihnen bilden kann, zu explizieren. Noch irreführender ist jedoch die Vorstellung, daß „Konzeptspezifikation“ stets darauf hinauslaufen müsse, Merkmalsräume für statistische Variablen zu definieren.[7] Dies ist zwar erforderlich, wenn *statistische* Sozialforschung betrieben werden soll. Aber die Forderung, daß sozialwissenschaftliche Überlegungen stets als Aussagen über Variablen bzw. Zusammenhänge zwischen Variablen formuliert werden müßten, führt sicherlich in eine theoretische Sackgasse.

[7]Besonders gedankenlos sagen C. und D. Nachmias 1981, S. 56: „Concepts are converted into variables by *mapping* them into a set of values.“ Es sei aber betont, daß dies Reden von ‘mapping’ nichts mit unserer Definition von Variablen als Abbildungen im Sinne mathematischer Funktionen zu tun hat.

6.3 Ein problematisches Beispiel

1. Um einige der im vorangegangenen Abschnitt angedeuteten Probleme zu illustrieren, besprechen wir einen Index, den E. K. Scheuch und H. Daheim (1961) vorgeschlagenen haben, um Menschen durch ihre Zugehörigkeit zu unterschiedlichen „sozialen Schichten“ zu charakterisieren.[8] Den Ausgangspunkt bilden umgangssprachliche Vorstellungen über „soziale Schichten“, „sozialen Status“ und „Sozialprestige“. Bei Scheuch und Daheim steht vor allem der Begriff ‘Sozialprestige’ im Mittelpunkt. Ihr Versuch, eine Begriffsklärung vorzunehmen, mündet zunächst in folgende Definition:

„Unter Sozialprestige verstehen wir den Besitz einer Konjunktion von Eigenschaften, welche die Partner in kategorialen Kontakten veranlassen, konsistent jenen Akteuren einen höheren Rang in Bezug auf die gemeinsamen Werte der Gesamtgesellschaft zuzuerkennen, welche diese Eigenschaften zu einem höheren Grade besitzen.“ (Scheuch und Daheim 1961, S. 72)

Folgt man dieser Definition, soll mit dem Wort ‘Sozialprestige’ darauf verwiesen werden, daß Menschen in manchen gesellschaftlichen Situationen sich wechselseitig durch „Rangunterschiede“ einstufen. Jeder kennt dafür Beispiele aus hierarchisch verfaßten Organisationen. Ob das auch außerhalb hierarchischer Organisationen, im gewöhnlichen öffentlichen und privaten Leben der Fall ist, ist zwar fragwürdig, wird jedoch von Scheuch und Daheim nicht untersucht. Sie nehmen vielmehr ihre Definition zum Ausgangspunkt, um sich auf folgende Weise auf soziale Situationen zu beziehen:

„Besonders wichtig ist die Spezifizierung: ‘Kategoriale Kontakte’. Hiermit meinen wir Beziehungen in einer Art Halbdistanz, wo die Partner in der Interaktion vornehmlich als Merkmalsträger wahrgenommen werden. [...] Mit dieser Spezifizierung des Bezugssystems der Akteure verbinden wir ein anschauliches Modell für die Zuerkennung von Prestige. Stellen wir uns eine Begegnung zwischen Fremden vor, bei der die Akteure beim Partner zunächst eine Reihe von Merkmalen wahrnehmen, die sie zur Definition der Situation zu benötigen glauben. Mit ‘Definition der Situation’ beziehen wir uns auf eine in Deutschland weithin unbeachtet gebliebene Konzeption, die von George H. Mead entwickelt worden

[8]Dieses Beispiel wird u.a. auch von Schnell, Hill und Esser 1999, S. 163, und von Bortz und Döring 1995, S. 135, behandelt.

Tabelle 6.3.1 Variable X_1: Beruflicher Status.

		$\tilde{X}_1$
Arbeiter:	ungelernte Arbeiter	1
	angelernte Arbeiter	4
	gewöhnliche Facharbeiter	9
	höchstqualifizierte Facharbeiter	13
Angestellte:	ausführende Angestellte	10
	qualifizierte Angestellte	16
	leitende Angestellte	27
Beamte:	untere Beamte	10
	mittlere Beamte	16
	leitende Beamte	23
Selbständige:	kleine Selbständige	15
	mittlere Selbständige	20
	führende Selbständige	30
	freie Berufe, intellektuelle Berufe	25
	kleine Landwirte (5ha / 20 Morgen)	15
	mittlere Landwirte (20 ha / 80 Morgen)	20
	große Landwirte	23

Bei Arbeitslosen wird von der jeweiligen Punktzahl für den Beruf 1 Punkt abgezogen.

ist. Es ist nun offensichtlich eine empirisch entscheidbare Frage, welche Informationen der Partner zu benötigen glaubt und welche Informationen der Akteur selbst gibt, damit gegenseitige Sicherheit in den Verhaltenserwartungen erreicht wird. [...] Den Mechanismus der Zuerkennung von Sozialprestige verstehen wir entsprechend als Konsequenz eines primären Bemühens um die Strukturierung von Situationen. Eine empirische Untersuchung dieses Prozesses erlaubt uns damit Aussagen über Sozialprestige. Zugleich können die beobachtete Auswahl von Informationen und deren Bewertung in 'Situationen der Halbdistanz' auch als Indiz dienen für das Schichtungssystem einer konkreten Gesellschaft.“ (S. 72f)

Man könnte vermuten, daß Scheuch und Daheim nach diesen Überlegungen zur Begriffsklärung sich in einem zweiten Schritt Gedanken darüber machen, wie sich der von ihnen unterstellte Prozeß der wechselseitigen Zuweisung von Rang- oder Prestigeunterschieden in konkreten „Situationen der Halbdistanz“ tatsächlich abspielt und ob das von ihnen angedeutete „Modell“ sich eignet, um solche Situationen zu verstehen. Das ist jedoch nicht der Fall, sondern die Autoren machen etwas anderes: Sie konstruieren einen Schichtindex, der mit der Vorstellung einer wechselseitigen Einschätzung von Statusunterschieden in sozialen Situationen nichts zu tun hat.

2. Sehen wir uns an, wie Scheuch und Daheim ihren Schichtindex konstruieren. Sie definieren zunächst drei Variablen, die wir in unserer Notation mit X_1, X_2, X_3 bezeichnen. Es handelt sich um empirische Variablen, deren Werte durch eine Befragung ermittelbar sind. Die Tabellen 6.3.1–3 zeigen, wie die Variablen durch Scheuch und Daheim (1961, S. 102f) definiert worden sind. Aus den drei Variablen wird zunächst ein einfacher additiver Index $X^* := X_1 + X_2 + X_3$ gebildet; und dieser Index wird schließlich verwendet, um die Personen, die in der Befragung Angaben zu den drei Variablen gemacht haben, in jeweils eine von sechs Schichten einzuteilen. Wie Scheuch und Daheim diese Schichten gebildet haben, ersieht man aus Tabelle 6.3.4.

Tabelle 6.3.2 Variable X_2: Einkommen.

				$\tilde{X}_2$
Nettoeinkommen:			unter 150 DM	1
	150	—	299 DM	3
	300	—	399 DM	5
	400	—	499 DM	6
	500	—	599 DM	8
	600	—	699 DM	9
	700	—	799 DM	10
	800	—	899 DM	13
	900	—	999 DM	13
	1000	—	1499 DM	16
	1500	—	1999 DM	19
	2000 und mehr			20

Der Interviewer hatte durch Nachfragen das Nettoeinkommen des Hauptenährers festzustellen.

3. Wenn man diesen Schichtindex ernstnehmen will, muß man zunächst feststellen, daß er mit der oben skizzierten theoretischen Vorstellung – daß Rang- oder Prestigeunterschiede situationsspezifisch wahrgenommen bzw. erzeugt werden – nichts zu tun hat. Wenn dieser Schichtindex irgendetwas zeigt, dann nur, wie sich Scheuch und Daheim eine Gliederung der Gesellschaft in Schichten vorstellen. Tatsächlich nicht nur eine Gliederung, sondern eine Hierarchie, denn ihr Schichtindex soll ja dazu dienen, die Gesellschaftsmitglieder in eine soziale Rangordnung einzuordnen.

4. Die Kritik kann noch etwas präzisiert werden. Folgt man nämlich den vorgängigen theoretischen Überlegungen von Scheuch und Daheim, dann *entstehen* soziale Rangunterschiede dadurch, daß sie in konkreten sozialen Situationen von den Beteiligten *gemacht* werden; es handelt sich um wechselseitige Zuschreibungen (soweit solche Zuschreibungen tatsächlich stattfinden). Bei „sozialem Status“ oder „Prestige“ oder „Schichtzugehörigkeit“ handelt es sich also nicht um empirisch ermittelbare Eigenschaften von Menschen, so wie man zum Beispiel ihren Beruf, ihr Einkommen und ihre Schulbildung ermitteln kann, sondern um situationsspezifisch vorgenommene Bewertungen. Was sie zum Ausdruck bringen, besteht also darin, wie Menschen sich wechselseitig bewerten; sie zeigen, wie bewertet wird,

Tabelle 6.3.3 Variable X_3: Schulbildung.

		$\tilde{X}_3$
Schulbildung:	Volksschule, unvollständig	0
	Volksschule ohne Lehre	2
	Volksschule mit Lehre	4
	Handelsschule oder Mittelschule ohne Abschluß	5
	Höhere Schule bis Obertertia	7
	Mittlere Reife	9
	Höhere Schule länger als Untersekunda, ohne Abitur	11
	Höhere Fachschule mit Abschluß	12
	Abitur	14
	Hochschule ohne Abschluß	18
	Hochschule mit Abschluß	20

und liefern mithin bestenfalls eine Information über denjenigen, der die Bewertung jeweils vornimmt. Mit der Konstruktion eines Schichtindex wird jedoch dieser Sachverhalt nicht nur unkenntlich gemacht, sondern verdreht. Es wird der Eindruck erweckt, als könne man „Schichtzugehörigkeit“ – oder „sozialen Status“ oder „Prestige“ – als eine Eigenschaft von Menschen konzipieren, die sich dann so ermitteln läßt, wie man z.B. ihren Beruf, ihr Einkommen und ihre Schulbildung feststellen kann. Sobald man den Fehler erkennt, wird jedoch deutlich, daß der Schichtindex ebenfalls nur als ein Instrument der Bewertung aufgefaßt werden kann. Er zeigt, wie Scheuch und Daheim oder diejenigen, die diesen oder ähnliche Indizes verwenden, Menschen im Hinblick auf Statusunterschiede *bewerten* wollen.

5. Indizes werden oft als „Meßinstrumente“ bezeichnet. Auch Scheuch und Daheim (1961, S. 67) deuten dies an, wenn sie sagen, daß ihr Schichtindex „als standardisiertes Maß bei Umfragen eingesetzt“ werden sollte. Aber die zuvor angestellte Überlegung zeigt, daß – jedenfalls in diesem Beispiel – die Vorstellung, daß mit einem Index eine „Schichtzugehörigkeit“ von Menschen *gemessen* werden kann, irreführend ist. Die Konstruktion eines Schichtindex liefert kein Meßverfahren, sondern zeigt, wie seine Konstrukteure die Gesellschaftsmitglieder in Schichten einteilen möchten. Oder etwas anders gesagt: Schichtindizes liefern keine Informationen über die Beschaffenheit der Gesellschaft, sondern bestenfalls über Gesellschaftsbilder ihrer Konstrukteure. Schnell, Hill und Esser (1999, S. 165) sprechen in diesem Zusammenhang von „willkürlicher Messung“. Aber das ist nicht nur eine merkwürdige Formulierung, sondern ebenfalls irreführend. Denn es handelt sich hier nicht um eine „willkürliche“, sondern überhaupt nicht um eine Messung, vielmehr um eine Bewertung.

6. Faßt man den Schichtindex von Scheuch und Daheim als einen Vorschlag zur „Operationalisierung“ des Begriffs ‘soziale Schicht’ (oder ‘sozialer Sta-

Tabelle 6.3.4 Bildung von Schichten (Scheuch und Daheim 1961, S. 103).

0	$\leq$	X^*	$\leq$	14	untere Unterschicht
15	$\leq$	X^*	$\leq$	22	obere Unterschicht
23	$\leq$	X^*	$\leq$	29	untere Mittelschicht
30	$\leq$	X^*	$\leq$	39	mittlere Mittelschicht
40	$\leq$	X^*	$\leq$	49	obere Mittelschicht
50	$\leq$	X^*	$\leq$	63	Oberschicht

tus’ oder ‘Prestige’) auf, kann man anhand des Beispiels noch eine etwas allgemeinere Einsicht gewinnen: daß nämlich Operationalisierung keineswegs garantiert, daß ein Begriff entsteht, mit dem man sich sinnvoll auf empirische Realität beziehen kann. Hätte man in diesem Beispiel die erste der beiden auf S. 79 im Anschluß an Schnell, Hill und Esser unterschiedenen Aufgaben ernsthaft verfolgt, hätte man sich sogleich klarmachen können, daß die Begriffe ‘soziale Schicht’, ‘sozialer Status’ und ‘Prestige’ überhaupt nicht operationalisiert werden können, weil sie gar nicht auf empirisch ermittelbare Eigenschaften von Menschen oder Situationen, in denen sie sich befinden, verweisen. Wenn man jedoch auf die durch das Wort ‘Operationalisierung’ suggerierte Rhetorik verzichtet, wird ganz unklar, welchen Sinn es haben könnte, die drei Merkmale Beruf, Einkommen und Schulbildung in der von Scheuch und Daheim vorgeschlagenen oder in irgendeiner anderen Weise zu einem Index zusammenzufassen. Das Problem ist ja nicht nur, daß es beliebig viele unterschiedliche Möglichkeiten gibt, aus diesen drei Merkmalen einen Index zu bilden, sondern daß es überhaupt keinen Grund gibt, einen solchen Index zu konstruieren.

7. Hier sei noch einmal an die Unterscheidung zwischen statistischen und Skalierungsverfahren erinnert. Unsere Kritik richtet sich dagegen, daß von Scheuch und Daheim ein Schichtindex zur willkürlichen Skalierung der Mitglieder einer Gesellschaft konstruiert wird. Es gibt keinerlei Begründung dafür, daß Sozialwissenschaftler den Gesellschaftsmitgliedern ausgehend von ihrem Einkommen, ihrem Beruf und ihrer Schulbildung Rangunterschiede überhaupt *geben* sollten. Ein wesentlich anderer Kontext entsteht jedoch, wenn sich das Erkenntnisinteresse darauf richtet, statistische Zusammenhänge zwischen Variablen zu ermitteln. Scheuch und Daheim deuten diesen alternativen Kontext an, wenn sie davon sprechen, daß ihr Schichtindex auch „als unabhängige Variable in mehrdimensionalen Aufgliederungen“ dienen soll (S. 67). Aber dann gibt es tatsächlich keinerlei Grund, einen Index zu verwenden, sondern man kann direkt mit den jeweils interessierenden empirischen Variablen operieren. Zum Beispiel könnte man untersuchen, ob es einen Zusammenhang zwischen der Dauer der Arbeitslosigkeit und dem Beruf, der Schulbildung und dem zuletzt verdienten Einkommen gibt. Selbst bei einem linearen Regressionsansatz ergäbe sich eine größere Flexibilität bei der Ermittlung von Variablenzusammenhängen, wenn man die Regressorvariablen nicht zuvor zu einem Index zusammenfaßt.[9] Aus einem solchen Erkenntnisinteresse kann also kein Argument zur Bildung eines Schichtindex gewonnen werden.

6.4 Datenreduktion durch Indizes

1. Mit Indexbildung ist stets eine „Datenreduktion" verbunden, d.h. aus den Werten eines Index können im allgemeinen nicht die Werte der Variablen rekonstruiert werden, aus denen der Index gebildet worden ist. In diesem Abschnitt soll besprochen werden, ob und ggf. in welcher Weise das auf ein Problem verweist, das bei der Indexbildung berücksichtigt werden sollte. Zunächst ein allgemeiner Hinweis zum Begriff der „Datenreduktion", der oft im Zusammenhang mit statistischen Verfahren verwendet wird. Folgende Bemerkung des Statistikers R. A. Fisher zeigt die Intention:

„Briefly, and in its most concrete form, the object of statistical methods is the reduction of data. A quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole, or which, in other words, shall contain as much as possible, ideally the whole, of the relevant information contained in the original data." (Fisher 1922, S. 311)

Obwohl Fisher nicht in erster Linie daran interessiert war, kann dies Zitat auch helfen, den Sinn statistischer Methoden in der Sozialforschung zu verstehen. Will man sich ein Bild gesellschaftlicher Verhältnisse machen, soweit sie durch unmittelbare Erfahrungen nicht zugänglich sind, müssen zunächst Daten über soziale Sachverhalte erhoben werden. Diese Daten ergeben jedoch nicht unmittelbar ein Bild der gesellschaftlichen Verhältnisse, aus denen sie stammen. Ein solches Bild muß vielmehr durch geeignete Aufbereitung und „Reduktion" der Daten konstruiert werden. Die für die Statistik grundlegende Form der Datenreduktion besteht darin, daß man anstelle der individuellen Daten ihre statistischen Verteilungen betrachtet, wie dies in Abschnitt 4.2 erläutert worden ist.

[9] Bereits ein einfacher linearer Regressionsansatz

$$\mathrm{M}(Y|X_1 = x_1, \ldots, X_m = x_m) = \beta_0 + x_1\beta_1 + \cdots + x_m\beta_m$$

macht das deutlich. Bei der Indexbildung werden die Gewichte $\beta_1, \ldots, \beta_m$ vor der Durchführung der Regressionsrechnung mehr oder weniger willkürlich festgesetzt. Dagegen kann die Regressionsrechnung zeigen, wie die einzelnen Indikatoren $X_1, \ldots, X_m$ mit der abhängigen Variablen Y zusammenhängen. Gelegentlich zeigt sich dann auch, daß die Koeffizienten sogar ihre Vorzeichen wechseln können; man vgl. z.B. die Tabellen 2 und 3 bei Birchfield und Crepaz 1998, S. 189f. Auch die Autoren einer neueren Arbeit, die sich mit Weiterentwicklungen des Schichtindex von Scheuch und Daheim beschäftigt, bemerken: „Eine Aggregation zu dimensionsübergreifenden Schichtindizes kann nur mit Einschränkungen empfohlen werden. Immer dann, wenn die genauen Zusammenhänge zwischen sozialer Ungleichheit und dem zu untersuchenden Phänomen im Vordergrund stehen, wird die getrennte Analyse der Einzeldimensionen empfohlen." (Jöckel u.a. 1998, S. 20)

2. Datenreduktion findet aber auch bei der Konstruktion von Indizes statt. Hat man einen Index $X^* := g(X_1, \ldots, X_m)$ konstruiert, können aus den Werten von X^* im allgemeinen die Werte der Variablen $X_1, \ldots, X_m$ nicht rekonstruiert werden. Man denke an das Klausurbeispiel aus Abschnitt 6.1. Kennt man die Anzahl der Punkte, die insgesamt erzielt worden sind, kann man daraus im allgemeinen nicht entnehmen, wieviele Punkte in den einzelnen Aufgaben erzielt worden sind. Um zu zeigen, worin die Datenreduktion besteht, können zwei Äquivalenzrelationen verwendet werden.[10] Erstens impliziert ein Index eine Äquivalenzrelation für die Mitglieder der Gesamtheit Ω. Zu jedem möglichen Indexwert $\tilde{x}^* \in \tilde{\mathcal{X}}^*$ gibt es eine Äquivalenzklasse

$$K(\tilde{x}^*) := \{\omega \in \Omega \,|\, X^*(\omega) = \tilde{x}^*\}$$

Zwei Mitglieder werden durch den Index genau dann als „für den Indexbegriff äquivalent" angesehen, wenn sie derselben Äquivalenzklasse angehören. Zweitens impliziert ein Index auch eine Äquivalenzrelation für den Merkmalsraum der Indikatorvariablen, also für $\tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$ Die Äquivalenzklassen haben in diesem Fall die Form

$$\tilde{K}(\tilde{x}^*) := \{(\tilde{x}_1, \ldots, \tilde{x}_m) \,|\, g(\tilde{x}_1, \ldots, \tilde{x}_m) = \tilde{x}^*\}$$

bestehen also aus allen Merkmalskombinationen der Indikatorvariablen, die zum gleichen Indexwert führen.

3. Wichtig ist nun, auf den unterschiedlichen Verwendungszweck zu achten. Statistische Verfahren dienen einer Charakterisierung von Variablen bzw. der Gesamtheiten, für die die Variablen definiert sind, und ggf. einer Charakterisierung von Zusammenhängen zwischen Variablen. Die wesentliche Datenreduktion resultiert aus diesem Zweck und findet statt, indem man statistische Verteilungen betrachtet. Indizes dienen jedoch der Klassifikation oder Skalierung von Individuen bzw. Objekten. Als Beispiel kann der Schichtindex von Scheuch und Daheim dienen. Jeder mögliche Indexwert faßt tatsächlich zahlreiche sehr unterschiedliche Konstellationen von Beruf, Einkommen und Ausbildung zusammen. Wie in Tabelle 6.4.1 gezeigt wird, gibt es z.B. zum Indexwert 24 insgesamt 30 verschiedene Wertekonstellationen der drei Variablen, auf denen der Index beruht. Und hinter den von Scheuch und Daheim gebildeten „Schichten" (Tabelle

[10] Der Begriff einer *Äquivalenzrelation* kann allgemein für beliebige Mengen definiert werden. Zur Erläuterung beziehen wir uns auf den Merkmalsraum $\tilde{\mathcal{X}}$ einer statistischen Variablen. Zur Notation verwenden wir das Symbol $\equiv$. Man spricht von einer Äquivalenzrelation, wenn folgende Bedingungen erfüllt sind: (a) Für alle $\tilde{x} \in \tilde{\mathcal{X}} : \tilde{x} \equiv \tilde{x}$ (Reflexivität); (b) für alle $\tilde{x}, \tilde{x}' \in \tilde{\mathcal{X}} : \tilde{x} \equiv \tilde{x}' \iff \tilde{x}' \equiv \tilde{x}$ (Symmetrie); (c) für alle $\tilde{x}, \tilde{x}'\tilde{x}'' \in \tilde{\mathcal{X}} : \tilde{x} \equiv \tilde{x}'$ und $\tilde{x}' \equiv \tilde{x}'' \implies \tilde{x} \equiv \tilde{x}''$ (Transitivität). Gibt es für eine Menge eine Äquivalenzrelation, kann man sie in Äquivalenzklassen partitionieren, die jeweils aus all denjenigen Elementen bestehen, die untereinander äquivalent sind.

Tabelle 6.4.1 Merkmalskombinationen für den Wert $\tilde{x}^* = 24$ des Schichtindex.

$\tilde{X}_1$	$\tilde{X}_2$	$\tilde{X}_3$	$\tilde{X}_1$	$\tilde{X}_2$	$\tilde{X}_3$	$\tilde{X}_1$	$\tilde{X}_2$	$\tilde{X}_3$
1	3	20	4	6	14	9	1	14
1	5	18	4	8	12	9	3	12
1	9	14	4	9	11	9	6	9
1	16	7	4	13	7	9	8	7
1	19	4	4	16	4	9	10	5
			4	20	0	9	13	2
10	3	11	13	6	5	15	5	4
10	5	9	13	9	2	15	9	0
10	9	5						
10	10	4						
16	1	7	23	1	0			
16	3	5						
16	6	2						
16	8	0						

6.3.3) verbergen sich natürlich noch viel mehr unterschiedliche Kombinationen. Aber es gibt keinerlei Begründung, warum die Klassifizierung auf diese und nicht auf irgendeine andere Weise vorgenommen werden sollte. Es gibt tatsächlich nicht einmal eine Begründung dafür, warum überhaupt eine Klassifizierung zweckmäßig sein könnte.

4. Daraus kann zwar keine grundsätzliche Kritik von Indexbildungen abgeleitet werden, die Überlegung verweist jedoch auf ein Begründungsproblem. Wer einen Index konstruiert, muß einen Grund für die Zweckmäßigkeit und den Sinngehalt der durch den Index induzierten Äquivalenzrelationen angeben können. Zum Beispiel wird bei dem in Abschnitt 6.1 besprochenen Index zur Bewertung von Statistikklausuren eine solche Begründung durch das Argument geliefert, daß jede Punktzahl gleichermaßen zum Gesamtergebnis beitragen soll. Eine sich anschließende Überlegung könnte hier ansetzen und dies unter Hinweis auf die Konzeption einzelner Punkte in Frage stellen. Immerhin wird eine Begründung gegeben, an die eine Diskussion anschließen kann. Bei dem Schichtindex von Scheuch und Daheim wird jedoch keinerlei Begründung für die implizierten Äquivalenzrelationen gegeben; infolgedessen kann auch über den möglichen Sinn des Index nicht diskutiert werden.

6.5 Quantifizierende Indizes

1. C. und D. Nachmias (1981, S. 267) geben folgende Gründe an, die Indexkonstruktionen in der Sozialforschung sinnvoll erscheinen lassen:

„Scales and indexes are employed in the social sciences for several reasons. First, they enable several variables to be represented by a single score that reduces the complexity of the data. Second, scales and indexes provide quantitative measures that are amenable to greater precision and statistical manipulation. Finally, indexes and scales increase the reliability of measurement.“

Daß die erste dieser drei Begründungen fragwürdig ist, wurde bereits in den vorangegangenen Abschnitten besprochen. Das dritte Argument stellen wir zunächst zurück. Wie aber verhält es sich mit dem Anspruch, daß man durch Indizes quantitative Variablen (sozialwissenschaftliche Größenbegriffe) konstruieren kann? Um diese Frage zu diskutieren, beziehen wir uns im folgenden auf *additive* Indizes, also Indizes der Form

$$X^* := \sum_{j=1}^{m} w_j X_j$$

wobei $w_1, \ldots, w_m$ irgendwie gewählte Gewichte sind. Sicherlich kann man X^* bestenfalls dann als eine quantitative Variable interpretieren, wenn bereits die Indikatoren $X_1, \ldots, X_m$ quantitative Variablen sind. Dies ist jedoch nur eine notwendige, keine hinreichende Bedingung.

2. Betrachten wir zwei Beispiele, um das Problem zu verdeutlichen. Zunächst den Index für die Ergebnisse einer Statistik-Klausur, der in Abschnitt 6.1 besprochen worden ist. In diesem Fall scheint es keine Schwierigkeiten zu machen, den Index quantitativ zu interpretieren. Bei den Indikatoren handelt es sich um Zählgrößen, mit denen Punkte gezählt werden, wobei jeder Punkt per Definition die gleiche Bedeutung hat, gleichgültig in welcher der fünf verschiedenen Aufgaben er erzielt worden ist. Infolgedessen erscheint es unproblematisch, die Einzelergebnisse additiv zu einer neuen Zählgröße zusammenzufügen. Anders verhält es sich jedoch bei dem Schichtindex, der in Abschnitt 6.3 besprochen worden ist. Hier stellt sich zunächst das Problem, daß es sich bei zwei der drei Indikatoren (beruflicher Status und Ausbildung) nicht um quantitative Variablen handelt, so daß schon deshalb daraus kein quantifizierender Index gebildet werden kann.[11] Aber sehen wir von diesem Problem ab und nehmen wir an, daß

[11] Für beruflichen Status ist das unmittelbar deutlich. Bei der Konzeption von „Bildung“ als Variable wird oft von der Vorstellung eines „Bildungsniveaus“ ausgegangen, die eine Quantifizierbarkeit suggeriert. Daß es sich jedoch auch in diesem Fall um eine fragwürdige Vorstellung handelt, zeigen folgende Bemerkungen von W. Müller 1979, S. 198: „Wenn wir uns nochmal einige Ausführungen in Erinnerung rufen, die oben gemacht wurden zur historischen Entwicklung des Bildungswesens, seiner qualitativen Veränderungen aufgrund gesetzlicher Regelungen, seiner institutionellen Verfesti-

ein Ausbildungsgrad quantifiziert werden könnte. Dennoch gäbe es Schwierigkeiten, aus der additiven Zusammenfassung von Ausbildungsgrad und Einkommenshöhe einen neuen quantifizierenden Index zu bilden. Denn es ist evident, daß diese Variablen eine ganz unterschiedliche Bedeutung haben und deshalb nicht additiv zusammengefaßt werden können.

3. In der Methodenliteratur wird das Problem manchmal so formuliert: Quantitative Indikatoren können nur dann zu einem quantifizierenden Index additiv zusammengefaßt werden, wenn sie sich auf die gleiche „Dimension" beziehen. Schnell, Hill und Esser (1999, S. 174f) sagen das so:

„Alle Items einer Skala sollten nur eine Dimension (z.B. eine zu messende Einstellung) erfassen, die resultierende Skala sollte also ‘eindimensional’ sein. [...] Daß eine Skala eindimensional sein muß, wenn sie brauchbar sein soll, läßt sich an einem Beispiel von Holm (1976) verdeutlichen. Würde eine Skala zwei verschiedene Dimensionen A und B messen, so ergäbe ein Skalenwert von ‘100’ weder eine Information über die Dimension A noch über die Dimension B. Holm (1976: 125) schreibt dazu: ‘Es ist, wie wenn man die Frage stellt: Auf einem Schiff stellt der Kapitän insgesamt 100 Grad fest. Auf dem wievielten Längengrad befindet sich das Schiff und wieviel Grad Celsius herrschen an Bord (wenn beide zusammen 100 ergeben)?’"

Um diesen Gedanken zu verfolgen, muß man zunächst verstehen, was mit dem Wort ‘Dimension’ gemeint sein soll.[12] Leider sind die in der Methodenliteratur dazu gegebenen Erläuterungen ziemlich unklar. Z.B. erklären Schnell, Hill und Esser (1999, S. 122) „Dimensionen" als „theoretische Aspekte eines bestimmten Gegenstandsbereichs" (zum Kontext vgl. man die in Abschnitt 6.2 angeführten Zitate); und bei Kromrey (1994, S. 68) handelt es sich um „Aspekte des empirischen Untersuchungsgegenstands". Etwas ausführlicher heißt es dann:

„Unter ‘Dimensionen’ sind also diejenigen Einzelheiten zu verstehen, die an einem empirischen Sachverhalt unterschieden werden können. Je nach Fragestellung kann diese dimensionale Unterscheidung weit vorangetrieben werden (im Beispielsfall etwa bis hin zu Individualmerkmalen wie Alter, Familienstand, Kinderzahl), oder man wird sich auf einer höheren Abstraktionsstufe bewegen (z.B. wenn Dimensionen der Lebensqualität in verschiedenen Regionen zu untersuchen sind). Auf derjenigen Ebene, bis zu der die dimensionale Unterscheidung vorangetrieben wird, setzt der Forscher für die zu unterscheidenden Dimensionen Begriffe ein, die zu definieren und für die gegebenenfalls Indikatoren festzulegen sind." (Kromrey 1994, S. 73)

gung und seiner quantitativen Entwicklung, die zumindest ebensosehr auf gesetzliche Veränderungen und administrative Maßnahmen wie individuelle Verhaltensänderungen zurückzuführen ist, wenn wir dies bedenken, dann müssen alle Versuche einer einheitlichen metrischen Skalierung von erworbener Bildung mehr als zweifelhaft erscheinen."

[12]Hierbei geht es natürlich nicht um die rein formale Bedeutung, in der von mehrdimensionalen Variablen oder Merkmalsräumen gesprochen wird.

Diese Erläuterungen sind schwer verständlich. Denkt man jedoch daran, daß diese Autoren stets auch betonen, daß „Dimensionen" durch statistische Variablen erfaßt werden sollen, könnte man versuchen, den Sprachgebrauch so zu verstehen: daß man sich mit dem Wort ‘Dimension’ auf die Begriffe beziehen möchte, die zur Explikation von Merkmalsräumen statistischer Variablen verwendet werden. Dann könnte man auch folgende Bemerkung von Laatz (1993, S. 27) verstehen:

„Sowohl bei deskriptiven als auch explanativen Studien werden die Einzelfälle und die Gesamtpopulation unter ausgewählten interessierenden Gesichtspunkten betrachtet. Diese Gesichtspunkte nennen wir *Dimensionen.*"

Es ist jedoch nicht sinnvoll, bei jedem beliebigen Merkmalsraum (oder gar jedem beliebigen Begriff) von einer ‘Dimension’ zu sprechen. Denkt man an die Geschichte des Begriffs in der Mathematik und Naturphilosophie,[13] muß eine wesentliche Bedingung hinzukommen: Es muß sich um einen Größenbegriff handeln. Halten wir vorläufig an diesem Verständnis fest, sollte also das Wort ‘Dimension’ nur bei quantitativen Merkmalsräumen verwendet werden, deren Bedeutung sich durch die Bezugnahme auf einen Größenbegriff explizieren läßt.[14]

4. Hält man sich an diesen Dimensionsbegriff, ergibt sich folgende Überlegung: Eine wesentliche Sinnvoraussetzung, um aus quantitativen Indikatorvariablen $X_1, \ldots, X_m$ einen quantitativ interpretierbaren additiven Index zu bilden, besteht darin, daß sich alle Indikatoren auf eine gleiche Dimension beziehen, d.h. daß jeder ihrer Merkmalsräume durch eine Bezugnahme auf den gleichen Größenbegriff expliziert werden kann. Wir nennen dies das *Prinzip der dimensionalen Homogenität.*[15] Von dimensionaler Homogenität kann man z.B. bei dem Index für die Klausurergebnisse und bei Preisindizes sprechen, nicht jedoch bei dem Schichtindex von Scheuch und Daheim.

5. Akzeptiert man dieses Prinzip der dimensionalen Homogenität, folgt daraus, daß additive Indizes im allgemeinen nicht als quantifizierende Indizes aufgefaßt werden können; denn im allgemeinen beziehen sich die zur Konstruktion verwendeten Indikatorvariablen auf unterschiedliche Dimensionen (soweit man überhaupt voraussetzen kann, daß mit den Indikatoren auf Größenbegriffe Bezug genommen wird). Eine Ausnahme wären nur

[13]Vgl. dazu die ausführlichen Darlegungen von Gosztonyi 1976, insb. S. 1046ff.

[14]In Abschnitt 14.5 werden wir uns noch einmal etwas ausführlicher mit dem Dimensionsbegriff beschäftigen.

[15]Es wurde in der Physik durch die Entwicklung von Verfahren zur „Dimensionsanalyse" durch J. Fourier explizit gemacht. Sich darauf beziehend schrieb W. S. Jevons (1958, S. 327): „According to the *Principle of Homogeneity*, all the quantities *added* together, and equated in any equation, must have the same dimensions. Hence if, on estimating the dimensions of the terms in any equation, they be not homogeneous, some blunder must have been committed." Zur Geschichte der Dimensionsanalyse vgl. Roche 1998, Kap. 10.

diejenigen Indizes, bei deren Indikatoren es sich um Zählgrößen handelt, die vergleichbare Einheiten zählen, und Indizes, bei deren Indikatoren es sich um monetäre Größen handelt.[16] Es ist deshalb verständlich, daß an Quantifizierung interessierte Sozialforscher immer wieder nach Möglichkeiten gesucht haben, das Kriterium der dimensionalen Vergleichbarkeit zu unterlaufen. Die Möglichkeit dafür liegt darin, daß es sich bei diesem Kriterium um eine *Sinnvoraussetzung* additiver Indizes handelt, über die nicht formal entschieden werden kann. Denn ob sich zwei oder mehr Indikatorvariablen auf die gleiche Dimension beziehen oder nicht, ist in gewisser Weise auch eine Interpretationsfrage. So könnten Scheuch und Daheim zum Beispiel sagen, daß sich Ausbildung und Einkommen durchaus auf eine gleiche Dimension beziehen, die sie dann als „sozialen Status“ oder wie auch immer interpretieren.

6. Es gibt hauptsächlich zwei – komplementäre – Strategien, um das Sinnkriterium der dimensionalen Vergleichbarkeit zu unterlaufen. Die eine Strategie besteht darin, dimensionale Vergleichbarkeit als eine Interpretationsfrage aufzufassen. Dies kommt zum Beispiel in der häufig verwendeten Rhetorik zum Ausdruck, daß mit Indikatoren „Aspekte“ oder „Teildimensionen“ einer „übergreifenden Dimension“ erfaßt werden können, die dann die Bedeutung des aus den Indikatoren gebildeten Index ausdrückt. Es ist klar, daß schon das Wort ‘Indikator’ hier einen großen Spielraum schafft. Denn man braucht sich ja nur eine beliebige „übergreifende Dimension“ (einen beliebigen „theoretischen Begriff“) auszudenken, um dann über „Indikatoren“ spekulieren zu können, die sich wie auch immer auf die „übergreifende Dimension“ beziehen. Es sollte auch deutlich sein, daß die in Abschnitt 6.2 kritisierte Vorstellung, daß Begriffe „operationalisiert“ werden können, ziemlich genau zu dieser Rhetorik paßt; tatsächlich wird „Operationalisierung“ oft durch die Aufgabe definiert, daß es gelte, „Indikatoren“ für „theoretische Begriffe“ zu finden.

7. Eine komplementäre Strategie besteht darin, die inhaltliche Analyse der Dimensionen von Indikatoren durch formal-statistische Verfahren zu substitutieren. Z.B. schreiben Bortz und Döring (1995, S. 201):

„Bei eindimensionalen Tests werden die Itemwerte in der Regel additiv zu einem gewichteten oder ungewichteten Gesamtwert (bzw. Index, ...) zusammengefaßt [...]. Welche dieser Vorgehensweisen gerechtfertigt ist, zeigt die *Dimensionalitätsüberprüfung*, die üblicherweise mit explorativen oder konfirmativen Faktorenanalysen durchgeführt wird ... Faktorenanalysen produzieren u.a. für jedes Item eine sog. *Faktorladung*. Eindimensionalität liegt vor, wenn die Item-Interkorrelationen auf einen Faktor (sog. Generalfaktor) reduziert werden können, auf dem sie hoch ‘laden’ (d.h. mit dem sie hoch korrelieren). Der Faktor repräsentiert inhaltlich das ‘Gemeinsame’, das in allen Items ausgedrückt wird, und steht für das zu messende Konstrukt.“

[16] Auch Indizes, die auf Zählgrößen beruhen, sind nicht immer unproblematisch, wie folgendes Beispiel von Schnell, Hill und Esser 1999, S. 166, zeigt: „Ein Beispiel für einen additiven Index stellt ein ‘Wohnungsausstattungs-Index’ dar, bei dem für eine Reihe von Einrichtungen (Bad, Zentralheizung, etc.) oder auch eine Reihe von Geräten (Telefon, TV, PC etc.), deren Vorhandensein erhoben und für jedes Vorhandensein ein Punkt vergeben wird. Ein additiver Index besteht dann lediglich aus der Summe der vergebenen Punkte.“ Im allgemeinen wird aus einer solchen Vorgehensweise sicherlich kein sinnvoller Index resultieren.

Schon ohne ein genaueres Verständnis der angesprochenen Methoden (mit der Faktorenanalyse beschäftigen wir uns erst in Kapitel 14) wird bereits deutlich, daß Bortz und Döring der Auffassung sind, daß die Dimensionsanalysen, die zur Begründung von Indizes erforderlich sind, mit statistischen Methoden vorgenommen werden können. Eine einfachere Variante dieser Strategie kann man sich auch anhand folgender Ausführungen von Kromrey (1994, S. 121) verdeutlichen:

„Generell gilt für die Indexbildung – was die Zusammenhänge zwischen den Teildimensionen eines Sachverhalts und die Beziehungen dieser Teildimensionen zu der abhängigen Variablen betrifft –: Eine Zusammenfassung der Teildimensionen zu einem Index ist nur dann zulässig, wenn die Korrelationen *zwischen* den Index-Teildimensionen (= den Indikatoren) alle ein positives Vorzeichen haben und wenn die Korrelationen der Teildimensionen (der Indikatoren) mit der abhängigen Variablen alle das gleiche Vorzeichen haben (also entweder alle positiv oder alle negativ sind). Andernfalls wird durch die Zusammenfassung zu einem Gesamtindex die Struktur der Realität nicht korrekt abgebildet; das Ergebnis wird verfälscht.“

Der Gedanke besteht also darin, daß man die Frage der dimensionalen Vergleichbarkeit von Indikatoren dadurch beantworten könne, daß man ihre Korrelationen untersucht.[17]

8. Man kann sich jedoch leicht klarmachen, daß die Korrelation von zwei Variablen mit der Frage ihrer dimensionalen Vergleichbarkeit überhaupt nichts zu tun hat. Zum Beispiel sind Ausbildung (wenn man sie irgendwie quantifiziert) und Einkommen ziemlich hoch korreliert, aber es handelt sich offensichtlich um zwei unterschiedliche Dimensionen. Ebenso gibt es z.B., wenn man sich auf Menschen bezieht, zwischen der Körpergröße und dem Körpergewicht eine ziemlich große Korrelation; aber Größe und Gewicht sind natürlich unterschiedliche Dimensionen. Andererseits sind in unserem Klausurbeispiel die in den einzelnen Aufgaben erzielten Punktzahlen nur ziemlich schwach miteinander korreliert (vgl. Tabelle 6.5.1); gleichwohl beziehen sich alle 5 Indikatoren auf die gleiche Dimension. Schließlich ist auch darauf hinzuweisen, daß Korrelationskoeffizienten statistische und mithin verteilungsabhängige Informationen liefern. Ein Korrelationskoeffizient charakterisiert eine Gesamtheit von Objekten, unser Ω. Es wäre

[17] Kromreys Gedankengang bewegt sich in einem Kontext, in dem es um die Bildung eines Index zur Erklärung einer abhängigen Variablen geht. Wir werden in Abschnitt 7.2 die sog. Likert-Skalierung besprechen, bei der der gleiche Gedankengang unmittelbar zur Begründung von Indexbildungen dient.

Tabelle 6.5.1 Korrelationen der Punktzahlen beim Klausur-Index.

	X_1	X_2	X_3	X_4	X_5
X_1	1.00				
X_2	-0.03	1.00			
X_3	0.20	-0.13	1.00		
X_4	0.15	0.44	0.12	1.00	
X_5	0.15	0.11	0.12	0.20	1.00
X^*	0.42	0.43	0.41	0.67	0.74

Zur Definition der Variablen und zu den Daten vgl. Abschnitt 6.1.

unsinnig, mit Korrelationskoeffizienten auch Indizes charakterisieren zu wollen. Bestenfalls kann man davon sprechen, daß mit ihrer Hilfe die gemeinsame statistische Verteilung der für die Indexbildung verwendeten Indikatoren charakterisiert werden kann. Dies liefert dann aber eine Aussage über die Werte des Index in einer gegebenen Gesamtheit, also über eine Menge kontingenter Daten. Das zeigt sich schon darin, daß Korrelationskoeffizienten einfach dadurch verändert werden können, daß man sich auf willkürlich ausgewählte Teilgesamtheiten bezieht.

Kapitel 7

Indizes für Einstellungen

Fragt man Menschen nach ihren Meinungen, insbesondere nach Präferenzen und Bewertungen, geht es darum, *was* sie zu sagen haben. Wie bereits in Abschnitt 3.4 ausgeführt worden ist, interessiert sich die Umfrageforschung meistens gar nicht für die Meinungen der jeweils befragten Personen, sondern möchte Einstellungen konstruieren, die sich möglicherweise dafür eignen, Verhaltensweisen voraussagbar zu machen. Diesem Zweck dienen Rating-Skalen und die sog. Likert-Skalierung, die das Thema dieses Kapitels bilden.

7.1 Rating-Skalen

1. Folgendes Zitat von L. L. Thurstone und E. J. Chave (1929) soll noch einmal daran erinnern, wie in der sozialwissenschaftlichen Literatur von Einstellungen gesprochen wird.

„The concept 'attitude' will be used here to denote the sum-total of a man's inclinations and feelings, prejudice or bias, preconceived notions, ideas, fears, threats, and convictions about any specific topic. Thus a man's attitude about pacifism means here all that he feels and thinks about peace and war. It is admittedly a subjective and personal affair. The concept 'opinion' will here mean a verbal expression of attitude. If a man said that we made a mistake in entering the war against Germany, that statement would be called his opinion. The term 'opinion' will be restricted to verbal expression. But it is an expression of what? It expresses an attitude, supposedly. [...] An opinion symbolizes an attitude. [...] We shall use opinions as the means for measuring attitudes." (S. 6f) "In our entire study we shall be dealing, then, with opinions, not primarily because of their cognitive content but rather because they serve as the carriers or symbols of the attitudes of the people who express or indorse these opinions.“ (S. 13)

Man sollte sich nicht dadurch täuschen lassen, daß es auf den ersten Blick so aussieht, als ob es sich bei Einstellungen insbesondere auch um Präferenzen und Bewertungen handeln könnte. Es besteht tatsächlich ein wesentlicher Unterschied. Bei Präferenzen und Bewertungen geht es darum, *was* von den Befragungspersonen zum Ausdruck gebracht wird; das Erkenntnisinteresse richtet sich darauf, *was* sie sagen. Bei der Ermittlung von Einstellungen geht es dagegen um die „psychische Verfassung“ der Befragungspersonen.

2. Dann stellt sich die Frage, ob und ggf. wie von der „psychischen Verfassung“ von Menschen sinnvoll gesprochen werden kann. Den Versuchen, Einstellungen zu ermitteln und sie mit Skalierungsverfahren „meßbar“ zu machen, liegt die Vorstellung zugrunde, daß es sich um „beschreibbare

Sachverhalte“ handelt. Thurstone und Chave (1929, S. 64) sagen das so:

„It is to be expected that many of the users of an attitude scale will have as their motive the influencing of people toward a chosen end of the scale. Our object is to produce a tool, as objective as possible, by which to describe attitudes.“

Aber einen Sachverhalt, der *beschrieben* werden könnte, gibt es in diesem Fall nicht. Schon deshalb nicht, weil das Reden von Einstellungen wegen seiner konzeptionellen Unklarheiten gar nicht auf irgendeinen Sachverhalt verweist. Man kann die Überlegung aber auch noch fortsetzen und in Frage stellen, ob man überhaupt sinnvoll über „psychische Verfassungen“ in der gleichen sprachlichen Form reden kann, wie dies bei objektivierbaren Sachverhalten möglich erscheint. Einige Überlegungen wurden bereits in Abschnitt 3.4 dargelegt und sollen hier nicht wiederholt werden. Aber selbst wenn wir diese Frage offen lassen, zeigt sich doch schon daran, wie Einstellungen ermittelt werden, daß in keiner Weise von einer Beschreibung von Sachverhalten gesprochen werden kann, sondern daß es sich um reine Konstruktionen handelt.

3. Man erkennt das zunächst daran, wie die Daten erzeugt werden, die hinterher als Indikatoren zur Konstruktion von Einstellungen dienen sollen. Verwendet werden Rating-Skalen, wie wir sie schon in Abschnitt 3.4 kennengelernt haben. Allgemein besteht die Vorgehensweise darin, daß den Befragungspersonen ein „Statement“ oder ein beliebiges „Einstellungsobjekt“ vorgegeben wird, zu dem sie sich in einer vorgeschriebenen Weise äußern sollen, die die Form einer „Skala“ hat. Bortz und Döring (1995, S. 165f) geben u.a. folgende Beispiele.

„*Instruktion:* Im folgenden wird Ihnen eine Reihe von Behauptungen vorgelegt. Bitte entscheiden Sie, ob das jeweils Behauptete Ihrer Ansicht nach eher richtig oder falsch ist (pro Aussage bitte nur ein Kreuz).

In der Mode kehrt alles wieder:

○ stimmt gar nicht
○ stimmt wenig
○ stimmt teils-teils
○ stimmt ziemlich
○ stimmt völlig“

Hier ist ein weiteres Beispiel:

„*Instruktion:* Im folgenden zeige ich Ihnen Video-Aufnahmen gruppentherapeutischer Sitzungen. Beurteilen Sie bitte die Gruppenatmosphäre bzgl. des Merkmals ‘gespannt–gelöst’. Hierfür steht Ihnen eine 5-stufige Skala zur Verfügung. Die einzelnen Skalenwerte haben folgende Bedeutung: 1 = Gruppenatmosphäre ist gespannt; 2 = Gruppenatmosphäre ist eher gespannt als gelöst; 3 = Gruppenatmosphäre ist weder gespannt noch gelöst; 4 = Gruppenatmosphäre ist eher gelöst als gespannt; 5 = Gruppenatmosphäre ist gelöst. Bitte urteilen Sie möglichst spontan! Uns interessiert Ihre persönliche Meinung, d.h. es gibt keine ‘richtigen’

oder ‘falschen’ Antworten. Noch ein Hinweis: Bitte achten Sie darauf, daß die Abstände zwischen den einzelnen Stufen gleich groß sind.“

Die Beispiele deuten schon an, daß Rating-Skalen in vielen verschiedenen Variationen verwendet werden. Die Unterschiede beziehen sich einerseits auf die Anzahl der unterschiedlichen „Skalenwerte“; oft 5, manchmal mehr, manchmal weniger, manchmal bipolar um einen „Nullpunkt“ herum geordnet usw. Und sie unterscheiden sich außerdem darin, welche Art von „sprachlicher Reaktion“ von den Befragungspersonen erwartet wird; etwa Grade des Zustimmens oder Ablehnens, Grade des Empfindens, Grade des Etwas-schön-finden, Grade der Ähnlichkeit usw.

4. Der wesentliche Zweck von Rating-Skalen besteht offenbar darin, daß man durch sie das Problem, ob und ggf. wie etwas quantifiziert werden kann, scheinbar auf Befragungs- bzw. Testpersonen abschieben kann. Die eigentliche Leistung der Quantifizierung soll bei dieser Vorgehensweise von ihnen erbracht werden. Dazu heißt es bei Upshaw (1971, S. 74):

„In supplying data for scale construction, a respondent may be called upon to perform a task which resembles nominal, ordinal, interval, or ratio scaling. That is, he may be asked to sort elements into classes, to report perceived order relations among classes, to estimate the amount of difference among classes, or to judge the absolute magnitude or the ratios of stimuli. It is an appealing simple idea that numerical responses can be interpreted as measures of the variable described in the instructions to the respondents. If such an interpretation could be justified, then virtually the entire labor of scale construction could be assigned to respondents. The only real deterrents to the wholesale adoption of the practice of having respondents do the entire scaling job are fears that they cannot or will not do it.“

Diejenigen, die Rating-Skalen verwenden, teilen solche Befürchtungen vermutlich nicht. Sie nutzen vielmehr die Tatsache aus, daß man sich in der Umgangssprache oft graduell abgestufter Redeweisen bedient.

5. Allerdings steht diese Sprachform in der umgangssprachlichen Kommunikation stets in einem spezifischen Kontext und kann je nachdem ganz unterschiedlichen Zwecken dienen. Sie dient etwa dazu, sich nicht genau festzulegen oder um mögliche Differenzierungen anzudeuten oder zur Formulierung von Unter- oder Übertreibungen. Innerhalb umgangssprachlicher Kommunikation ist das meistens nicht problematisch, denn Formulierungen können modifiziert und erläutert werden. Dieser Kontext muß jedoch systematisch unterdrückt werden, wenn Psychologen und Sozialwissenschaftler Rating-Skalen innerhalb standardisierter Interviews oder Tests verwenden, um – wie Bortz und Döring (1995, S. 143) sagen – die Befragungs- bzw. Testpersonen „als Meßinstrumente“ verwenden zu können.[1] Aber Menschen werden sicherlich nicht dadurch zu Meßinstru-

[1]Dieser Gedanke wurde bereits von Likert verfolgt. Gegen Thurstone, der seine Skalenbildung auf „Urteile von Experten“ gründen wollte, schrieb Likert 1932, S. 24: „Further,

menten, daß man ihnen standardisierte Rating-Skalen vorgibt. Was sollten sie denn messen? Was sie angesichts eines „Einstellungsobjekts“ denken, fühlen, glauben, sich wünschen, sich ausdenken können, oder was ihnen dabei „in den Sinn“ kommt?

6. Man sollte darauf achten, *wie* bei der Verwendung von Rating-Skalen zur Konstruktion von Einstellungen an das umgangssprachliche Reden angeknüpft wird. Einerseits wird es – scheinbar – so verwendet, wie es der alltagspraktischen Kommunikation dient; jedenfalls wird dies den Befragungspersonen suggeriert. Andererseits wird es jedoch für einen vollständig anderen Zweck mißbraucht. Man kann sich das anhand folgender Ausführungen von Thurstone und Chave (1929) verdeutlichen:[2]

„The main argument so far has been to show that since in ordinary conversation we readily and understandably describe individuals as more and less pacifistic or more and less militaristic in attitude, we may frankly represent this linearity in the form of a unidimensional scale.“ (S. 14) „As a matter of fact, we get along quite well with the concept of a linear scale in describing traits even so qualitative as education, social and economic status, or beauty. A scale or linear continuum is implied when we say that a man has more education than another, or that a woman is more beautiful than another, even though, if pressed, we admit that perhaps the pair involved in each of the comparisons have little in common. It is clear that the linear continuum which is implied in a ‘more and less’ judgment may be conceptual, that it does not necessarily have the physical existence of a yardstick.“ (S. 10f)

Die Autoren übersehen jedoch einen wichtigen Unterschied. Es ist zwar richtig, daß sich umgangssprachliche Kommunikation in vielen Fällen komparativer und quantifizierender Formulierungen bedient. Dies geschieht jedoch in den meisten Fällen nicht nur kontextabhängig, sondern bleibt auch in seinem Aussagegehalt auf einen situationsabhängigen Diskurskontext beschränkt. Wenn sich umgangssprachliche Kommunikation komparativer und quantifizierender Formulierungen bedient, setzt dies keineswegs eine „lineare Skala“ oder ein „eindimensionales Kontinuum“ voraus; nicht einmal ist ein Größenbegriff impliziert, für den eine *situationsunabhängige* Explikation angegeben werden kann.[3] Man kann sich deshalb zur Begründung

it [d.h. das von Likert vorgeschlagene Skalierungsverfahren] avoids the difficulties encountered when using a judging group to construct the scale. [...] Previously attempts have been made to find the scale value of each particular statement along a continuum; a person’s score being then determined by the scale value of the statement that he accepts. In this study, however, each statement becomes a scale in itself and a person’s reaction to each statement is given a score. These scores are then combined by using a median or mean.“ Es sei erwähnt, daß die von Likert vorgeschlagene Methode zur Konstruktion von Einstellungen auch heute noch in der empirischen Sozialforschung verwendet wird. Wir besprechen sie im nächsten Abschnitt.

[2]Ganz entsprechende Aussagen gibt es bei Thurstone 1928, S. 535.

[3]Wir sehen hier natürlich von denjenigen Fällen ab, in denen auch in der umgangssprachlichen Kommunikation Größenbegriffe verwendet werden, die bereits als Zählgrößen, Meßgrößen oder monetäre Größen institutionalisiert sind.

sozialwissenschaftlicher Größenbegriffe nicht auf umgangssprachliche Redensarten berufen.[4]

7.2 Likert-Skalierung

1. In diesem Abschnitt besprechen wir ein Verfahren, das von R. Likert (1932) vorgeschlagen worden ist, um Einstellungen zu konstruieren. Es beruht auf Indikatoren, die durch Rating-Verfahren erzeugt werden, und ist auch heute noch weit verbreitet (Spector 1992). Die Vorgehensweise ist leicht zu verstehen. Laatz (1993, S. 277) erläutert sie so:

„Einstellungsmessungen nach dem Likert-Verfahren gehen so vor, daß den einzelnen Befragten verschiedene Aussagen ‘statements’ vorgelegt werden, zu denen sie in abgestufter Form auf einer rating-Skala Stellung nehmen. Ein Beispiel für ein solches Statement und die dazugehörige rating-Skala aus der deutschen Version der A-Skala lautet:

	Zustimmung			Ablehnung		
	stark	mittel	schwach	schwach	mittel	stark
Nicht auf Gesetze und Verfassungen kommt es an, sondern einzig und allein auf den Menschen	+3	+2	+1	−1	−2	−3

Die Reaktion auf solche Statements wird als Einzelmeßwert verbucht. Der eigentliche Skalenwert der Gesamtskala (der Gesamtscore) ergibt sich aus der Addition dieser Einzelmeßwerte.“

Die Likert-Skalierung besteht also einfach darin, daß ein additiver Index gebildet wird, wobei die jeweils verwendeten Indikatoren durch Rating-Skalen

[4]Der gleiche Fehler wird auch in folgenden Ausführungen von Gigerenzer 1981, S. 199, deutlich: „Gleich der Bildung von Äquivalenzklassen ist die Konstruktion von Rangordnungen ein grundlegendes Muster der Sprache. Ordnungserzeugende grammatikalische Formen wie der Komparativ und der Superlativ (gut – besser – am besten) sowie entsprechende Adverbien wie wenig, ziemlich, sehr und äußerst zeugen davon. Die alltägliche Sprache und das Denken über einen Gegenstandsbereich tragen häufig die Charakteristika einer eindimensionalen schwachen Ordnung, wie z.B. die Orientierung von Parteien, Persönlichkeiten und anderen Bedeutungsträgern aus dem Bereich der Politik anhand einer Links-Rechts-Skala.“ Die wesentliche Differenz kommt hier gar nicht zur Sprache: daß zur Begründung eines (sozialwissenschaftlichen) Größenbegriffs eine situationsunabhängige und situationsinvariante Explikation – typischerweise durch ein Meßverfahren – entwickelt werden muß; daß jedoch die umgangssprachliche Kommunikation gerade nicht situationsunabhängig und situationsinvariant funktioniert. — Hierzu gibt es bei Herkner 1991, S. 197, folgenden Hinweis: „De Soto stellte fest, daß eine starke Tendenz zu kognitiven linearen Ordnungen besteht. [...] Nach De Soto neigt man dazu, lineare Ordnungen auch dort zu sehen, wo die Realität nur partiell oder noch schwächer geordnet ist. Man projiziert quasi lineare, d.h. eindimensionale Ordnungen in die Realität. De Soto ist der Meinung, daß auch (und nicht zuletzt) Sozialwissenschaftler dieser Tendenz nachgeben.“

definiert sind. Sie wird deshalb auch „technique of summated rating" oder „Methode der summierten Ratings" genannt.

2. Daß hier überhaupt von einem besonderen Verfahren gesprochen wird, liegt daran, daß mit der Likert-Skalierung Überlegungen bzw. Techniken verbunden werden, die einer Prüfung und Begründung des resultierenden Index dienen sollen. Zur Darstellung der Vorgehensweise unterscheidet Laatz (1993, S. 279ff) drei Schritte:

„Der *erste Schritt* bei der Konstruktion einer Likert-Skala besteht in der *Formulierung einer großen Zahl von Statements*. Beim Entwurf der Statements muß neben der für die Likert-Skala charakteristischen Polung vor allem beachtet werden, daß die gesamte Bedeutungsdomäne der Dimension möglichst differenziert abgedeckt wird. [...]

Im *zweiten Schritt* wird danach eine große Zahl als geeignet angesehener Statements an einer Testpopulation auf ihre Brauchbarkeit empirisch getestet. Auf Basis einer statistischen Analyse werden die geeignetsten für eine verbesserte Version ausgewählt (*Itemauswahl*). [...]

Als *dritter Schritt* nach der Sammlung und Auswahl von items folgt nun die Beurteilung der Gesamtskala. [...]"

Weitgehend ähnlich wird die Likert-Skalierung von anderen Autoren dargestellt. Z.B. gibt Bailey (1982, S. 365f) folgende Erklärung:

„The basic procedure for Likert scaling is as follows:

1. Write a large number of questions thought to measure the dimension to be scaled.
2. Select a sample of respondents representative of the population on which the scale will be used.
3. Code all responses so that a higher score on a particular item indicates stronger agreement with the attitude being scaled (code 5 for either strong agreement with a positive statement or strong disagreement with a negative one, and code 1 for strong disagreement with a positive statement or strong agreement with a negative one); compute a scale score for each person by summing her or his scores on all questions.
4. Analyze responses and select for the scale the items most clearly differentiating between the highest and lowest scores."

3. Anhand der angeführten Zitate läßt sich bereits eine wichtige Eigentümlichkeit der Likert-Skalierung erkennen. Die inhaltliche Bedeutung der Indikatoren (Items) spielt keine oder bestenfalls eine untergeordnete Rolle; es genügt, daß die einzelnen „Statements" irgendeinen vagen Bezug zu einer „Einstellungsdimension" haben, nach der Menschen skaliert werden sollen. Items werden nicht nach ihrem potentiellen Aussagegehalt ausgewählt, sondern – wie wir noch genauer sehen werden – mit formal-statistischen Verfahren. Deshalb sagen zum Beispiel Bortz und Döring (1995, S. 203),

daß zunächst „möglichst viele Behauptungen, die unterschiedliche Ausprägungen des untersuchten Merkmals repräsentieren, gesammelt" werden sollten. Und bei Schnell, Hill und Esser (1999, S. 181) heißt es:

„Die Konstruktion einer Likert-Skala beginnt mit der Sammlung einer großen Zahl von Items (üblich sind ca. 100). Die Items stellen hierbei Aussagen dar, von denen angenommen wird, daß sie die interessierende Einstellung wiedergeben."

Aber kann geprüft werden, ob solche Annahmen zutreffend sind? In dem oben angeführten Zitat behauptet Laatz, daß dies „empirisch getestet" werden könne. Man wird jedoch sehen, daß es sich um eine irreführende Aussage handelt. Tatsächlich werden diese Annahmen nicht geprüft, und sie können mit den vorgeschlagenen Verfahren auch nicht geprüft werden.

4. Um die Vorgehensweisen bei der Likert-Skalierung zu diskutieren, verwenden wir die bereit eingeführte Notation und nehmen als Ausgangspunkt an, daß es eine „große Zahl von Items" $X_1, \ldots, X_m$ gibt und daß daraus ein additiver Index

$$X^* := \sum_{j=1}^{m} X_j$$

gebildet werden soll. Um zu prüfen, welche Items sich als „Indikatoren" der „Einstellungsdimension" eignen, sind hauptsächlich zwei Verfahren gebräuchlich. Das erste Verfahren beruft sich auf Korrelationen zwischen den Indikatoren X_j und dem Index X^*, also

$$\mathrm{Corr}(X^*, X_j) = \frac{\sum_{i=1}^{n}(x_i^* - \mathrm{M}(X^*))(x_{ij} - \mathrm{M}(X_j))}{\sqrt{\sum_{i=1}^{n}(x_i^* - \mathrm{M}(X^*))^2 \sum_{i=1}^{n}(x_{ij} - \mathrm{M}(X_j))^2}}$$

wobei hier angenommen wird, daß Daten für $i = 1, \ldots, n$ Personen zur Verfügung stehen; x_{ij} ist der Wert von Person i beim Indikator X_j und $\mathrm{M}(X_j)$ ist der Mittelwert; entsprechend ist x_i^* der Wert von Person i für den Index X^* und $\mathrm{M}(X^*)$ der Mittelwert. Als Beispiel kann man an die in Tabelle 6.5.1 angegebenen Korrelationen für unseren Klausur-Index denken. Zwar beruht dieser Index nicht auf Rating-Skalen, aber die hier und im folgenden zu besprechenden Verfahren sind davon ganz unabhängig.

5. Das Argument lautet nun, daß die Korrelation von X_j mit X^* als ein Kriterium verwendet werden kann, um zu beurteilen, wie gut sich Item X_j zur Erfassung der vermuteten „Einstellungsdimension" eignet.[5] Zugrunde liegt die Vorstellung: Je größer die Korrelation zwischen X_j und X^* ist, desto besser eignet sich Item X_j für die Indexbildung. Dementsprechend lautet die Empfehlung, für den schließlich zu bildenden Index nur solche

[5]Gelegentlich wird vorgeschlagen, daß man die Korrelationen zwischen X_j und einem modifizierten Index, der jeweils den Indikator X_j nicht enthält, verwenden soll; zum Beispiel von McIver und Carmines 1994, S. 162.

Items zu verwenden, die eine hohe Korrelation mit dem Index aufweisen.[6] Aber was ist von dieser Überlegung zu halten? Sie verwechselt offenbar Bedeutung mit Korrelation. Die Bedeutung eines Items – wenn es überhaupt eine sinnvolle Bedeutung hat – ergibt sich aus seiner Formulierung und daraus, wie es von Befragungspersonen verstanden wird und wie sie Stellung nehmen. Aber dies hat offenbar nichts damit zu tun, wie ein Item – als statistische Variable – mit einem aus einer Menge von Items gebildeten Index korreliert.

6. Tatsächlich geht es um ein vollständig anderes Kriterium. Das wird deutlich, wenn man sich ein komplementäres Verfahren ansieht, mit dem die Beurteilung von Items erfolgt. Schnell, Hill und Esser (1999, S. 183) erläutern die Vorgehensweise folgendermaßen:

„Insbesondere in der älteren Literatur basieren Itemanalysen auf *'Trennschärfe-Indizes'*. Hierbei werden aus den Befragten zwei Gruppen ausgewählt. Eine Gruppe besteht aus den 25 % der Versuchspersonen mit den niedrigsten Skalenwerten; die andere Gruppe besteht aus den 25 % der Versuchspersonen mit den höchsten Skalenwerten. Für jedes Item wird der Mittelwert und die Varianz des Items in beiden Gruppen berechnet. Die Differenz dieser Mittelwerte wird mit den Varianzen zum 'Trennschärfe-Index' dieses Items umgerechnet. Ein hoher Trennschärfe-Koeffizient besagt nichts anderes, als daß die beiden Gruppen dieses Item unterschiedlich beantworteten. Für die endgültige Likert-Skala werden meist Items mit einem Trennschärfeindex größer als 1.65 ausgewählt. Es werden also diejenigen Items ausgewählt, die zwischen Personen mit einer negativen Einstellung und Personen mit einer positiven Einstellung zu einem gegebenen Objekt zu unterscheiden (diskriminieren) vermögen."

Es geht also nicht um die Frage, ob mit den Items irgendein sinnvoller, in seiner Bedeutung nachvollziehbarer Inhalt erfaßt werden kann, sondern um etwas anderes: wie gut Items in der Lage sind, zwischen Befragungspersonen zu diskriminieren. In einer Formulierung von C. und D. Nachmias (1981, S. 280): „those items with the largest DP [discriminative power] are selected."[7]

[6]Bei Likert 1932, S. 48, heißt es dazu: „An objective check ought then be applied to see (1) if the numerical values are properly assigned and (2) whether the statements are 'differentiating'. One possible check is item analysis which calls for calculating the correlation coefficient of each statement with the battery. If a negative correlation coefficient is obtained, it indicates that the numerical values are not properly assigned and that the ONE and FIVE ends [der 5-Punkte Rating-Skala] should be reversed. If a zero or very low correlation coefficient is obtained, it indicates that the statement fails to measure that which the rest of the statements measure. Such statements will be called 'undifferentiating'. Thurstone (38) refers to them as irrelevant or ambiguous. By 'undifferentiating' we merely mean that the statement does not measure what the battery measures and hence to include it contributes nothing to the scale." Daraus zieht Likert dann u.a. die Schlußfolgerung, daß keine „Statements" verwendet werden sollten, zu denen fast alle Befragungspersonen in der gleichen Weise Stellung nehmen.

[7]Eine ganz ähnliche Formulierung verwendet Spector 1992, S. 31. Infolge dieses Auswahlkriteriums für Items kann auch der schließlich gebildete Index nur so verstanden werden, daß er dazu dienen soll, möglichst gut zwischen den „Versuchspersonen" zu diskriminieren; oder in statistischer Terminologie formuliert: Die Varianz des schließlich zu bildenden Index soll möglichst groß sein. Wie das auf direktem Wege erreicht werden kann, besprechen wir in Kap. 13.

7. Zur Illustration der Berechnung von Trennschärfe-Indizes[8] verwenden wir die in Abschnitt 6.1 angegebenen Klausurdaten, aus denen der additive Index X^* gebildet wird (vgl. Tabelle 6.1.1). Zunächst werden die Klausurteilnehmer nach der Größe der Werte des Index X^* geordnet. Dann werden das untere und das obere Quartil betrachtet (wir verwenden jeweils 11 Fälle mit den niedrigsten und höchsten Werten) und für jede Gruppe gesondert die Mittelwerte und Varianzen berechnet; M_j^u und V_j^u bezeichnen Mittelwert und Varianz von Aufgabe A_j im unteren, M_j^o und V_j^o bezeichnen die entsprechenden Größen im oberen Quartil. Folgende Tabelle zeigt die Werte für unser Beispiel:

Aufgabe	M^u	V^u	M^o	V^o	T
A_1	7.45	3.24	10.00	0.00	2.61
A_2	7.91	3.33	10.00	0.00	2.08
A_3	2.27	2.61	5.00	0.00	3.47
A_4	4.27	3.88	9.64	0.81	4.49
A_5	7.45	6.95	14.91	0.30	3.56

Schließlich werden aus diesen Werten die Trennschärfe-Indizes berechnet. Dafür kann z.B. folgende Formel verwendet werden:[9]

$$T_j := \frac{M_j^o - M_j^u}{\sqrt{V_j^o/n^o + V_j^u/n^u}}$$

wobei n^u und n^o die Anzahl der Personen im unteren bzw. oberen Quartil bezeichnen. Die Tabelle zeigt in der letzten Spalte die so berechneten Trennschärfe-Indizes. Am besten diskriminiert offenbar Aufabe 4, die von den wenigsten Klausurteilnehmern erfolgreich gelöst worden ist; und am schlechtesten diskriminiert Aufgabe 2, die die meisten Teilnehmer richtig gelöst haben.

8. Schließlich erfolgt bei der Likert-Skalierung in der Regel noch eine Prüfung der „Dimensionalität" des Index, denn das mit dem Verfahren angestrebte Ziel besteht darin, einen *quantifizierenden* Index zu konstruieren. Die Voraussetzung, daß man es mit quantitativen Indikatoren zu tun hat,

[8]Man beachte, daß es unterschiedliche Benennungen gibt. Z.B. sprachen Schnell, Hill und Esser in einer früheren Auflage ihres Lehrbuchs von einem „Trennschärfe-Koeffizienten", und Laatz 1993, S. 280, spricht von einem „Trennwert". Andererseits spricht auch Diekmann 1995, S. 212, von einem „Trennschärfekoeffizienten", meint aber damit die „Item-Summenscore-Korrelation".

[9]Vgl. Schnell, Hill und Esser 1992, S. 204.

wird damit begründet, daß die Indikatoren durch Rating-Skalen zustande kommen.[10] Zwar ist schon diese Voraussetzung fragwürdig. Aber auch wenn man sie akzeptieren würde, bliebe noch zu prüfen, ob die weitere Voraussetzung erfüllt ist, daß sich alle Indikatoren auf die gleiche Dimension beziehen. Wie bereits in Abschnitt 6.5 besprochen worden ist, wird darin jedoch von den meisten Autoren kein inhaltliches Problem gesehen, für das die Bedeutung der Items reflektiert werden müßte. Stattdessen wird vorgeschlagen, die „Dimensionalität" des schließlich gebildeten Index mit Hilfe formal-statistischer Verfahren zu untersuchen. Das am häufigsten verwendete Verfahren ist die Faktorenanalyse. Da zu ihrem Verständnis etwas ausführlichere Überlegungen erforderlich sind, werden wir uns mit diesem, die Likert-Skalierung abschließenden Schritt erst in Kap. 14 beschäftigen.

[10] Zum Beispiel sagen Bortz und Döring (1995, S. 163), daß mit „Rating-Skalen auf unkomplizierte Weise direkt intervallskalierte Urteile erzeugt werden" können.

Kapitel 8

Zur Rhetorik des Messens

In vielen der in vorangegangenen Kapiteln angeführten Zitate, insbesondere im Kontext von Skalierungsverfahren, wird von „Messen" gesprochen. In diesem Kapitel versuchen wir, diese Rhetorik etwas genauer zu verstehen.

8.1 Bemerkungen zum Sprachgebrauch

1. Zunächst kann festgestellt werden, daß sich in der empirischen Sozialforschung ein weitgehend inhaltsleeres Reden von „Messen" verbreitet hat. In einer Einführung in Methoden der Sozialforschung von R. Mayntz, K. Holm und P. Hübner (1969, S. 36) heißt es zum Beispiel:

„Wenn Daten über gesellschaftliche Phänomene das Ergebnis begrifflich strukturierter, systematischer und kontrollierter Beobachtungen der Merkmalsausprägungen von Untersuchungseinheiten auf den Merkmalsdimensionen darstellen, dann ist Messen in einem sehr allgemeinen Sinn zunächst nichts anderes als der Prozeß dieses Beobachtens selbst; also die Datenerhebung."

Ähnlich allgemein sagt F. Sixtl (1976, S. 11): „Messen besteht allgemein im Zuordnen von Prädikaten zu Objekten." Die meisten Autoren verlangen allerdings, daß von „Messen" nur dann gesprochen werden sollte, wenn numerische Aussagen gemacht werden. Hier sind einige exemplarische Formulierungen:

Campbell 1921, S. 1797: Measurement „may be defined, in general, as the assignment of numbers to represent properties." Stevens 1946, S. 677: „Paraphrasing N.R. Campbell [...], we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules." Savage und Ehrlich 1992, S. 1: „Today, measurement in general is taken to be the assignment of numbers (numerals, say the nominalists) to entities and events to represent their properties." Orth 1974, S. 13: „Messen ist die Bestimmung der Ausprägung einer Eigenschaft eines Dinges. Messen erfolgt durch eine Zuordnung von Zahlen zu Dingen, die Träger der zu messenden Eigenschaft sind."

In der neueren Methodenliteratur wird meistens an diesen sehr allgemeinen Sprachgebrauch angeknüpft.

2. Es ist allerdings offensichtlich, daß dieser Sprachgebrauch problematisch ist, denn man kann sich beliebige und auch ganz unsinnige Verfahren ausdenken, um Objekten und Sachverhalten nach gewissen Regeln Zahlen zuzuordnen. Zwar könnte man versuchen, den Sprachgebrauch mit der Überlegung zu verteidigen, daß es keine harten Kriterien für eine Unterscheidung zwischen sinnvollen und nicht sinnvollen Verfahren zur Charakterisierung von Objekten und Sachverhalten durch in ihrer Bedeutung explizierbare Zahlen gibt; und daß infolgedessen nur im Einzelfall diskutiert und entschieden werden kann, ob ein sinnvolles Verfahren vorliegt. Der Sprachgebrauch, wie er in den angeführten Zitaten zum Ausdruck kommt, hat jedoch noch einen weiteren Mangel. Er verschleiert eine wichtige Unterscheidung: zwischen *numerischer Repräsentation* und *Quantifizierung*. Wir haben diese Unterscheidung bereits in Abschnitt 4.1 diskutiert und dort festgestellt, daß numerische Repräsentation im allgemeinen keinerlei Quantifizierung impliziert und daß demgegenüber von Quantifizierung nur dann gesprochen werden sollte, wenn man sich auf einen Größenbegriff beziehen kann. Versteht man unter „Messen" nur eine „Zuordnung von Zahlen nach gewissen Regeln", geht jedoch der Zusammenhang zur Idee der Quantifizierung verloren und wäre bereits jede Verwendung von Zahlen, die sich irgendeiner numerischen Repräsentation verdanken, als „Messen" zu bezeichnen. Hält man sich an den außerhalb der sozialwissenschaftlichen Methodenliteratur üblichen Sprachgebrauch, geht es aber beim „Messen" in erster Linie darum, Eigenschaften von Objekten oder Sachverhalten zu quantifizieren. — In diesem Text folgen wir dem außerhalb der sozialwissenschaftlichen Methodenliteratur üblichen Sprachgebrauch und betrachten deshalb den Anspruch, daß mit einem Meßverfahren eine Eigenschaft *quantifiziert* werden soll, als konstitutiv. Dieses Verständnis weist auch auf das wesentliche Problem hin: Es muß zunächst ein Größenbegriff begründet werden, um dann von einem Meßverfahren zur Messung der durch den Größenbegriff definierbaren Größen bei jeweils bestimmten Objekten oder Sachverhalten sprechen zu können.

8.2 Repräsentieren und Messen

1. Einige ergänzende Bemerkungen sollen dazu dienen, die Unterscheidung zwischen „Messen" und „Repräsentieren" zu verdeutlichen. Orientiert man sich daran, welche Auffassungen in der Geschichte des Nachdenkens über Meßverfahren vertreten worden sind, kann man etwas vereinfacht hauptsächlich zwei Leitideen unterscheiden. Erstens die „klassische" Leitidee, daß Messen als eine Verallgemeinerung von Zählen (nämlich Zählen sinnvoll definierbarer Einheiten) verstanden werden kann. Zweitens die Leitidee der „modernen" Meßtheorie, daß Messen als *numerische Repräsentation* (der Beziehungen zwischen Objekten) verstanden werden kann. Um auf diese unterschiedlichen Leitideen zu verweisen, sprechen wir von der *traditionellen* bzw. der *repräsentationalen* Meßtheorie.[1]

[1]Einführungen in die repräsentationale Meßtheorie findet man u.a. bei Orth 1974 und bei Coombs, Dawes und Tversky 1975, Kap. 2, die den Grundgedanken folgendermaßen formuliert haben (S. 44): „Messen ist als Repräsentation eines empirischen Relativs durch ein numerisches Relativ zu betrachten." Einen historischen Überblick gibt Diez 1997; kritische Überlegungen finden sich u.a. bei Adams 1966, Schönemann 1994. Die erste systematische Darstellung der traditionellen Meßtheorie findet sich bei H. von Helmholtz 1887; sehr ähnliche Überlegungen dann auch bei Duhem 1908, Kap. 5. Die Formulierung ‘traditionelle Meßtheorie’ soll hier nur einen Kontrast zur modernen Meßtheorie bezeichnen, die im wesentlichen axiomatisch vorgeht, allerdings in der wissenschaftstheoretischen Diskussion meistens mit einem empiristischen Theorieverständnis verknüpft wird. Von Vertretern der modernen Meßtheorie wird Helmholtz gelegentlich als einer ihrer Ausgangspunkte genannt, dessen Ansatz durch die moderne Theorie „verallgemeinert" worden sei; vgl. z.B. Narens und Luce 1986. Wir finden es demgegenüber aufschlußreicher, den konzeptionellen Unterschied zwischen „Messen" (als Verallgemeinerung von Zählen) und „Repräsentieren" (einer irgendwie gegebenen Empirie) zu verfolgen.

2. Die neuere Methodenliteratur zur empirischen Sozialforschung orientiert sich überwiegend an der repräsentationalen Meßtheorie. Exemplarisch sei auf Bortz und Döring (1995, S. 65) verwiesen:

> „Zur Darlegung des Repräsentationsproblems gehen wir von einem *empirischen Relativ* (oder Relationensystem) aus, das aus einer Menge von Objekten sowie einer oder mehreren Relationen besteht, welche die Art der Beziehung der Objekte untereinander charakterisieren. Dieses empirische Relativ wird in ein *numerisches Relativ* abgebildet, deren Zahlen so geartet sein müssen, daß sie die Objektrelationen des empirischen Relativs korrekt repräsentieren. Eine Abbildung mit dieser Eigenschaft bezeichnet man als *homomorph* bzw. strukturerhaltend."

Daran schließt sich folgende Definition an:

> „Messen ist eine Zuordnung von Zahlen zu Objekten oder Ereignissen, sofern diese Zuordnung eine homomorphe Abbildung eines empirischen Relativs in ein numerisches Relativ ist."

Ähnliche Aussagen findet man sehr oft in der Methodenliteratur zur empirischen Sozialforschung.[2]

3. Um besser zu verstehen, was gemeint sein könnte, betrachten wir ein einfaches Beispiel für ein „empirisches Relativ": drei Objekte, die in der Realität in einer bestimmten Weise geordnet auftreten: $\omega_1 \prec \omega_2 \prec \omega_3$. Zum Beispiel drei Dosen, die nebeneinander stehen, zuerst kommt ω_1, dann ω_2 und schließlich ω_3; das Zeichen '$\prec$' soll diese empirische Relation symbolisieren. Man kann dann diese Relation durch eine Variable repräsentieren, indem man den Objekten Zahlen zuordnet, zum Beispiel: $X(\omega_1) = 1, X(\omega_2) = 2, X(\omega_3) = 3$. In diesem Fall entspricht die $<$-Relation zwischen den Werten der Variablen der $\prec$-Relation zwischen den Objekten. Aber man kann auch andere Zahlenwerte wählen, etwa $X(\omega_1) = 7, X(\omega_2) = 5, X(\omega_3) = 9$. In diesem Fall entspricht die Relation zwischen den Zahlen nicht mehr der Relation zwischen den Objekten. In der Terminologie des oben angeführten Zitats von Bortz und Döring ist die Abbildung (Variable) X also nur im ersten Fall „strukturerhaltend".

4. Aber was hat die Frage, wie man für ein „empirisches Relativ" eine „strukturerhaltende" numerische Repräsentation finden kann, mit der Idee des Messens zu tun? Offenbar kommt es darauf an, was mit dem Wort

[2]Z.B. bei Schnell, Hill und Esser 1999, S. 132ff; Laatz 1993, S. 51ff; Kromrey 1994, S. 151ff; Diekmann 1995, S. 244ff.

'Meßverfahren' gemeint sein soll. Folgen wir zunächst dem Ansatz der repräsentationalen Meßtheorie mit der Definition: Ein Meßverfahren ist ein Verfahren, mit dem einer Menge von Objekten Zahlen zugeordnet werden können, und zwar so, daß gewisse empirische Relationen zwischen den Objekten eine Entsprechung in den numerischen Relationen zwischen den ihnen zugeordneten Zahlen haben. Geht man von dieser Definition aus, sind zur Begründung eines Meßverfahrens zwei Schritte erforderlich:

a) Zunächst müssen die Objekte und ihre empirischen Relationen fixiert werden, die durch das Meßverfahren erfaßt werden sollen.

b) Dann muß eine numerische Repräsentation der Objekte, also in unserer Terminologie eine statistische Variable mit einem numerischen Wertebereich, gefunden werden, so daß sich die empirischen Beziehungen zwischen den Objekten und die numerischen Beziehungen zwischen den ihnen zugeordneten Zahlen entsprechen.

Zu überlegen ist, ob mit empirischen Argumenten entschieden werden kann, wie hierbei vorgegangen werden sollte. Beginnen wir mit dem zweiten Schritt. Hier sind die Objekte und ihre empirischen Beziehungen schon gegeben, und es bleibt ein rein mathematisches Problem übrig: wie eine bereits *gegebene* Menge von Objekten und Relationen numerisch repräsentiert werden kann.[3] Um zu einer Antwort zu gelangen, ist empirisches Wissen offenbar nicht erforderlich. Es bleibt also der erste Schritt, in dem die Objekte und ihre empirischen Relationen fixiert werden, die durch das Meßverfahren erfaßt werden sollen. Um was für eine Art von Problem handelt es sich? Jedenfalls nicht um ein mathematisches Problem, und es wird deshalb in der Regel von der modernen Meßtheorie (insoweit sie eine rein mathematische Theorie ist) auch gar nicht behandelt. Es handelt sich aber auch nicht um ein empirisches Problem, das durch Hinweise auf Beobachtungen gelöst werden könnte.

5. In unserem Beispiel kann sicherlich vorausgesetzt werden, daß jeder weiß, was Dosen sind und was es bedeutet, daß Dosen nebeneinander oder nicht nebeneinander stehen. Deshalb kann auch empirisch geprüft werden, ob die drei Dosen wirklich nebeneinander stehen. Aber dazu soll das Meßverfahren nicht dienen, vielmehr soll es die Reihenfolge erfassen, in der die Dosen nebeneinander stehen. Zu diesem Zweck müssen den Dosen zunächst Namen gegeben werden, etwa D_1, D_2, D_3. Außerdem muß man verstehen, was es heißt, von der Reihenfolge von Dosen zu sprechen, und dafür bedarf es einiger Festlegungen (z.B. muß festgelegt werden, daß die Dosen von links nach rechts zu betrachten sind). Hat man sich auf diese Weise mit der erforderlichen Sprache ausgerüstet, kann man schließlich (zum Beispiel) sagen: zuerst kommt D_1, dann D_2 und dann D_3. Nun, unser Meßverfahren soll die Dosen und ihre Reihenfolge numerisch repräsentierten. Also verwenden wir folgende Abbildung (Variable): $X(D_1) = 1, X(D_2) = 2, X(D_3) = 3$. Warum? Die Antwort scheint selbstverständlich lauten zu müssen: Weil dann die räumliche Ordnung der Dosen der $<$-Relation zwischen den sie repräsentierenden Zahlen entspricht. Aber ist diese Antwort ausreichend? Vielleicht ist sie ausreichend, um das Wort 'repräsentieren' zu verstehen. Sie ist aber sicherlich nicht ausreichend, um zu verstehen, was „Messen" heißt. Sie ist schon dann nicht ausreichend, wenn man von dem minimalen Vorverständnis ausgeht, daß mit Hilfe von Messungen festgestellt werden soll, wie Sachverhalte beschaffen sind und wie sie sich verändern. Betrachten wir nochmal die drei Dosen. Aber jetzt kommt ein Windstoß und bringt sie durcheinander. Wir können hingehen und sie wieder in einer Reihe aufstellen, zum Beispiel: $D_1 \prec D_3 \prec D_2$. Wird dann die numerische Repräsentation, die zuvor vorgenommen wurde, falsch? Sicherlich nicht; sie liefert sogar immer noch eine brauchbare Information (nämlich darüber, wie die Dosen vor dem Windstoß aufgestellt waren). Aber es kann aus dieser Repräsentation nicht ermittelt werden, in welcher Ordnung sie nach dem Windstoß stehen. — Muß also das Meßverfahren immer dann geändert werden, wenn in den zu repräsentierenden Relationen eine Veränderung eintritt? Natürlich nicht; ein Meßverfahren soll gerade dazu dienen, Unterschiede und Veränderungen zu erfassen. Man muß deshalb Meßverfahren *als Verfahren* definieren und versuchen, sie gegenüber Veränderungen in der zu messenden Realität relativ konstant zu halten. Dies verlangt aber, daß sich die Repräsentation der Objekte durch Zahlen immer dann verändern muß, wenn sich ihre Eigenschaften oder Beziehungen zu anderen Objekten (soweit sie durch das Meßverfahren erfaßt werden sollen) verändern.

[3] Wie bereits bemerkt worden ist, beschäftigt sich die repräsentationale Meßtheorie nur mit dieser Frage. Dazu heißt es bei Kyburg 1992, S. 75: „What is now considered classical measurement theory – which has nothing to do with the theory of making measurements – is concerned with specifying the homomorphisms of some 'qualitative (or empirical) structure into a numerical one' (Narens 1985, p. 5). This view of classical measurement theory is referred to as the *representational theory*, because we are concerned with how to characterize the ways in which a *given* empirical structure can be *represented* in a numerical structure."

6. Überlegen wir uns, wie das Meßverfahren beschaffen sein sollte. Es soll zeigen, in welcher Reihenfolge zu zwei Zeitpunkten t_1 und t_2 die drei Dosen stehen. Es soll also Meßergebnisse liefern, die sich als eine Information über die Rangplätze der Dosen interpretieren lassen; in unserem Beispiel soll es die folgenden Rangzahlen liefern:

	D_1	D_2	D_3
t_1	1	2	3
t_2	1	3	2

Wenn man annehmen kann, daß das Meßverfahren so funktioniert, kann man die Zahlenwerte, die es liefert, verwenden, um Aussagen über die Reihenfolge der Dosen zu den Zeitpunkten t_1 und t_2 zu treffen. Das hat aber mit der Repräsentation der empirischen Relation $\prec$ durch die numerische Relation $<$ eigentlich nichts zu tun. Die numerische Relation $<$ besteht zwischen Zahlen, und an der Ordnung der Zahlen durch diese Relation ändert sich nichts, wie auch immer sich die Realität verändert. Die Relation $2 < 3$ gilt immer, egal wie die Dosen aufgestellt werden. Andererseits gelten bestimmte empirische Relationen zwischen Objekten keineswegs immer, sondern dies hängt davon ab, wie die empirische Realität beschaffen ist *und sich verändert*. Die wesentliche Aufgabe, die ein Meßverfahren zu lösen hat, besteht infolgedessen nicht darin, eine *als gegeben* voraussetzbare empirische Relation numerisch zu repräsentieren, sondern zu gewährleisten, daß mit seiner Hilfe Zahlen gewonnen werden können, die Aussagen über kontingente Beschaffenheiten und Veränderungen der empirischen Realität erlauben. D.h. in unserem Beispiel, daß ein Meßverfahren gewährleisten muß, daß den Dosen die Zahlen stets so zugeordnet werden, daß aus dem Meßergebnis „$X(D_i) < X(D_j)$“ die Schlußfolgerung gezogen werden kann, daß in der zum Zeitpunkt der Messung bestehenden Dosenreihe die Dose D_i vor der Dose D_j steht.

7. Kann geprüft werden, ob das Meßverfahren so funktioniert? Die Antwort hängt wieder davon ab, wie man das Wort ‘Meßverfahren’ verwenden möchte. Stellen wir uns zunächst vor, daß der Meßvorgang folgendermaßen abläuft: In einem ersten Schritt beobachtet man die Dosen und stellt ihre Reihenfolge fest, etwa: Zuerst kommt D_1, dann D_3, dann D_2. Dabei folgt man den Regeln, die für das Feststellen der Reihenfolge von Dosen vorab fixiert worden sind. In einem zweiten Schritt repräsentiert man diese Feststellung durch Zahlen, d.h. gibt der Variablen X die Werte $X(D_1) = 1, X(D_3) = 2$ und $X(D_2) = 3$. Sagt man nun, daß Messen darin besteht, empirisch gewonnene Beobachtungen numerisch zu repräsentieren, bezieht sich das Meßverfahren nur auf den zweiten Schritt. Und wenn das Ergebnis des ersten Schrittes schon feststeht, kann natürlich geprüft werden, ob der zweite Schritt richtig durchgeführt worden ist; man braucht nur zu prüfen, ob die Regeln zur numerischen Kodierung der Variablen X richtig angewendet worden sind. Das wäre dann allerdings ein merkwürdiger Gebrauch der Worte ‘Messen’ und ‘Meßverfahren’, denn der zweite Schritt ist im Prinzip völlig überflüssig. Die gesamte empirisch relevante Information steht bereits am Ende des ersten Schrittes fest. Die numerische Repräsentation, die der zweite Schritt hinzufügt, ändert nur die sprachliche Form, in der diese Information zum Ausdruck gebracht wird.

8. Wenn man sich mit dem Wort ‘Meßverfahren’ jedoch auf beide Schritte bezieht, besteht das Verfahren aus zwei Arten von Regeln: erstens aus Regeln, die festlegen, wie die Reihenfolge von Dosen empirisch zu bestimmen ist, und zweitens aus Kodierungsregeln für die numerische Repräsentation der Ergebnisse des ersten Schritts. Den zweiten Schritt haben wir gerade besprochen. Im ersten Schritt handelt es sich darum, die hierfür festgelegten Regeln richtig zu befolgen, und das kann wiederum in gewissen

Grenzen (z.B. durch einen zweiten Beobachter) überprüft werden. Aber an dieser Stelle muß auf eine wichtige Unterscheidung geachtet werden. Geprüft werden kann, ob das Meßverfahren richtig *angewendet* worden ist. Das Meßverfahren selbst, d.h. die das Verfahren konstituierenden Regeln, werden hierdurch nicht geprüft. Es wäre auch ganz sinnlos, sie prüfen zu wollen, denn durch sie wird zuallererst festgelegt, was es bedeuten soll, die Reihenfolge der Dosen richtig festzustellen. Hat man ein Meßverfahren definiert, kann geprüft werden, ob damit richtig – nämlich entsprechend seiner Definition – umgegangen wird. Aber es bleibt immer noch die Frage übrig, ob die Definition des Meßverfahrens „richtig“ ist. Da die empirischen Daten, die man schließlich gewinnt, von der Definition der Meßverfahren abhängen, kommt dieser Frage offenbar eine wichtige Bedeutung zu. Sie kann aber nicht empirisch geprüft werden und muß deshalb anders formuliert werden. Die Frage ist hier nicht, ob Meßverfahren richtig oder falsch sind, sondern ob sie sinnvoll sind und ob durch sie wissenswerte und verläßliche Daten gewonnen werden können.[4]

9. Die Kritik kann so zusammengefaßt werden. Die Schwäche der repräsentationalen Meßtheorie besteht darin, daß sie sich auf einen zweitrangigen Aspekt von Meßverfahren konzentriert: wie *gegebene* empirische Relationen numerisch repräsentiert werden können. Das eigentlich schwierige Problem wird dadurch ausgeblendet. Dies Problem besteht zunächst allgemein darin, wie über Beschaffenheiten und Veränderungen der empirischen Realität sinnvoll gesprochen werden kann. Insoweit geht es allgemein um Fragen der Begriffsbildung. Eine durchaus spezielle Variante dieses allgemeinen Problems besteht darin, ob sich auch *Größenbegriffe* begründen lassen, um einige Aspekte der empirischen Realität zu quantifizieren.

10. Es ist nützlich, sich diese Kritik noch einmal anhand unseres Variablenbegriffs zu verdeutlichen. Eine statistische Variable ist eine Abbildung

$$X : \Omega \longrightarrow \tilde{\mathcal{X}}$$

Damit eine solche Variable definiert werden kann, benötigt man einen Merkmalsraum $\tilde{\mathcal{X}}$. Dieser Merkmalsraum faßt eine Menge von Attributen oder Merkmalsausprägungen zusammen, durch die die Objekte in Ω charakterisiert werden können. Zunächst muß also ein Merkmalsraum durch eine Explikation von Attributen oder Merkmalsausprägungen definiert werden. Ob und ggf. wie diese Attribute oder Merkmalsausprägungen

[4]Zum besseren Verständnis ist vielleicht folgende Überlegung hilfreich. Wir stellen uns vor, daß ein Meßapparat konstruiert werden soll, der die Ordnung der drei Dosen automatisch erfaßt. Man könnte vielleicht einen Computer mit geeigneten Sensoren versehen und dann so programmieren, daß er stets drei Zahlen liefert, die der momentanen Ordnung der Dosen entsprechen. Kann dann geprüft werden, ob dieser Meßapparat richtig funktioniert? Sicherlich kann man prüfen, ob sich der Apparat so verhält, wie er es aufgrund unserer Festlegungen tun sollte. Aber wie auch immer die Antwort ausfällt, sie sagt uns nur etwas über den Meßapparat, nicht jedoch darüber, ob unsere Festlegungen ein sinnvolles und zweckmäßiges Meßverfahren begründen.

durch Zahlen *repräsentiert* werden können, ist eine zweitrangige Frage,[5] die sich darauf bezieht, eine Abbildung

$$\phi : \tilde{\mathcal{X}} \longrightarrow \mathbf{R}$$

zu finden, so daß jeder Merkmalswert $\tilde{x} \in \tilde{\mathcal{X}}$ durch eine Zahl $\phi(\tilde{x}) \in \mathbf{R}$ repräsentiert werden kann (vgl. Abschnitt 4.1). Wenn es für den Merkmalsraum $\tilde{\mathcal{X}}$ bereits bestimmte Relationen gibt, sollte die numerische Repräsentation natürlich so gewählt werden, daß sie auch mit Hilfe der die Merkmalswerte repräsentierenden Zahlen formuliert werden können. Dies ist die Fragestellung der repräsentationalen Meßtheorie: ob und ggf. wie man numerische Repräsentationen ϕ finden kann, die für $\tilde{\mathcal{X}}$ gegebene Relationen erhalten.[6] Dabei wird aber vorausgesetzt, daß es diese Relationen für $\tilde{\mathcal{X}}$ bereits gibt. Die primär wichtige Frage, wie man solche Relationen begründen kann, wird von der repräsentationalen Meßtheorie gar nicht behandelt.

11. Es sei betont, daß sich diese Kritik nicht gegen die repräsentationale Meßtheorie richtet, soweit es sich um eine rein mathematische Theorie handelt; sie richtet sich vielmehr gegen die Annahme, daß eine solche mathematische Theorie zum Verständnis und zur Beurteilung von Meßverfahren helfen kann. Mit dieser Annahme beginnt indessen die in der sozialwissenschaftlichen Methodenliteratur oft vertretene, aber falsche Vorstellung: daß, indem man zunächst Relationen zwischen Objekten empirisch ermittelt, durch eine numerische Repräsentation dieser Relationen Meßverfahren begründet werden können. Der Fehler besteht darin, daß man übersieht, daß ein Meßverfahren dazu dienen soll, einen quantitativen Merkmalsraum zu begründen, daß jedoch diese Begründungsaufgabe zunächst wahrgenommen werden muß, bevor man anfangen kann, Eigenschaften von Objekten durch die Werte eines Merkmalsraums zu quantifizieren.[7]

[5] Man kann auch infrage stellen, ob beim Messen stets Zahlen verwendet werden müssen; vgl. die Hinweise bei Roche 1998, S. 4f.

[6] So kann man folgende Aussage von Coombs, Dawes und Tversky 1975, S. 45, verstehen: „Die [repräsentationale] Meßtheorie beschäftigt sich mit den Bedingungen, unter denen verschiedene Skalentypen konstruiert werden können.“ Dabei bezieht sich das Wort ‘Skalentyp’ auf numerische Repräsentationen, die wir durch ϕ symbolisiert haben.

[7] Sehr viele Skalierungsverfahren beruhen auf diesem Fehler; z.B. die Skalogramm-Analyse von L. Guttman, mit der wir uns in Abschnitt 15.2 beschäftigen werden. Ausgehend von der repräsentationalen Meßtheorie hat Gigerenzer 1981, S. 168, das Problem, das durch dieses Verfahren gelöst werden soll, folgendermaßen dargestellt:

„Wir [...] bezeichnen mit P eine Menge von Individuen, die mit einer Menge A von Testaufgaben, Einstellungsaussagen, Verhaltensweisen oder anderen Bedeutungsträgern konfrontiert wird. Das empirische System hat die Form

$$E = \langle A, P, \succ \rangle$$

Die Dominanzrelation $\succ$ ist zwischen den Mengen A und P definiert; gilt etwa $p \succ a$, dann ist $p \in P$ und $a \in A$. Beispielsweise kann $p \succ a$ als ‘p löst die Aufgabe a’ und die Ungültigkeit von $p \succ a$ als ‘p löst die Aufgabe a nicht’ bestimmt werden. Im ersten Fall dominiert das Individuum über die Aufgabe, im zweiten die Aufgabe über das Individuum. Das Repräsentationsproblem, welches sich Guttman stellte, bestand darin, ob den Individuen und den Aufgaben Zahlen so zugeordnet werden können, daß jedes Individuum genau dann eine Aufgabe löst, wenn die ihm zugeordnete Zahl größer als die der Aufgabe ist.“

Dies ist ein gutes Beispiel für ein *Repräsentations*problem. Ausgangspunkt ist eine *gegebene* Menge von Individuen bzw. Testergebnissen. Dann erst kann man feststellen, ob die von Guttman gewünschte Repräsentation möglich ist. Vielleicht ist sie in einigen Fällen möglich, sicherlich nicht immer. Man kann sich insbesondere nicht auf eine „Gesetzmäßigkeit“ berufen, um ein Verfahren zur numerischen Repräsentation von Fähigkeiten zum Lösen von Testaufgaben zu begründen. Die eigentlich wichtige Frage, ob überhaupt und ggf. wie Fähigkeiten quantifiziert werden können, wird offenbar durch das Repräsentationsproblem überhaupt nicht berührt.

12. Ergänzend sei bemerkt, daß sich die mathematischen Überlegungen, die es im Hintergrund der repräsentationalen Meßtheorie gibt, gar nicht unmittelbar auf die Frage beziehen, wie ein „empirisches Relativ“ numerisch repräsentiert werden kann. Denn die in der Methodenliteratur sogenannten „empirischen Relative“ bestehen stets aus endlichen Mengen von Objekten, für die eine oder mehrere Relationen gegeben sind; typischerweise handelt es sich um irgendeine Datenmatrix. Eine solche Datenmatrix kann jedoch meistens auf beliebig viele unterschiedliche Weisen numerisch repräsentiert werden, und zwar selbst dann, wenn man (soweit dies möglich ist) fordert, daß die numerische Repräsentation „strukturerhaltend“ sein soll. Deshalb beschäftigt sich der mathematische Zweig der repräsentationalen Meßtheorie gar nicht mit „empirischen Relativen“, sondern mit mathematischen Strukturen, die aus unendlich vielen mathematischen Objekten bestehen und deren Eigenschaften durch Axiome vorab festgelegt worden sind. Die daran anschließenden Überlegungen sind nicht nur für die Begründung von Meßverfahren irrelevant; sie helfen auch nicht bei der Frage, wie man Daten auf zweckmäßige Weise numerisch repräsentiert.

8.3 Die Idee der Objektivität

1. In der Methodenliteratur werden meistens drei Kriterien zur Beurteilung von Meßverfahren angeführt: Objektivität, Reliabilität und Validität. Diese Begriffe sind uns bereits in Abschnitt 2.2 begegnet. In einem dort angeführten Zitat werden sie von Lienert und Raatz als Gütekriterien für psychologische Tests bezeichnet. In der Methodenliteratur zur Sozialforschung werden die gleichen Begriffe mit weitgehend ähnlichen Erläuterungen ganz generell als Gütekriterien für Meßverfahren ausgegeben. Wie schon erwähnt worden ist, zeigt sich darin – und in vielen weiteren Aspekten der jeweils verwendeten Rhetorik –, daß sozialwissenschaftliche Meßverfahren meistens in Analogie zu (psychologischen) Tests verstanden werden. Dem entspricht, daß diese Rhetorik hauptsächlich dort verbreitet ist, wo sich Sozialforschung nicht mit sozialen Sachverhalten beschäftigt, sondern mit (psychologischen) Konstrukten, insbesondere dispositionalen Konstrukten. Erst dann entsteht auch eine besondere Problematik. Denn wenn sich Sozialforschung mit sozialen Sachverhalten beschäftigt, ist in der Regel durch die soziale Praxis, die die Sachverhalte konstituiert, vorgegeben, ob und ggf. wie sie quantifiziert werden können. Quantifizierende Begriffsbildung als eine eigenständige Tätigkeit der Sozialforschung findet hauptsächlich nur dort statt, wo sie Dispositionen sozialer Akteure als Eigenschaftskonstrukte „meßbar" zu machen versucht.

2. Beginnen wir mit dem Wort 'Objektivität'. Meßverfahren sollen objektiv sein. In gewisser Weise ist das leicht zu verstehen: Das Ergebnis einer Messung soll nicht davon abhängen, welches Exemplar eines reproduzierbaren Meßinstruments verwendet wird und von wem es angewendet wird. Diese Forderung ergibt sich daraus, daß man sich wünscht, daß die aus Meßverfahren gewonnenen Ergebnisse eine gewisse Autorität für soziale Kommunikations- und Interaktionsprozesse haben. Folgende Überlegung von L. Wittgenstein (1967, S. 4) macht deutlich, daß es dabei um Zweckmäßigkeitsfragen geht:

„Wie würden wir mit der Wahrheit in Konflikt geraten, wenn unsere Zollstäbe aus sehr weichem Gummi wären, anstatt aus Holz oder Stahl? – 'Nun, wir würden nicht das richtige Maß des Tisches kennenlernen.' – Du meinst, wir würden nicht, oder nicht zuverlässig, *die* Maßzahl erhalten, die wir mit unsern harten Maßstäben erhalten. *Der* wäre also im Unrecht, der den Tisch mit dem dehnbaren Maßstab gemessen hätte und behauptete, er mäße 1.80 m nach unserer gewöhnlichen Meßart; sagte er aber, der Tisch mißt 1.80 m nach der seinen, so ist das richtig. – 'Aber das ist dann doch überhaupt kein Messen!' – Es ist unserem Messen ähnlich und kann unter Umständen 'praktische Zwecke' erfüllen. (Ein Kaufmann könnte auf diese Weise verschiedene Kunden verschieden behandeln.)"

Wittgenstein weist darauf hin, daß es zum Verständnis von Meßverfahren erforderlich ist, die Zwecke zu verstehen, die mit ihnen verfolgt werden sollen, so daß sich auch Anforderungen an die Objektivität von Meßverfahren nur daraus verstehen lassen. Dieses Argument ist auch von O. D. Duncan (1984) betont und mit vielen Beispielen belegt worden. Denkt man daran, wie sich unsere heutigen Meßverfahren erst in einem langen, gelegentlich konfliktreichen Prozeß durchgesetzt haben, bis sie schließlich durch *staatliche* Garantien – nicht durch Wahrheitsansprüche – abgesichert worden sind, ist das sicherlich richtig.[8]

3. Die zu Beginn dieses Kapitels angeführten Definitionen von 'Messen' sollten deshalb auch unter dem Gesichtspunkt kritisiert werden, daß sie einen Bezug auf Zwecksetzungen nicht oder nur in wesentlich verkürzter Form vornehmen. J. Roche (1998, S. 4) erläutert diese Kritik am Beispiel der folgenden Definition von N. Campbell: „Measurement is the process of

[8]Über diesen Prozeß gibt es bereits zahlreiche historische Studien; einen guten Einstieg vermitteln die Beiträge in Wise 1995.

assigning numbers to represent quantities; the object of measurement is to enable the powerful weapon of mathematical analysis to be applied to the subject matter of science." Dazu schreibt Roche (ebda.):

„Campbell's definition of the 'object of measurement' also seems inadequate because it excludes many ordinary reasons for measurement which have always existed. Throughout history the main goal of measurement has been the control of materials and circumstances through information which allows a property to be accurately reproduced or recognized. With this function it is needed in fair trading, for example, and in manufacture. In physics and engineering it has been and is used in this manner to design, build and repair artifacts. It also, of course, has the more sophisticated uses which Campbell draws attention to."

4. Ähnlich wie bei Meßverfahren kann man auch von Tests fordern, daß sie objektiv sein sollten. Die Ergebnisse eines Tests sollten nicht davon abhängen, wer den Test durchführt. Aber tatsächlich ist es in diesem Fall schon schwieriger, die Forderung nach Objektivität zu begründen. Denn folgt man der üblichen Definition, wäre ein Test schon dann nicht objektiv, wenn in das Testergebnis auch von Menschen abzugebende Urteile einfließen. Objektivität wäre vielmehr erst erreicht, wenn Testergebnisse vollständig automatisch von nicht-menschlichen Akteuren erzeugt werden könnten, gewissermaßen als automatisch erzeugbare Orakel. — Die Forderung nach Objektivität kann tatsächlich leicht irreführend werden. Zum Beispiel heißt es bei Bortz und Döring:

„Ein Test oder Fragebogen ist objektiv, wenn verschiedene Testanwender bei denselben Personen zu den gleichen Resultaten gelangen, d.h. ein objektiver Test ist vom konkreten Testanwender unabhängig." (1995, S. 180)

Man kann wohl verstehen, was es heißen soll, daß ein Test objektiv sein sollte. Aber diese Forderung auch auf Fragebögen (Interviews) zu beziehen, erscheint, zumindet in der von Bortz und Döring gewählten Formulierung, ziemlich unsinnig. Denn diese Forderung wäre bei schriftlichen Interviews (wenn die Fragebögen mit der Post verschickt werden) immer, bei mündlichen Interviews jedoch niemals erfüllt.[9] Sie berührt in diesem Fall auch gar nicht das Problem: was man mit Hilfe von Interviews herausbekommen möchte. Definiert man Interviews, wie in Abschnitt 2.1 vorgeschlagen wurde, als kommunikative Verfahren der Informationsgewinnung, können Interviews – abgesehen von ihren nicht-kommunikativen Grenzfällen – niemals in derjenigen Weise „objektiv" sein, wie dies von Bortz und Döring nahegelegt wird.

[9]Im Zweifelsfall könnte man sich an folgender Bemerkung von Thurstone 1932, S. 251, orientieren: „... if, instead of printing our questions on a piece of paper we should rig up an elaborate automatic contraption for exposing these questions, running the said contraption by an electric motor, and spreading plenty of kymographs and telegraph keys and speech keys and time markers all around the room, then I am sure that our studies would qualify as experimental psychology."

5. Die Diskussion über die „Objektivität“ von Interviews verdankt sich großenteils der falschen Vorstellung, daß es sich bei Interviews um Meßverfahren oder Tests handelt, von denen man auf eine explizierbare Weise Objektivität fordern kann. Folgendes Zitat von Kerlinger (1964, S. 468) kann diese falsche Vorstellung noch einmal illustrieren:

„For the most part, interviews and schedules have been used simply for gathering so-called facts. Little thought seems to have been given to their potentialities as instruments designed to measure the variables of research problems. The most important use of interviews should be to study relations and to test hypotheses. The interview, in other words, is a psychological measurement instrument. Perhaps more accurately, the products of interviews, respondent's answers to carefully contrived questions, can be translated into measures of psychological variables. Interviews and interview schedules are therefore subject to the same criteria of reliability, validity, and objectivity as any other measuring instruments.“

Tatsächlich verhält es sich umgekehrt. Erst durch die Zurückweisung der falschen Auffassung, daß Interviews als Meßverfahren verstanden werden können, kann man ein Verständnis für die Frage entwickeln, in welcher Weise und in welchem Sinn durch Interviews zutreffende Informationen gewonnen werden können. Denn dies setzt voraus, daß man sich zunächst überlegt, von welcher Art die Informationen sind, die man durch Interviews gewinnen möchte: entweder Informationen über empirisch fixierbare Sachverhalte oder Meinungsäußerungen, mit denen Präferenzen oder Bewertungen zum Ausdruck gebracht werden. Der Wunsch, richtige Informationen zu gewinnen, kann bestenfalls formuliert werden, wenn sich Fragen auf Sachverhalte beziehen. Aber auch dann hängt es von der Art der Sachverhalte ab, ob und ggf. wie ein solcher Wunsch formuliert werden kann und welche Kompetenzvoraussetzungen bei den Befragungspersonen erforderlich sind.

6. Glaubt man indessen, Interviews als „psychologische Meßinstrumente“ auffassen zu können, wird schon die Forderung nach „Objektivität“ zu einer sinnlosen Phrase. Man denke etwa an die Frage, ob man mit Rating-Skalen „objektive Meßergebnisse“ gewinnen kann. Man könnte die Frage einerseits trivialerweise mit „ja“ beantworten, wenn man sich an folgender Definition von Kerlinger (1964, S. 479) orientiert: „Objective methods of observation are those in which anyone following the prescribed rules will assign the same numerals to objects and sets of objects as anyone else.“ Denn wenn bei einem Interview oder Test eine Rating-Skala verwendet wird, bleibt es offenbar den „Testpersonen“ überlassen, irgendetwas zu „messen“.[10] Andererseits kann man auch – wie z.B. Bortz und Döring

[10]Die hierzu in der Literatur zur „Einstellungsmessung“ vertretenen Auffassungen sind gelegentlich äußerst merkwürdig. Zum Beispiel schreibt Witte (1989, S. 364): „Bei der Messung einer Einstellungsintensität über einen Fragebogen (Einstellungsskala) muß

1995, S. 143 – die „Testpersonen“ verdächtigen, unzuverlässige „Meßinstrumente“ zu sein, so daß dann scheinbar folgendes Problem entsteht:

„Ein zentrales Problem aller auf menschlichen Urteilen basierenden Meßverfahren ist deshalb die Frage, wie die Unsicherheiten im menschlichen Urteil minimiert oder doch zumindest kalkulierbar gemacht werden können.“ (ebda.)

Ähnliche Ansichten haben gelegentlich auch andere Autoren geäußert;[11] verbreitet sind sie insbesondere in der Literatur über „reaktive Meßverfahren“.[12] Aber die Vorstellung, daß sich Menschen, wenn sie in Interviews nach Meinungen, Präferenzen und Bewertungen gefragt werden, als „Meßinstrumente“ betätigen, erscheint absurd.[13] Nicht nur kann die Forderung,

ebenfalls ein Gesetz unterstellt werden. Im einfachsten Fall kann man annehmen, daß durch die einzelnen Einstellungsitems eine standardisierte Form dessen vorliegt, was intern in der Person gemessen werden soll, was dann der Gewichtsmessung mit einer Balkenwaage entspräche. Dieses Gesetz kann aber doch nicht ganz so einfach sein, weil mindestens zwei Wahrnehmungsprozesse tangiert sind: Der Meßvorgang besteht nämlich in der Wahrnehmung der Einstellung $[W(E)]$ bzw. deren Elemente, in der Wahrnehmung des Meßinstrumentes in Form von einzelnen Items, und schließlich der Verbindung zur Wahrnehmung der Items $[W(I)]$. Man verändert das Meßgerät durch Auswahl soweit, bis beide Wahrnehmungen gleich sind. Setzt man nun voraus, daß jede Wahrnehmung nur bis auf lineare Transformationen eindeutig ist, dann gilt für jedes Individuum folgendes Gesetz:

$$aW(E) + b = cW(I) + d, \; W(E) = \frac{c}{a}W(I) + \frac{d-b}{a}$$

Jede Einstellung ist also in jedem Item nur eindeutig bis auf lineare Transformationen bestimmbar.“ So funktioniert also ein Mensch bei der „Messung“ seiner Einstellungen wie eine Balkenwaage? Aber schon die erste Voraussetzung, daß Einstellungen beobachtet werden könnten, ist falsch.

[11]Zum Beispiel Holm 1970, S. 369: „Wie entsteht nun der Fehler, der sozialwissenschaftlichem Skalieren anhaftet? Bei naturwissenschaftlicher Forschung kann der Meßfehler in der Regel auf zwei Quellen zurückgeführt werden, auf das Meßinstrument (etwa die optische Linse, die komplizierte Mechanik eines Instruments) und auf den Beobachter (der beispielsweise einen Ablesefehler begeht). In der Sozialforschung kommt noch eine dritte Quelle hinzu: der Forschungsgegenstand selbst, der befragte Mensch.“ Etwas später (S. 371) wird dann von einem „Fehler im Befragten“ und „einem Fehler im Untersuchungsgegenstand“ gesprochen. Aber schon bei der Frage, wer denn eigentlich eine „Messung“ vornimmt, gibt es Unklarheiten; denn an anderer Stelle sagt derselbe Autor 1975, S. 11: „Die Stimuli bilden also die Meßdimensionen. Die Befragten bilden die Meßobjekte. Und unsere [?] Aufgabe besteht darin, den Meßobjekten, d.h. den Befragten, einen 'Ort' auf der Meßdimension der Stimuli zuzuweisen.“

[12]Z.B. schreibt Albrecht 1975, S. 15: „Teil des Meß- und Erhebungsinstruments ist sehr oft eine Person, deren Merkmale den Response verändern können [...].“ Gleichermaßen merkwürdige Formulierungen finden sich bei Schnell, Hill und Esser 1999, S. 330: „Erhebungstechniken, bei denen sich die 'Untersuchungsobjekte' bewußt darüber sind, daß sie Gegenstand einer Untersuchung sind und die Möglichkeit haben, auf den Datenerhebungsvorgang selbst zu reagieren, werden als *'reaktive'* Meßverfahren bezeichnet. Die Tatsache, daß Personen in Erhebungssituationen z.B. auf die Art der Fragestellung oder auf den Interviewer reagieren, wird *'Reaktivität'* genannt.“ Soll man sich also zum Beispiel vorstellen, daß „Meßfehler“ dadurch entstehen, daß Befragungspersonen „auf die Art der Fragestellung oder auf den Interviewer reagieren“?

[13]So wie die von Cattell 1973, S. 99f, propagierten Tests, „die als vollständig objektiv

daß das Ergebnis einer Messung von der Person, die die Messung vornimmt, unabhängig sein sollte, nicht erfüllt werden. In diesem Fall beziehen sich nicht einmal die Fragestellungen auf Sachverhalte, die in irgendeiner objektivierbaren Weise beschrieben werden könnten.

8.4 Reliabilität und Validität

1. Ein zweites von vielen Lehrbuchautoren für wichtig gehaltenes Kriterium für Meßverfahren und Tests wird „Reliabilität" genannt. Schnell, Hill und Esser (1999, S. 145) geben folgende Erläuterung:

„Als ‘Reliabilität’ oder ‘Zuverlässigkeit’ kann das Ausmaß bezeichnet werden, in dem wiederholte Messungen eines Objektes mit einem Meßinstrument die gleichen Werte liefern. Offensichtlich ist ein Meßinstrument, das bei wiederholten Messungen desselben Objektes völlig verschiedene Meßwerte liefert, nicht zuverlässig."

Das ist allerdings eine merkwürdige Erläuterung. Angenommen, wir messen mit einem Thermometer die Temperatur des Kaffees, der sich in unserer Tasse befindet, und stellen einen Wert von 30° C fest. Dann wiederholen wir die Messung einige Minuten später und stellen einen Wert von 28° C fest. Sollte man dann die Schlußfolgerung ziehen, daß das Thermometer nicht zuverlässig mißt? Es erscheint doch plausibler anzunehmen, daß sich der Kaffee etwas abgekühlt hat. Das Beispiel zeigt bereits, daß die Forderung nach „Reliabilität" in der von Schnell, Hill und Esser gegebenen Definition nicht sinnvoll ist.[14]

2. Man erhält vielleicht näheren Aufschluß, wenn man herauszufinden versucht, wie die „Reliabilität" eines Meßverfahrens festgestellt werden soll. In der Methodenliteratur werden hauptsächlich zwei Vorschläge gemacht. Der erste Vorschlag wird als *Test-Retest-Methode* bezeichnet. Schnell, Hill und Esser (1999, S. 145) geben folgende Erläuterung:

„Die zeitliche Stabilität von Meßergebnissen als ein Aspekt der Reliabilität ließe sich am einfachsten dadurch messen, daß dasselbe Meßinstrument zweimal auf dasselbe Objekt angewendet würde. Die Korrelation zwischen beiden Messungen wäre dann eine Schätzung der Reliabilität. Diese ‘Test-Retest-Methode’ basiert auf der Annahme, daß die wahren Werte zwischen den beiden Messungen unverändert bleiben."

Aber das ist natürlich keine Lösung des Problems.[15] Nicht nur kann man nicht wissen, ob sich das, was gemessen werden soll, verändert hat oder nicht. Vielmehr soll das Meßinstrument (wenn es eines ist) gerade dazu dienen, um festzustellen, ob und ggf. welche Veränderungen eingetreten sind.[16] Es gibt zwar in vielen Fällen ein mehr oder weniger großes Erfahrungswissen darüber, wie schnell oder langsam sich Sachverhalte verändern. Zum Beispiel verändert sich die menschliche Körpergröße nur langsam, und wenn sie zunächst heute und dann morgen gemessen wird, sollten – wenn man richtig gemessen hat – ungefähr die gleichen Werte herauskommen. Aber wie verhält es sich bei sozialen Sachverhalten? Stellt man heute fest, daß jemand noch einen Arbeitsplatz hat, kann er morgen schon arbeitslos sein. Und wie verhält es sich schließlich bei Meinungen, Einstellungen und Fähigkeiten? Eine wichtige Differenz besteht hier offenbar darin, daß Menschen lernfähig sind und sich dementsprechend ihre Meinungen, Einstellungen und Fähigkeiten verändern können. Man könnte zum Beispiel die gleiche Klausur an zwei aufeinanderfolgenden Tagen durchführen lassen; könnte dann ein Vergleich der Klausurergebnisse Aufschluß über die „Reliabilität" des Verfahrens geben?

3. Eine zweite Methode, die vorgeschlagen wurde, um „Reliabilität" festzustellen, wird von A. Diekmann (1995, S. 217f) so erläutert:

„Bei der *Methode der Testhalbierung* (‘Split-half-Reliabilität’) wird ein Meßinstrument mit multiplen Indikatoren in zwei Hälften aufgeteilt. Aus der Korrelation zwischen beiden Testhälften kann die Split-half-Reliabilität (Testhalbierungs-Reliabilität) errechnet werden."

Dies Verfahren bezieht sich auf quantifizierende Indizes, denen eine Mehrzahl von Indikatoren zugrundeliegt. In unserer Terminologie besteht der Ausgangspunkt in einem Index $X^* := g(X_1, \ldots, X_m)$. Dann werden die

bezeichnet werden können. [...] Da diese Tests so konstruiert sind, daß die Vp möglichst nicht weiß, nach welchem Gesichtspunkt ihr Verhalten wirklich bewertet wird, kann man diese Tests als objektiv bezeichnen."

[14] Besonders unsinnig erscheint das in der Methodenliteratur verbreitete Reden von „Reliabilität", wenn es auf Interviews angewendet wird. Ausgehend von der falschen Vorstellung, daß es sich bei Interviews um Meßverfahren handelt, sagt zum Beispiel Atteslander 1975, S. 125: „Unter *Verläßlichkeit* oder Zuverlässigkeit (reliability) versteht man die Stabilität eines Ergebnisses bei wiederholten Messungen. Ein Befragungsinstrument ist also dann verläßlich, wenn es exakt mißt, so daß bei einer Wiederholung unter gleichen Bedingungen das gleiche Ergebnis erzielt wird." Ganz entsprechend heißt es bei McKennell 1977, S. 196, über Tests: „A test is perfectly reliable if individuals obtain exactly the same score on it on two separate occasions."

[15] Nur am Rande sei erwähnt, daß man bei zwei Messungen keine Korrelation feststellen kann, sondern daß dieser Begriff nur im Hinblick auf zwei oder mehr statistische Variablen sinnvoll verwendet werden kann.

[16] Insbesondere dann, wenn ein verhältnismäßig langes Zeitintervall zwischen den beiden Zeitpunkten liegt. Bortz und Döring 1995, S. 182, sagen: „in der Regel sind es mehrere Wochen"; bei Borg und Staufenbiel 1993, S. 46, sind es „meist einige Monate". Likert 1932, S. 31, wiederholte seine Befragungen einen Monat später, um „Test-Retest-Reliabilitäten" berechnen zu können. Soweit es um „Einstellungsmessung" geht, kann man sich indessen mit einer „theoretischen Annahme" helfen, die Herkner 1991, S. 189, so formuliert: „Es ist eine (meistens implizite) Annahme so gut wie aller Einstellungstheorien, daß die Einstellung einer Person zu einem Objekt konstant ist und nur durch ziemlich massive Eingriffe von außen verändert werden kann."

Indikatoren in zwei Gruppen eingeteilt,[17] und man bildet (analog zur Funktion g) zwei Teilindizes:

$$X_1^* := g_1(X_{k_1},\ldots,X_{k_p}) \quad \text{und} \quad X_2^* := g_2(X_{k_{p+1}},\ldots,X_{k_m})$$

Hat man Daten, mit denen die Indizes berechnet werden können, kann man die Korrelation von X_1^* und X_2^* ermitteln. Die Behauptung besteht schließlich darin, daß mit Hilfe dieser Korrelation beurteilt werden kann, ob der Index X^* „reliabel" ist. — Diese Methode läuft darauf hinaus, ein Vorgehen zu wiederholen, das bereits am Beispiel der Likert-Skalierung besprochen und kritisiert worden ist: Man wählt die Indikatoren für einen Index so aus, daß es eine möglichst hohe Korrelation zwischen den Indikatoren untereinander und der Indikatoren mit dem resultierenden Index gibt. Das Verfahren liefert aber nicht nur keine sinnvollen Messungen, sondern hat auch mit „Reliabilität" überhaupt nichts zu tun. Es zeigt bestenfalls, daß man mit der einen Hälfte der Indikatoren zu ähnlichen Zahlenwerten kommt wie mit der anderen Hälfte der Indikatoren. Aber ob die resultierenden Zahlenwerte in irgendeinem nachvollziehbaren Sinn des Wortes „reliabel" sind, hat damit offenbar nichts zu tun.[18]

4. Unsere Kritik an den Lehrbuchdefinitionen von „Reliabilität" soll natürlich nicht unkenntlich machen, daß das Problem, wie verläßlich Daten

[17] Man kann auch alle möglichen Einteilungen in jeweils zwei Gruppen betrachten. Man spricht dann gelegentlich von einer „Konsistenzanalyse" der Items. Aber zum Verständnis und zur Kritik des Grundgedankens braucht diese Komplikation nicht berücksichtigt zu werden.

[18] Auf der gleichen irreführenden Idee beruht eine Kennziffer, die von Cronbach (1951) vorgeschlagen worden ist, um die Reliabilität eines Index zu quantifizieren, und die seither *Cronbachs Alpha* genannt wird. Sie bezieht sich auf einen additiven Index $X^* = X_1 + \cdots + X_m$. Die Formel, um Cronbachs Alpha zu berechnen, ist

$$\alpha := \frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^m \mathrm{V}(X_j)}{\mathrm{V}(X^*)}\right)$$

Erinnert man sich an die Gleichung

$$\mathrm{V}(X^*) = \mathrm{V}(X_1) + \cdots + \mathrm{V}(X_m) + \sum_{j \neq k} \mathrm{Cov}(X_j, X_k)$$

(man vgl. z.B. Rohwer und Pötter 2001, S. 95), erkennt man, daß die Kennziffer nur von den Kovarianzen (oder bei standardisierten Indikatoren von den Korrelationen) zwischen den Indikatoren abhängt. Je mehr die Indikatoren untereinander korrelieren, desto größer wird Cronbachs Alpha. Schnell, Hill und Esser 1999, S. 147, schreiben: „Alpha kann Werte zwischen Null und Eins annehmen; empirische Werte über 0.8 können als akzeptabel betrachtet werden. In der Praxis werden meist weit niedrigere Koeffizienten noch akzeptiert." Als ein Beispiel aus der Forschungspraxis, in dem mit Cronbachs Alpha argumentiert wird, vgl. man z.B. Jakoby und Jacob 1999. Für unseren Klausurindex findet man den Wert $\alpha = 0.42$. Würde man der Rhetorik Glauben schenken, müßte man an der Reliabilität dieses Index zweifeln. In Wirklichkeit zeigt sich nur, daß Cronbachs Alpha – wie alle Überlegungen, die sich auf Korrelationen zwischen Indikatoren berufen – mit Fragen der Reliabilität eines Index nichts zu tun hat.

Tabelle 8.4.1 Prozentualer Anteil inkonsistener Antworten bei einer Wiederholungsbefragung (Marquardt u.a. 1998).

Variable	Anzahl	Fehlerquote (%)
Familienstand	3738	0.5
Lebt Vater noch?	3736	0.4
Lebt Mutter noch?	3737	0.0
Anzahl Herzinfarkte	3750	0.3
Hatten Sie schon einen Schlaganfall?	3750	0.2
Wann war Ihr letzter Schlaganfall?	15	46.7
Schulabschluß	3715	2.7
Schulabschluß des Ehepartners	2534	3.7
Ausbildungsdauer	3719	24.7
Heiratsjahr	2848	9.4
Anzahl Schwangerschaften	1891	4.7

sind, durchaus wichtig sein kann. Dieses Problem stellt sich zum Beispiel bei Daten aus standardisierten Interviews, wenn sie sich der Intention nach auf Sachverhalte beziehen, über die von Befragungspersonen mehr oder weniger genau berichtet werden kann (also nicht z.B. bei Meinungsfragen und insbesondere nicht bei einer Verwendung von Rating-Skalen). Wenn sich eine Interviewfrage auf Sachverhalte bezieht, z.B. auf das Geburtsjahr und den Geburtsmonat, kann man sinnvoll annehmen, daß die Antworten, die man erhält, mehr oder weniger fehlerhaft sein können. Hinweise erhält man aus Wiederholungsbefragungen, in denen die gleichen Personen in einem gewissen zeitlichen Abstand über die gleichen Sachverhalte befragt werden. Als Beispiel beziehen wir uns auf eine Untersuchung, die von M. Marquardt u.a. (1998) durchgeführt worden ist. Es wurden zwei Befragungen (Gesundheitssurveys) mit dem gleichen Fragebogen durchgeführt. Die erste Befragung 1984/85 hatte 4022 Teilnehmer im Alter von 25 bis 74 Jahren; dann konnten 3753 von ihnen 1987/88 noch einmal befragt werden. Den Autoren ist natürlich bewußt, daß sich einige der von ihnen erfragten Sachverhalte in der Zeit zwischen den beiden Umfragen verändert haben können, und sie beschränken ihre Untersuchung deshalb auf Inkonsistenzen. D.h., sie definieren fehlerhafte Angaben dadurch, daß in der zweiten Befragung eine Antwort gegeben wird, die mit der in der ersten Befragung gegebenen Antwort nicht vereinbar ist. Tabelle 8.4.1 zeigt einige der von den Autoren ermittelten Ergebnisse. Man erkennt, daß es bei einigen Fragen erhebliche Inkonsistenzen gibt. Z.B. besagt die Fehlerquote von 24.7 bei der Ausbildungsdauer, daß in der zweiten Befragung eine kürzere Ausbildungsdauer angegeben wurde als in der ersten Befragung. Das Ausmaß der tatsächlichen Fehler ist vermutlich noch größer.

5. Schließlich gibt es noch ein drittes Kriterium für Meßverfahren und Tests, das in der Lehrbuchliteratur regelmäßig angeführt wird: Validität.

Schnell, Hill und Esser (1999, S. 148) geben folgende Erläuterung:

„Unter ‘Validität’ (Gültigkeit) eines Meßinstruments versteht man das Ausmaß, in dem das Meßinstrument tatsächlich das mißt, was es messen sollte. Neben Reliabilität ist Validität das zentrale Gütekriterium einer Messung.“

Ähnlich heißt es bei Bortz und Döring (1995, S. 185): „Die Validität eines Tests gibt an, wie gut ein Test in der Lage ist, genau das zu messen, was er zu messen vorgibt.“ Der Unterschied besteht nur darin, daß Schnell, Hill und Esser von Meßinstrumenten, Bortz und Döring von Tests sprechen. Die Formulierung ist allerdings in beiden Fällen gleichermaßen merkwürdig. Wie kann man zum Beispiel feststellen, ob ein Thermometer das mißt, was es messen soll? Einen sinnvollen Gedankengang liefert vielleicht die Überlegung, daß man sich in diesem Fall auf Erfahrungen beziehen kann, die man zunächst auch ohne ein Thermometer machen kann. Dann kann man sich überlegen, ob mit Hilfe eines Thermometers solche Erfahrungen besser kalkulierbar gemacht werden können. Aber ohne auf einen vorgängigen Erfahrungshintergrund Bezug nehmen zu können, erscheint die Frage, ob ein Meßinstrument „das mißt, was es messen soll“, irreführend.[19] — Der Begriff ‘Validität’ ist also irreführend. Es geht nicht darum, ob ein Meßinstrument „das mißt, was es messen soll“ (diese Frage könnte auch gar nicht beantwortet werden), sondern es geht darum, ob bzw. wie gut sich ein Meßinstrument eignet, bestimmte Zwecke zu erreichen. Ganz unabhängig davon, ob man psychologische und sozialwissenschaftliche Tests als Meßverfahren bezeichnen möchte, kann diese Frage nach der Zweckmäßigkeit natürlich auch für Tests gestellt werden.[20]

6. Daß es nicht um „Gültigkeit“, sondern um Fragen der Zweckmäßigkeit geht, wird auch sichtbar, wenn man sich einige der Vorschläge ansieht, die gemacht worden sind, um Validität praktisch zu ermitteln. Oft werden drei Arten der Validität unterschieden:

„Unter ‘empirischer Validität’ eines Instrumentes versteht man die Korrelation zwischen diesem Instrument und einer anderen beobachteten Variablen, die als Kriterium dient. Es ist also sinnlos, von ‘der’ Validität eines Instrumentes

[19] Hierzu bemerkt Roche 1998, S. 4f: „Measurement itself is always a form of indirect knowledge of a target quantity, but requires the direct knowledge of some auxiliary quantity. Even in numerical measurement direct knowledge is required. This is how we know what a metre is, for example. Without the direct knowledge of a metre, obtained through the possession of a metre, a distance expressed in metres is meaningless. Much if not all numerical measurement is based ultimately on the direct sensory knowledge of physical units. These units, moreover, preserve an irreducible qualitative element in the concepts of physical quantities. The unit of length, for example, is qualitatively different from the unit of mass. Numerical measurement requires, therefore, both numerical and qualitative knowledge.“

[20] In diesem Sinn wird auch gelegentlich in der Literatur über die Validität psychologischer Tests gesprochen, z.B. von Mehrens und Lehmann 1975, S. 109: „Validity can be best defined as the degree to which a test is capable of achieving certain aims.“

zu sprechen: Die Validität kann nur in bezug auf bestimmte andere Messungen beurteilt werden. Damit stellt sich die Frage danach, wie die Validität eines Instrumentes in der Praxis beurteilt werden kann. Die ‘American Psychological Association’ unterschied 1974 drei Formen der Validität; diese Unterscheidung nach Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität ist in der Literatur allgemein verbreitet.“ (Schnell, Hill und Esser 1999, S. 148f)

Der Hinweis auf die American Psychological Association zeigt schon, daß sich die weitere Diskussion auf Tests beziehen soll bzw. auf sozialwissenschaftliche „Meßverfahren“, die als Tests verstanden werden können.

7. Zur „Inhaltsvalidität“ sagen Schnell, Hill und Esser (1999, S. 163):

„‘Inhaltsvalidität’ bezieht sich darauf, daß möglichst alle Aspekte der Dimension, die gemessen werden sollte, berücksichtigt wurden. Diese schwammige Definition kann etwas verdeutlicht werden, wenn noch einmal auf die Konzeptualisierung eines theoretischen Begriffs Bezug genommen wird. Eine ‘gültige’ Messung kann natürlich nur erfolgen, wenn jeder Aspekt des theoretischen Begriffs in den Operationalisierungen berücksichtigt wird. Würde z.B. der Begriff ‘Intelligenz’ als ‘Problemlösungskapazität’ aufgefaßt, aber dann nur über einen Test, der lediglich ‘Rechenfertigkeit’ mißt, operationalisiert, so besäße der Test in Hinsicht auf ‘Intelligenz’ keine Inhaltsgültigkeit. Inhaltsvalidität sollte deshalb nicht als Validitätskriterium aufgefaßt werden, sondern als Idee, die bei der Konstruktion eines Instrumentes nützlich sein kann.“

Lassen wir hier die in Abschnitt 6.2 bereits kritisierte Vorstellung, daß Begriffe (durch Meßverfahren) operationalisiert werden können, beiseite. Dem letzten Satz kann man dennoch zustimmen. Bei der Konstruktion eines Tests sollte man sich überlegen, *was* getestet werden soll und wie dies durch geeignete Fragestellungen und Aufgaben fixiert werden kann.

8. Daß es bei der Frage der Validität um Zweckmäßigkeitsfragen geht, wird noch deutlicher beim Begriff der „Kriteriumsvalidität“. Schnell, Hill und Esser unterscheiden zwei Varianten, zunächst „prädiktive Validität“:

„Prädiktive Validität besitzt ein Instrument dann, wenn Voraussagen, die auf einer ersten Messung mit dem Instrument beruhen, durch spätere Messungen mit einem anderen Instrument bestätigt werden können. Z.B. kann ein Test, der ‘Wissenschaftliches Interesse’ von Gymnasiasten messen soll, dann prädiktive Validität besitzen, wenn aufgrund des Tests die Aufnahme eines Studiums erfolgreich vorhergesagt werden kann.“ (S. 149)

Es ist offensichtlich, daß man auf diese Weise nicht herausfinden kann, ob ein Test das erfaßt, was er „messen“ soll, sondern bestenfalls etwas anderes: ob er sich eignet, gewisse Voraussagen zu machen. Die zweite Variante wird von Schnell, Hill und Esser als „concurrent validity“ bezeichnet:

„Der Unterschied zwischen prädiktiver Validität und ‘concurrent validity’ liegt darin, daß ‘Vorhersagen’ bei der ‘concurrent validity’ auf Messungen zum selben Zeitpunkt bezogen sind. Die zu beurteilende Messung erfolgt also zum selben Zeitpunkt wie die Messung des Kriteriums.“ (S. 149f)

Das setzt jedoch voraus, daß zwei verschiedene Tests „zum selben Zeitpunkt“ durchgeführt werden können, was natürlich im allgemeinen nicht möglich ist. Aber selbst wenn das möglich sein sollte, bliebe rätselhaft, warum es für die Validität eines Tests sprechen sollte, wenn man mit seiner Hilfe die Ergebnisse eines anderen Tests voraussagen könnte.

9. Wie irreführend die Vorstellung einer „Kriteriumsvalidität“ ist, bemerkt man, wenn man versucht, sie ernst zu nehmen. Sei nämlich der Test durch eine Indexfunktion $X^* := g(X_1, \ldots, X_m)$ definiert. $X_1, \ldots, X_m$ sind die Indikatoren, aus deren Werten sich entsprechend der Funktion g das Testergebnis X^* zusammensetzt. Sei nun Y eine Kriteriumsvariable, deren Werte mit Hilfe des Tests vorausgesagt werden sollen. Dann handelt es sich also darum, eine Funktion f zu finden, so daß sich $f(X^*)$ möglichst gut zur Vorhersage von Y eignet. Warum aber sollte man, wenn man an diesem Ziel interessiert ist, dafür die durch den Test konstruierte Variable X^* verwenden? Einen viel flexibleren Ansatz zur Konstruktion einer Vorhersagefunktion könnte man verfolgen, wenn man nicht den hier ganz überflüssigen Umweg über den Test verfolgt, sondern direkt versucht, Y mit Hilfe der Werte der Indikatoren $X_1, \ldots, X_m$ vorauszusagen, also z.B. eine Regressionsfunktion

$$\mathrm{M}(Y \mid X_1 = x_1, \ldots, X_m = x_m) \;=\; f^*(x_1, \ldots, x_m)$$

berechnet. Man wird sich erinnern, daß dieses Argument bereits in Abschnitt 6.3 verwendet worden ist.

10. Bleibt schließlich noch die Vorstellung einer „Konstruktvalidität“. Dazu sagen Schnell, Hill und Esser (1999, S. 150):

> „Inhaltsvalidität und Kriteriumsvalidität sind nach dem bisher gesagten entweder kaum aussagekräftig oder nur selten anwendbar. In den Sozialwissenschaften ist daher die ‘Konstruktvalidierung’ von großer Bedeutung. Unter dem Begriff ‘Konstrukt’ werden theoretische Eigenschaftsdimensionen (latente Variablen) verstanden. ‘Konstruktvalidität’ liegt dann vor, wenn aus dem Konstrukt empirisch überprüfbare Aussagen über Zusammenhänge dieses Konstruktes mit anderen Konstrukten theoretisch hergeleitet werden können und sich diese Zusammenhänge empirisch nachweisen lassen.“

Versuchen wir, diesem Gedankengang zu folgen. Ausgangspunkt sind zwei oder mehr Indizes (Tests oder Meßverfahren): $X_j^* := g_j(X_{j1}, \ldots, X_{jm_j})$. $X_1^*, X_2^*, \ldots$ sind die Indizes, die von Schnell, Hill und Esser als „theoretische Konstrukte“ oder auch als „latente Variablen“ bezeichnet werden; $g_1, g_2, \ldots$ sind die entsprechenden Konstruktionsverfahren. „Konstruktvalidität“ soll vorliegen, wenn man über Zusammenhänge zwischen den Indizes $X_1^*, X_2^*, \ldots$ Hypothesen bilden und „empirisch nachweisen“ kann. Die Möglichkeit, Hypothesen zu bilden, besteht natürlich immer; es ist ja nur eine Frage der Rhetorik, für $X_1^*, X_2^*, \ldots$ passende Namen zu finden und sich dann Aussagen über Zusammenhänge auszudenken. Zum

Beispiel könnte X_1^* der Name ‘Frustration’ und X_2^* der Name ‘Aggression’ gegeben werden, und die Hypothese könnte dann lauten: „Frustrierte Menschen reagieren aggressiv.“ (Das Beispiel findet sich bei Bortz und Döring 1995, S. 7.) Der wichtige Punkt ist jedoch, daß „Konstruktvalidität“, wenn man der Definition von Schnell, Hill und Esser folgt, im Kern das gleiche meint wie die bereits kritisierte Idee einer „Kriteriumsvalidität“. Der einzige Unterschied liegt in der Rhetorik. Bei „Kriteriumsvalidität“ wird unmittelbar davon gesprochen, daß mit einem Index ein anderer Index erfolgreich vorausgesagt werden soll; bei der „Konstruktvalidität“ wird dagegen von „theoretischen Zusammenhängen“ gesprochen, um auf Beziehungen zwischen Indizes zu verweisen.[21] Die unterschiedliche Rhetorik ändert allerdings nichts daran, daß es sich bei der so verstandenen „Konstruktvalidität“ nur um eine Variante der „Kriteriumsvalidität“ handelt. Auch Schnell, Hill und Esser bestätigen diese Schlußfolgerung, wenn sie schließlich folgendes Beispiel geben:

> „Beispielsweise könnten die Konstrukte ‘Entfremdung vom politischen System’ und ‘Unzufriedenheit mit dem politischen System’ erhoben werden. Es ist zu vermuten, daß bei vorhandenem politischen Interesse zwischen den beiden Konstrukten ein Zusammenhang besteht. Kann ein solcher Zusammenhang empirisch nachgewiesen werden, so kann die Hypothese der Konstruktvalidität akzeptiert werden.“ (Schnell, Hill und Esser 1999, S. 151)

11. Damit endet unsere Diskussion wie schon bei der „Reliabilität“ in einer Sackgasse. Man hat vielleicht eine vage Vorstellung bekommen, was mit „Reliabilität“ und „Validität“ gemeint sein könnte, aber keinerlei Hinweise, wie diese Vorstellungen präzisiert und praktisch verfolgt werden könnten. Das ist sicherlich auch eine Folge der in der Methodenliteratur verbreiteten Neigung, Überlegungen, die sich auf den potentiellen Sinn von Begriffen beziehen müßten, durch formal-statistische Betrachtungen zu substituieren. Anstatt über den möglichen Sinn eines Begriffs der Reliabilität nachzudenken, wird über Korrelationen zwischen Indikatoren geredet; und statt dem Begriff der Validität durch eine Reflexion von Zwecken nachzugehen, die mit Tests oder Meßverfahren verfolgt werden können, wird die gleiche Idee – jetzt in der Form von Regressionsfunktionen – nur noch einmal ausgebreitet. So wundert es nicht, daß das eigentliche Problem, wenn es um Meßverfahren und nicht um Tests geht, bei der Diskussion von „Gütekriterien“ gar nicht thematisiert wird: wie nämlich sinnvolle *Größenbegriffe* begründet werden können.

[21] Folgende Äußerung von Diekmann 1995, S. 224, paßt ersichtlich gut zu diesem Verständnis: „Mit der Konstruktvalidität [...] wird die Aufmerksamkeit auf die Brauchbarkeit von Meßinstrumenten für die Entwicklung von Theorien gelenkt. Konstruktvalidität verlangt, daß das von einem Meßinstrument erfaßte Konstrukt mit möglichst vielen anderen Variablen in theoretisch begründbaren Zusammenhängen steht und hieraus Hypothesen ableitbar sind, die einer empirischen Prüfung standhalten.“

Kapitel 9

Größen und Metriken

Wie die Ausführungen im vorangegangenen Kapitel gezeigt haben, ist die in der sozialwissenschaftlichen Methodenliteratur verbreitete Rhetorik des Messens in vielerlei Hinsicht fragwürdig. Demgegenüber halten wir an der grundlegenden Unterscheidung von qualitativen und quantitativen Merkmalsräumen für Variablen fest, wie sie in Abschnitt 4.1 eingeführt worden ist. Dabei haben wir uns bisher an folgender Überlegung orientiert: Wenn es einen sinnvollen Größenbegriff gibt, kann durch ihn ein quantitativer Merkmalsraum definiert werden (der dann seinerseits einen Ausgangspunkt liefert, um quantitative Variablen zu definieren). In diesem Kapitel soll besprochen werden, daß man sich der Problematik des Quantifizierens noch aus einer anderen Perspektive nähern kann, bei der nicht unmittelbar von der Idee quantifizierbarer Eigenschaften ausgegangen wird.

9.1 Größen und Skalen

1. Die Hauptfrage, die wir in diesem Text verfolgen, richtet sich darauf, wie man auf sinnvolle Weise zu statistischen Variablen gelangen kann. Deshalb sei zunächst noch einmal an die Definition erinnert. Eine statistische Variable ist eine Funktion $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, die jedem Objekt ω aus einer Menge Ω einen Wert $X(\omega)$ in einem Merkmalsraum $\tilde{\mathcal{X}}$ zuordnet. Zur Begründung einer Variablen muß also zunächst ein Merkmalsraum definiert werden. Wie in Abschnitt 4.1 besprochen worden ist, kann eine solche Definition stets so vorgenommen werden, daß ein numerischer Merkmalsraum entsteht, der als eine Teilmenge der reellen Zahlen aufgefaßt werden kann. Ganz unabhängig von der jeweils verwendeten numerischen Repräsentation wurden zwei Arten von Merkmalsräumen unterschieden. Einerseits qualitative Merkmalsräume, bei denen die Elemente eines Merkmalsraums (die Merkmalswerte oder Merkmalsausprägungen) qualitativ unterschiedliche Attribute sind. Andererseits quantitative Merkmalsräume, bei denen die Elemente des Merkmalsraums als Werte einer Größe aufgefaßt werden können. Zur Begründung eines quantitativen Merkmalsraums muß also auf einen Größenbegriff Bezug genommen werden.

2. Um herauszufinden, was für den Begriff einer Größe wesentlich ist, bietet sich folgende Überlegung an: die möglichen Werte eines Größenbegriffs können in eine lineare Ordnung gebracht werden. Anders gesagt: von jeweils zwei Werten einer Größe kann man sagen, ob sie gleich sind oder ob der eine größer ist als der andere. Diese Überlegung entspricht der Definition von H. von Helmholtz, auf die wir uns in Abschnitt 4.1 bezogen haben. Sie liefert auch eine allgemeine Definition eines quantitativen

Merkmalsraums. Dafür kann vorausgesetzt werden, daß eine numerische Repräsentation schon zur Verfügung steht, daß also der Merkmalsraum $\tilde{\mathcal{X}}$ als eine Teilmenge der reellen Zahlen betrachtet werden kann. Infolgedessen stehen auch schon die üblichen Relationen für numerische Größen zur Verfügung, nämlich 'gleich', 'kleiner' und 'größer'; wir nennen sie die *numerischen Standardrelationen.* Somit kann ein quantitativer Merkmalsraum durch die Forderung definiert werden, daß sich die numerischen Standardrelationen, wie sie für die numerische Repräsentation des Merkmalsraums definiert sind, durch Verweis auf einen Größenbegriff explizieren lassen. Die erforderliche Explikation hängt dabei von der Art des Größenbegriffs ab. Bei Zählgrößen muß dargelegt werden, was gezählt werden soll; bei Meßgrößen muß sich die Explikation auf ein Meßverfahren beziehen; und bei monetären Größen kann man sich auf ihre gesellschaftlich institutionalisierte Verwendung beziehen. Wenn daraus weitere Größenbegriffe abgeleitet werden, muß natürlich auch das Konstruktionsverfahren erläutert werden.

3. Aus dieser Definition eines quantitativen Merkmalsraums kann man den Begriff einer Skala gewinnen, wie er in Abschnitt 4.3 eingeführt worden ist: Das Wort bezeichnet die numerische Repräsentation eines (quantitativen) Merkmalsraums. Die Festlegung einer Skala bildet somit einen wesentlichen Teil der Definition eines quantitativen Merkmalsraums. Denn man muß sich auf eine jeweils bestimmte numerische Repräsentation beziehen, um die einzelnen Merkmalswerte im Hinblick auf einen Größenbegriff zu explizieren. Wenn man z.B. einen Merkmalsraum für monetäre Größen definieren will, muß man eine bestimmte numerische Repräsentation fixieren, so daß man sagen kann: die Zahl 1 soll 1 DM bedeuten, oder: die Zahl 1 soll 1000 DM bedeuten, je nachdem wie man es haben möchte.

4. Wie schon in Abschnitt 4.3 angemerkt wurde, wird das Wort 'Skala' in der sozialwissenschaftlichen Methodenliteratur in unterschiedlichen Bedeutungen verwendet. Noch etwas verwirrender wird es dadurch, daß oft von unterschiedlichen „Skalenniveaus“ oder „Meßniveaus“ gesprochen wird. Versuchen wir zu verstehen, was damit gemeint ist. Bei Schnell, Hill und Esser (1999, S. 136ff) finden sich folgende Erläuterungen:

„Skalen können danach klassifiziert werden, welche Transformationen für sie zulässig sind. Solche Klassifikationen werden als 'Skalenniveaus' oder 'Meßniveaus' bezeichnet. Da es unendlich viele zulässige Transformationen einer bestimmten Skala gibt, können prinzipiell auch unendlich viele verschiedene Skalenniveaus unterschieden werden. Die meistverwendete Klassifikation geht auf Stevens (1946) zurück; Stevens unterscheidet Nominal-, Ordinal-, Intervall- und Ratioskalen.

'*Nominalskalen*' erfordern auf der Ebene der Objekte lediglich die Möglichkeit, die Objekte auf Gleichheit in bezug auf die interessierende Dimension zu unterscheiden. Die Zuordnung der Zahlen stellt hier dann lediglich eine Benennung dar. Nominale Messung besteht also in der Erstellung einer einfachen Klasseneinteilung, die *jedes* Objekt genau *einer* Klasse zuordnet: Weder darf ein bestimmtes Objekt überhaupt nicht zugeordnet werden, noch darf ein Objekt mehreren Klassen zugeordnet werden. [...]

'*Ordinalskalen*' erfordern zusätzlich zur Möglichkeit, Objekte auf ihre Gleichheit zu untersuchen, die Möglichkeit der Rangordnung der Objekte in bezug auf die interessierende Dimension. Die entsprechend zugeordneten Zahlen müssen diese Rangordnung wiedergeben. Demzufolge sind keine mathematischen Transformationen zulässig, die die Abfolge der Zahlen verändern. Beispiele für Ordinalskalen können z.B. Rangordnungen von Lehrbüchern nach ihrer Langweiligkeit sein. Das traditionelle Beispiel sind Schulnoten, bei denen man (angeblich) sagen kann, daß eine '2' besser als eine '4', jedoch nicht, daß eine '2' doppelt so gut wie eine '4' ist. (Es muß natürlich erst nachgewiesen werden, daß die Axiome der Ordinalskala in diesem Fall tatsächlich erfüllt sind.)

'*Intervallskalen*' erfordern zusätzlich zur Unterscheidungsmöglichkeit und zur Rangordnungsmöglichkeit die Möglichkeit der Angabe, daß die Unterschiede zwischen zwei beliebigen aufeinanderfolgenden Objekten jeweils gleich groß sind; mit anderen Worten: Die Intervalle müssen die gleiche Größe besitzen. Die entsprechenden Zahlen müssen daher stets dieselbe Differenz besitzen. Damit sind nur mathematische Transformationen zulässig, die die Größe der Intervalle nicht oder alle Intervalle im selben Maß verändern. Das beliebteste Beispiel für Intervallskalen sind Temperaturmessungen in Celsius, Fahrenheit oder Reaumur.

'*Ratioskalen*' erfordern zusätzlich zur Unterscheidungs- und Rangordnungsmöglichkeit und der gleichen Intervallgröße die Möglichkeit, die Existenz eines '*natürlichen Nullpunktes*' zu zeigen: Der Meßwert 'Null' entspricht der tatsächlichen Abwesenheit des gemessenen Merkmals. Bei einer Ratioskala werden Aussagen, wie 'Objekt A besitzt doppelt soviel X wie das Objekt B', sinnvoll, mit anderen Worten, es sind Aussagen über Quotienten jeweils zweier beliebiger Objekte möglich. Damit ist als einzige mathematische Transformation nur noch die Multiplikation mit einer Konstanten (ungleich Null) zulässig. Bei einer Ratioskala sind die Meßwerte bis auf die Wahl der Einheit festgelegt. Beispiele sind Messungen der Temperatur in Kelvin und des elektrischen Widerstands in Ohm."

Ähnliche Ausführungen finden sich sehr oft in der Methodenliteratur zur Sozialforschung. Es gibt jedoch eine Reihe von Problemen.[1]

5. Ein erstes Problem betrifft schon das Reden von einer „Nominalskala". Die Erläuterungen zeigen, daß tatsächlich die numerische Repräsentation eines *qualitativen* Merkmalsraums gemeint ist. Dann sollte indessen nicht von „Messen" (Quantifizieren) gesprochen werden, sondern von „Klassifizieren". Es handelt sich nicht um ein „Meßniveau". Man denke als Beispiel an den Schichtindex von Scheuch und Daheim, der in Abschnitt 6.3 besprochen wurde. Es handelt sich um ein Verfahren, mit dem Menschen in „Schichten" eingeteilt werden können. Man stellt ihr Einkommen, ihre Bildung und ihren Beruf fest und verwendet dann den Schichtindex, um sie einer bestimmten „Schicht" zuzuordnen. Es wird dadurch nicht irgendeine Eigenschaft der Menschen erfaßt – insbesondere keine quantifizierbare

[1] Lesenswerte kritische Überlegungen finden sich u.a. bei Duncan 1984, Kap. 4.

Eigenschaft –, sondern es handelt sich um ein von Scheuch und Daheim ausgedachtes Klassifikationsverfahren.

6. Ein weiteres Problem besteht darin, daß die Unterscheidung von Skalenniveaus auf einer Verwechslung von „Repräsentieren" und „Quantifizieren" beruht.[2] Betrachten wir die Erläuterungen, die Schnell, Hill und Esser zum Begriff einer „Ordinalskala" geben. Eine solche Skala beruht, wie sie sagen, auf der Möglichkeit, Objekte „in bezug auf die interessierende Dimension" in eine Rangordnung zu bringen. Das kann man allerdings immer machen, und jeder kann es machen, wie er möchte. Man denke zum Beispiel an die Lehrbücher von Schnell, Hill und Esser, Bortz und Döring, Diekmann, Kromrey und Laatz. Jeder kann sie auf seine Weise in eine Rangordnung bringen. Wenn man es mit einer bestimmten Menge von Objekten zu tun hat, kann man ihnen immer nach irgendwelchen Gesichtspunkten Zahlen zuordnen und dann die numerischen Standardrelationen zwischen den Zahlen als Ausdruck einer Rangordnung interpretieren. Auf dieser Beliebigkeit beruhen insbesondere Rating-Skalen.

7. Betrachten wir das von Schnell, Hill und Esser (1999, S. 137) angeführte Beispiel der Schulnoten. Dazu gibt es den Zusatz: „Es muß natürlich erst nachgewiesen werden, daß die Axiome der Ordinalskala in diesem Fall tatsächlich erfüllt sind." Um zu verstehen, was damit gemeint sein könnte, muß man untersuchen, wie die Autoren auf die repräsentationale Meßtheorie Bezug nehmen:

„Unter 'Messen' wird allgemein die Zuordnung von Zahlen ('Meßwerten') zu Objekten gemäß festgelegten Regeln verstanden. Die Mängel dieser Definition werden bereits dadurch deutlich, daß die Zuweisung von Schulnoten durch Würfeln eine definitionsgemäße Messung wäre." (S. 132) „Objekte können zwar mit völlig verschiedenen Regeln auf eine Zahlenmenge abgebildet werden, aber nur bestimmte Regeln liefern eine Abfolge der Zahlen, die der Abfolge der Objekte hinsichtlich der zu messenden Eigenschaft entspricht: Nur ganz bestimmte Regeln liefern 'strukturtreue Abbildungen'. Am Beispiel der Schulnoten läßt sich das deutlich zeigen: Der theoretische Begriff 'Leistung in einem bestimmten Fach' erfordert die Möglichkeit, Schüler nach ihrer 'Leistung' zu ordnen. Die Abbildung dieser Reihenfolge der Schüler in eine Zahlenmenge durch Würfeln erbringt aber keine Abfolge der Zahlen, die der Reihenfolge der Schüler nach Leistung entspricht (das ist zwar möglich, aber dann ein zufälliges und sehr seltenes Ereignis). Wäre es hingegen möglich, eine Regel (eine Meßtechnik) anzugeben, die eine Abfolge der Zahlen erbringt, die der Reihenfolge der 'Leistung' der Schüler entspricht (so daß z.B. gilt: hohe 'Leistung' = große Zahl, kleine 'Leistung' = kleine Zahl), so wäre eine strukturtreue Abbildung, eben eine 'Messung' gelungen." (S. 133) „Wie das Beispiel der gewürfelten Schulnoten zeigt, ist nicht jede Abbildung eine homomorphe [strukturtreue] Abbildung: Die Existenz einer Skala muß erst bewiesen werden. Dieser Beweis wird mit einem 'Repräsentationstheorem' geführt. Ein Repräsentationstheorem gibt die Bedingungen im

[2] Im wesentlichen ist dies eine Folge der Orientierung an einer repräsentationalen Meßtheorie, vgl. Abschnitt 8.2.

empirischen Relativ an, die erfüllt sein müssen, damit eine homomorphe Abbildung möglich ist. Diese Bedingungen werden in diesem Zusammenhang als 'Axiome' bezeichnet." (S. 134)

Aber wie kann geprüft werden, ob im Beispiel der Schulnoten die „Axiome der Ordinalskala" erfüllt sind? In gewisser Weise ist das ganz leicht: denn *jedes* Verfahren, mit dem man Schüler in eine Rangordnung bringen kann, ist offenbar hinreichend, um die Rangfolge durch eine „Ordinalskala" numerisch zu repräsentieren. Insbesondere liefert auch das Auswürfeln von Schulnoten ein Verfahren, um Schüler in eine Rangordnung zu bringen. Dann hat man ein „empirisches Relativ": eine Menge von Schülern und eine durch Auswürfeln zustandegekommene „empirische Relation". Verwendet man die durch das Auswürfeln erzeugten Zahlen, hat man auch gleich eine passende numerische Repräsentation. Es ist also nicht richtig, wenn Schnell, Hill und Esser sagen, daß das Auswürfeln von Schulnoten die „Axiome einer Ordinalskala" verletzt. Aus der Perspektive der repräsentationalen Meßtheorie ist vielmehr das Auswürfeln von Schulnoten genausogut wie jedes andere Verfahren, um Schüler in eine Rangordnung zu bringen. Aber darin liegt gerade die Unzulänglichkeit dieser Konzeption von „Messen". Das eigentliche Problem besteht in diesem Beispiel vielmehr darin, sinnvolle von nicht sinnvollen Verfahren zur Ermittlung von Schulnoten unterscheidbar zu machen. Das würde erfordern, entsprechende Sinnkriterien zu explizieren. Die „Axiome" der repräsentationalen Meßtheorie helfen jedoch dafür überhaupt nicht. Insofern ist auch der Satz: „Es muß natürlich erst nachgewiesen werden, daß die Axiome der Ordinalskala in diesem Fall tatsächlich erfüllt sind" ganz irreführend.[3]

8. Das nächste Problem betrifft die Unterscheidung zwischen „Ordinalskala" und „Intervallskala". Folgt man den oben zitierten Erläuterungen von Schnell, Hill und Esser, ist eine „Intervallskala" eine „Ordinalskala", bei der zusätzlich gilt: „daß die Unterschiede zwischen zwei beliebigen aufeinanderfolgenden Objekten jeweils gleich groß sind;[4] mit anderen Worten: Die Intervalle müssen die gleiche Größe besitzen. Die entsprechenden Zahlen müssen daher stets dieselbe Differenz besitzen." Aber *das* kann man bei *jeder* „Ordinalskala" leicht erreichen. Es genügt schon, die natürlichen Zahlen zu verwenden, um eine Rangfolge zu repräsentieren. So wird es zum Beispiel bei den Schulnoten gemacht: die Differenz zwischen einer 1 und einer 2 ist genauso groß wie die zwischen einer 4 und einer 5. Aber warum sagen dann Schnell, Hill und Esser, daß es sich bei Schulnoten um eine „Ordinalskala", nicht um eine „Intervallskala" handelt? Sie scheinen nicht zu bemerken, daß man ihren Erläuterungen zufolge jede „Ordinalskala" auch als eine „Intervallskala" betrachten kann.

9. Sehen wir uns an, wie andere Autoren die Unterscheidung erläutern. Bei Laatz (1993, S. 353) heißt es:

> „Das Meßniveau ist davon abhängig, welche der vier Eigenschaften reeller Zahlen tatsächlich erreicht werden: Kann man Fälle als gleich oder ungleich unterscheiden? Kann man sie in eine eindeutige Ordnung bringen? Ist es möglich, gleiche Abstände zwischen Kategorien anzunehmen? Existiert ein absoluter Nullpunkt? Diese vier Kriterien bauen hierarchisch aufeinander auf, so daß ein höheres immer die Existenz des niedrigeren Kriteriums voraussetzt."

Hier bezieht sich das Wort 'Kategorien' auf Merkmalsausprägungen, nicht auf die zu ihrer Repräsentation verwendeten Zahlen. Aber wie kann man feststellen, ob es einen „gleichen Abstand" zwischen Merkmalsausprägungen gibt? In die gleiche Schwierigkeit führt auch folgende Erläuterung von Bortz und Döring (1995, S. 67):

> „Für eine Intervallskala gilt, daß gleich große Merkmalsunterschiede durch äquidistante Zahlen abgebildet werden, d.h. identische Meßwertunterschiede zwischen Objektpaaren entsprechen identischen Merkmalsunterschieden."

Denken wir wieder an Schulnoten: Der Unterschied zwischen der Note 1 und der Note 2 ist offenbar genauso groß wie der Unterschied zwischen der Note 4 und der Note 5; der Unterschied beträgt jeweils eine Note. Man kann natürlich sagen: Der Unterschied *bedeutet* in den beiden Fällen etwas anderes.[5] Aber Argumente, die sich auf Bedeutungen beziehen, kann man gleichermaßen auch bei „Intervallskalen" anführen. Zum Beispiel wird Einkommen meistens als eine „intervallskalierte" Größe betrachtet. Aber

[3] Es ist auch ziemlich unklar, warum an dieser Stelle das Wort 'Axiom' verwendet wird. In dem oben angeführten Zitat wird auf „Bedingungen im empirischen Relativ" Bezug genommen. Aber solche Bedingungen können sich verändern; zum Beispiel kann sich eine durch Schulnoten hergestellte Rangfolge von Schülern verändern. Eine bestimmte Rangfolge kann sicherlich nicht „axiomatisch" festgelegt werden. Oder könnte etwas anderes gemeint sein? Gigerenzer 1981, S. 46, verwendet folgende Formulierung: „Gegenstand der *Meßtheorie* sind die Gesetzmäßigkeiten, welche notwendig und hinreichend für eine homomorphe Abbildung eines empirischen Systems auf ein bestimmtes numerisches System sind. Die notwendigen und hinreichenden Bedingungen sind Eigenschaften der empirischen Relationen. Sie heißen *Axiome*." Aber empirische Relationen, also empirisch feststellbare Beziehungen zwischen Objekten, Individuen und Situationen, können sich grundsätzlich immer verändern. Es ist unverständlich, warum Gigerenzer solche empirischen Relationen „Gesetzmäßigkeiten" nennt.

[4] [An dieser Formulierung zeigt sich erneut die Verwechslung von Objekten und Merkmalsräumen. Wenn man die Schüler einer Schulklasse entsprechend ihren Schulnoten in eine Reihenfolge bringt, müssen keineswegs zwischen je zwei aufeinanderfolgenden Schülern die gleichen Notenunterschiede bestehen. Es sollte also betont werden, daß sich Skalen auf Merkmalsräume, nicht auf Objekte beziehen; in diesem Beispiel auf einen Merkmalsraum, bei dessen Merkmalswerten es sich um Schulnoten handelt.]

[5] Dem entsprechen Äußerungen wie z.B. die folgende von Mehrens und Lehmann 1975, S. 67: „With interval data we can interpret the distances between scores." Bei Diekmann 1995, S. 252, heißt es: „Bei einer Ordinalskala sind die Abstände zwischen den Skalenwerten nicht sinnvoll interpretierbar." Aber: „Mit *Intervallskalen* ist nicht nur eine Aussage über die Rangordnung der Meßwerte möglich. Zusätzlich informieren die Skalenwerte auch über die Abstände (Intervalle) zwischen den Meßwerten."

sicherlich *bedeutet* der Unterschied zwischen 500 und 600 DM etwas anderes als der Unterschied zwischen 5000 und 5100 DM.

10. Dies Beispiel führt auch sogleich zu einem weiteren Problem. Schnell, Hill und Esser (wie auch viele andere Autoren) sagen, daß man bei „Ordinalskalen“ beliebige monotone Transformationen vornehmen dürfe, daß bei „Intervallskalen“ jedoch „nur mathematische Transformationen zulässig [sind], die die Größe der Intervalle nicht oder alle Intervalle im selben Maß verändern.“ Aber denken wir daran, daß man bei Einkommensfunktionen oft das logarithmierte Einkommen als abhängige Variable verwendet. Darf man das nicht tun, weil sich durch das Logarithmieren Einkommensdifferenzen nicht-linear verändern? Natürlich darf man es. Die Verwirrung verdankt sich der irreführenden Vorstellung, daß Skalenniveaus durch „zulässige Transformationen“ charakterisierbar gemacht werden könnten.[6] Hält man sich dagegen an den Begriff einer Skala, wie er zu Beginn dieses Abschnitts noch einmal erläutert worden ist, kann die Verwirrung leicht aufgeklärt werden. Das Wort ‘Skala’ (im engeren Sinn) bezieht sich auf die numerische Repräsentation eines quantitativen Merkmalsraums. Um eine Skala zu erläutern, muß man also erklären, *wie* man sich mit ihr auf einen Größenbegriff beziehen möchte. Es ist auch evident, daß man für denselben Größenbegriff, z.B. ‘Einkommen’, ganz unterschiedliche Skalen verwenden kann. Hat man eine Skala, kann man aus ihr durch beliebige Transformationen neue Skalen bilden. Hat man zum Beispiel eine Skala $\tilde{\mathcal{X}}$ für Einkommen, kann man durch eine Funktion

$$\phi : \tilde{\mathcal{X}} \longrightarrow \tilde{\mathcal{Y}}$$

eine neue Skala $\tilde{\mathcal{Y}}$ bilden. Beide Skalen beziehen sich auf den Größenbegriff ‘Einkommen’. Und versteht man die Skala $\tilde{\mathcal{X}}$ und kennt man die Funktion (Skalentransformation) ϕ, versteht man auch die Skala $\tilde{\mathcal{Y}}$. Es gibt keinen erkennbaren Gesichtspunkt, um an dieser Stelle „zulässige“ von „unzulässigen“ Skalentransformationen zu unterscheiden.[7]

11. Es bleibt die Frage, ob und ggf. wie eine Unterscheidung zwischen „Ordinalskalen“ und „Intervallskalen“ gemacht werden kann. Wie deutlich geworden sein sollte, ist es hierfür erforderlich, daß man über „Unterschiede“ oder „Abstände“ zwischen den Merkmalsausprägungen eines

[6] Diese Vorstellung ist zuerst von Stevens (1946) verbreitet worden und dann aus schwer nachvollziehbaren Gründen in die Methodenliteratur zur empirischen Sozialforschung eingedrungen.

[7] Die Vorstellung, daß je nach Skalenniveau nur bestimmte Skalentransformationen „zulässig“ sind, ist insbesondere in der psychometrischen Literatur verbreitet. Das praktische Verhalten steht allerdings meistens im Widerspruch dazu. Einerseits werden psychologische Tests als Meßverfahren aufgefaßt, die „intervallskalierte“ Daten liefern, so daß der Ideologie zufolge nur lineare Skalentransformationen „zulässig“ wären. Andererseits werden die „Meßergebnisse“ zum Beispiel „normalisiert“, so wie dies in Abschnitt 5.4 besprochen worden ist, also einer nichtlinearen Skalentransformation unterzogen. Man vgl. z.B. Anastasi 1982, S. 76ff.

quantitativen Merkmalsraums sprechen kann. Das Problem besteht offenbar darin, daß man dies auf ganz unterschiedliche Weisen tun kann. Für eine Bank sind 100 DM stets der gleiche Betrag; aber wenn man an Einkommensbezieher denkt, sind 100 DM manchmal 10 %, manchmal weniger als 1 % ihres Einkommens. Somit erscheint die Vorstellung, in irgendeiner absoluten Weise von „Unterschieden“ oder „Abständen“ zwischen Merkmalswerten sprechen zu können, von vornherein verfehlt. Erforderlich ist vielmehr ein Begriffsrahmen, mit dessen Hilfe unterschiedliche Vorstellungen expliziert werden können und ihr Zusammenhang reflektierbar wird. Zu diesem Zweck kann der Begriff einer Metrik herangezogen werden.

9.2 Zum Begriff der Metrik

1. Zu überlegen ist, wie über Unterschiede und Abstände gesprochen werden kann. Zunächst ist zu überlegen, worauf man sich beziehen möchte. Denken wir an unseren Begriff einer statistischen Variablen: $X : \Omega \longrightarrow \tilde{\mathcal{X}}$. Man könnte versuchen, sich direkt auf die Mitglieder von Ω zu beziehen, um über deren Unterschiede oder Abstände nachzudenken. Dafür benötigt man jedoch zunächst einen Merkmalsraum, der eine Menge von Attributen (Merkmalsausprägungen) liefert, durch die sich die Mitglieder von Ω charakterisieren lassen. Dementsprechend muß sich auch das Reden von Unterschieden und Abständen zunächst auf die im Merkmalsraum zusammengefaßten Merkmalsausprägungen beziehen. Wir erinnern deshalb an unsere anfänglich gewählte Schreibweise für Merkmalsräume:

$$\tilde{\mathcal{X}} := \{\tilde{x}_1, \tilde{x}_2, \ldots\}$$

so daß man sich auf die einzelnen Merkmalsausprägungen $\tilde{x}_1, \tilde{x}_2, \ldots$ beziehen kann. Die Frage ist, wie man über Unterschiede und Abstände *dieser Merkmalswerte* sprechen kann.

2. Hier setzt nun eine neue Begriffsbildung ein, bei der man sich auf Funktionen bezieht, mit denen jeweils zwei Merkmalswerten in $\tilde{\mathcal{X}}$ ein Abstand zugeordnet werden kann. In formaler Schreibweise wird also eine Funktion

$$d : \tilde{\mathcal{X}} \times \tilde{\mathcal{X}} \longrightarrow \mathbf{R}$$

betrachtet, die jeweils zwei Merkmalswerten $\tilde{x}, \tilde{x}' \in \tilde{\mathcal{X}}$ eine reelle Zahl $d(\tilde{x}, \tilde{x}')$ zuordnet. Wir nennen d eine *Abstandsfunktion* (oder kurz einen *Abstand*), wenn für alle $\tilde{x}, \tilde{x}' \in \tilde{\mathcal{X}}$ folgende Bedingungen erfüllt sind:

(a) $d(\tilde{x}, \tilde{x}') \geq 0$

(b) $d(\tilde{x}, \tilde{x}') = d(\tilde{x}', \tilde{x})$

(c) $d(\tilde{x}, \tilde{x}) = 0$

Diese drei Bedingungen bilden Minimalanforderungen. Denkt man z.B. an räumliche Abstände, kann man eine weitere Bedingung ins Auge fassen; daß für jeweils drei Merkmalswerte $\tilde{x}, \tilde{x}', \tilde{x}'' \in \tilde{\mathcal{X}}$ gelten soll:

(d) $d(\tilde{x}, \tilde{x}'') \leq d(\tilde{x}, \tilde{x}') + d(\tilde{x}', \tilde{x}'')$

Diese Bedingung wird *Dreiecksungleichung* genannt. Wenn bei einer Abstandsfunktion auch diese Bedingung erfüllt ist, wird sie eine *Semi-* oder *Quasi-Metrik* genannt. Schließlich spricht man von einer *Metrik*, wenn auch noch die folgende Bedingung erfüllt ist:

(e) $d(\tilde{x}, \tilde{x}') = 0 \implies \tilde{x} = \tilde{x}'$

3. Der Begriff einer Metrik (oder abgeschwächt Semi-Metrik oder Abstand) ist wichtig, weil man sich mit seiner Hilfe klarmachen kann, daß man über „Abstände“ auf ganz unterschiedliche Weisen reden kann. Für einen Merkmalsraum können beliebig viele unterschiedliche Metriken eingeführt werden. Hier sind zur Illustration drei Beispiele:

$$d(\tilde{x}, \tilde{x}') := \begin{cases} 0 & \text{wenn } \tilde{x} = \tilde{x}' \\ 1 & \text{andernfalls} \end{cases}$$

$$d(\tilde{x}, \tilde{x}') := |\,\tilde{x} - \tilde{x}'\,|$$

$$d(\tilde{x}, \tilde{x}') := (\tilde{x} - \tilde{x}')^2$$

Es gibt also keine „natürliche“ Definition von Abständen, sondern man muß explizit sagen, wie man das Wort verwenden will.

4. Ein klassisches Beispiel ist die *euklidische Metrik*. Sie kann allgemein für einen n-dimensionalen Vektorraum $\mathbf{R}^n$ definiert werden. Hat man zwei Vektoren $(x_1, \ldots, x_n), (x'_1, \ldots, x'_n) \in \mathbf{R}^n$, ist ihr euklidischer Abstand durch

$$d((x_1, \ldots, x_n), (x'_1, \ldots, x'_n)) := \left(\sum_{j=1}^{n} (x_j - x'_j)^2 \right)^{1/2}$$

definiert. Für 1-, 2- oder 3-dimensionale Räume liefert diese Definition die aus der Elementargeometrie bekannten Abstände. Auch für das praktische Leben ist dies jedoch nur eine der möglichen Abstandsdefinitionen. Man denke z.B. daran, daß man auf durchaus unterschiedliche Weisen über Abstände (Entfernungen) zwischen Städten sprechen kann. Der euklidische Abstand verwendet die „Luftlinie“; stattdessen kann man zur Definition von Abständen auch Straßen- oder Zugverbindungen verwenden oder einfach nur zwischen benachbarten und nicht-benachbarten Städten unterscheiden. Dann wird jeweils eine unterschiedliche Metrik verwendet.

5. Dieser Hinweis ist wichtig, weil es bei Größenbegriffen oft „natürlich“ erscheint, eine (eindimensionale) euklidische Metrik zu verwenden, also die Definition $d(\tilde{x}, \tilde{x}') := |\,\tilde{x} - \tilde{x}'\,|$. Fragt man zum Beispiel nach der Differenz zwischen 100 DM und 150 DM, ist die „natürliche“ Antwort: 50 DM. Das ist zwar nicht falsch, setzt jedoch eine bestimmte Metrik als Konvention voraus. Andererseits haben wir auch schon gesehen, daß man Einkommensunterschieden ganz unterschiedliche Bedeutungen geben kann. Mit Hilfe des Metrikbegriffs kann das explizit gemacht werden.

6. Man kann sich auch leicht überlegen, wie Skalentransformationen zu unterschiedlichen Metriken führen. Sei nämlich $\tilde{\mathcal{X}}$ ein Merkmalsraum, für den eine Metrik d eingeführt worden ist. Nimmt man dann eine Skalentransformation $\phi : \tilde{\mathcal{X}} \longrightarrow \tilde{\mathcal{Y}}$ vor und ist ϕ eine injektive Funktion, kann man für $\tilde{\mathcal{Y}}$ eine *induzierte Metrik*

$$d_\phi(\tilde{y}, \tilde{y}') := d(\phi^{-1}(\tilde{y}), \phi^{-1}(\tilde{y}'))$$

definieren. Als Beispiel kann man etwa daran denken, daß $\tilde{\mathcal{X}}$ ein Merkmalsraum für Einkommen und ϕ die Logarithmus-Funktion ist. Verwendet man zur numerischen Repräsentation von $\tilde{\mathcal{X}}$ und $\tilde{\mathcal{Y}}$ jeweils die reellen Zahlen, kann man natürlich in beiden Fällen auf die gleiche Weise rechnen; stets ist $1 + 1 = 2$ und $3 - 1$ ist ebenfalls 2. D.h. für das bloße Rechnen ist die Tatsache, daß mit einer Skalentransformation der Übergang zu einer neuen Metrik verbunden ist, ganz unerheblich. Sie wird erst dadurch wichtig, daß man sich mit Skalen und Metriken in jeweils spezifischer Weise auf Größen, in diesem Beispiel auf Einkommen, beziehen möchte. — Umgekehrt kann man auch den Fall betrachten, daß eine Skalentransformation $\phi : \tilde{\mathcal{X}} \longrightarrow \tilde{\mathcal{Y}}$ gerade deshalb vorgenommen wird, weil es inhaltliche Gesichtspunkte gibt, mit deren Hilfe sich für den Merkmalsraum $\tilde{\mathcal{Y}}$ eine Metrik definieren läßt. Dann liefert die Skalentransformation auch eine induzierte Metrik für $\tilde{\mathcal{X}}$. Wenn nämlich d eine Metrik für $\tilde{\mathcal{Y}}$ ist, erhält man durch

$$d_\phi(\tilde{x}, \tilde{x}') := d(\phi(\tilde{x}), \phi(\tilde{x}'))$$

eine (Semi-) Metrik d_ϕ für $\tilde{\mathcal{X}}$.

7. Die Idee einer induzierten Metrik kann man auch verwenden, um Metriken für Objekte zu definieren, so daß man über Unterschiede und Abstände zwischen Objekten (oder Individuen oder Situationen) sprechen kann. Dies setzt allerdings voraus, daß eine statistische Variable X und eine Metrik d für ihren Merkmalsraum $\tilde{\mathcal{X}}$ schon definiert worden sind. Dann kann man eine Semi-Metrik

$$d_o : \Omega \times \Omega \longrightarrow \mathbf{R}$$

durch $d_o(\omega, \omega') := d(X(\omega), X(\omega'))$ definieren. Wichtig ist, daß man zuerst einen Merkmalsraum und eine für ihn definierte Metrik hat. Erst das verschafft die begrifflichen Voraussetzungen, um – durch die Bildung einer induzierten Metrik – dann auch über Unterschiede und Abstände zwischen Objekten sprechen zu können. Man kann sich auch überlegen, daß man auf diese Weise für Objekte im allgemeinen nur eine Semi-Metrik erhält. Das heißt: Wenn für zwei Objekte $d_o(\omega, \omega') = 0$ gilt, kann daraus im allgemeinen nicht die Schlußfolgerung gezogen werden, daß es sich um identische Objekte handelt. Wenn zum Beispiel zwei Personen das gleiche Einkommen beziehen, ist ihr Einkommensunterschied Null, aber es bleiben natürlich zwei unterschiedliche Personen.

9.3 Metrische Größen

1. Mit Hilfe des Metrikbegriffs kann jetzt noch einmal über die Unterscheidung zwischen Ordinal- und Intervallskalen nachgedacht werden. Die in Abschnitt 9.1 angeführten Zitate legen folgende Interpretation nahe: Bei einer Ordinalskala wird angenommen, daß es für die Merkmalswerte eines Merkmalsraums nur eine lineare Ordnung gibt; bei einer Intervallskala wird angenommen, daß es für den Merkmalsraum außerdem eine Metrik gibt, so daß von Abständen zwischen Merkmalswerten gesprochen werden kann. Wir sprechen im ersten Fall von *nicht-metrischen*, im zweiten Fall von *metrischen Größen* (oder *Skalen* oder *Merkmalsräumen*). Die Worte 'Ordinalskala' bzw. 'Intervallskala' werden wir nicht verwenden, da sie in ihrer üblichen Verwendung mit der irreführenden Vorstellung unterschiedlicher Skalenniveaus verbunden sind.

2. Hier muß man allerdings etwas aufpassen. Nicht nur hat die Frage, welches Skalenniveau *Daten* haben, keinen bestimmten Sinn. Auch die Frage, ob eine Größe metrisch ist oder nicht, macht in dieser Form keinen Sinn. Metriken sind keine Eigenschaften von Größenbegriffen, die man feststellen könnte. Man kann zum Beispiel nicht prüfen, ob es für Schulnoten eine Metrik gibt. Größenbegriffe sind, wie auch andere Begriffe, Bestandteile einer Sprache und sollen helfen, sinnvolle und verständliche Aussagen zu bilden. Bei der Erläuterung eines Größenbegriffs muß zunächst gezeigt werden, wie man mit seiner Hilfe Objekte (oder Individuen oder Situationen) „entsprechend ihrer Größe“ ordnen kann. Dann kann man in einem zweiten Schritt daran denken, die möglichen Werte einer Größe durch Zahlen zu repräsentieren. Dieser zweite Schritt führt zu einer Skala. Schließlich kann man ausgehend von einer Skala eine Vielzahl unterschiedlicher Metriken einführen. So gibt es für jeden Größenbegriff nicht nur eine bestimmte Skala und eine bestimmte Metrik, sondern eine Vielzahl unterschiedlicher Skalen und Metriken. Wichtig ist, daß sie explizit eingeführt werden und ihr Sinn erläutert wird.[8]

[8]Man kann den Gedankengang auch durch die Überlegung verfolgen, daß uns die Gegenstände, die wir in der Welt vorfinden, nicht von sich aus sagen, ob und ggf. wie gewisse ihrer Eigenschaften gemessen werden können. Insofern kann man A. Kaplan zustimmen, wenn er in seiner „Methodology for Behavioral Science“ 1964, S. 176, schreibt: „that whether we can measure something depends, not on that thing, but on how we have conceptualized it, on our knowledge of it, above all on the skill and ingenuity which we can bring to bear on the process of measurement which our inquiry can put to use [...].“ Allerdings geht es dann folgendermaßen weiter: „To say of something that it is uncapable of being measured is like saying of it that it is knowable only up to a point, that our ideas of it must inevitably remain indeterminate.“ Darüber kann sicherlich gestritten werden. Eine oft wiederholte, aber auch bestrittene Formulierung dieser Position stammt von dem Physiker Kelvin aus dem Jahr 1889: „I often say that when you can measure what you are speaking about, you know something about it; when you cannot measure it [...] your knowledge is of a meagre and unsatisfactory kind.“ (Hier zititiert nach Hacking 1983, S. 242.) Das Programm eines rein quantitativen (physikalischen) Weltbilds ist jedoch wesentlich älter und geht insbesondere auf Descartes zurück. Vgl. hierzu die kurzen, aber sehr informativen Ausführungen von Duhem 1908, Kap. 5 und 6; eine kritische historische Darstellung gibt auch Kline 1980. Die Feststellung qualitativer Unterschiede ist grundsätzlich nicht weniger legitim oder wissenswert als die Feststellung quantitativer Unterschiede. Gerade wenn man erkannt hat, daß Eigenschaften nicht an und für sich meßbar oder nicht meßbar sind, stellt sich nicht nur die Frage, wie Eigenschaften meßbar gemacht werden können, sondern primär wichtig wird die Frage, welche theoretischen und praktischen Interessen die Erfindung und Institutionalisierung neuer Meßtechniken sinnvoll machen könnten.

3. Zur Illustration des Gedankengangs kann die in der Methodenliteratur diskutierte Frage dienen, welches Skalenniveau für Pseudo-Daten angenommen werden kann, die aus Rating-Skalen gewonnen werden. Zum Beispiel sagen Bortz und Döring (1995, S. 163):

> „Während bei Rangordnungen und Paarvergleichen von den Untersuchungsteilnehmern ordinale Urteile abzugeben sind, können mittels Rating-Skalen auf unkomplizierte Weise direkt intervallskalierte Urteile erzeugt werden.“

Dagegen sagt Laatz (1993, S. 353f):

> „Die meisten mit Rating-Skalen gewonnenen Daten haben Ordinalskalenniveau. Wenn jemand sagt, er sei mit seiner Arbeit 'sehr zufrieden', kann man wohl unterstellen, daß er sich von jemandem unterscheidet, der behauptet er sei 'etwas zufrieden' und ebenso von einem, der angibt, 'etwas unzufrieden' zu sein. Auch ergibt sich eine eindeutige Ordnung, wenn man für diese drei Stufen die Werte 2; 3; 4 vergeben wollte, könnte man 2 für 'etwas unzufrieden', 4 für 'sehr zufrieden' vergeben (umgekehrt wäre es auch denkbar). Auf jeden Fall müßte dann 'etwas zufrieden' den Wert 3 bekommen. Sicher kann man aber keine gleichen Abstände der Werte annehmen. Die Werte 2; 3; und 4 haben auf der Zahlengeraden für reelle Zahlen gleiche Abstände. Aber man kann nicht gesichert behaupten, daß zwischen den drei Werten des empirischen Relativs ebenfalls gleiche Abstände existieren. Vielleicht sind sie ungleich und es wäre angemessener, die Werte 2; 3,5 und 6 zu vergeben.“

Die Spekulationen, die hier von Laatz vorgenommen werden, sind zwar merkwürdig, verdeutlichen aber, daß zunächst vollständig unklar ist, wie – insbesondere bei Rating-Skalen – über Abstände zwischen Merkmalswerten gesprochen werden kann. Vielmehr setzt ein Reden von Abständen voraus, daß man sich – explizit oder implizit – auf eine Metrik beziehen kann, durch die ein Abstandsbegriff entsteht. In der Alltagskommunikation mag das oft implizit bleiben; Sozialwissenschaftler sollten aber die von ihnen verwendeten Metriken explizit machen und begründen. Sicherlich kann man auch bei Rating-Skalen eine Metrik *definitorisch einführen*. Aber das kann man auf viele unterschiedliche Weisen machen. Deshalb ist auch die Aussage von Bortz und Döring irreführend. Denn ihr Reden von einer Intervallskala liefert tatsächlich überhaupt keinen Hinweis auf irgendeine bestimmte Metrik.

4. Die Schlußfolgerung ist insoweit sehr einfach: *Wenn* mit Rating-Skalen sinnvoll konzipierbare Größen erfaßt werden könnten, könnte man auch Metriken einführen und hätte dann metrische Größen. Alternativ könnte man auf die Definition einer Metrik verzichten und Daten als Werte nicht-metrischer Größen behandeln. Eine Diskussion kann sich bestenfalls darauf beziehen, ob eine Metrik explizit eingeführt werden sollte und wie dies begründet werden könnte. Das Hauptproblem ist in diesem Fall jedoch ein ganz anderes: daß Daten, die mit Rating-Skalen erzeugt werden, überhaupt nicht als Werte von Größen verstanden werden können. Denn das würde voraussetzen, daß zunächst ein Größenbegriff gebildet werden kann, auf den man sich bei der Ermittlung von Daten beziehen kann. Aber Rating-Verfahren sind dadurch definiert, daß jeder sagen kann, was er möchte, und daß es keinerlei objektivierbare Kriterien gibt, mit denen diese Beliebigkeit eingeschränkt werden könnte. Wollte man in diesem Fall der Rhetorik des Messens folgen, müßte man sich einen Menschen, der angesichts einer Rating-Skala eine Antwort produziert, als Subjekt eines Meßvorgangs vorstellen, dem jedoch gestattet ist, ein beliebiges Meßergebnis abzuliefern. Aber wie beim Auswürfeln von Schulnoten sollte man dann nicht von Messen sprechen.

5. Schließlich sollte man sich auch klarmachen, daß Metriken keineswegs nur dem Zweck dienen, metrische Größen zu definieren. Insbesondere in der Statistik dienen Metriken in erster Linie der Aufgabe, Varianten für Begriffsbildungen und Methoden explizit zu machen. Als Beispiel kann man an den Begriff eines Mittelwerts denken. Definiert man ihn durch

$$\mathrm{M}(X) := \frac{1}{n} \sum_{\omega \in \Omega} X(\omega)$$

ist offenbar eine gedankliche Bezugnahme auf irgendeine Metrik nicht erforderlich. Andererseits kann man jedoch auch die Frage stellen, wie sich Mittelwerte durch ihre Abstände zu den Werten einer Variablen charakterisieren lassen. Dann ist es erforderlich, explizit zu sagen, wie man in diesem Zusammenhang von Abständen sprechen möchte. Z.B. kann man die Metrik $d_2(x, x') := (x - x')^2$ verwenden und dann feststellen, daß der Mittelwert $\mathrm{M}(X)$ die Eigenschaft hat, die Summe der so definierten Abstände zu den Werten der Variablen X zu minimieren. Oder anders

formuliert: Wenn man nach einer Zahl x sucht, so daß die Funktion

$$f(x) := \frac{1}{n} \sum_{\omega \in \Omega} d_2(X(\omega), x) = \frac{1}{n} \sum_{\omega \in \Omega} (X(\omega) - x)^2$$

einen möglichst kleinen Wert annimmt, findet man als Lösung gerade $x = \mathrm{M}(X)$. Dieser Gedankengang liefert auch sogleich einen Hinweis, um über Alternativen nachzudenken; denn anstelle der Metrik d_2 könnte man auch andere verwenden. Z.B. könnte man eine Metrik $d_1(x, x') := |x - x'|$ verwenden. Gesucht ist dann eine Zahl x, die die absoluten Abstände minimiert, also ein Minimum der Funktion

$$f^*(x) := \frac{1}{n} \sum_{\omega \in \Omega} d_1(X(\omega), x) = \frac{1}{n} \sum_{\omega \in \Omega} |\, X(\omega) - x \,|$$

liefert. Dann gelangt man nicht zum Mittelwert, sondern zum Median der Variablen X. So wird deutlich, wie der Begriff der Metrik helfen kann, um über unterschiedliche Möglichkeiten zur Begriffsbildung nachzudenken.

Kapitel 10

Lineare und partielle Ordnungen

Um zu sinnvollen Größenbegriffen zu gelangen, kommt es vor allem darauf an, daß man für die Elemente eines Merkmalsraums eine lineare Ordnung begründen kann. Wie am Beispiel des Schichtindex von Scheuch und Daheim gezeigt worden ist, gelingt das nicht immer, sondern man gelangt nur zu partiellen Ordnungen. Damit beschäftigen wir uns in diesem Kapitel. Zuerst werden lineare und partielle Ordnungen unterschieden, dann wird besprochen, wie partielle Ordnungen dargestellt und analysiert werden können.

10.1 Ordnungsrelationen

1. Die Worte ‘Ordnung’ und ‘ordnen’ werden in vielen unterschiedlichen Bedeutungen verwendet. Oft werden bereits Klassifikationen und Typenbildungen als elementare Formen des Herstellens von Ordnung verstanden (J. v. Kempski 1952). Hier beziehen wir uns jedoch auf Ordnungen in einem spezielleren Sinn: Gemeint ist, daß die Elemente einer Menge in eine *lineare Ordnung* gebracht werden können. Das prototypische Beispiel sind Zahlen, die entsprechend ihrer Größe linear geordnet werden können. In diesem Fall kann man sich auf eine durch Konstruktionsverfahren gegebene Menge von Objekten (hier die Menge der reellen Zahlen $\mathbf{R}$) beziehen, und die Bedingungen für eine lineare Ordnung können folgendermaßen angegeben werden:

a) Die Ordnungsrelation ist *vollständig*, d.h. für je zwei Zahlen $x, y \in \mathbf{R}$ trifft genau eine der folgenden Aussagen zu:

$$x < y \quad \text{oder} \quad y < x \quad \text{oder} \quad y = x$$

b) Die Ordnungsrelation ist *transitiv*, d.h. für je drei Zahlen $x, y, z \in \mathbf{R}$ gilt: $x < y$ und $y < z \implies x < z$.

Bezieht man sich auf diese Bedingungen, spricht man auch von einer *strikten linearen Ordnung*. Man kann sie verwenden, um auch eine *schwache* lineare Ordnung zu definieren: $x \leq y \iff x < y$ oder $x = y$. Umgekehrt kann man ausgehend von einer schwachen linearen Ordnung eine strikte lineare Ordnung definieren.

2. Diese aus der Welt der Zahlen gewonnenen Vorstellungen über lineare Ordnungen können in analoger Weise für beliebige Mengen konzipiert werden. Ein für uns besonders wichtiges Beispiel sind Merkmalsräume statistischer Variablen. Wenn für einen Merkmalsraum eine lineare Ordnung

gegeben ist, können die Merkmalswerte ihrer Größe nach geordnet werden. Schon diese Formulierung zeigt den begrifflichen Zusammenhang zwischen linearen Ordnungen und dem Reden von Größen. Hier muß man allerdings unterscheiden: eine lineare Ordnung für den Merkmalsraum und eine lineare Ordnung für seine numerische Repräsentation. Da für Merkmalsräume stets eine numerische Repräsentation verwendet werden kann, gibt es auch stets eine lineare Ordnung für die zur Repräsentation verwendeten Zahlen, die durch die numerischen Standardrelationen definiert wird (vgl. Abschnitt 9.1). Dieser linearen Ordnung entspricht aber nicht unbedingt eine sinnvoll interpretierbare Ordnung für die Merkmalswerte. Das ist nur bei quantitativen Merkmalsräumen der Fall. Sie sind gerade dadurch definiert, daß eine sinnvolle Interpretation für die numerische Ordnungsrelation angegeben werden kann.

3. Zu überlegen ist, wie Ordnungsrelationen zustande kommen. In der Mathematik geht man meistens davon aus, daß man sich auf eine irgendwie *gegebene* Menge von Objekten beziehen kann. Dies erlaubt es, beliebige Relationen mit den begrifflichen Mitteln der Mengenlehre einzuführen. Ist nämlich Ω irgendeine gegebene Menge, kann man jede Teilmenge

$$R \subseteq \Omega \times \Omega$$

als eine Relation zwischen den Elementen von Ω auffassen. Man kann auch explizit ein Relationssymbol, zum Beispiel $\sim_R$, einführen, indem man folgende Definition vornimmt:

$$\omega \sim_R \omega' \iff (\omega, \omega') \in R$$

Die Beschaffenheit der Teilmenge R bestimmt dann, welche Eigenschaften die Relation hat. Für eine strikte Ordnungsrelation muß die Menge R folgenden Bedingungen genügen:

a) Für alle $\omega \in \Omega$: $(\omega, \omega) \notin R$

b) Für alle $\omega, \omega' \in \Omega$: $(\omega, \omega') \in R \implies (\omega', \omega) \notin R$

c) Für alle $\omega, \omega' \in \Omega$: $\omega \neq \omega' \implies (\omega, \omega') \in R$ oder $(\omega', \omega) \in R$

d) Für alle $\omega, \omega', \omega'' \in \Omega$: $(\omega, \omega'), (\omega', \omega'') \in R \implies (\omega, \omega'') \in R$

Eine schwache lineare Ordnung würde entstehen, wenn man auf die Bedingungen (a) und (b) verzichtet.

4. Eine weitere Überlegung bezieht sich auf Merkmalsräume für mehrdimensionale statistische Variablen der Form

$$(X_1, \ldots, X_m) : \Omega \longrightarrow \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$$

Selbst wenn man annehmen kann, daß es sich bei $\tilde{\mathcal{X}}_1, \ldots, \tilde{\mathcal{X}}_m$ um quantitative Merkmalsräume handelt, so daß es für jeden Teilmerkmalsraum $\tilde{\mathcal{X}}_j$

eine lineare Ordnung gibt, erhält man für den gemeinsamen Merkmalsraum

$$\tilde{\mathcal{X}} := \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$$

auf eindeutige Weise nur eine *partielle Ordnung*, die folgendermaßen definiert ist:[1]

$$\text{Für } (\tilde{x}_1,\ldots,\tilde{x}_m),(\tilde{x}'_1,\ldots,\tilde{x}'_m) \in \tilde{\mathcal{X}}: \qquad (10.1.1)$$
$$(\tilde{x}_1,\ldots,\tilde{x}_m) < (\tilde{x}'_1,\ldots,\tilde{x}'_m) \iff \tilde{x}_1 < \tilde{x}'_1,\ldots,\tilde{x}_m < \tilde{x}'_m$$

Die so für $\tilde{\mathcal{X}}$ definierte Ordnungsrelation ist offenbar transitiv, sie ist jedoch nicht vollständig. Oder anders gesagt: Es gibt im allgemeinen mindestens zwei Elemente in $\tilde{\mathcal{X}}$, von denen man nicht sagen kann, daß eines der beiden kleiner ist als das andere. In Abschnitt 10.2.1 wird besprochen, daß diese partielle Ordnung im allgemeinen auf unterschiedliche Weisen zu einer linearen Ordnung fortgesetzt werden kann.

5. Ist man an Fragen der Begriffsbildung interessiert, muß man sich unabhängig von jeweils gegebenen kontingenten Daten zunächst auf Merkmalsräume beziehen. Gibt es für einen Merkmalsraum eine lineare oder partielle Ordnung, kann man sie natürlich auch zur Datenanalyse verwenden. Sei nämlich

$$X : \Omega \longrightarrow \tilde{\mathcal{X}}$$

eine statistische Variable, für deren Merkmalsraum $\tilde{\mathcal{X}}$ eine lineare oder partielle Ordnung definiert ist. Dann hat man automatisch auch eine lineare oder partielle Ordnung für den realisierten Merkmalsraum $X(\Omega)$, d.h. für die Menge der bei den Mitgliedern von Ω beobachteten Merkmalswerte. Sie besteht einfach aus der Restriktion der zunächst für $\tilde{\mathcal{X}}$ definierten Ordnung auf den realisierten Merkmalsraum $X(\Omega) \subseteq \tilde{\mathcal{X}}$. Ebenfalls erhält man eine induzierte Ordnung für die Elemente von Ω, indem man folgende Definitionen verwendet:[2]

$$\omega \prec \omega' \iff X(\omega) < X(\omega')$$
$$\omega \preceq \omega' \iff X(\omega) \leq X(\omega')$$
$$\omega \equiv \omega' \iff X(\omega) = X(\omega')$$

Infolgedessen müssen Fragen der Begriffsbildung und Fragen der Datenanalyse unterschieden werden. Im ersten Fall geht es darum, ob und ggf. wie sich die Elemente eines Merkmalsraums ordnen lassen. Im zweiten Fall wird untersucht, ob es bei jeweils bestimmten Daten Ordnungen gibt und wie sie sich darstellen lassen.

[1] Dies ist eine strikte partielle Ordnung. Wenn man $\leq$ anstelle von $<$ verwendet, kann man analog eine schwache partielle Ordnung definieren.

[2] Wir verwenden die Symbole $\prec$, $\preceq$ und $\equiv$ anstelle von $<$, $\leq$ und $=$, wenn sich die Ordnungsrelation auf eine Menge bezieht, deren Elemente keine Zahlen sind. Da wir annehmen, daß für Merkmalsräume stets eine numerische Repräsentation gegeben ist, können für sie die numerischen Standardrelationen $<$, $\leq$ und $=$ verwendet werden.

10.2 Partielle Ordnungen

Wenn es für eine Menge nur eine partielle Ordnung gibt, bedeutet dies, daß mindestens zwei ihrer Elemente mit Hilfe der gegebenen Ordnungsrelation nicht sinnvoll vergleichbar sind. Insbesondere kann man dann die Elemente der Menge nicht (ohne weiteres) in eine lineare Ordnung bringen. Somit entsteht die Frage, wie man partielle Ordnungen darstellen und analysieren kann.

10.2.1 Vervollständigungen

1. Für die folgenden Überlegungen beziehen wir uns auf einen mehrdimensionalen Merkmalsraum $\tilde{\mathcal{X}} := \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$ und nehmen an, daß es für jeden Teilmerkmalsraum $\tilde{\mathcal{X}}_j$ eine lineare Ordnung gibt. Daraus entsteht für den gemeinsamen Merkmalsraum $\tilde{\mathcal{X}}$ zwar nur eine partielle Ordnung, man kann jedoch die Frage stellen, ob man diese partielle Ordnung zu einer linearen Ordnung fortsetzen kann. Eine minimale Anforderung besteht darin, daß die lineare Ordnung mit der zunächst gegebenen partiellen Ordnung verträglich sein soll. Wir verwenden folgende Definition: Eine lineare Ordnung $<_*$ heißt eine *konsistente Fortsetzung* der partiellen Ordnung $<$, wenn gilt:

$$\text{Für alle } \tilde{x},\tilde{x}' \in \tilde{\mathcal{X}}: \quad \tilde{x} < \tilde{x}' \implies \tilde{x} <_* \tilde{x}'$$

Analog nennen wir eine schwache lineare Ordnung $\leq_*$ eine *konsistente Fortsetzung* der partiellen Ordnung $\leq$, wenn gilt:

$$\text{Für alle } \tilde{x},\tilde{x}' \in \tilde{\mathcal{X}}: \quad \tilde{x} \leq \tilde{x}' \implies \tilde{x} \leq_* \tilde{x}'$$

2. Meistens gibt es mehrere Möglichkeiten, um aus einer partiellen Ordnung eine konsistente Fortsetzung zu bilden. Als Beispiel betrachten wir

$$\tilde{\mathcal{X}} := \tilde{\mathcal{X}}_1 \times \tilde{\mathcal{X}}_2 \quad \text{wobei} \quad \tilde{\mathcal{X}}_1 := \{1,2,3\} \text{ und } \tilde{\mathcal{X}}_2 := \{1,2,3\}$$

Mit Hilfe der partiellen Ordnung kann man z.B. sagen: $(1,2) < (2,3)$, aber die Elemente $(1,2)$ und $(2,1)$ können nicht miteinander verglichen werden. Um sich zu überlegen, wie man eine konsistente Fortsetzung finden kann, ist es hilfreich, die Elemente von $\tilde{\mathcal{X}}$ in einem zweidimensionalen Schema anzuordnen:

	1	2	3
1	(1, 1)	(1, 2)	(1, 3)
2	(2, 1)	(2, 2)	(2, 3)
3	(3, 1)	(3, 2)	(3, 3)

Anstelle der Elemente von $\tilde{\mathcal{X}}$ kann man in dieses Schema auch auf beliebige

Weisen die Zahlen von 1 bis 9 eintragen und dadurch beliebige lineare Ordnungen für $\tilde{\mathcal{X}}$ erzeugen. Aber nicht jede dieser linearen Ordnungen ist eine konsistente Fortsetzung der zunächst gegebenen partiellen Ordnung. Die folgenden drei Beispiele liefern jedoch konsistente Fortsetzungen:

	1	2	3
1	1	2	3
2	4	5	6
3	7	8	9

	1	2	3
1	1	4	7
2	2	5	8
3	3	6	9

	1	2	3
1	1	2	6
2	3	5	7
3	4	8	9

Offenbar erhält man stets dann eine konsistente Fortsetzung der partiellen Ordnung, wenn man die Ordnungszahlen so wählt, daß sie in jeder Zeile und Spalte größer (nicht kleiner) werden.

3. Obwohl man leicht konsistente Fortsetzungen einer partiellen Ordnung finden kann, entsteht ein Problem: Wie kann man eine Ordnungsbeziehung zwischen zwei Elementen von $\tilde{\mathcal{X}}$ begründen, wenn man sich dafür nicht auf die partielle Ordnung, die durch die Teilmerkmalsräume von $\tilde{\mathcal{X}}$ geliefert wird, berufen kann? Betrachten wir das erste der drei Beispiele. In diesem Fall impliziert die lineare Ordnung $<_*$ z.B. die Beziehungen

$$(1,2) <_* (2,1) \quad \text{und} \quad (1,3) <_* (3,1)$$

Aber würde man stattdessen die im zweiten Beispiel angegebene lineare Ordnung verwenden, würden sich beide Beziehungen umkehren. Die Vorgehensweise erscheint somit sehr willkürlich.

10.2.2 Hierarchische Partitionen

1. Aus den Überlegungen des vorangegangenen Abschnitts folgt, daß man nicht ohne weiteres sinnvolle Fortsetzungen oder Vervollständigungen partieller Ordnungen finden kann. Aber inwiefern entsteht daraus ein Problem? Zunächst kann man sicherlich feststellen, daß für statistische Analysen kein Problem entsteht. Bezieht man sich in der Statistik auf eine mehrdimensionale Variable

$$X := (X_1, \ldots, X_m) : \Omega \longrightarrow \tilde{\mathcal{X}} := \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$$

ist man daran interessiert, Einsichten in die gemeinsame Häufigkeitsverteilung von X zu gewinnen. Der Begriff einer Häufigkeitsverteilung setzt jedoch nicht voraus, daß für den gemeinsamen Merkmalsraum $\tilde{\mathcal{X}}$ eine lineare oder irgendeine andere Ordnung definiert ist. Das ist auch dann nicht erforderlich, wenn man, z.B. mit den Methoden der Regressionsrechnung, untersuchen möchte, welche Beziehungen es zwischen den Komponenten von X gibt. Da sich das Erkenntnisinteresse auf gemeinsame und bedingte

Verteilungen richtet, kann man sich auch problemlos auf die jeweils gegebenen numerischen Repräsentationen für die Merkmalsräume beziehen.

2. Probleme entstehen, wenn man für einen mehrdimensionalen Merkmalsraum eine quantitativ interpretierbare Skala konstruieren möchte. Denn diese Absicht impliziert den Anspruch, daß man für den gemeinsamen Merkmalsraum eine lineare Ordnung begründen kann. Wie bereits besprochen wurde, besteht eine oft verwendete Methode darin, einen additiven Index zu bilden. Man verschafft sich dann eine numerische Repräsentation für den gemeinsamen Merkmalsraum $\tilde{\mathcal{X}}$ durch eine Funktion

$$g : \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m \longrightarrow \mathbf{R}$$

die durch $g(\tilde{x}_1, \ldots, \tilde{x}_m) := \tilde{x}_1 + \cdots + \tilde{x}_m$ definiert ist. Somit erhält man auch eine lineare Ordnung für den gemeinsamen Merkmalsraum:

$$(\tilde{x}_1, \ldots, \tilde{x}_m) < (\tilde{x}'_1, \ldots, \tilde{x}'_m) \iff g(\tilde{x}_1, \ldots, \tilde{x}_m) < g(\tilde{x}'_1, \ldots, \tilde{x}'_m)$$

Als Beispiel kann man an den Schichtindex von Scheuch und Daheim denken, der in Abschnitt 6.3 besprochen wurde. Selbst wenn man zugestehen würde, daß sich für die drei beteiligten Merkmalsräume (Einkommen, Bildung und beruflicher Status) sinnvolle lineare Ordnungen definieren lassen, erhielte man für den gemeinsamen Merkmalsraum nur eine partielle Ordnung, also keine hinreichende Grundlage für die Definition eines neuen Größenbegriffs. Andererseits impliziert jedoch der von Scheuch und Daheim verwendete additive Index zugleich eine lineare Ordnung für den gemeinsamen Merkmalsraum und suggeriert einen Größenbegriff.

3. Auf die Probleme, die mit der Verwendung additiver Indizes verbunden sind, soll an dieser Stelle nicht erneut eingegangen werden. Bemerkenswert ist jedoch, daß es sich um eine spezielle Variante einer Methode handelt, die allgemein darin besteht, für einen Merkmalsraum (oder eine beliebige Menge), für den nur eine partielle Ordnung verfügbar ist, eine hierarchische Partition zu bilden, so daß die Elemente der Partition linear geordnet sind. Um die Methode zu erläutern, betrachten wir einen Merkmalsraum $\tilde{\mathcal{X}} := \{\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4, \tilde{x}_5, \tilde{x}_6, \tilde{x}_7, \tilde{x}_8\}$, der acht Elemente enthält und für den es eine partielle Ordnung gibt, die durch folgende Beziehungen definiert ist:

$$\tilde{x}_1 < \tilde{x}_2,\ \tilde{x}_1 < \tilde{x}_3,\ \tilde{x}_2 < \tilde{x}_4,\ \tilde{x}_2 < \tilde{x}_5,\ \tilde{x}_3 < \tilde{x}_6,$$
$$\tilde{x}_5 < \tilde{x}_8,\ \tilde{x}_6 < \tilde{x}_8,\ \tilde{x}_7 < \tilde{x}_3$$

Wir beginnen also mit einem Beispiel einer partiellen Ordnung, die nicht durch lineare Ordnungen auf Teilräumen definiert ist. Eine solche partielle Ordnung kann man anschaulich in Form eines gerichteten Graphen darstellen, der keine Zyklen aufweist.[3] Für unser Beispiel sieht dieser Graph

[3]Hier und an einigen späteren Stellen ist es nützlich, einige Begriffe aus der Theorie der

folgendermaßen aus:[4]

$$\begin{tikzcd}
 & & \tilde{x}_4 & \\
\tilde{x}_1 \arrow[r] \arrow[dr] & \tilde{x}_2 \arrow[r] \arrow[ur] & \tilde{x}_5 \arrow[r] & \tilde{x}_8 \\
\tilde{x}_7 \arrow[r] & \tilde{x}_3 \arrow[r] & \tilde{x}_6 \arrow[ur] &
\end{tikzcd}$$

In diesem Bild wird durch $\tilde{x} \rightarrow \tilde{x}'$ angedeutet, daß $\tilde{x}$ ein *unmittelbarer Vorgänger* von $\tilde{x}'$ bzw. $\tilde{x}'$ ein *unmittelbarer Nachfolger* von $\tilde{x}$ ist, d.h. es gilt: $\tilde{x} < \tilde{x}'$, und es gibt kein $\tilde{x}''$, so daß $\tilde{x} < \tilde{x}''$ und $\tilde{x}'' < \tilde{x}'$ ist. Somit kann man auch für jedes Element $\tilde{x} \in \tilde{\mathcal{X}}$ die Menge seiner unmittelbaren Nachfolger bzw. die Menge seiner unmittelbaren Vorgänger definieren:

$$N(\tilde{x}) := \{\tilde{x}' \in \tilde{\mathcal{X}} \,|\, \tilde{x} \rightarrow \tilde{x}'\}, \quad V(\tilde{x}) := \{\tilde{x}' \in \tilde{\mathcal{X}} \,|\, \tilde{x}' \rightarrow \tilde{x}\}$$

4. Diese Begriffsbildungen können verwendet werden, um eine Hierarchie zu definieren: Eine Partition[5] $\tilde{\mathcal{X}} = P_1 \cup \cdots \cup P_q$ wird eine *hierarchische Partition* des Merkmalsraums $\tilde{\mathcal{X}}$ genannt, wenn für alle $\tilde{x} \in \tilde{\mathcal{X}}$ folgende Bedingungen erfüllt sind:

a) $\tilde{x} \in P_l \Longrightarrow N(\tilde{x}) \subseteq P_{l+1} \quad (l = 1, \ldots, q-1)$

b) $\tilde{x} \in P_l \Longrightarrow V(\tilde{x}) \subseteq P_{l-1} \quad (l = 2, \ldots, q)$

Die Zahlen $1, \ldots, q$ heißen die *Level* der Hierarchie. In unserem Beispiel findet man folgende hierarchische Partition:

$$P_1 = \{\tilde{x}_1, \tilde{x}_7\}, \; P_2 = \{\tilde{x}_2, \tilde{x}_3\}, \; P_3 = \{\tilde{x}_4, \tilde{x}_5, \tilde{x}_6\}, \; P_4 = \{\tilde{x}_8\}$$

5. Hat man eine hierarchische Partition gebildet, kann man sie zur Definition einer schwachen linearen Ordnung für $\tilde{\mathcal{X}}$ verwenden. Ist nämlich

Graphen zu verwenden. Es genügt jedoch ein informelles Verständnis: Ein Graph besteht aus einer Menge von *Knoten*, wobei jeweils zwei dieser Knoten durch eine Linie oder einen Pfeil verbunden sein können. Sind alle Knoten paarweise verbunden, spricht man von einem *vollständigen*, andernfalls von einem *unvollständigen Graphen*. Die Linien bzw. Pfeile, durch die die Knoten verbunden werden können, werden als *Kanten* des Graphen bezeichnet. Man unterscheidet *ungerichtete Kanten* (Linien) und *gerichtete Kanten* (Pfeile) und dementsprechend ungerichtete und gerichtete Graphen. Somit kann man entlang der Linien ungerichtete und entlang der Pfeile gerichtete *Wege* verfolgen (wobei angenommen wird, daß keine Linie bzw. kein Pfeil mehrfach durchlaufen wird). Ein *Zyklus* ist ein gerichteter Weg, der zu seinem Anfangsknoten zurückführt.

[4] Ein solcher Graph wird in der Mathematik ein *Hasse-Diagramm* genannt. Daß der Graph eine partielle Ordnung repräsentiert, sieht man daran, daß es mindestens zwei Knoten gibt, die nicht durch einen gerichteten Weg verbunden sind.

[5] Allgemein versteht man unter einer Partition einer beliebigen Menge M eine Zerlegung von M in paarweise disjunkte Teilmengen, deren Vereinigungsmenge gleich M ist.

$\tilde{x} \in P_j$, kann man sagen:

a) $\tilde{x} <_* \tilde{x}'$ für alle $\tilde{x}' \in P_l$, wenn $l > j$ ist;

b) $\tilde{x}' <_* \tilde{x}$ für alle $\tilde{x}' \in P_l$, wenn $l < j$ ist;

c) $\tilde{x} \leq_* \tilde{x}'$ und $\tilde{x}' \leq_* \tilde{x}$ für alle $\tilde{x}' \in P_j$.

Die so definierte lineare Ordnung ist zugleich eine konsistente Fortsetzung der partiellen Ordnung, die zur Definition der Hierarchie verwendet worden ist. Es ist jedoch bemerkenswert, daß nicht für jede partielle Ordnung eine hierarchische Partition gebildet werden kann. Gäbe es in unserem Beispiel zusätzlich die Beziehung $\tilde{x}_6 \rightarrow \tilde{x}_5$, wäre es nicht mehr möglich, eine hierarchische Partition zu bilden; denn $\tilde{x}_5$ müßte dann einerseits der gleichen Äquivalenzklasse angehören wie $\tilde{x}_6$, andererseits müßten jedoch $\tilde{x}_5$ und $\tilde{x}_6$ unterschiedlichen Äquivalenzklassen angehören.[6]

6. Wenn jedoch $\tilde{\mathcal{X}}$ ein mehrdimensionaler Merkmalsraum ist, für dessen Komponenten es lineare Ordnungen gibt, kann stets eine hierarchische Partition gebildet werden. Um das deutlich zu machen, nehmen wir an, daß es m Komponenten $\tilde{\mathcal{X}}_j := \{1, 2, 3, 4, \ldots\}$ gibt. $\tilde{\mathcal{X}}$ besteht somit aus allen m-Tupeln der Form $(\tilde{x}_1, \ldots, \tilde{x}_m)$, und man kann eine hierarchische Partition durch die Äquivalenzklassen

$$P_l := \{(\tilde{x}_1, \ldots, \tilde{x}_m) \,|\, \tilde{x}_1 + \cdots + \tilde{x}_m = m + l - 1\}$$

definieren. Abbildung 10.2.1 illustriert dies für $m = 2$. Man erkennt, daß die durch eine solche hierarchische Partition induzierte lineare Ordnung für $\tilde{\mathcal{X}}$ mit derjenigen identisch ist, die man durch einfache (ungewichtete) additive Indizes definieren kann.[7]

7. Auch wenn eine partielle Ordnung zu einer schwachen linearen Ordnung erweitert werden kann, erhält man nicht unbedingt einen sinnvollen Größenbegriff. Zur Verdeutlichung der Überlegung beziehen wir uns auf zwei verschiedene Elemente $\tilde{x}, \tilde{x}' \in \tilde{\mathcal{X}}$, die mit einer für $\tilde{\mathcal{X}}$ definierten partiellen Ordnung nicht verglichen werden können, so daß also weder $\tilde{x} < \tilde{x}'$ noch $\tilde{x}' < \tilde{x}$ ist; z.B. auf zwei nicht-vergleichbare Merkmalskombinationen im Merkmalsraum des Schichtindex von Scheuch und Daheim. Konstruiert man eine hierarchische Partition, erhält man zwar eine Fortsetzung der partiellen Ordnung, die wir wie bisher mit $\leq_*$ bezeichnen, so daß man sagen kann: $\tilde{x} \leq_* \tilde{x}'$ oder $\tilde{x}' \leq_* \tilde{x}$ oder beides. Aber mit welcher Begründung? Eine Begründung müßte sich mit inhaltlichen Überlegungen

[6] Es sei auch darauf hingewiesen, daß hierarchische Partitionen nicht unbedingt eindeutig sind. Dies ist z.B. dann der Fall, wenn das Hasse-Diagramm aus zwei oder mehr Komponenten besteht, die untereinander nicht verbunden sind.

[7] Man beachte, daß gewichtete additive Indizes nicht unbedingt zu konsistenten Fortsetzungen partieller Ordnungen führen.

$$\begin{tikzcd}[column sep=small, row sep=small]
 & & (1,1) \arrow[dl] \arrow[dr] & & & & P_1 \\
 & (2,1) \arrow[dl] \arrow[dr] & & (1,2) \arrow[dl] \arrow[dr] & & & P_2 \\
(3,1) \arrow[dr] & & (2,2) \arrow[dl] \arrow[dr] & & (1,3) \arrow[dl] & & P_3 \\
 & (3,2) \arrow[dl] \arrow[dr] & & (2,3) \arrow[dl] \arrow[dr] & & & P_4 \\
(4,2) & & (3,3) & & (2,4) & & P_5
\end{tikzcd}$$

Abb. 10.2.1 Beispiel für ein Hasse-Diagramm.

auf einen Größenbegriff beziehen, der durch eine lineare Ordnung für $\tilde{\mathcal{X}}$ repräsentiert werden kann. Wie in Kapitel 6 besprochen worden ist, stehen solche Überlegungen jedoch sehr oft nicht zur Verfügung.

10.2.3 Datenbezogene Überlegungen

1. Bisher bezog sich die Fragestellung darauf, wie man sich ein Bild der partiellen Ordnung machen kann, die für einen mehrdimensionalen Merkmalsraum definiert ist. Bei dieser Fragestellung ist es erforderlich, sich auf alle Merkmalskombinationen zu beziehen, die innerhalb des Merkmalsraums vorkommen können. Wenn man sich dagegen auf bestimmte Daten, d.h. auf eine mehrdimensionale Variable

$$X := (X_1, \dots, X_m) : \Omega \longrightarrow \tilde{\mathcal{X}} := \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$$

bezieht, kann es für Zwecke der Datenanalyse sinnvoll sein, nur den realisierten Merkmalsraum $X(\Omega) \subseteq \tilde{\mathcal{X}}$ zu betrachten; denn die Anzahl der realisierten Merkmalskombinationen ist oft wesentlich kleiner als die Anzahl der Elemente in $\tilde{\mathcal{X}}$.

2. Zur Illustration verwenden wir die Klausurdaten aus Abschnitt 6.1. Da es 5 Klausuraufgaben gibt, ist X in diesem Beispiel eine 5-dimensionale Variable. Für jeden Teilnehmer ω_i kann man die Ergebnisse in Form eines Datenvektors $X(\omega_i) = \mathbf{x}_i = (x_{i1}, \dots, x_{i5})$ darstellen, wobei x_{ij} die Anzahl der in der j-ten Aufgabe erzielten Punkte angibt. Bezieht man sich auf die Numerierung der Klausurteilnehmer in Tabelle 6.1.1 (Abschnitt 6.1), können z.B. die Ergebnisse des ersten und sechsten Teilnehmers verglichen werden: $\mathbf{x}_1 = (2, 10, 0, 8, 15) \leq \mathbf{x}_6 = (8, 10, 0, 9, 15)$. Aber man findet natürlich auch Beispiele, bei denen ein Vergleich nicht möglich ist. Z.B. können bereits die ersten beiden Klausurteilnehmer nicht verglichen werden. Man erhält also nur eine partielle Ordnung, und es entsteht die Frage, wie man diese partielle Ordnung darstellen kann. Der realisierte

Tabelle 10.2.1 Ergebnismuster in den Klausurdaten.

K	X_1	X_2	X_3	X_4	X_5	X^*	H
1	8	5	0	4	1	18	1
2	4	10	0	5	0	19	1
3	10	5	5	0	2	22	1
4	8	10	5	6	0	29	1
5	10	0	5	1	15	31	1
6	8	7	0	10	7	32	1
7	10	10	0	0	13	33	1
8	10	10	0	0	14	34	1
9	2	10	0	8	15	35	1
10	2	10	5	3	15	35	1
11	10	10	5	10	0	35	1
12	10	10	0	4	12	36	2
13	10	10	3	6	7	36	1
14	10	10	0	10	7	37	1
15	10	10	2	6	10	38	2
16	10	10	5	4	9	38	1
17	10	10	3	6	10	39	1
18	10	5	5	5	15	40	2
19	10	10	0	6	14	40	1
20	10	10	1	10	9	40	2
21	10	10	5	4	12	41	1
22	8	10	0	9	15	42	1
23	10	10	0	8	15	43	1
24	10	10	5	9	9	43	2
25	10	10	0	9	15	44	1
26	10	10	0	10	15	45	2
27	10	10	5	8	13	46	1
28	10	10	5	8	15	48	3
29	10	10	5	10	13	48	1
30	10	10	5	10	14	49	1
31	10	10	5	10	15	50	8

Merkmalsraum $X(\Omega)$ umfaßt 31 verschiedene Ergebnismuster, die in Tabelle 10.2.1 dargestellt sind. Die erste Spalte (K) liefert eine Numerierung, die letzte Spalte (H) gibt die Häufigkeiten an, mit der die Ergebnismuster auftreten, und die Spalte X^* enthält die Summe der jeweils erzielten Punktzahlen. Abbildung 10.2.2 zeigt ein Hasse-Diagramm für die partielle Ordnung, die es zwischen den Ergebnismustern gibt.

3. Das Bild wurde so gezeichnet, daß sich bereits eine gewisse „hierarchische Struktur“ andeutet.[8] Allerdings kann man in diesem Beispiel keine

[8] Für die Zeichnung wurde das Programm `dot` aus dem von E. Koutsofios und S. C. North entwickelten Programmpaket `graphviz` verwendet. Das Programm ist über die

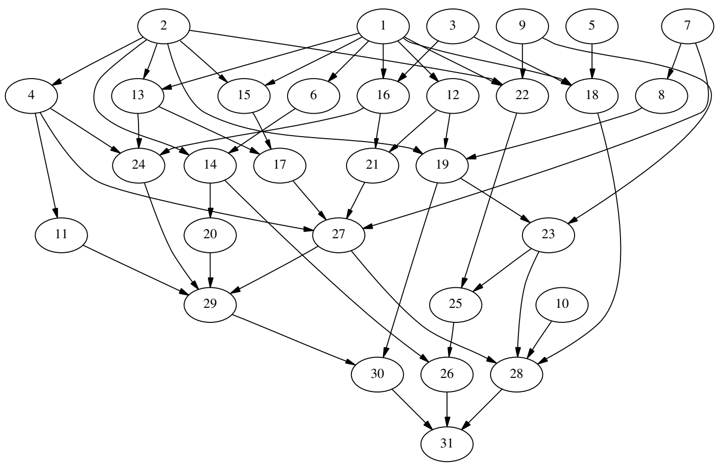


Abb. 10.2.2 Hasse-Diagramm für die partielle Ordnung in den Ergebnismustern der Kausurdaten aus Tabelle 10.2.1.

Hierarchie konstruieren, die der in Abschnitt 10.2.2 angegebenen Definition genügt. Man kann sich aber fragen, wie man zu einer hierarchie-ähnlichen Darstellung gelangen kann. Eine einfache Möglichkeit besteht darin, sich an den Weglängen im Hasse-Diagramm zu orientieren. Um das Verfahren allgemein zu beschreiben, sei $\mathcal{N}$ die Menge der Knoten des Hasse-Diagramms, und für jeden Knoten $k \in \mathcal{N}$ sei $V(k)$ die Menge seiner unmittelbaren Vorgänger und $N(k)$ die Menge seiner unmittelbaren Nachfolger (vgl. Abschnitt 10.2.2). Gesucht ist eine Partition von $\mathcal{N}$ in Teilmengen P_l $(l = 0, \ldots, q)$, die sich als Level einer Hierarchie interpretieren lassen. Dafür kann folgendes Verfahren verwendet werden:

a) P_0 mit dem niedrigsten Level besteht aus allen Knoten, die keine Vorgänger haben, also $P_0 := \{k \in \mathcal{N} \mid V(k) = \emptyset\}$, in unserem Beispiel: $P_0 = \{1, 2, 3, 5, 7, 9, 10\}$.[9]

b) Für jeden Knoten $k \in \mathcal{N}$ wird der längste Weg ermittelt, durch den er von einem Knoten aus P_0 erreichbar ist. Dabei ist die Länge eines Weges als die Anzahl der Pfeile definiert, die man durchlaufen muß. Die Länge dieser längsten Wege bezeichnen wir mit $L(k)$. In unserem Beispiel findet man etwa $L(4) = 1$ und $L(25) = 4$.

c) Schließlich werden die Teilmengen P_l durch $P_l := \{k \in \mathcal{N} \mid L(k) = l\}$ definiert.

Internetadresse `www.research.att.com/~north/graphviz` erhältlich.

[9]Knoten, die keine Vorgänger haben, werden als *minimale Elemente* eines Hasse-Diagramms bezeichnet; analog spricht man von *maximalen Elementen*, wenn es keine Nachfolger gibt.

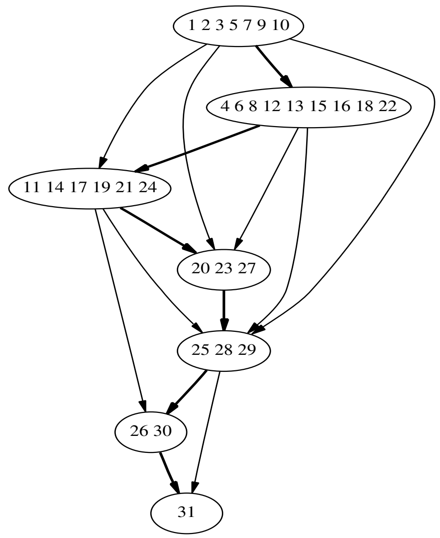


Abb. 10.2.3 Reduziertes Hasse-Diagramm.

Die Idee besteht also einfach darin, von den minimalen Elementen des Hasse-Diagramms auszugehen und dann den Level jedes Knotens durch die Länge des längsten Weges zu bestimmen, der ihn von einem minimalen Element trennt. Dann kann man ein reduziertes Hasse-Diagramm zeichnen, dessen Knoten aus den Teilmengen P_l bestehen. Zwei Teilmengen P_l und $P_{l'}$ werden durch einen Pfeil verbunden, wenn mindestens ein Element in P_l einen unmittelbaren Nachfolger in $P_{l'}$ hat. Abbildung 10.2.3 zeigt das auf diese Weise reduzierte Hasse-Diagramm für die Klausurdaten.

10.2.4 Intervallwertige Positionen

1. Wenn man für eine Menge, für die eine partielle Ordnung definiert ist, eine hierarchische Partition finden kann, bedeutet dies, daß man die partielle Ordnung durch eine schwache lineare Ordnung repräsentieren kann. Das ist jedoch nur ausnahmsweise möglich. Im allgemeinen findet man keine hierarchische Partition, die der in Abschnitt 10.2.2 gegebenen Definition entspricht. Man kann dann zwar quasi-hierarchische Darstellungen erzeugen, z.B. mit der im vorangegangenen Abschnitt beschriebenen Methode, um zu einer übersichtlichen Darstellung des Hasse-Diagramms zu gelangen; man erhält aber keine lineare Ordnung.

2. Eine hierarchische Partition kann immer dann nicht gebildet werden, wenn sich mindestens ein Element nicht eindeutig einem bestimmten Level der Hierarchie zuordnen läßt. Wenn man also daran festhält, Hierarchien

Tabelle 10.2.2 Intervallwertige Positionen für die Ergebnismuster in den Klausurdaten.

Muster	Intervall	Muster	Intervall
$\tilde{x}_1$	[1,4]	$\tilde{x}_{16}$	[2,3]
$\tilde{x}_2$	[1,4]	$\tilde{x}_{17}$	[3,4]
$\tilde{x}_3$	[1,4]	$\tilde{x}_{18}$	[2,5]
$\tilde{x}_4$	[2,4]	$\tilde{x}_{19}$	[2,5]
$\tilde{x}_5$	[1,4]	$\tilde{x}_{20}$	[3,4]
$\tilde{x}_6$	[2,4]	$\tilde{x}_{21}$	[3,4]
$\tilde{x}_7$	[1,4]	$\tilde{x}_{22}$	[2,4]
$\tilde{x}_8$	[2,4]	$\tilde{x}_{23}$	[2,5]
$\tilde{x}_9$	[1,4]	$\tilde{x}_{24}$	[3,4]
$\tilde{x}_{10}$	[1,5]	$\tilde{x}_{25}$	[3,5]
$\tilde{x}_{11}$	[3,4]	$\tilde{x}_{26}$	[3,6]
$\tilde{x}_{12}$	[2,4]	$\tilde{x}_{27}$	[2,5]
$\tilde{x}_{13}$	[2,3]	$\tilde{x}_{28}$	[2,6]
$\tilde{x}_{14}$	[2,5]	$\tilde{x}_{29}$	[3,5]
$\tilde{x}_{15}$	[2,3]	$\tilde{x}_{30}$	[3,6]

durch hierarchische Partitionen zu definieren, wird man im allgemeinen keinen Erfolg haben. Hier kann jedoch eine alternative Überlegung einsetzen: Man kann versuchen, jedem Element einen Bereich möglicher Level zuzuordnen, die ihrerseits durch eine unterstellte lineare Ordnung definierbar sind. Im folgenden beschreiben wir ein Verfahren, mit dem dies erreicht werden kann.

3. Ausgangspunkt ist eine Menge $\tilde{\mathcal{X}}$, für die eine partielle Ordnung definiert ist. Der erste Schritt besteht darin, diese Menge durch ein minimales Element $\tilde{x}^u$ und ein maximales Element $\tilde{x}^o$ zu ergänzen, so daß für alle $\tilde{x} \in \tilde{\mathcal{X}}$ gilt: $\tilde{x}^u \leq \tilde{x}$ und $\tilde{x} \leq \tilde{x}^o$. Möglicherweise sind solche Elemente in der Menge $\tilde{\mathcal{X}}$ bereits enthalten; andernfalls werden sie zu $\tilde{\mathcal{X}}$ hinzugefügt. Verwendet man für $\tilde{\mathcal{X}}$ z.B. die in Tabelle 10.2.1 angegebenen Ergebnismuster der Klausurdaten, enthält diese Menge bereits das maximale Element $\tilde{x}^o = \tilde{x}_{31} = (10, 10, 5, 10, 15)$. Es gibt jedoch kein minimales Element, so daß man ein solches Element explizit bilden und zu $\tilde{\mathcal{X}}$ hinzufügen muß; z.B. das Element $\tilde{x}^u = (2, 0, 0, 0, 0)$, das in jeder Komponente den kleinsten realisierten Merkmalswert enthält.

4. Die Elemente $\tilde{x}^u$ und $\tilde{x}^o$ können nun verwendet werden, um den niedrigsten und höchsten Level der Hierarchie zu definieren. Die Überlegung besteht einfach darin: Es gibt in $\tilde{\mathcal{X}}$ kein kleineres Element als $\tilde{x}^u$ und kein größeres Element als $\tilde{x}^o$. Weiterhin kann man auch die Anzahl der Level fixieren, indem man den längsten Weg berechnet, der innerhalb des Hasse-Diagramms von $\tilde{x}^u$ zu $\tilde{x}^o$ führt. Bei den Klausurdaten kann man maximal

Tabelle 10.2.3 Arztdichte (A), Bettendichte (B) und intervallwertige Positionen (P). (OECD-Länder ca. 1982. Daten für A und B aus Alber 1988, S. 122.)

Land	A	B	P	Land	A	B	P
Italien	32	77	[3,5]	Norwegen	21	65	[2,3]
Spanien	28	54	[2,4]	Niederlande	21	120	[2,4]
Belgien	27	95	[3,5]	Australien	20	110	[3,3]
Griechenland	25	62	[2,4]	Canada	20	69	[2,3]
Schweden	24	140	[2,5]	USA	19	59	[1,3]
Deutschland	24	111	[3,4]	Österreich	17	108	[2,2]
Dänemark	24	74	[2,4]	Neuseeland	17	99	[1,1]
Portugal	22	51	[1,3]	Schweiz	16	115	[1,3]
Finnland	22	156	[2,5]	Japan	14	121	[1,4]
Frankreich	22	116	[2,4]	Grossbritannien	30	81	[3,5]

7 Pfeile durchlaufen, um von $\tilde{x}^u$ zu $\tilde{x}^o$ zu gelangen; und somit kann man für $\tilde{x}^u$ den Level 0 und für $\tilde{x}^o$ den Level 7 festsetzen.

5. Es bleibt also nur die Frage, wie man für die Elemente $\tilde{x} \in \tilde{\mathcal{X}}$, die weder mit $\tilde{x}^u$ noch mit $\tilde{x}^o$ identisch sind, passende Level definieren kann. Wie bereits bemerkt worden ist, wird man diesen Elementen im allgemeinen keinen eindeutigen Level zuordnen können. Man kann aber versuchen, ihnen jeweils einen Bereich (ein Intervall) möglicher Level zuzuordnen. Man betrachte z.B. das Element mit der Nr. 10 in Abbildung 10.2.2. Einerseits hat dieses Element innerhalb von $\tilde{\mathcal{X}}$ (ohne $\tilde{x}^u$) keinen Vorgänger, was dafür sprechen würde, diesem Element einen sehr niedrigen Level zuzuordnen. Bildlich gesprochen: Ein Klausurteilnehmer mit dem Ergebnis $\tilde{x}_{10} = (2, 10, 5, 3, 15)$ kann auf keinen anderen Teilnehmer hinweisen, der schlechter abgeschnitten hätte. Andererseits wird dieses Ergebnismuster jedoch nur durch zwei Schritte vom besten Ergebnis $\tilde{x}^o = \tilde{x}_{31}$ getrennt, so daß man auch sagen könnte, daß das Ergebnis $\tilde{x}_{10}$ fast so gut ist wie das beste und deshalb einen relativ hohen Level in der Hierarchie bekommen sollte. Es bietet sich also folgende Vorgehensweise an. Jedem Element $\tilde{x} \in \tilde{\mathcal{X}}$ wird ein Intervall möglicher Level $[l^u(\tilde{x}), l^o(\tilde{x})]$ zugeordnet, dessen Intervallgrenzen folgendermaßen definiert sind: $l^u(\tilde{x})$:= Länge des kürzesten Weges, der von $\tilde{x}^u$ zu $\tilde{x}$ führt; und $l^o(\tilde{x})$:= Level von $\tilde{x}^o$ minus die Länge des kürzesten Weges, der von $\tilde{x}$ zu $\tilde{x}^o$ führt. Somit kann man sagen, daß der Level von $\tilde{x}$ jedenfalls nicht kleiner als $l^u(\tilde{x})$ und nicht größer als $l^o(\tilde{x})$ sein kann. Tabelle 10.2.2 zeigt diese Intervallgrenzen für die Klausurdaten.

6. Um zu zeigen, daß es sich um eine sinnvolle Alternative zur gewöhnlichen Indexbildung handelt, besprechen wir noch ein zweites Beispiel. Die Daten entnehmen wir einer Untersuchung von J. Alber (1988), in der er sich mit den Gesundheitssystemen der OECD-Länder beschäftigt. Um zu

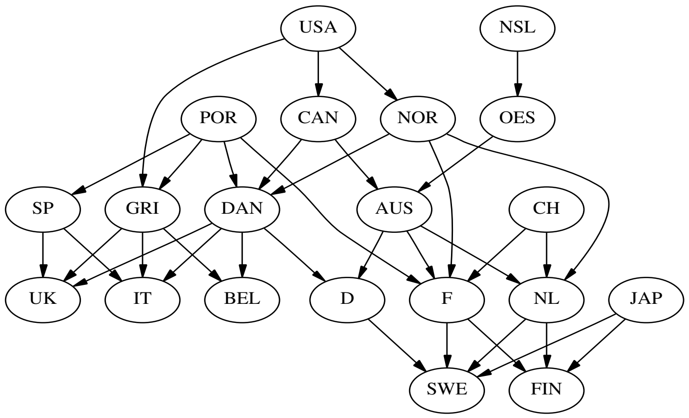


Abb. 10.2.4 Hasse-Diagramm für die partielle Ordnung der OECD-Länder nach Arzt- und Bettendichte.

einem „Index der Gesundheitsversorgung" zu gelangen, bezieht sich Alber auf Angaben über die Arzt- und Bettendichte (jeweils pro 10000 Einwohner). Tabelle 10.2.3 zeigt seine Daten. Alber gewinnt seinen Index, indem er Arzt- und Bettendichte multipliziert. Dann kann er die OECD-Länder entsprechend dem so definierten Grad ihrer Gesundheitsversorgung in eine lineare Ordnung bringen. Ein Problem besteht jedoch darin, daß die beiden Indikatoren nur eine partielle Ordnung liefern. Einige der Länder sind vergleichbar; z.B. war in Deutschland sowohl die Arzt- als auch die Bettendichte größer als in Portugal. Andererseits können jedoch viele Länder auf diese Weise nicht in eine Rangfolge gebracht werden. Das wird anschaulich anhand des Hasse-Diagramms in Abbildung 10.2.4 deutlich.

7. Mit der zuvor beschriebenen Methode ist es jedoch möglich, den Ländern intervallwertige Positionen zuzuordnen. Eine Analyse der Daten zeigt zunächst, daß es ausreicht, fünf hierarchische Level zu unterscheiden.[10] Dann kann man jedes Land durch eine intervallwertige Menge von Leveln charakterisieren, die die Position des Landes in der Hierarchie anzeigt (s. Tabelle 10.2.3). Dies Beispiel zeigt auch, daß man in einigen Fällen eindeutige Positionen erhält. Z.B. hat Neuseeland die Position 1, Österreich die Position 2 und Australien die Position 3. Mit Ausnahme von Neuseeland und Österreich enthalten jedoch alle Intervalle den Level 3, so daß die Indexbildung kaum eindeutige Informationen liefert.

[10] In diesem Beispiel muß dem Datensatz sowohl ein minimales als auch ein maximales Element hinzugefügt werden. Wir haben $\tilde{x}^u = (14, 51)$ und $\tilde{x}^o = (32, 156)$ verwendet.

Kapitel 11

Rangordnungen

In diesem Kapitel beschäftigen wir uns mit Rangordnungsdaten. Zunächst werden Begriffsbildungen und Notationen erläutert, dann wird überlegt, wie man Rangordnungen vergleichen und aggregieren kann.

11.1 Rangordnungsdaten

1. Bei dem Wort 'Rangordnung' kann man zunächst an ordinale Variablen denken, also an Variablen $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, für deren Merkmalsraum eine lineare Ordnung definiert ist. Je nachdem, ob es sich um eine strikte oder um eine schwache lineare Ordnung handelt, erhält man eine strikte oder schwache Rangordnung für die Elemente des Merkmalsraums. Im Unterschied zu ordinalen Variablen sprechen wir von *Rangordnungsdaten*, wenn Befragungspersonen Rangordnungen erzeugen. Es gibt dann zwei Objektmengen. Einerseits eine Menge Ω, die aus den Befragungspersonen besteht, und außerdem eine Menge Ω^*, für deren Elemente eine Rangordnung gebildet werden soll. Man kann sich vorstellen, daß die Elemente von Ω^* entweder Alternativen repräsentieren, die von den Befragungspersonen entsprechend ihrer Präferenzen in eine Rangordnung gebracht werden sollen, oder auf Objekte oder Situationen verweisen, die von den Befragungspersonen bewertet werden sollen. In beiden Fällen erzeugt jede Befragungsperson $\omega \in \Omega$ eine spezifische Rangordnung für die Elemente in Ω^*. Definiert man also eine statistische Variable $R : \Omega \longrightarrow \tilde{\mathcal{R}}$, besteht der Merkmalsraum $\tilde{\mathcal{R}}$ aus Rangordnungen für die Elemente von Ω^*. Variablen dieser Art nennen wir *Rangordnungsvariablen*. Zur numerischen Repräsentation von $\tilde{\mathcal{R}}$ werden Vektoren verwendet. Enthält Ω^* m Elemente, haben diese Vektoren m Komponenten und können als Zeilenvektoren in der Form $\mathbf{r} = (r_1, \ldots, r_m)$ geschrieben werden. Jeder Vektor dieser Art liefert eine bestimmte Rangordnung für die Elemente von Ω^*, indem man von den folgenden Korrespondenzen ausgeht:

$$\omega_i^* \prec \omega_j^* \iff r_i < r_j$$

$$\omega_i^* \equiv \omega_j^* \iff r_i = r_j$$

$$\omega_i^* \preceq \omega_j^* \iff r_i \leq r_j$$

Im ersten Fall wird das Objekt ω_j^* dem Objekt ω_i^* vorgezogen (oder höher bewertet), im zweiten Fall besteht Indifferenz, und im dritten Fall gibt es eine Präferenz oder eine Indifferenz.

2. Orientiert man sich an dieser Notation, kann man zur numerischen Repräsentation des Merkmalsraums $\tilde{\mathcal{R}}$ die Gesamtheit der Vektoren verwenden, die aus m Komponenten bestehen, also den Zahlenraum $\mathbf{R}^m$. Allerdings können unterschiedliche Vektoren die gleiche Rangordnung repräsentieren. Um dem Rechnung zu tragen, verwenden wir folgende Definition:

> Zwei Vektoren $\mathbf{r}, \mathbf{r}' \in \mathbf{R}^m$ heißen *strikt äquivalent*, wenn für alle $1 \leq i < j \leq m$ gilt:
>
> $$r_i < r_j \iff r'_i < r'_j \quad \text{und} \quad r_i = r_j \iff r'_i = r'_j$$

Somit kann man sagen: Zwei Vektoren aus $\mathbf{R}^m$ repräsentieren genau dann die gleiche Rangordnung (für die Elemente von Ω^*), wenn sie strikt äquivalent sind. Da es sich um eine Äquivalenzrelation handelt, kann man auch Äquivalenzklassen bilden, die jeweils aus all denjenigen Vektoren aus $\mathbf{R}^m$ bestehen, die untereinander strikt äquivalent sind. Jede Äquivalenzklasse entspricht genau einer Rangordnung. Wenn $m = 2$ ist, gibt es z.B. drei Rangordnungen bzw. Äquivalenzklassen:

$$\{(r_1, r_2) \,|\, r_1 < r_2\}, \;\; \{(r_1, r_2) \,|\, r_1 > r_2\}, \;\; \{(r_1, r_2) \,|\, r_1 = r_2\}$$

In einem zweidimensionalen Koordinatensystem sind dies die Mengen der Punkte oberhalb, unterhalb und auf der Hauptdiagonalen.

3. Im folgenden verwenden wir zur numerischen Repräsentation des Merkmalsraums $\tilde{\mathcal{R}}$ einer Rangordnungsvariablen R stets den Zahlenraum $\mathbf{R}^m$. Für jedes $\omega_i \in \Omega$ liefert die Rangordnungsvariable R einen Vektor $R(\omega_i) = \mathbf{r}_i = (r_{i1}, \ldots, r_{im}) \in \mathbf{R}^m$, der die von ω_i hergestellte (oder ihm zugeschriebene) Rangordnung für die Elemente von Ω^* angibt. Die Gesamtheit der Werte einer Rangordnungsvariablen können also auch in Gestalt einer Datenmatrix

$$\mathbf{R} = \begin{pmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & & \vdots \\ r_{n1} & \cdots & r_{nm} \end{pmatrix}$$

dargestellt werden, wobei n die Anzahl der Mitglieder von Ω ist.

4. Zur Illustration der Begriffsbildungen verwenden wir ein einfaches Beispiel, das auch zeigt, daß es keineswegs immer erforderlich ist, Rangordnungsdaten zu erheben. $\Omega := \{\omega_1, \ldots, \omega_{100}\}$ bezieht sich auf 100 Teilnehmer einer Vorlesung, und das Problem besteht darin, einen zusätzlichen Termin zu finden, den möglichst viele Teilnehmer wahrnehmen können. Zur Auswahl stehen m mögliche Termine, die durch die Menge $\Omega^* = \{\omega_1^*, \ldots, \omega_m^*\}$ repräsentiert werden. Offenbar ist es nicht unbedingt erforderlich, daß jeder Teilnehmer für alle m Termine eine Rangordnung herstellt. Eine einfache Methode, um zu einem Ergebnis zu gelangen, könnte

z.B. darin bestehen, jeden Teilnehmer zu bitten, den von ihm am meisten präferierten Termin anzugeben. Verwendet man einen Fragebogen, könnte man bei $m = 4$ Terminen folgende Frage stellen:

> Bitte kreuzen Sie denjenigen Termin an, der von Ihnen bevorzugt wird:
>
> ω_1^* ☐ ω_2^* ☐ ω_3^* ☐ ω_4^* ☐

Man könnte sich dann für denjenigen Termin entscheiden, der die meisten Stimmen bekommen hat.

5. Das Beispiel kann jedoch auch verwendet werden, um zu illustrieren, wie Rangordnungsdaten erhoben werden können. Jeder Teilnehmer wird dann gebeten, die m möglichen Termine entsprechend seiner Präferenzen in eine Rangordnung zu bringen:

> Bitte schreiben Sie in jedes Kästchen eine Zahl, die Ihre Präferenz für den jeweiligen Termin angibt. Je größer die Zahl, desto größer Ihre Präferenz. Gleiche Zahlen bedeuten, daß Sie zwischen den Terminen indifferent sind.
>
> ω_1^* ☐ ω_2^* ☐ ω_3^* ☐ ω_4^* ☐

Stellt man die Frage in dieser Form, erhält man als Ergebnis Rangordnungsdaten, die bereits unmittelbar den vorab beschriebenen Darstellungskonventionen entsprechen. Hat z.B. ein Teilnehmer im ersten Kästchen eine 3, im zweiten eine 17 und im dritten und vierten jeweils eine 5 angegeben, bedeutet dies, daß der Termin ω_2^* am meisten, der Termin ω_1^* am wenigsten präferiert wird; dazwischen liegen die Präferenzen für die Termine ω_3^* und ω_4^*, bei denen Indifferenz besteht.[1]

6. Unabhängig davon, wie Rangordnungsdaten erhoben werden, kann man für ihre numerische Repräsentation stets die oben erläuterte Notation verwenden. Allerdings ist die Darstellung nicht eindeutig. Für jede Rangordnung gibt es beliebig viele strikt äquivalente Darstellungen. Z.B. könnte man die Rangordnung $\mathbf{r} = (3, 17, 5, 5)$ auch durch die Vektoren $\mathbf{r} =$

[1] Unser Beispiel soll natürlich nicht nahelegen, daß man Fragen zur Ermittlung von Rangordnungen stets in dieser Weise formulieren sollte. Es gibt viele unterschiedliche Möglichkeiten, die je nach Kontext besser oder schlechter verständlich sein können. Eine andere Erhebungsform, die im SOEP verwendet worden ist, besprechen wir in Abschnitt 11.4. Eine weitere Variante besteht darin, Befragungspersonen zu bitten, für jeweils zwei der insgesamt m Alternativen eine Rangordnung herzustellen. Man spricht dann von *Paarvergleichsdaten*. Insgesamt muß man dann $m\,(m-1)/2$ Paare vergleichen lassen. Ein offensichtlicher Nachteil dieser Methode besteht darin, daß die Ergebnisse der separaten Paarvergleiche nicht unbedingt in eine Rangordnung gebracht werden können.

$(-1,1,0,0)$ und $\mathbf{r} = (50,100,51,51)$ repräsentieren. Eine Standarddarstellung erhält man durch die Verwendung von *Rangzahlen*, die folgendermaßen gebildet werden:

a) Zunächst werden die Komponenten von $\mathbf{r} = (r_1, \ldots, r_m)$ in aufsteigender Reihenfolge geordnet: $r_{(1)} \leq \cdots \leq r_{(m)}$.

b) Dann wird für jede Komponente $r_{(i)}$ eine Menge T_i ermittelt, die die Indizes derjenigen Komponenten umfaßt, die den gleichen Wert wie $r_{(i)}$ haben; also $T_i := \{j \mid r_{(j)} = r_{(i)}\}$. Diese Mengen werden *Bindungsblöcke* genannt.[2]

c) Dann werden aus den Durchschnitten der Indizes in den Bindungsblöcken Rangzahlen gebildet:

$$r^*_{(i)} := \frac{1}{|T_i|} \sum_{j \in T_i} j$$

d) Schließlich werden die Rangzahlen in die ursprüngliche Reihenfolge gebracht, so daß man die Darstellung $\mathbf{r}^* = (r^*_1, \ldots, r^*_m)$ erhält.

Zur Illustration verwenden wir das Beispiel $\mathbf{r} = (3,17,5,5)$. Zunächst liefert eine Ordnung in aufsteigender Reihenfolge: $r_{(1)} = 3$, $r_{(2)} = 5$, $r_{(3)} = 5$ und $r_{(4)} = 17$. Dann bildet man die Bindungsblöcke

$$T_1 = \{1\},\ T_2 = \{2,3\},\ T_3 = \{2,3\},\ T_4 = \{4\}$$

berechnet die Rangzahlen

$$r^*_{(1)} = 1,\ r^*_{(2)} = r^*_{(3)} = \frac{2+3}{2} = 2.5,\ r^*_{(4)} = 4$$

und erhält die Darstellung $\mathbf{r}^* = (1,4,2.5,2.5)$.

7. Schließlich sollte überlegt werden, wieviele unterschiedliche Rangordnungen für m Alternativen bzw. Objekte gebildet werden können. Wenn man zunächst nur Rangordnungen in Betracht zieht, bei denen es keine Indifferenzen gibt, ist die Antwort einfach: Man kann aus den m Alternativen eine wählen, die an die erste Stelle kommen soll, dann aus den verbleibenden $m-1$ Alternativen eine, die an die zweite Stelle kommen soll, usw. Insgesamt kann man auf diese Weise $m(m-1)\cdots 1 = m!$ unterschiedliche Rangordnungen erzeugen. Etwas schwieriger ist es herauszufinden, wieviele Rangordnungen es gibt, in denen auch Indifferenzen vorkommen können. In der Darstellung von Rangordnungen erscheinen Indifferenzen in Form von Bindungen. Also muß man sich überlegen, wieviele unterschiedliche

[2] Im Englischen werden Bindungen 'Ties' genannt und man spricht von Tie-Blöcken.

Bindungsgruppen möglich sind. Am besten orientiert man sich an der Anzahl der Bindungsgruppen, die von $k = 1, \ldots, m$ variieren kann. Wenn $k = 1$ ist, gibt es nur eine Bindungsgruppe und die entsprechende Rangordnung ist zwischen allen Alternativen indifferent. Wenn $k = 2$ ist, gibt es zwei Bindungsgruppen, d.h. die Menge der Alternativen zerfällt in zwei Gruppen, so daß innerhalb jeder Gruppe Indifferenz besteht. Usw. Wenn $k = m$ ist, besteht jede Bindungsgruppe nur aus einem Element, so daß es keine Indifferenzen gibt. Also muß man zunächst für jedes k feststellen, wieviele Möglichkeiten es gibt, die Indexmenge $\{1, \ldots, m\}$ in genau k Teilmengen zu zerlegen. Diese Anzahl wird durch die Stirlingschen Zahlen zweiter Art gegeben, für die das Symbol $S(m,k)$ verwendet wird. Z.B. ist $S(4,2) = 7$, d.h. es gibt 7 unterschiedliche Zerlegungen:

$$\{\{1\},\{2,3,4\}\},\ \{\{2\},\{1,3,4\}\},\ \{\{3\},\{1,2,4\}\},\ \{\{4\},\{1,2,3\}\},$$
$$\{\{1,2\},\{3,4\}\},\ \{\{1,3\},\{2,4\}\},\ \{\{1,4\},\{2,3\}\}$$

Innerhalb jeder Bindungsgruppe besteht Indifferenz. Allerdings muß berücksichtigt werden, daß zwischen den Bindungsgruppen unterschiedliche Präferenzen bestehen können, d.h. bei k Bindungsgruppen gibt es $k!$ unterschiedliche Rangordnungen für die Bindungsgruppen. Somit gelangt man zu folgender Formel für die Anzahl der Rangordnungen, wenn Bindungen zugelassen sind:

$$n_r(m) = \sum_{k=1}^{m} S(m,k)\, k!$$

Diese Anzahl wird schnell sehr groß, wie die folgende Tabelle zeigt.

m	$m!$	$n_r(m)$	m	$m!$	$n_r(m)$
3	6	13	6	720	4683
4	24	75	7	5040	47293
5	120	541	8	40320	545835

Man erkennt auch, daß die Zahl $n_r(m)$ sehr viel schneller größer wird als $m!$, also die Anzahl der Rangordnungen ohne Indifferenzen.

11.2 Vergleich von Rangordnungen

1. Offensichtlich liefert eine Rangordnung für die Elemente einer Menge Ω^* zugleich eine lineare Ordnung. Wenn es keine Indifferenzen gibt, erhält man eine strikte, andernfalls eine schwache lineare Ordnung. Insofern treten die im Zusammenhang mit partiellen Ordnungen diskutierten Probleme bei Rangordnungen zunächst nicht auf. Stattdessen treten zwei andere Fragen in den Vordergrund: wie man zwei oder mehr Rangordnungen vergleichen und wie man mehrere Rangordnungen zu einer „gemeinsamen Rangordnung“ zusammenfassen kann. In diesem Abschnitt besprechen wir eine Metrik zum Vergleich von Rangordnungen, die von J. G. Kemeny (1959/1961) vorgeschlagen wurde.[3]

2. Um Kemenys Metrik zu erläutern, beziehen wir uns auf Rangordnungen für die Elemente einer Menge $\Omega^* = \{\omega_1^*, \dots, \omega_m^*\}$. Wie im vorangegangenen Abschnitt besprochen wurde, kann jede Rangordnung durch einen Vektor $\mathbf{r} = (r_1, \dots, r_m) \in \mathbf{R}^m$ dargestellt werden, wobei zusätzlich vereinbart wird, daß strikt äquivalente Vektoren die gleiche Rangordnung repräsentieren. Also kann eine Metrik für Rangordnungen formal als eine Funktion

$$\mathrm{d} : \mathbf{R}^m \times \mathbf{R}^m \longrightarrow \mathbf{R}$$

konzipiert werden, die jeweils zwei Rangordnungen $\mathbf{r}, \mathbf{r}' \in \mathbf{R}^m$ eine Zahl $\mathrm{d}(\mathbf{r}, \mathbf{r}')$ zuordnet, die ihren Abstand angibt. Kemenys Vorschlag besteht darin, zunächst paarweise alle Komponenten der beiden Vektoren zu betrachten. Wenn also $\mathbf{r} = (r_1, \dots, r_m)$ und $\mathbf{r}' = (r'_1, \dots, r'_m)$ zwei Rangordnungen sind, werden zunächst alle Paare (r_i, r_j) und (r'_i, r'_j) verglichen. Zum Vergleich werden folgende Relationen verwendet:

a) Zunächst kann Identität bestehen:

$$\begin{array}{llllll} (r_i, r_j) =_r (r'_i, r'_j) & \text{wenn} & r_i < r_j & \text{und} & r'_i < r'_j \\ & \text{oder} & r_i > r_j & \text{und} & r'_i > r'_j \\ & \text{oder} & r_i = r_j & \text{und} & r'_i = r'_j \end{array}$$

In diesem Fall liefern (r_i, r_j) und (r'_i, r'_j) die gleiche Rangordnung für ω_i^* und ω_j^*.

b) Weiterhin gibt es die Möglichkeit, daß in einer der beiden Paare eine Indifferenz auftritt:

$$\begin{array}{llllll} (r_i, r_j) \simeq_r (r'_i, r'_j) & \text{wenn} & r_i < r_j & \text{und} & r'_i = r'_j \\ & \text{oder} & r_i > r_j & \text{und} & r'_i = r'_j \\ & \text{oder} & r_i = r_j & \text{und} & r'_i < r'_j \\ & \text{oder} & r_i = r_j & \text{und} & r'_i > r'_j \end{array}$$

c) Schließlich können beide Paare auch eine entgegengesetzte Reihenfolge ausdrücken:

$$\begin{array}{llllll} (r_i, r_j) \neq_r (r'_i, r'_j) & \text{wenn} & r_i < r_j & \text{und} & r'_i > r'_j \\ & \text{oder} & r_i > r_j & \text{und} & r'_i < r'_j \end{array}$$

[3]Eine ausführlichere Darstellung findet man bei Kemeny und Snell 1962, S. 9ff.

3. Dementsprechend gibt es drei Abstufungen:

$$\delta_{ij}(\mathbf{r}, \mathbf{r}') := \begin{cases} 0 & \text{wenn } (r_i, r_j) =_r (r'_i, r'_j) \\ 1 & \text{wenn } (r_i, r_j) \simeq_r (r'_i, r'_j) \\ 2 & \text{wenn } (r_i, r_j) \neq_r (r'_i, r'_j) \end{cases}$$

$\delta_{ij}(\mathbf{r}, \mathbf{r}')$ ist also Null, wenn $\mathbf{r}$ und $\mathbf{r}'$ für die Objekte ω_i^* und ω_j^* die gleiche Rangfolge ausdrücken, der Ausdruck ist 1, wenn bei einer der Rangfolgen Indifferenz besteht, und er ist 2, wenn es eine unterschiedliche Rangfolge gibt. Kemenys Vorschlag besteht nun darin, die Gesamtheit der Übereinstimmungen und Abweichungen zu addieren. Wegen der Symmetrie $\delta_{ij}(\mathbf{r}, \mathbf{r}') = \delta_{ji}(\mathbf{r}, \mathbf{r}')$ und weil $\delta_{ii}(\mathbf{r}, \mathbf{r}') = 0$ ist, genügt es allerdings, nur die Indizes $1 \leq j < i \leq m$ zu betrachten. Somit gelangt man zu folgender Definition:[4]

$$\mathrm{d}_r(\mathbf{r}, \mathbf{r}') := \sum_{j<i} \delta_{ij}(\mathbf{r}, \mathbf{r}') \tag{11.2.1}$$

Hat man z.B. die Rangordnungen $\mathbf{r} = (1, 2, 3, 4)$ und $\mathbf{r}' = (2, 1, 3, 3)$, liefert ein Vergleich der Komponenten für $1 \leq j < i \leq 4$

i	j	$\delta_{ij}(\mathbf{r}, \mathbf{r}')$
2	1	2
3	1	0
3	2	0
4	1	0
4	2	0
4	3	1

und man findet $\mathrm{d}_r(\mathbf{r}, \mathbf{r}') = 3$.

4. Man kann sich überlegen, daß die Funktion d_r allen Bedingungen für eine Metrik genügt.

a) Zunächst ist $\mathrm{d}_r(\mathbf{r}, \mathbf{r}')$ sicherlich immer größer oder gleich Null. Man sieht auch, daß $\mathrm{d}_r(\mathbf{r}, \mathbf{r}')$ einen maximalen Wert annimmt, wenn $\delta_{ij}(\mathbf{r}, \mathbf{r}')$ für alle Indexpaare den Wert 2 hat. Das ist genau dann der Fall, wenn $\mathbf{r}$ und $\mathbf{r}'$ *inverse* Rangordnungen sind.[5] $\mathrm{d}_r(\mathbf{r}, \mathbf{r}')$ nimmt dann den maximalen Wert $m\,(m-1)$ an.

[4]Der Ausdruck $\Sigma_{j<i}$ soll bedeuten, daß über alle Indexpaare $1 \leq j < i \leq m$ summiert wird.

[5]Eine Rangordnung $\mathbf{r}'$ heißt zu einer Rangordnung $\mathbf{r}$ invers, wenn $\mathbf{r}$ keine Bindungen aufweist und $\mathbf{r}'$ für alle Indexpaare eine entgegengesetzte Rangfolge liefert. Um aus $\mathbf{r}$ eine inverse Rangordnung zu erzeugen, kann man die Definition $r'_i := r_{m-i+1}$ verwenden.

b) Weiterhin ist $\mathrm{d}_r(\mathbf{r},\mathbf{r}') = \mathrm{d}_r(\mathbf{r}',\mathbf{r})$, da $\delta_{ij}(\mathbf{r},\mathbf{r}') = \delta_{ij}(\mathbf{r}',\mathbf{r})$ ist, wie aus der Definition unmittelbar folgt.

c) Ebenfalls folgt die Dreiecksungleichung für d_r daraus, daß sie auch für δ_{ij} gilt, nämlich:

$$\delta_{ij}(\mathbf{r},\mathbf{r}') + \delta_{ij}(\mathbf{r}',\mathbf{r}'') \geq \delta_{ij}(\mathbf{r},\mathbf{r}'')$$

da folgende Kombinationen, die die Ungleichung verletzen könnten, nicht vorkommen können:

$\delta_{ij}(\mathbf{r},\mathbf{r}')$	$\delta_{ij}(\mathbf{r}',\mathbf{r}'')$	$\delta_{ij}(\mathbf{r},\mathbf{r}'')$
0	0	1
0	0	2
0	1	2
1	0	2

Somit bleibt nur die Frage, ob aus $\mathrm{d}_r(\mathbf{r},\mathbf{r}') = 0$ folgt, daß $\mathbf{r} = \mathbf{r}'$ ist. Im allgemeinen ist das zwar nicht der Fall; aus $\mathrm{d}_r(\mathbf{r},\mathbf{r}') = 0$ folgt jedoch, daß $\mathbf{r}$ und $\mathbf{r}'$ strikt äquivalent sind, also die gleiche Rangordnung repräsentieren. Man kann deshalb sinnvoll von einer *Metrik für Rangordnungen* sprechen. Wir nennen d_r die *Kemeny-Metrik*.

11.3 Kendalls Rangkorrelation

1. Eine andere Möglichkeit zum Vergleich von Rangordnungen besteht darin, sich am statistischen Korrelationsbegriff zu orientieren. In diesem Abschnitt besprechen wir einen Vorschlag von M. G. Kendall (1938, 1962), der auf ähnlichen Überlegungen beruht wie die Kemeny-Metrik. Den Ausgangspunkt bilden wieder zwei Rangordnungen $\mathbf{r} = (r_1,\ldots,r_m)$ und $\mathbf{r}' = (r'_1,\ldots,r'_m)$. Wie bei der Kemeny-Metrik werden zunächst alle Indexpaare betrachtet und folgende Hilfsgrößen definiert:

$$\tau_{ij}(\mathbf{r},\mathbf{r}') = \begin{cases} 1 & \text{wenn} \quad r_i < r_j \quad \text{und} \quad r'_i < r'_j \\ & \text{oder} \quad r_i > r_j \quad \text{und} \quad r'_i > r'_j \\ -1 & \text{wenn} \quad r_i < r_j \quad \text{und} \quad r'_i > r'_j \\ & \text{oder} \quad r_i > r_j \quad \text{und} \quad r'_i < r'_j \\ 0 & \text{wenn} \quad r_i = r_j \quad \text{oder} \quad r'_i = r'_j \end{cases}$$

Offenbar gilt $\tau_{ij}(\mathbf{r},\mathbf{r}') = \tau_{ji}(\mathbf{r},\mathbf{r}')$ und $\tau_{ii}(\mathbf{r},\mathbf{r}') = 0$, so daß man sich für die Summation auf die Indizes $1 \leq j < i \leq m$ beschränken kann. Somit erhält man zunächst

$$\tau^*(\mathbf{r},\mathbf{r}') := \sum_{j<i} \tau_{ij}(\mathbf{r},\mathbf{r}')$$

Diese Größe liegt zwischen $\pm\, m(m-1)/2$, und man definiert deshalb

$$\tau(\mathbf{r},\mathbf{r}') := \frac{2\,\tau^*(\mathbf{r},\mathbf{r}')}{m\,(m-1)} = \frac{2\sum_{j<i} \tau_{i,j}(\mathbf{r},\mathbf{r}')}{m\,(m-1)} \tag{11.3.1}$$

um zu erreichen, daß $\tau(\mathbf{r},\mathbf{r}')$ nur Werte zwischen -1 und 1 annehmen kann. Dies ist Kendalls Definition einer *Rangkorrelation*, oft auch „Kendalls τ" genannt.

2. Offenbar beruht diese Definition einer Rangkorrelation in ähnlicher Weise auf einem Vergleich von Rangordnungen wie die Kemeny-Metrik. Man sieht das besonders deutlich bei einem Vergleich von Rangordnungen, die keine Bindungen aufweisen, denn dann gilt $\delta_{ij}(\mathbf{r},\mathbf{r}') = |\,\tau_{ij}(\mathbf{r},\mathbf{r}') - 1\,|$. Als Beispiel betrachten wir die Rangordnungen $\mathbf{r} = (1,2,3,4)$ und $\mathbf{r}' = (2,3,4,1)$. Zur Berechnung von Kendalls Rangkorrelation kann die Tabelle

i	j	$\tau_{ij}(\mathbf{r},\mathbf{r}')$
2	1	1
3	1	1
3	2	1
4	1	−1
4	2	−1
4	3	−1

dienen, aus der man $\tau(\mathbf{r},\mathbf{r}') = 0$ findet. Allerdings ist nicht ohne weiteres klar, wie man dieses Ergebnis interpretieren kann.

3. Eine weitere Überlegung von Kemeny kann helfen, den Sachverhalt besser zu verstehen. Kemeny hat nämlich eine Definition vorgeschlagen, mit der man davon sprechen kann, daß eine Rangordnung *zwischen* zwei anderen Rangordnungen liegt.[6] Die Definition erfolgt in zwei Schritten. Gegeben sind drei Rangordnungen $\mathbf{r} = (r_1,\ldots,r_m)$, $\mathbf{r}' = (r'_1,\ldots,r'_m)$ und $\mathbf{r}'' = (r''_1,\ldots,r''_m)$. Zunächst wird für jedes Indexpaar definiert: (r_i,r_j) *liegt zwischen* (r'_i,r'_j) und (r''_i,r''_j), wenn gilt:

$$(r_i,r_j) =_r (r'_i,r'_j) \quad \text{oder} \quad (r_i,r_j) =_r (r''_i,r''_j)$$

Dann kann in einem zweiten Schritt definiert werden: $\mathbf{r}$ *liegt zwischen* $\mathbf{r}'$ und $\mathbf{r}''$, wenn für alle Indexpaare gilt, daß (r_i,r_j) zwischen (r'_i,r'_j) und (r''_i,r''_j) liegt. Mit dieser Definition ist zugleich eine interessante Eigenschaft der Kemeny-Metrik verbunden. Wenn $\mathbf{r}$ zwischen $\mathbf{r}'$ und $\mathbf{r}''$ liegt, gilt für alle Indexpaare

$$\delta_{ij}(\mathbf{r}',\mathbf{r}) + \delta_{ij}(\mathbf{r},\mathbf{r}'') = \delta_{ij}(\mathbf{r}',\mathbf{r}'')$$

[6] Kemeny 1959/1961, S. 46; Kemeny und Snell 1962, S. 12f.

und infolgedessen gilt dann die Dreiecksungleichung exakt:

$$\mathrm{d}_r(\mathbf{r}', \mathbf{r}) + \mathrm{d}_r(\mathbf{r}, \mathbf{r}'') \;=\; \mathrm{d}_r(\mathbf{r}', \mathbf{r}'')$$

4. Als Beispiel betrachten wir noch einmal die beiden Rangordnungen $\mathbf{r} = (1,2,3,4)$ und $\mathbf{r}' = (2,3,4,1)$. Man findet fünf Rangordnungen, die zwischen ihnen liegen:

	Rangordnungen				Abstandsmatrix						
$\mathbf{r}$	1	2	3	4	0	1	2	3	4	5	6
$\mathbf{r}_1$	1	2	3	3	1	0	1	2	3	4	5
$\mathbf{r}_2$	1	2	4	3	2	1	0	1	2	3	4
$\mathbf{r}_3$	1	2	4	2	3	2	1	0	1	2	3
$\mathbf{r}_4$	1	3	4	2	4	3	2	1	0	1	2
$\mathbf{r}_5$	1	2	3	1	5	4	3	2	1	0	1
$\mathbf{r}'$	2	3	4	1	6	5	4	3	2	1	0

Die Abstandsmatrix zeigt nicht nur, in welcher Weise die Dreiecksungleichung exakt gilt, sondern auch, daß man die Rangordnungen in Form einer linearen Kette anordnen kann, wobei $\mathbf{r}$ und $\mathbf{r}'$ die äußeren Pole bilden.

5. Wieviele Rangordnungen es zwischen je zwei vorgegebenen Rangordnungen $\mathbf{r}$ und $\mathbf{r}'$ gibt, hängt von der Beschaffenheit der beiden Rangordnungen ab. Es kann sein, daß es überhaupt keine Rangordnung zwischen ihnen gibt;[7] andererseits liegen stets alle möglichen Rangordnungen zwischen $\mathbf{r}$ und $\mathbf{r}'$, wenn $\mathbf{r}'$ die zu $\mathbf{r}$ inverse Rangordnung ist. Daraus folgt auch, daß man jede Rangordnung $\mathbf{r}$ verwenden kann, um die Gesamtheit aller möglichen Rangordnungen in Form einer linearen Kette zu ordnen, deren Pole durch $\mathbf{r}$ und die zu ihr inverse Rangordnung gebildet werden.

6. Diese Einsicht kann man verwenden, um auch die Kendallsche Rangkorrelation etwas besser zu verstehen. Bezieht man sich auf Rangordnungen $\mathbf{r}$ und $\mathbf{r}'$, kann man z.B. $\mathbf{r}$ als Referenz verwenden und dann alle möglichen Rangordnungen zwischen $\mathbf{r}$ und ihrer inversen Rangordnung $-\mathbf{r}$ anordnen. Insbesondere liegt dann auch $\mathbf{r}'$ zwischen $\mathbf{r}$ und $-\mathbf{r}$, und die Dreiecksungleichung ist exakt erfüllt: $\mathrm{d}_r(\mathbf{r},\mathbf{r}') + \mathrm{d}_r(\mathbf{r}',-\mathbf{r}) = \mathrm{d}_r(\mathbf{r},-\mathbf{r}) = m\,(m-1)$. Bei den Rangordnungen $\mathbf{r} = (1,2,3,4)$ und $\mathbf{r}' = (2,3,4,1)$ findet man z.B.: $\mathrm{d}_r(\mathbf{r},\mathbf{r}') = \mathrm{d}_r(\mathbf{r}',-\mathbf{r}) = 6$ und $\mathrm{d}_r(\mathbf{r},-\mathbf{r}) = 12$. So kann man sich vorstellen, daß durch die Kendallsche Rangkorrelation gewissermaßen die Lage der Rangordnung $\mathbf{r}'$ in einer Kette, die von $\mathbf{r}$ bis $-\mathbf{r}$ reicht, angegeben wird. Allerdings ist der Zusammenhang zu den metrischen Vorstellungen nicht linear. Bezieht man sich z.B. auf die Rangordnungen $\mathbf{r} = (1,2,3,4)$ und $\mathbf{r}' = (2,1,3,4)$, findet man als Rangkorrelation $\tau(\mathbf{r},\mathbf{r}') = 2/3$; jedoch ist $\mathrm{d}_r(\mathbf{r},\mathbf{r}') = 2$ und $\mathrm{d}_r(\mathbf{r}',-\mathbf{r}) = 10$.

[7] Z.B. gibt es keine weitere Rangordnung zwischen $(1,1,2,3)$ und $(1,1,1,3)$.

11.4 Aggregation von Rangordnungen

1. Eine weitere Frage bei der Beschäftigung mit Rangordnungsdaten bezieht sich darauf, wie man, wenn mehrere Rangordnungen für die gleiche Menge von Objekten Ω^* gegeben sind, eine mittlere Rangordnung finden kann, die die gegebenen Rangordnungen möglichst gut repräsentiert. Diese Frage stellt sich z.B. bei Abstimmungen oder wenn aus mehreren Bewertungen eine gemeinsame Bewertung gebildet werden soll. Ausführliche Diskussionen gibt es in der Literatur über „social choice“ (man vgl. z.B. Kelly 1988). In diesem Abschnitt besprechen wir einen Ansatz, der von Kemeny vorgeschlagen wurde und auf seiner Metrik für Rangordnungen beruht.[8]

2. Den Ausgangspunkt bilden n Rangordnungen für die Alternativen in einer Menge $\Omega^* = \{\omega_1^*, \ldots, \omega_m^*\}$, die wir durch $\mathbf{r}_i = (r_{i1}, \ldots, r_{im}) \in \mathbf{R}^m$ (für $i = 1, \ldots, n$) vergegenwärtigen. Gesucht ist eine neue Rangordnung, die die Rangordnungen $\mathbf{r}_1, \ldots, \mathbf{r}_n$ möglichst gut repräsentiert. Kemenys Vorschlag orientiert sich daran, wie in der Statistik Mittelwerte gebildet werden. Die neue Rangordnung soll so gebildet werden, daß sie im Durchschnitt möglichst geringe Abstände zu den gegebenen Rangordnungen $\mathbf{r}_1, \ldots, \mathbf{r}_n$ aufweist. Dafür kann Kemenys Metrik für Rangordnungen verwendet werden; und man kann analog zur Unterscheidung zwischen Mittelwert und Median zwei Definitionen in Betracht ziehen. Im ersten Fall betrachtet man Rangordnungen, die die Funktion

$$\bar{f}(\mathbf{r}) \;:=\; \sum_{i=1}^{n} \mathrm{d}_r(\mathbf{r}_i, \bar{\mathbf{r}})^2 \qquad (11.4.1)$$

minimieren. Die Menge der Lösungen bezeichnen wir mit $\bar{L}$. Jedes Element von $\bar{L}$ liefert eine *mittlere Rangordnung* für $\mathbf{r}_1, \ldots, \mathbf{r}_n$. Im zweiten Fall betrachtet man Rangordnungen, die die Funktion

$$\tilde{f}(\mathbf{r}) \;:=\; \sum_{i=1}^{n} \mathrm{d}_r(\mathbf{r}_i, \tilde{\mathbf{r}}) \qquad (11.4.2)$$

minimal machen. Es werden also die absoluten, nicht die quadrierten Abstände minimiert, und die Lösungen werden dementsprechend *Median-Rangordnungen* für $\mathbf{r}_1, \ldots, \mathbf{r}_n$ genannt. Die Menge der Lösungen bezeichnen wir analog mit $\tilde{L}$.

3. Sowohl mittlere als auch Median-Rangordnungen sind nicht unbedingt eindeutig, d.h. sowohl $\bar{L}$ als auch $\tilde{L}$ können mehrere Elemente enthalten, die die Zielfunktionen $\bar{f}$ bzw. $\tilde{f}$ gleichermaßen minimieren. Als Beispiel

[8] Kemeny 1959/1961, S. 45ff, sowie Kemeny und Snell 1962, S. 9ff.

betrachten wir die vier Rangordnungen

$$\mathbf{r}_1 = (3,2,3,3,3) \quad \mathbf{r}_3 = (3,3,2,2,1)$$
$$\mathbf{r}_2 = (1,2,2,3,3) \quad \mathbf{r}_4 = (3,1,1,2,2)$$

Man findet als Lösungen jeweils zwei mittlere und Median-Rangordnungen, die in diesem Fall identisch sind: $\bar{\mathbf{r}}_1 = \tilde{\mathbf{r}}_1 = (2,1,1,2,2)$ und $\bar{\mathbf{r}}_2 = \tilde{\mathbf{r}}_2 = (3,1,1,2,2)$. Meistens unterscheiden sich jedoch mittlere und Median-Rangordnungen.

4. Zur Illustration verwenden wir Rangordnungsdaten, die 1996 im Rahmen des SOEP (Welle 13) erhoben worden sind, indem folgende Frage gestellt wurde:

Auch in der Politik kann man nicht alles auf einmal haben. Wir nennen Ihnen jetzt einige Ziele, die man in der Politik verfolgen kann. Wenn Sie zwischen diesen verschiedenen Zielen wählen müßten:

Welches Ziel erschiene Ihnen persönlich am wichtigsten? Welches am zweitwichtigsten? Welches käme an dritter Stelle? Und welches an vierter Stelle?

In jede Spalte *eine* Nennung!

	1. Stelle	2. Stelle	3. Stelle	4. Stelle
Aufrechterhaltung von Ruhe und Ordnung in diesem Land	☐	☐	☐	☐
Mehr Einfluß der Bürger auf die Entscheidungen der Regierung	☐	☐	☐	☐
Kampf gegen steigende Preise	☐	☐	☐	☐
Schutz des Rechts auf freie Meinungsäußerung	☐	☐	☐	☐

Nur 172 der insgesamt 13511 Personen, denen diese Frage gestellt worden ist, haben sie ignoriert. Bei den übrigen gibt es in 246 Fällen bei mindestens einer der vier Teilfragen keine Angabe. Diese Fälle schließen wir aus, so daß 13093 Personen verbleiben, die in jeder Teilfrage eine Angabe gemacht haben.

5. Obwohl durch den Wortlaut der Frage Indifferenzen eigentlich ausgeschlossen werden sollten, haben viele der Befragungspersonen dies ignoriert. Das sollte bei der Interpretation berücksichtigt werden. Z.B. haben 270 Personen bei jeder Teilfrage die erste Stelle angegeben, was man so interpretieren kann, daß diesen Personen alle vier Ziele gleichermaßen wichtig sind. Und sechs Personen haben z.B. bei den ersten beiden Teilfragen die erste Stelle, bei der dritten die dritte und bei der vierten die

Tabelle 11.1 Rangordnungen in den SOEP-Daten.

r_1	r_2	r_3	r_4	h	r_1	r_2	r_3	r_4	h	r_1	r_2	r_3	r_4	h
1	2	4	3	146	4	4	4	3	63	1	3	3	4	6
3	4	2	1	822	4	1	3	2	732	1	4	3	4	12
2	4	1	3	402	1	2	3	4	151	3	2	4	3	6
3	1	2	4	149	2	4	2	4	20	3	3	4	4	14
2	1	3	4	104	2	4	3	1	404	3	4	4	2	6
1	4	2	3	291	4	4	3	3	36	4	4	2	3	6
4	3	2	1	1968	4	3	4	4	81	1	2	4	2	5
4	2	3	1	1524	4	2	3	3	16	2	3	2	4	3
1	3	2	4	276	4	3	4	3	69	3	4	3	3	7
3	4	1	2	528	3	2	4	1	540	4	1	2	4	24
4	4	4	4	270	4	3	3	3	56	3	4	3	2	7
1	4	3	2	249	2	1	4	3	127	2	2	3	4	2
3	1	4	2	252	1	3	4	4	1	3	2	4	4	3
3	2	1	4	327	3	3	4	3	12	3	4	2	2	3
4	1	2	3	517	4	2	4	3	17	2	4	3	2	2
2	3	4	1	284	3	2	4	2	2	1	4	1	3	2
1	3	4	2	123	3	3	3	4	5	4	4	3	2	12
4	3	1	2	1090	4	3	2	4	18	2	3	4	2	5
4	2	1	3	694	4	2	3	2	18	3	2	3	4	2
4	4	2	4	52	4	3	3	2	28	4	2	2	3	2
4	3	1	1	16	4	3	2	3	27	1	4	4	3	2
2	3	1	4	324	3	4	2	4	10	1	1	3	2	1
4	3	3	4	53	3	3	2	4	1	3	3	4	2	1
3	4	4	3	11	4	3	4	2	22					
3	4	2	3	11	3	4	4	4	21					

zweite Stelle angekreuzt, was in analoger Interpretation bedeuten würde, daß ihnen die ersten beiden Ziele gleichermaßen am wichtigsten sind, daß in der Rangfolge dann das vierte und schließlich das dritte Ziel kommt. Es handelt sich also um Rangordnungsdaten, die auch direkt in unsere Standardform gebracht werden können. Man erhält dann für alle Personen $i = 1, \ldots, 13093$ Rangordnungen $\mathbf{r}_i = (r_{i1}, \ldots, r_{i4})$. Z.B. bei den eben erwähnten sechs Personen gibt es die Rangordnung $(4,4,2,3)$. Natürlich sind nicht alle Rangordnungen unterschiedlich, denn bei $m = 4$ Alternativen gibt es maximal 75 unterschiedliche Rangordnungen. Tatsächlich kommen 73 dieser 75 Rangordnungen auch mindestens einmal vor. Tabelle 11.1 zeigt die Rangordnungen und ihre Häufigkeiten.

6. Berechnet man mittlere und Median-Rangordnungen, findet man, daß sie in diesem Fall eindeutig sind, und zwar:

$$\bar{\mathbf{r}} = (4,3,2,2) \quad \text{und} \quad \tilde{\mathbf{r}} = (4,3,2,1)$$

Dies sind natürlich Mittelwerte; z.B. haben nur 15 % der Befragungspersonen tatsächlich die Median-Rangordnung angegeben. Ein Politiker könnte vielleicht dennoch sagen, daß die Median-Rangordnung am besten die Gesamtheit der unterschiedlichen Rangordnungen repräsentiert.

Kapitel 12

Implizite Indikatoren (Conjoint Analysis)

Im Kapitel 10 stand die Frage im Mittelpunkt, ob und ggf. wie sich für einen mehrdimensionalen Merkmalsraum eine lineare Ordnung begründen läßt. In diesem Kapitel besprechen wir einen komplementären Ansatz, bei dem angenommen wird, daß man sich irgendwie (typischerweise durch Rating-Verfahren) eine lineare Ordnung für einen mehrdimensionalen Merkmalsraum verschaffen kann und die Fragestellung darin besteht, wie man daraus lineare Ordnungen für die Teilmerkmalsräume konstruieren kann. In der englischsprachigen Literatur wird von „conjoint analysis“ oder „conjoint measurement“ gesprochen, im Deutschen manchmal von „(additiv) verbundener Messung“. Da das Verfahren darauf hinausläuft, Indikatoren zu konstruieren, sprechen wir von einem *Verfahren der impliziten Indikatoren.*

12.1 Der konzeptionelle Ansatz

1. Ausgangspunkt ist ein mehrdimensionaler Merkmalsraum

$$\tilde{\mathcal{X}} := \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$$

bei dessen Komponenten es sich um qualitative oder um quantitative Merkmalsräume handeln kann. Die weitere Vorgehensweise besteht aus zwei Schritten: In einem ersten Schritt verschafft man sich durch ein Rating-Verfahren eine lineare Ordnung für den gemeinsamen Merkmalsraum $\tilde{\mathcal{X}}$. Dann wird in einem zweiten Schritt versucht, lineare Ordnungen für die Teilmerkmalsräume $\tilde{\mathcal{X}}_j$ zu konstruieren, so daß sich daraus die im ersten Schritt ermittelte Ordnung für den gemeinsamen Merkmalsraum rekonstruieren läßt. — Daraus wird bereits deutlich, daß mit diesem Verfahren keine sozialwissenschaftlichen Größenbegriffe begründet werden können. Man kann es aber verwenden, um Rangordnungen zu analysieren, die sich durch Befragungen ermitteln lassen. Praktische Anwendungen gibt es bisher hauptsächlich in der Marktforschung, wenn man herausfinden möchte, wie Präferenzen für (potentielle) Konsumgüter von deren Eigenschaften abhängig sein können.[1]

[1]Eine an Anwendungen in der Marktforschung orientierte Einführung in das Verfahren findet man bei Backhaus, Erichson, Plinke, Schuchard-Ficher und Weiber 1987, Kap. 8. Eine neuere Übersicht über Varianten der Anwendung des Verfahrens in der Marktforschung geben Gustafsson, Herrmann und Huber 2000.

2. Einer der wenigen Versuche, das Verfahren auch in der Sozialforschung anzuwenden, findet sich in einer Arbeit von A. Diekmann (1982), in der er eine „additiv verbundene Messung des Sozialprestiges“ vorgeschlagen hat. In seinen Überlegungen verwendet Diekmann einen 2-dimensionalen Merkmalsraum $\tilde{\mathcal{X}} := \tilde{\mathcal{X}}_1 \times \tilde{\mathcal{X}}_2$. Die erste Komponente bezieht sich auf drei Arten von Ausbildung:

$$\tilde{\mathcal{X}}_1 := \{\tilde{x}_{11}, \tilde{x}_{12}, \tilde{x}_{13}\}$$

die so definiert sind: $\tilde{x}_{11}$ = 'Volksschule', $\tilde{x}_{12}$ = 'Abitur', $\tilde{x}_{13}$ = 'Universität'. Die zweite Komponente bezieht sich auf vier Einkommensklassen:

$$\tilde{\mathcal{X}}_2 := \{\tilde{x}_{21}, \tilde{x}_{22}, \tilde{x}_{23}, \tilde{x}_{24}\}$$

die so definiert sind:

$$\begin{aligned} \tilde{x}_{21} &= \text{'10000 (öst. Schilling)'} & \tilde{x}_{23} &= \text{'30000 (öst. Schilling)'} \\ \tilde{x}_{22} &= \text{'20000 (öst. Schilling)'} & \tilde{x}_{24} &= \text{'40000 (öst. Schilling)'} \end{aligned}$$

Es ist klar, daß sich für den Merkmalsraum $\tilde{\mathcal{X}}$ keine sinnvolle lineare Ordnung begründen läßt, so daß man von einer „Größe des Sozialprestiges“ sprechen könnte. Tatsächlich hat Diekmann das auch nicht versucht, sondern stattdessen eine Reihe von Personen gebeten, nach ihrem jeweiligen Gutdünken die 12 Elemente von $\tilde{\mathcal{X}}$ in eine Reihenfolge zu bringen. Einige dieser Rangordnungen werden wir später als Beispiele verwenden.

3. Hier geht es zunächst um den konzeptionellen Ansatz. Der erste Schritt besteht also darin, daß man sich auf irgendeine Weise eine Rangordnung r^* für den mehrdimensionalen Merkmalsraum $\tilde{\mathcal{X}}$ verschafft; dabei ist $r^*(\tilde{x})$ die Rangnummer für den Merkmalswert $\tilde{x} \in \tilde{\mathcal{X}}$.[2] Dann beginnt der zweite Schritt im Verfahren der impliziten Indikatoren. Es wird nach *Bewertungsfunktionen* für die Komponenten von $\tilde{\mathcal{X}}$ gesucht, so daß aus ihnen die vorgegebene Rangordnung r^* rekonstruiert werden kann. Diese Bewertungsfunktionen können allgemein so konzipiert werden:

$$\phi_j : \tilde{\mathcal{X}}_j \longrightarrow \mathbf{R} \qquad (\text{für } j = 1, \ldots, m) \tag{12.1.1}$$

Eine Bewertungsfunktion besteht also in einer numerischen Repräsentation für die in einem Merkmalsraum zusammengefaßten Attribute. Gelegentlich spricht man auch davon, daß durch Bewertungsfunktionen „Scores“ für diese Attribute erzeugt werden.

4. Wären nun solche Bewertungsfunktionen gegeben, könnte daraus auch eine Bewertungsfunktion

$$\phi : \tilde{\mathcal{X}} \longrightarrow \mathbf{R}$$

[2]Die Notation folgt den Ausführungen über Rangordnungen in Abschnitt 11.1. Die Schreibweise $r^*(\tilde{x}) < r^*(\tilde{x}')$ soll bedeuten, daß $\tilde{x}'$ höher bewertet wird als $\tilde{x}$ (oder daß $\tilde{x}'$ gegenüber $\tilde{x}$ präferiert wird), und $r^*(\tilde{x}) = r^*(\tilde{x}')$ soll bedeuten, daß Indifferenz besteht. Es wird also nur eine schwache lineare Ordnung vorausgesetzt.

für den gemeinsamen Merkmalsraum $\tilde{\mathcal{X}}$ gebildet werden. Dafür gibt es viele verschiedene Möglichkeiten. In den Standardvarianten des Verfahrens der impliziten Indikatoren wird ϕ additiv gebildet:[3]

$$\phi(\tilde{x}_1, \ldots, \tilde{x}_m) := \sum_{j=1}^{m} \phi_j(\tilde{x}_j) \tag{12.1.2}$$

Die Vorgehensweise ist dann analog zur Bildung additiver Indizes, allerdings in umgekehrter Reihenfolge. Bei der Bildung von Indizes wird von gegebenen Indikatoren ausgegangen, um daraus einen Index zu bilden. Bei dem Verfahren, das wir hier besprechen, wird nach einer additiven Zerlegung eines vorgegebenen Index in Indikatoren gesucht. Deshalb sprechen wir von einem Verfahren der impliziten Indikatoren.

5. Allerdings kennt man zunächst auch die Bewertungsfunktion ϕ nicht; sondern als Voraussetzung ist nur die Rangordnung r^* für die Elemente von $\tilde{\mathcal{X}}$ gegeben. Hier setzt nun folgende Idee ein: Es wird nach Bewertungsfunktionen $\phi_1, \ldots, \phi_m$ gesucht, so daß für eine daraus entsprechend (12.1.2) gebildete Bewertungsfunktion ϕ gilt:

Für alle $\tilde{x}, \tilde{x}' \in \tilde{\mathcal{X}}$: (12.1.3)

$$\begin{aligned} \phi(\tilde{x}) = \phi(\tilde{x}') &\iff r^*(\tilde{x}) = r^*(\tilde{x}') \\ \phi(\tilde{x}) < \phi(\tilde{x}') &\iff r^*(\tilde{x}) < r^*(\tilde{x}') \end{aligned}$$

Diese Bedingung wird *strikte (monotone) Äquivalenz* genannt. Wenn es gelingt, diese Bedingung zu erfüllen, kann die vorgegebene Rangordnung r^* aus den Bewertungsfunktionen für die Komponenten des Merkmalsraums $\tilde{\mathcal{X}}$ rekonstruiert werden. Allerdings ist das oft nicht möglich, und man begnügt sich dann mit einer schwächeren Bedingung:

Für alle $\tilde{x}, \tilde{x}' \in \tilde{\mathcal{X}}$: (12.1.4)

$$\begin{aligned} \phi(\tilde{x}) < \phi(\tilde{x}') &\Longrightarrow r^*(\tilde{x}) \leq r^*(\tilde{x}') \\ r^*(\tilde{x}) < r^*(\tilde{x}') &\Longrightarrow \phi(\tilde{x}) \leq \phi(\tilde{x}') \end{aligned}$$

die wir *schwache (monotone) Äquivalenz* nennen.

12.2 Berechnungsverfahren

In den folgenden Abschnitten wird besprochen, ob und ggf. wie Bewertungsfunktionen konstruiert werden können, die das Kriterium (12.1.3) oder die schwächere Variante (12.1.4) erfüllen. Zunächst werden Varianten der Regressionsrechnung behandelt, dann wird ein Verfahren der linearen Programmierung skizziert, mit dem strikt monoton äquivalente Lösungen gefunden werden können (wenn es sie gibt).

[3]In der Literatur wird bei dieser Vorgehensweise von „additiv verbundener Messung“ oder „additive conjoint measurement“ gesprochen.

12.2.1 Einfache Regression

1. Um zu einer Lösung zu gelangen, müssen die Werte der vorgegebenen Rangordnung $r^*(\tilde{x})$ und die Werte einer zu konstruierenden Bewertungsfunktion $\phi(\tilde{x})$ verglichen werden. Ein einfacher Ansatz besteht darin, die Werte der Rangordnung einfach als Zahlen zu verwenden, so wie sie vorliegen. Man kann dann das Ziel darin sehen, eine Bewertungsfunktion zu finden, die diese Zahlen möglichst gut annähert. Verwendet man zur Beurteilung der Näherung die quadrierten Abstände, gewinnt man folgende zu minimierende Zielfunktion:

$$\sum_{\tilde{x} \in \tilde{\mathcal{X}}} \left(r^*(\tilde{x}) - \phi(\tilde{x}) \right)^2$$

Um sich auf die Bewertungsfunktionen für die Teilmerkmalsräume $\tilde{\mathcal{X}}_j$ zu beziehen, kann man diese Zielfunktion auch so schreiben:

$$\sum_{(\tilde{x}_1, \ldots, \tilde{x}_m) \in \tilde{\mathcal{X}}} \left(r^*(\tilde{x}_1, \ldots, \tilde{x}_m) - \sum_{j=1}^{m} \phi_j(\tilde{x}_j) \right)^2 \tag{12.2.1}$$

2. Wenn es in jedem Teilmerkmalsraum nur eine endliche Anzahl unterschiedlicher Merkmalswerte gibt,[4] kann man die Bewertungsfunktionen als Linearkombinationen der Merkmalswerte schreiben. Es sei $\tilde{\mathcal{X}}_j = \{\tilde{x}_{j1}, \ldots, \tilde{x}_{jq_j}\}$, so daß q_j die Anzahl der Merkmalswerte in $\tilde{\mathcal{X}}_j$ ist und Indikatorvariablen

$$I_{jk} := \begin{cases} 1 & \text{wenn } \tilde{x}_j = \tilde{x}_{jk} \\ 0 & \text{andernfalls} \end{cases}$$

definiert werden können. Jede Bewertungsfunktion ϕ_j läßt sich dann als Linearkombination dieser Indikatorvariablen darstellen:

$$\phi_j(\tilde{x}_j) := \sum_{k=1}^{q_j} I_{jk}\,\beta_{jk} \tag{12.2.2}$$

Aus den Koeffizienten β_{jk} gewinnt man die Bewertungsfunktionen ϕ_j bzw. die numerischen Repräsentationen für die Teilmerkmalsräume.

3. Die Aufgabe, die Funktionen ϕ_j zu bestimmen, besteht somit darin, die Koeffizienten β_{jk} zu finden. Aus der Darstellung (12.2.2) gewinnt man auch eine entsprechende Darstellung für die Bewertungsfunktion ϕ, nämlich

$$\phi(\tilde{x}_1, \ldots, \tilde{x}_m) = \sum_{j=1}^{m} \phi_j(\tilde{x}_j) = \sum_{j=1}^{m} \sum_{k=1}^{q_j} I_{jk}\,\beta_{jk}$$

[4]Dies muß schon deshalb vorausgesetzt werden, um Rangordnungen für die Elemente von $\tilde{\mathcal{X}}$ mit Rating-Verfahren erzeugen zu können.

Tabelle 12.2.1 Exemplarische Daten für implizite Indikatoren.

$\tilde{x}_1$	$\tilde{x}_2$	$r^*(\tilde{x}_1, \tilde{x}_2)$	I_{11}	I_{12}	I_{13}	I_{21}	I_{22}	I_{23}	I_{24}	$\phi(\tilde{x}_1, \tilde{x}_2)$
$\tilde{x}_{11}$	$\tilde{x}_{21}$	1	1	0	0	1	0	0	0	0.00
$\tilde{x}_{11}$	$\tilde{x}_{22}$	3	1	0	0	0	1	0	0	3.00
$\tilde{x}_{11}$	$\tilde{x}_{23}$	6	1	0	0	0	0	1	0	5.67
$\tilde{x}_{11}$	$\tilde{x}_{24}$	9	1	0	0	0	0	0	1	8.00
$\tilde{x}_{12}$	$\tilde{x}_{21}$	2	0	1	0	1	0	0	0	1.25
$\tilde{x}_{12}$	$\tilde{x}_{22}$	5	0	1	0	0	1	0	0	4.25
$\tilde{x}_{12}$	$\tilde{x}_{23}$	7	0	1	0	0	0	1	0	6.92
$\tilde{x}_{12}$	$\tilde{x}_{24}$	10	0	1	0	0	0	0	1	9.25
$\tilde{x}_{13}$	$\tilde{x}_{21}$	4	0	0	1	1	0	0	0	4.00
$\tilde{x}_{13}$	$\tilde{x}_{22}$	8	0	0	1	0	1	0	0	7.00
$\tilde{x}_{13}$	$\tilde{x}_{23}$	11	0	0	1	0	0	1	0	9.67
$\tilde{x}_{13}$	$\tilde{x}_{24}$	12	0	0	1	0	0	0	1	12.00

Ein zu (12.2.1) äquivalenter Regressionsansatz besteht also darin, folgenden Abstand zu minimieren (vgl. Abschnitt A.6):

$$\| r^*(\tilde{x}_1, \ldots, \tilde{x}_m) - \sum_{j=1}^{m} \sum_{k=1}^{q_j} I_{jk} \, \beta_{jk} \|^2 \tag{12.2.3}$$

Aus der Lösung dieser Regressionsaufgabe gewinnt man die Werte für die Koeffizienten β_{jk}.

4. Zur Illustration des Verfahrens verwenden wir den oben erläuterten Merkmalsraum von Diekmann. Tabelle 12.2.1 zeigt in den ersten beiden Spalten die möglichen Merkmalskombinationen. In der dritten Spalte steht eine Rangordnung, die von der Befragungsperson Nr. 7 stammt (Diekmann 1982, S. 27). Die weiteren Spalten enthalten die Indikatorvariablen. Dies sind die Daten für den Regressionsansatz. Allerdings sind die Indikatorvariablen innerhalb jeder Komponente des Merkmalsraums linear abhängig, so daß man auf jeweils eine von ihnen verzichten kann. Für unsere Illustration nehmen wir an: $\beta_{11} = \beta_{21} = 0$. Fügt man einen Interzept-Term hinzu, besteht die Regressionsaufgabe darin, folgenden Abstand zu minimieren:

$$\| r^*(\tilde{x}_1, \tilde{x}_2) - (\alpha + I_{12}\beta_{12} + I_{13}\beta_{13} + I_{22}\beta_{22} + I_{23}\beta_{23} + I_{24}\beta_{24}) \|^2$$

Daraus gewinnt man die Werte der Koeffizienten, die die Zielfunktion (12.2.1) minimieren:

$$\begin{array}{rrr} \hat{\alpha} = 0.58 & \hat{\beta}_{12} = 1.25 & \hat{\beta}_{13} = 4.00 \\ \hat{\beta}_{22} = 3.00 & \hat{\beta}_{23} = 5.67 & \hat{\beta}_{24} = 8.00 \end{array}$$

Sie liefern auch unmittelbar die gesuchten Bewertungsfunktionen bzw. numerischen Repräsentationen für die einzelnen Merkmalsräume, nämlich:

$$\begin{array}{llll} \phi_1(\tilde{x}_{11}) = 0.00 & \phi_1(\tilde{x}_{12}) = 1.25 & \phi_1(\tilde{x}_{13}) = 4.00 & \\ \phi_2(\tilde{x}_{21}) = 0.00 & \phi_2(\tilde{x}_{22}) = 3.00 & \phi_2(\tilde{x}_{23}) = 5.67 & \phi_2(\tilde{x}_{24}) = 8.00 \end{array}$$

5. Die letzte Spalte von Tabelle 12.2.1 zeigt die mit Hilfe der Regressionsrechnung ermittelte Rangordnung, also die Werte der Funktion

$$\phi(\tilde{x}_1, \tilde{x}_2) := \phi_1(\tilde{x}_1) + \phi_2(\tilde{x}_2)$$

Man erkennt, daß das Kriterium (12.1.3) erfüllt wird, d.h. es ist gelungen, Bewertungsfunktionen ϕ_1 und ϕ_2 zu finden, so daß ein daraus gebildeter additiver Index die vorgegebene Rangordnung reproduziert.

6. Die Bewertungsfunktionen sind nicht eindeutig. Insbesondere kann man lineare Transformationen $\phi'_j = a + b\,\phi_j$ $(a \geq 0, b > 0)$ vornehmen, ohne daß sich die Rangordnung verändert. Wählt man z.B. $a = 1$ und $b = 2$, erhält man die Rangordnung

$$(1, 7, 12.34, 18, 4.5, 10.5, 15.84, 20.5, 10, 16, 21.34, 26)$$

die wiederum zur vorgegebenen Rangordnung strikt äquivalent ist. Dies ist auch der Grund dafür, daß man bei der Bildung der Bewertungsfunktionen den Koeffizienten $\hat{\alpha}$, der dem Interzept-Term entspricht, ignorieren kann. In Abschnitt 12.3.1 werden wir uns noch einmal mit der Unbestimmtheit von Bewertungsfunktionen beschäftigen.

12.2.2 Regression mit Rangordnungen

1. Bevor wir uns mit weiteren Beispielen beschäftigen, muß besprochen werden, daß das bisher verwendete Regressionsverfahren unzureichend ist, weil es nicht berücksichtigt, daß durch das anfängliche Rating-Verfahren bestenfalls eine Rangordnung für die Elemente des kombinierten Merkmalsraums $\tilde{\mathcal{X}}$ zur Verfügung steht. Eine solche Rangordnung kann jedoch auf viele unterschiedliche Weisen numerisch repräsentiert werden; und es sollte deshalb ein Verfahren verwendet werden, daß nur die Rangordnungsinformationen ausnutzt. Um diesen Gedanken zu verfolgen, präzisieren wir in diesem Abschnitt die Problemstellung und besprechen im nächsten Abschnitt ein Rechenverfahren.

2. Ausgangspunkt ist ein kombinierter Merkmalsraum

$$\tilde{\mathcal{X}} = \{(\tilde{x}_1, \ldots, \tilde{x}_m) \,|\, \tilde{x}_1 \in \tilde{\mathcal{X}}_1, \ldots, \tilde{x}_m \in \tilde{\mathcal{X}}_m\}$$

Wie bereits im vorangegangenen Abschnitt nehmen wir an, daß der Teilmerkmalsraum $\tilde{\mathcal{X}}_j$ aus q_j Elementen besteht. Somit hat $\tilde{\mathcal{X}}$ insgesamt $q =$

$q_1 \cdots q_m$ Elemente. Diese q Elemente können in einer (q, m)-Matrix folgendermaßen dargestellt werden:

$$\tilde{\mathbf{X}} := \begin{pmatrix} \tilde{x}_{11} & \cdots & \tilde{x}_{m1} \\ \vdots & & \vdots \\ \tilde{x}_{11} & \cdots & \tilde{x}_{mq_m} \\ \vdots & & \vdots \\ \tilde{x}_{1q_1} & \cdots & \tilde{x}_{m1} \\ \vdots & & \vdots \\ \tilde{x}_{1q_1} & \cdots & \tilde{x}_{mq_m} \end{pmatrix}$$

Jede Zeile liefert eine der möglichen Merkmalskombinationen in $\tilde{\mathcal{X}}$. Zur Illustration kann man an die ersten beiden Spalten in Tabelle 12.2.1 denken. Eine Rangordnung entsteht dadurch, daß für die Zeilen von $\tilde{\mathbf{X}}$ Rangplätze vergeben werden. Indiziert man die Zeilen durch $i = 1, \ldots, q$, kann also eine Rangordnung als ein Spaltenvektor $\mathbf{r}^* := (r_1^*, \ldots, r_q^*)'$ dargestellt werden: r_i^* liefert die Rangnummer für die i-te Zeile von $\tilde{\mathbf{X}}$.

3. Nun gibt es jedoch beliebig viele äquivalente Darstellungen für eine solche Rangordnung, nämlich alle Vektoren $\mathbf{r} = (r_1, \ldots, r_q)'$, für die gilt:

$$r_i = r_{i'} \iff r_i^* = r_{i'}^* \quad \text{und} \quad r_i < r_{i'} \iff r_i^* < r_{i'}^*$$

Wenn diese Bedingungen erfüllt sind, nennen wir $\mathbf{r}$ und $\mathbf{r}^*$ *strikt (monoton) äquivalent*. Stattdessen nennen wir sie *schwach (monoton) äquivalent*, wenn die Bedingungen

$$r_i < r_{i'} \implies r_i^* \leq r_{i'}^* \quad \text{und} \quad r_i^* < r_{i'}^* \implies r_i \leq r_{i'}$$

erfüllt sind. In Übereinstimmung mit dem schwächeren Kriterium (12.1.4) verwenden wir im folgenden stets schwache Äquivalenz.[5] Die Menge aller Vektoren $\mathbf{r}$, die eine zu $\mathbf{r}^*$ schwach äquivalente Rangordnung liefern, bezeichnen wir mit $\mathcal{R}^*$.

4. Jetzt soll überlegt werden, wie der bisher verwendete Regressionsansatz mit Hilfe der Matrixschreibweise formuliert werden kann (man vgl. dazu auch Abschnitt A.6). Zunächst wurde für jedes Element $\tilde{x}_{jk}$, das in einem der Teilmerkmalsräume $\tilde{\mathcal{X}}_j$ vorkommt, eine Indikatorvariable I_{jk} gebildet. Für den Regressionsansatz wurde dann die jeweils erste Indikatorvariable weggelassen und ein Interzept-Term hinzugefügt.[6] Um das mit Hilfe der Matrixschreibweise auszudrücken, muß also eine Matrix

$$\mathbf{Z} := (\mathbf{1}_q, I_{12}, \ldots, I_{1q_1}, \ldots, I_{m2}, \ldots, I_{mq_m})$$

[5] Mit der Frage, wie man strikt äquivalente Repräsentationen finden kann (wenn es sie gibt), beschäftigen wir uns in Abschnitt 12.2.5.

[6] Dies entspricht der Annahme, daß $\beta_{11} = \cdots = \beta_{m1} = 0$ ist.

gebildet werden, die aus q Zeilen und $p := 1 + (q_1 - 1) \cdots (q_m - 1)$ Spalten besteht.[7] Somit besteht der Regressionsansatz darin, folgenden Abstand zu minimieren:

$$\| \mathbf{r}^* - \mathbf{Z}\boldsymbol{\beta} \|^2 \tag{12.2.4}$$

wobei der Vektor $\boldsymbol{\beta} := (\alpha, \beta_{12}, \ldots, \beta_{1q_1}, \ldots, \beta_{m2}, \ldots, \beta_{mq_m})'$ die Parameter zusammenfaßt. Auf diese Weise konnte im vorangegangenen Abschnitt ein Parametervektor $\hat{\boldsymbol{\beta}}$ gefunden werden, so daß die Rangordnung $\hat{\mathbf{r}} := \mathbf{Z}\hat{\boldsymbol{\beta}}$ zur vorgegebenen Rangordnung $\mathbf{r}^*$ strikt äquivalent ist.

5. Bei dieser Vorgehensweise wird allerdings nicht berücksichtigt, daß der vorgegebene Rangordnungsvektor $\mathbf{r}^*$ durch beliebige andere schwach äquivalente Rangordnungen ersetzt werden könnte. Das Regressionsproblem muß deshalb in einer allgemeineren Weise formuliert werden: Gesucht ist *irgendeine* Rangordnung $\mathbf{r} \in \mathcal{R}^*$ und ein Parametervektor $\boldsymbol{\beta}$, so daß der Ausdruck $\| \mathbf{r} - \mathbf{Z}\boldsymbol{\beta} \|^2$ minimal wird. Es muß also eine Regression mit Rangordnungen durchgeführt werden; man spricht auch von einem *monotonen Regressionsproblem* bzw. von einer *monotonen Regression*. Im Unterschied zu einer gewöhnlichen Regression besteht dann auch die numerische Repräsentation für die als abhängige Variable dienende Rangordnung aus Parametern des Regressionsansatzes. Eine Lösung des so formulierten Regressionsproblems besteht also aus einem Parametervektor $\hat{\boldsymbol{\beta}}$ und einer numerischen Repräsentation $\check{\mathbf{r}} \in \mathcal{R}^*$ für die vorgegebene Rangordnung, so daß für alle möglichen Rangordnungen $\mathbf{r} \in \mathcal{R}^*$ und beliebige Parametervektoren $\boldsymbol{\beta}$ gilt: $\| \check{\mathbf{r}} - \mathbf{Z}\hat{\boldsymbol{\beta}} \|^2 \leq \| \mathbf{r} - \mathbf{Z}\boldsymbol{\beta} \|^2$.

6. Es sei angemerkt, daß man nicht immer eine Rangordnung $\check{\mathbf{r}} \in \mathcal{R}^*$ und einen Parametervektor $\hat{\boldsymbol{\beta}}$ finden kann, so daß gilt: $\check{\mathbf{r}} = \hat{\mathbf{r}} := \mathbf{Z}\hat{\boldsymbol{\beta}}$. Als ein Beispiel kann man den in Abschnitt 12.1 erläuterten zweidimensionalen Merkmalsraum von Diekmann verwenden und sich folgende Rangordnung vorstellen:

	$\tilde{x}_{21}$	$\tilde{x}_{22}$	$\tilde{x}_{23}$	$\tilde{x}_{24}$
$\tilde{x}_{11}$	1	6	7	12
$\tilde{x}_{12}$	2	5	8	11
$\tilde{x}_{13}$	3	4	9	10

Man findet keine Bewertungsfunktion ϕ_1 für den Teilmerkmalsraum $\tilde{\mathcal{X}}_1$, mit der sich eine zumindest schwach äquivalente Rangordnung herstellen ließe. Oder genauer gesagt, die einzige Bewertungsfunktion, die mit der vorgegebenen Rangordnung vereinbart werden kann, hätte für alle Elemente von $\tilde{\mathcal{X}}_1$ den gleichen Wert.

[7] Um die Notation nicht unnötig kompliziert zu machen, nehmen wir an, daß die Symbole I_{jk} wahlweise für die Indikatorvariablen und die entsprechenden Spaltenvektoren verwendet werden können; man vgl. zur Illustration die Tabelle 12.2.1.

12.2.3 Monotone Regressionsrechnung

1. In diesem Abschnitt besprechen wir, wie man monotone Regressionsprobleme praktisch lösen kann. Wir besprechen das Rechenverfahren ausführlich, weil es auch die Grundlage für einige Varianten der multidimensionalen Skalierung bildet. Ausgangspunkt ist ein linearer Regressionsansatz, der darin besteht, folgenden Abstand zu minimieren:

$$\| \mathbf{y} - (\mathbf{x}_1\beta_1 + \cdots + \mathbf{x}_m\beta_m) \|^2 = \| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \|^2$$

wobei im Unterschied zur gewöhnlichen Regressionsrechnung angenommen wird, daß es sich bei der abhängigen Variablen $\mathbf{y} = (y_1, \dots, y_n)'$ um eine Rangordnung handelt. Außerdem nehmen wir in diesem Abschnitt an, daß es sich um eine strikte Rangordnung handelt und daß man die Zeilen der Datenmatrix so umgeordnet hat, daß $y_1 < y_2 < \cdots < y_n$ gilt. Infolgedessen liefern alle Vektoren in der Menge

$$\mathcal{R}_n := \{(r_1, \dots, r_n)' \mid r_1 \leq \cdots \leq r_n\}$$

(schwach) äquivalente Repräsentationen für die abhängige Variable,[8] und das Problem besteht darin, simultan einen Vektor $\check{\mathbf{r}} \in \mathcal{R}_n$ und einen Parametervektor $\hat{\boldsymbol{\beta}}$ zu finden, so daß $\| \check{\mathbf{r}} - \mathbf{X}\hat{\boldsymbol{\beta}} \|^2 = (\check{\mathbf{r}} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\check{\mathbf{r}} - \mathbf{X}\hat{\boldsymbol{\beta}})$ einen minimalen Wert annimmt.

2. Um das Problem zu lösen, kann man ein Verfahren verwenden, daß auf iterative Weise die Regressionsrechnung mit einer weiteren Rechenoperation verknüpft. Wir nennen das Verfahren *ALS-Algorithmus* (in Anlehnung an die englischsprachige Bezeichnung „alternating least squares algorithm").

(1) $\check{\mathbf{r}} := \mathbf{y}$ (wobei $\mathbf{y}$ die anfänglich gegebene Rangordnung ist).

(2) Berechnung von $\hat{\boldsymbol{\beta}} := (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\check{\mathbf{r}}$.

(3) Bestimmung der angenäherten Rangordnung: $\hat{\mathbf{r}} := \mathbf{X}\hat{\boldsymbol{\beta}}$.

(4) Wenn $\hat{\mathbf{r}} \in \mathcal{R}_n$ ist, hat man eine perfekte Lösung gefunden.

(5) Andernfalls wird ein Vektor $\check{\mathbf{r}} \in \mathcal{R}_n$ konstruiert, so daß $\| \check{\mathbf{r}} - \hat{\mathbf{r}} \|^2$ minimal wird. Dann Fortsetzung bei (2).

Allerdings ist nicht gesichert, daß man irgendwann einen Vektor $\hat{\mathbf{r}} := \mathbf{X}\hat{\boldsymbol{\beta}}$ findet, der ein Element von $\mathcal{R}_n$ ist und somit eine gültige Repräsentation

[8] Es sei angemerkt, daß die Menge $\mathcal{R}_n$ in folgendem Sinn abgeschlossen ist: Für alle $\mathbf{v}, \mathbf{w} \in \mathcal{R}_n$ und beliebige nicht-negative Zahlen a und b ist auch $a\mathbf{v} + b\mathbf{w} \in \mathcal{R}_n$.

der Rangordnung $\mathbf{y}$ liefert. Um den Ablauf des Algorithmus beobachten und ggf. abbrechen zu können, wird deshalb eine Größe

$$S := \sqrt{\frac{\sum_{i=1}^n (\hat{r}_i - \check{r}_i)^2}{\sum_{i=1}^n (\hat{r}_i - \bar{r})^2}} \tag{12.2.5}$$

verwendet, wobei $\bar{r} := \sum_{i=1}^n \hat{r}_i / n$ ist. Diese Größe wird als *Stress* bezeichnet. Je kleiner sie ist, desto besser wird die Rangordnung $\mathbf{y}$ durch $\hat{\mathbf{r}}$ repräsentiert. Somit kann man den Algorithmus abbrechen, wenn S nicht mehr nennenswert verringert werden kann.[9]

3. Es muß noch geklärt werden, wie beim letzten Schritt (5) des Algorithmus vorgegangen werden kann. Um das Problem allgemein zu formulieren, sei $\mathbf{z}$ irgendein Vektor der Länge n. Man kann dann nach einem Vektor $p[\mathbf{z}] \in \mathcal{R}_n$ suchen, so daß gilt:

$$\text{Für alle } \mathbf{r} \in \mathcal{R}_n\text{:} \quad \| p[\mathbf{z}] - \mathbf{z} \|^2 \leq \| \mathbf{r} - \mathbf{z} \|^2 \tag{12.2.6}$$

Der Vektor $p[\mathbf{z}]$ wird auch als Projektion von $\mathbf{z}$ auf $\mathcal{R}_n$ bezeichnet und ist für jeden Vektor eindeutig bestimmt.[10] Die in Schritt (5) des ALS-Algorithmus zu lösende Aufgabe besteht also darin, die Projektion $p[\hat{\mathbf{r}}]$ von $\hat{\mathbf{r}}$ auf $\mathcal{R}_n$ zu finden.

[9] Man erkennt, daß der Stress auch von der Varianz von $\hat{\mathbf{r}}$ abhängig ist. Um degenerierte Lösungen zu vermeiden, ist es deshalb zweckmäßig, den Algorithmus etwas zu verändern und vor der Ausführung von Schritt (3) den Parametervektor $\hat{\boldsymbol{\beta}}$ durch eine lineare Transformation so zu verändern, daß $\hat{\mathbf{r}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ den gleichen Mittelwert und die gleiche Varianz wie die vorgegebene Rangordnung $\mathbf{y}$ bekommt. Wir werden dies später anhand eines Beispiels illustrieren.

[10] Denn angenommen, es gäbe außer $\mathbf{p} := p[\mathbf{z}]$ noch einen anderen Vektor $\mathbf{v} \in \mathcal{R}_n$, der den Abstand zu $\mathbf{z}$ minimiert, so daß $\| \mathbf{v} - \mathbf{z} \|^2 = \| \mathbf{p} - \mathbf{z} \|^2$ ist; und angenommen, daß $\mathbf{v} \neq \mathbf{p}$ ist. Dann kann man den Vektor $(\mathbf{p} + \mathbf{v})/2$ betrachten, der ebenfalls ein Element von $\mathcal{R}_n$ ist. Nun findet man einerseits

$$\begin{aligned} \| (\mathbf{p} + \mathbf{v})/2 - \mathbf{z} \|^2 &= \| (\mathbf{p} - \mathbf{z})/2 + (\mathbf{v} - \mathbf{z})/2 \|^2 \\ &= \frac{1}{4} \| \mathbf{p} - \mathbf{z} \|^2 + \frac{1}{4} \| \mathbf{v} - \mathbf{z} \|^2 + \frac{1}{2} (\mathbf{p} - \mathbf{z})'(\mathbf{v} - \mathbf{z}) \end{aligned}$$

Andererseits ist

$$\begin{aligned} \| \mathbf{p} - \mathbf{v} \|^2 &= \| (\mathbf{p} - \mathbf{z}) - (\mathbf{v} - \mathbf{z}) \|^2 \\ &= \| \mathbf{p} - \mathbf{z} \|^2 + \| \mathbf{v} - \mathbf{z} \|^2 - 2\,(\mathbf{p} - \mathbf{z})'(\mathbf{v} - \mathbf{z}) \end{aligned}$$

Addiert man ein Viertel der zweiten zur ersten Gleichung, findet man:

$$\| (\mathbf{p} + \mathbf{v})/2 - \mathbf{z} \|^2 + \frac{1}{4} \| \mathbf{p} - \mathbf{v} \|^2 = \frac{1}{2} \| \mathbf{p} - \mathbf{z} \|^2 + \frac{1}{2} \| \mathbf{v} - \mathbf{z} \|^2 = \| \mathbf{p} - \mathbf{z} \|^2$$

Wenn $\mathbf{v} \neq \mathbf{p}$ und infolgedessen $\| \mathbf{p} - \mathbf{v} \|^2 > 0$ wäre, hätte man somit einen Vektor $(\mathbf{p} + \mathbf{v})/2$ gefunden, der den Abstand zu $\mathbf{z}$ kleiner macht als $\mathbf{p}$, was im Widerspruch zur Definition der Projektion steht.

4. Zu überlegen ist, wie man für einen beliebigen Vektor $\mathbf{z}$ die Projektion $p[\mathbf{z}]$ finden kann. Offenbar stellt sich diese Frage nur dann, wenn $\mathbf{z}$ kein Element von $\mathcal{R}_n$ ist, denn andernfalls wäre $\mathbf{z} = p[\mathbf{z}]$. Also nehmen wir für die folgende Überlegung an, daß $\mathbf{z} \notin \mathcal{R}_n$ ist. Dann läßt sich zunächst folgende Implikation zeigen:

$$z_k \geq z_{k+1} \implies p[\mathbf{z}]_k = p[\mathbf{z}]_{k+1} \tag{12.2.7}$$

und zwar für jeden Index $k = 1, \ldots, n-1$. Denn angenommen, daß zwar $z_k \geq z_{k+1}$, aber $p[\mathbf{z}]_k \neq p[\mathbf{z}]_{k+1}$ wäre. Da $p[\mathbf{z}] \in \mathcal{R}_n$ ist, müßte infolgedessen $p[\mathbf{z}]_k < p[\mathbf{z}]_{k+1}$ sein. Dann könnte man jedoch einen neuen Vektor $\mathbf{p}^*$ bilden, dessen Komponenten folgendermaßen definiert sind:

$$p_i^* := \begin{cases} (p[\mathbf{z}]_k + p[\mathbf{z}]_{k+1})/2, & \text{wenn } i = k \text{ oder } i = k+1 \text{ ist} \\ p[\mathbf{z}]_i & \text{andernfalls} \end{cases}$$

Sicherlich wäre dann auch dieser neue Vektor $\mathbf{p}^*$ ein Element von $\mathcal{R}_n$. Nun gilt jedoch:

$$\begin{aligned} \|\mathbf{p}^* - \mathbf{z}\|^2 &= \|\mathbf{p}^* - p[\mathbf{z}] + p[\mathbf{z}] - \mathbf{z}\|^2 \\ &= \|p[\mathbf{z}] - \mathbf{z}\|^2 + \|\mathbf{p}^* - p[\mathbf{z}]\|^2 + 2\,(\mathbf{p}^* - p[\mathbf{z}])'(p[\mathbf{z}] - \mathbf{z}) \end{aligned}$$

Definiert man zur Abkürzung $a := p[\mathbf{z}]_{k+1} - p_{k+1}^* = p_k^* - p[\mathbf{z}]_k$, findet man für den zweiten Term: $\|\mathbf{p}^* - p[\mathbf{z}]\|^2 = 2\,a^2$; und für den dritten Term:

$$\begin{aligned} &2\,(\mathbf{p}^* - p[\mathbf{z}])'(p[\mathbf{z}] - \mathbf{z}) = \\ &2\,\big((p_k^* - p[\mathbf{z}]_k)(p[\mathbf{z}]_k - z_k) + (p_{k+1}^* - p[\mathbf{z}]_{k+1})(p[\mathbf{z}]_{k+1} - z_{k+1})\big) = \\ &2\,\big(a\,(p[\mathbf{z}]_k - z_k) - a\,(p[\mathbf{z}]_{k+1} - z_{k+1})\big) = \\ &2\,a\,\big((z_{k+1} - z_k) - (p[\mathbf{z}]_{k+1} - p[\mathbf{z}]_k)\big) = \\ &2\,a\,\big((z_{k+1} - z_k) - 2\,a\big) = \\ &2\,a\,(z_{k+1} - z_k) - 4\,a^2 \end{aligned}$$

Somit erhält erhält man:

$$\begin{aligned} \|\mathbf{p}^* - \mathbf{z}\|^2 &= \|p[\mathbf{z}] - \mathbf{z}\|^2 + 2\,a^2 + 2\,a\,(z_{k+1} - z_k) - 4\,a^2 \\ &= \|p[\mathbf{z}] - \mathbf{z}\|^2 - 2\,a^2 + 2\,a\,(z_{k+1} - z_k) \end{aligned}$$

Nun ist jedoch nach Voraussetzung $z_{k+1} - z_k \leq 0$; und wenn die Annahme $p[\mathbf{z}]_k < p[\mathbf{z}]_{k+1}$ richtig wäre, müßte $a > 0$ sein. Dann hätte man jedoch einen Vektor $\mathbf{p}^*$ gefunden, der den Abstand zu $\mathbf{z}$ kleiner macht als die Projektion $p[\mathbf{z}]$. Dies stünde jedoch im Widerspruch zur Definition der Projektion, so daß die Annahme verworfen werden muß und man schließen kann, daß $p[\mathbf{z}]_k = p[\mathbf{z}]_{k+1}$ ist.

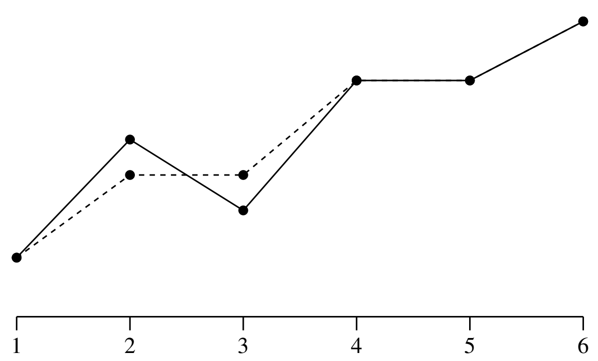


Abb. 12.2.1 Durchgezogene Linie: $\mathbf{z} = (1, 2, 1.4, 2.5, 2.5, 3)'$; gestrichelte Linie: $m[\mathbf{z}]$.

5. Die Implikation (12.2.7) bildet den Ausgangspunkt, um eine Methode zu gewinnen, mit der sich $p[\mathbf{z}]$ berechnen läßt. Um sie darzustellen, definieren wir einen Operator m, der aus dem Vektor $\mathbf{z}$ einen neuen Vektor $m[\mathbf{z}]$ macht. Und zwar sei $m[\mathbf{z}] := \mathbf{z}$, wenn $\mathbf{z} \in \mathcal{R}_n$ ist; andernfalls werden die Komponenten von $m[\mathbf{z}]$ durch

$$m[\mathbf{z}]_i := \begin{cases} (z_k + z_{k+1})/2, & \text{wenn } i = k \text{ oder } i = k+1 \text{ ist} \\ z_i & \text{andernfalls} \end{cases}$$

definiert, wobei k derjenige Index $(1 \leq k \leq n-1)$ ist, bei dem die Differenz $z_k - z_{k+1}$ ihren maximalen Wert annimmt. — Zur Illustration dieser Definition betrachten wir den Vektor $\mathbf{z} = (1, 2, 1.4, 2.5, 2.5, 3)'$, der offenbar kein Element von $\mathcal{R}_6$ ist. Wendet man den Operator m an, entsteht der neue Vektor $m[\mathbf{z}] = (1, 1.7, 1.7, 2.5, 2.5, 3)'$, der in Abbildung 12.2.1 dargestellt ist.

6. Mit Hilfe des Operators m kann nun eine praktische Methode zur Berechnung der Projektion $p[\mathbf{z}]$ gewonnen werden. Diese Methode beruht auf der Gleichung

$$p[\mathbf{z}] = p[m[\mathbf{z}]] \tag{12.2.8}$$

die zeigt, daß die Projektion von $\mathbf{z}$ mit derjenigen von $m[\mathbf{z}]$ übereinstimmt. Die Gleichung folgt daraus, daß für beliebige Vektoren $\mathbf{r} \in \mathcal{R}_n$ die Ungleichung

$$\|p[m[\mathbf{z}]] - \mathbf{z}\|^2 \leq \|\mathbf{r} - \mathbf{z}\|^2 \tag{12.2.9}$$

erfüllt ist. Davon kann man sich auf folgende Weise überzeugen. Zunächst gilt sicherlich für jeden Vektor $\mathbf{r} \in \mathcal{R}_n$:

$$\|\mathbf{r}-\mathbf{z}\|^2 = \sum_{i=1}^{n}(r_i - z_i)^2 = \sum_{i=1}^{n}\Big((r_i - m[\mathbf{z}]_i) + (m[\mathbf{z}]_i - z_i)\Big)^2 =$$

$$\sum_{i=1}^{n}(r_i - m[\mathbf{z}]_i)^2 + \sum_{i=1}^{n}(m[\mathbf{z}]_i - z_i)^2 +$$

$$2\Big((r_k - m[\mathbf{z}]_k)(m[\mathbf{z}]_k - z_k) + (r_{k+1} - m[\mathbf{z}]_{k+1})(m[\mathbf{z}]_{k+1} - z_{k+1})\Big)$$

Die Aufgabe, $\|\mathbf{r}-\mathbf{z}\|^2$ zu minimieren, ist somit äquivalent zur Aufgabe, den Ausdruck

$$\|\mathbf{r}-m[\mathbf{z}]\|^2 + \|m[\mathbf{z}]-\mathbf{z}\|^2 +$$

$$2\Big((r_k - m[\mathbf{z}]_k)(m[\mathbf{z}]_k - z_k) + (r_{k+1} - m[\mathbf{z}]_{k+1})(m[\mathbf{z}]_{k+1} - z_{k+1})\Big)$$

zu minimieren, wobei $\mathbf{r}$ irgendein Vektor in $\mathcal{R}_n$ sein kann. Von den drei Termen hängen jedoch nur der erste und der dritte von $\mathbf{r}$ ab. Außerdem folgt aus der Konstruktion von $m[\mathbf{z}]$, daß der dritte Term für alle $\mathbf{r} \in \mathcal{R}_n$ größer oder gleich Null ist. Denn definiert man $a := m[\mathbf{z}]_k - z_k$ und berücksichtigt, daß $m[\mathbf{z}]_k = m[\mathbf{z}]_{k+1}$ ist, kann man den dritten Term auch folgendermaßen schreiben:

$$2\left((r_k - m[\mathbf{z}]_k)\,a - (r_{k+1} - m[\mathbf{z}]_k)\,a\right) = 2\,(r_k - r_{k+1})\,a$$

Da $r_k \le r_{k+1}$ und $a < 0$ ist, kann dieser Term nicht negativ werden. Er wird sogar Null, wenn man anstelle von $\mathbf{r}$ die Projektion $p[m[\mathbf{r}]]$ einsetzt, denn aus (12.2.7) folgt, daß $p[m[\mathbf{z}]]_k = p[m[\mathbf{z}]]_{k+1}$ ist. Somit minimiert $p[m[\mathbf{z}]]$ den Ausdruck $\|\mathbf{r}-\mathbf{z}\|^2$, wodurch die Richtigkeit von (12.2.9) bestätigt wird.

7. Gestützt auf die Gleichung (12.2.8) kann man nun für jeden vorgegebenen Vektor $\mathbf{z}$ seine Projektion $p[\mathbf{z}]$ auf folgende Weise finden. Man bildet iterativ eine Folge von Vektoren

$$\mathbf{z} \longrightarrow m[\mathbf{z}] \longrightarrow m[m[\mathbf{z}]] \longrightarrow m[m[m[\mathbf{z}]]] \longrightarrow \cdots$$

Dann gibt es zwei Möglichkeiten. Entweder man findet nach endlich vielen Schritten einen Vektor, der ein Element von $\mathcal{R}_n$ ist (dies ist z.B. in Abb. 12.2.1 bereits nach einer Iteration der Fall); oder die Folge der Vektoren nähert sich immer mehr der gesuchten Projektion $p[\mathbf{z}]$, so daß man die Iterationen abbrechen kann, wenn eine weitere Anwendung des Operators m die bisher berechnete Näherung nur noch unwesentlich verändert.[11]

[11]Von der Konvergenz des Verfahrens kann man sich auf folgende Weise überzeugen. Es sei $m^0[\mathbf{z}] := \mathbf{z}$, $m^1[\mathbf{z}] := m[\mathbf{z}]$, $m^2[\mathbf{z}] := m[m[\mathbf{z}]]$, und ganz allgemein: $m^j[\mathbf{z}] :=$

8. Um den ALS-Algorithmus zu illustrieren, verwenden wir die folgenden einfach überschaubaren Daten:

$$\mathbf{y} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{pmatrix} \quad \text{und} \quad \mathbf{X} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \\ 1 & 4 \\ 1 & 7 \\ 1 & 6 \end{pmatrix}$$

Zunächst liefert eine einfache Regressionsrechnung das Ergebnis

$$\begin{pmatrix} -0.486 \\ 0.886 \end{pmatrix} = \hat{\boldsymbol{\beta}}_{(1)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{y}$$

wobei der Index (1) andeuten soll, daß es sich um die erste Iteration handelt. Daraus kann man den Vektor $\hat{\mathbf{r}}_{(1)} = \mathbf{X}\hat{\boldsymbol{\beta}}_{(1)}$ bilden (s. Tabelle 12.2.2). Es gibt folgende Mittelwerte und Standardabweichungen:

$$\mathrm{M}(\mathbf{y}) = 3.5,\quad \mathrm{S}(\mathbf{y}) = 1.708,\quad \mathrm{M}(\hat{\mathbf{r}}_{(1)}) = 3.5,\quad \mathrm{S}(\hat{\mathbf{r}}_{(1)}) = 1.513$$

$m[m^{j-1}[\mathbf{z}]]$. Zunächst findet man:

$$\|p[\mathbf{z}]-\mathbf{z}\|^2 = \|(p[\mathbf{z}]-m[\mathbf{z}]) + (m[\mathbf{z}]-\mathbf{z})\|^2$$

$$= \|p[\mathbf{z}]-m[\mathbf{z}]\|^2 + \|m[\mathbf{z}]-\mathbf{z}\|^2 + 2\,(p[\mathbf{z}]-m[\mathbf{z}])'(m[\mathbf{z}]-\mathbf{z})$$

Nun ist jedoch $(p[\mathbf{z}]-m[\mathbf{z}])'(m[\mathbf{z}]-\mathbf{z}) = 0$, denn $m[\mathbf{z}]-\mathbf{z}$ hat die Gestalt

$$m[\mathbf{z}]-\mathbf{z} = (0,\ldots,0,a,-a,0,\ldots,0)'$$

wobei a an der k-ten und $-a$ an der $k+1$-ten Stelle steht und $a := m[\mathbf{z}]_k - z_k$ ist; außerdem ist $p[\mathbf{z}]_k = p[\mathbf{z}]_{k+1}$ und $m[\mathbf{z}]_k = m[\mathbf{z}]_{k+1}$. Also gilt:

$$\|p[\mathbf{z}]-\mathbf{z}\|^2 = \|p[\mathbf{z}]-m[\mathbf{z}]\|^2 + \|m[\mathbf{z}]-\mathbf{z}\|^2$$

Daraus folgt durch Induktion und unter Berücksichtigung von $p[m[\mathbf{z}]] = p[\mathbf{z}]$:

$$\|p[\mathbf{z}]-\mathbf{z}\|^2 = \|p[\mathbf{z}]-m^j[\mathbf{z}]\|^2 + \sum_{i=1}^{j}\|m^i[\mathbf{z}]-m^{i-1}[\mathbf{z}]\|^2$$

Die Reihe auf der rechten Seite konvergiert, da sie durch $\|p[\mathbf{z}]-\mathbf{z}\|^2$ beschränkt wird und ihre Glieder nicht negativ sind. Also konvergiert auch die Folge $m^j[\mathbf{z}]$, und es sei $\mathbf{v}$ ihr Grenzwert. Dieser Grenzwert muß jedoch ein Element von $\mathcal{R}_n$ sein. Denn angenommen, daß $\mathbf{v} \notin \mathcal{R}_n$ wäre. Dann gäbe es einen Index k, so daß $v_k > v_{k+1}$ und infolgedessen $m[\mathbf{v}] \neq \mathbf{v}$ wäre. Das aber stünde im Widerspruch dazu, daß $\mathbf{v}$ Grenzwert der Folge $m^j[\mathbf{z}]$ ist. Also kann man schließen, daß $\mathbf{v} \in \mathcal{R}_n$ ist. Um schließlich einzusehen, daß $\mathbf{v} = p[\mathbf{z}]$ ist, kann man die Gleichung

$$\|p[\mathbf{z}]-m^j[\mathbf{z}]\|^2 = \|p[m^j[\mathbf{z}]]-m^j[\mathbf{z}]\|^2 \qquad (12.2.10)$$

betrachten. Wenn $j \to \infty$, erkennt man: $m^j[\mathbf{z}] \to \mathbf{v}$ und $p[m^j[\mathbf{z}]] \to p[\mathbf{v}] = \mathbf{v}$. Also wird die rechte Seite und infolgedessen auch die linke Seite der Gleichung (12.2.10) beliebig klein.

Tabelle 12.2.2 Ablauf einer monotonen Regression.

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{y}$	$\hat{\mathbf{r}}_{(1)}$	$\hat{\mathbf{r}}^*_{(1)}$	$\check{\mathbf{r}}_{(1)}$	$\hat{\mathbf{r}}_{(2)}$	$\hat{\mathbf{r}}^*_{(2)}$	$\check{\mathbf{r}}_{(2)}$
1	2	1	1.286	1	1.0	1.143	1	1.0
1	3	2	2.171	2	2.0	2.086	2	2.0
1	5	3	3.943	4	3.5	3.971	4	3.5
1	4	4	3.057	3	3.5	3.029	3	3.5
1	7	5	5.714	6	5.5	5.857	6	5.5
1	6	6	4.829	5	5.5	4.914	5	5.5

Durch eine lineare Transformation des Parametervektors

$$\hat{\boldsymbol{\beta}}_{(1)} = \begin{pmatrix} -0.486 \\ 0.886 \end{pmatrix} \longrightarrow \hat{\boldsymbol{\beta}}^*_{(1)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

kann man jedoch erreichen, daß $\hat{\mathbf{r}}^*_{(1)} = \mathbf{X}\hat{\boldsymbol{\beta}}^*_{(1)}$ den gleichen Mittelwert und die gleiche Standardabweichung wie $\mathbf{y}$ bekommt.[12] Die Werte sind in der Tabelle angegeben, und man erkennt, daß in diesem Beispiel eine einfache Permutation von $\mathbf{y}$ entsteht. Infolgedessen ist jedoch $\hat{\mathbf{r}}^*_{(1)}$ keine zulässige Repräsentation von $\mathbf{y}$ und man muß nach einem neuen Vektor $\check{\mathbf{r}}_{(1)} \in \mathcal{R}_6$ suchen, so daß

$$\| \check{\mathbf{r}}_{(1)} - \hat{\mathbf{r}}^*_{(1)} \|^2$$

minimal wird. Das kann mit der oben beschriebenen Projektionsmethode erreicht werden, wie die in der Tabelle angegebenen Werte illustrieren. Somit kann man den ersten Iterationsschritt abschließen, indem man zum Vergleich von $\check{\mathbf{r}}_{(1)}$ und $\hat{\mathbf{r}}^*_{(1)}$ den Stress berechnet, der in unserem Beispiel den Wert $S_{(1)} = 0.239$ hat. Dann kann der nächste Iterationsschritt beginnen. Man findet den neuen Parametervektor

$$\begin{pmatrix} -0.743 \\ 0.943 \end{pmatrix} = \hat{\boldsymbol{\beta}}_{(2)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\check{\mathbf{r}}_{(1)}$$

Führt man die bereits beschriebene Standardisierung durch, erhält man jedoch wiederum $\hat{\boldsymbol{\beta}}^*_{(2)} = (-1, 1)'$ und somit auch die gleichen Vektoren $\hat{\mathbf{r}}_{(2)}$ und $\check{\mathbf{r}}_{(2)}$, so daß sich der Stress nicht verändert. Somit kann man die Berechnungen bereits nach der zweiten Iteration abschließen.

12.2.4 Berücksichtigung von Bindungen

1. Im vorangegangenen Abschnitt wurde angenommen, daß die abhängige Variable $\mathbf{y}$ als eine strikte Rangordnung vorgegeben ist. Bei praktischen

[12]Man multipliziert den Parametervektor zunächst mit $\mathrm{S}(\mathbf{y})/\mathrm{S}(\hat{\mathbf{r}}_{(1)})$; in einem zweiten Schritt wird dann der Parameter für den Interzept modifiziert.

Anwendungen treten dagegen oft Indifferenzen auf. Man kann dann zwar immer noch eine schwache lineare Ordnung bilden, aber es gibt Bindungsblöcke, in denen es keine eindeutige Ordnung gibt (vgl. Abschnitt 11.1). Als Beispiel betrachten wir den Vektor

$$\mathbf{y} = (1,2,2,3,3,3,4,5,5,6,6,6)' \qquad (12.2.11)$$

Es gibt folgende Bindungsblöcke:

$$\begin{array}{lll} T_1 = \{1\} & T_2 = T_3 = \{2,3\} & T_4 = T_5 = T_6 = \{4,5,6\} \\ T_7 = \{7\} & T_8 = T_9 = \{8,9\} & T_{10} = T_{11} = T_{12} = \{10,11,12\} \end{array}$$

2. Zu überlegen ist, wie der ALS-Algorithmus sinnvoll durchgeführt werden kann, wenn es Bindungen gibt. Zwei verbreitete Varianten wurden von J. B. Kruskal vorgeschlagen.[13] Kruskal nannte sie den *primary* und den *secondary approach*; wir sprechen im folgenden von der P- bzw. S-Methode. Die Überlegungen betreffen den letzten Schritt (5) im ALS-Algorithmus, in dem nach einer geeigneten Projektion von $\hat{\mathbf{r}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ gesucht wird. Bei der im vorangegangenen Abschnitt besprochenen Projektionsmethode wird die Projektion $p[\hat{\mathbf{r}}]$ verwendet. Wenn es Bindungen gibt, führt diese Methode jedoch nicht zu einem eindeutigen Ergebnis. Als Beispiel beziehen wir uns auf den in (12.2.11) angegebenen Vektor $\mathbf{y}$ und nehmen an, daß zu Beginn von Schritt (5) des ALS-Algorithmus der folgende Vektor gegeben ist:[14]

$$\mathbf{z} = (1,3,2,3,4,5,4,5,6,5,4,4.5)'$$

Berechnet man mit der im vorangegangenen Abschnitt besprochenen Methode die Projektion $p[\mathbf{z}]$, findet man:

$$p[\mathbf{z}] = (1, 2.5, 2.5, 3, 4, 4.5, 4.5, 4.9, 4.9, 4.9, 4.9, 4.9)'$$

Man erkennt, auch anhand von Abbildung 12.2.2, daß die Komponenten von $p[\mathbf{z}]$ in den Bindungsblöcken nicht unbedingt gleiche Werte aufweisen.

3. Hier knüpft die S-Methode an. Die Idee besteht darin, die Menge der Vektoren, die für eine Projektion in Frage kommen, so einzuschränken, daß die Bindungen erhalten bleiben. Man bezieht sich also nicht auf die bisher verwendete Menge $\mathcal{R}_n$, sondern auf eine zusätzlich eingeschränkte Menge

$$\mathcal{R}^{**}_n[\mathbf{y}] := \{(r_1,\ldots,r_n)' \in \mathcal{R}_n \mid \text{gleiche Bindungen wie bei } \mathbf{y}\,\}$$

Gesucht ist jetzt eine Projektion, die den Abstand $\| \mathbf{r} - \mathbf{z} \|^2$ minimal macht, wobei $\mathbf{r}$ jedoch ein Element von $\mathcal{R}^{**}_n[\mathbf{y}]$ sein muß. Diese Projektion wird im folgenden $p^{**}[\mathbf{z}]$ genannt. Die Berechnung erfolgt in drei Schritten.

[13]Kruskal 1964a, S. 22, und 1964b, S. 128. Man vgl. auch de Leeuw 1977.

[14]Wir nennen diesen Vektor $\mathbf{z}$ anstelle von $\hat{\mathbf{r}}$, um die Notation zu vereinfachen.

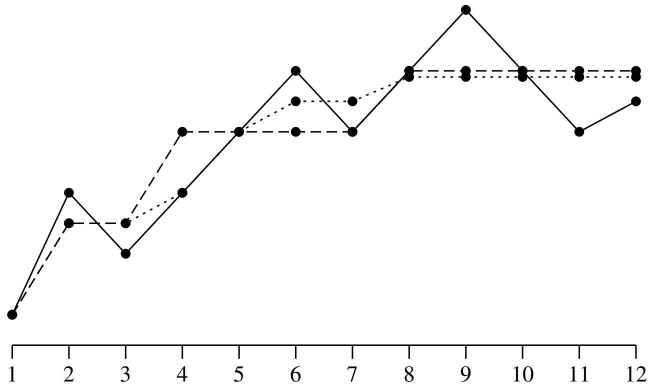


Abb. 12.2.2 Durchgezogene Linie: $\mathbf{z} = (1,3,2,3,4,5,4,5,6,5,4,4.5)'$; gestrichelte Linie: $p^{**}[\mathbf{z}]$; gepunktete Linie: $p[\mathbf{z}]$.

In einem ersten Schritt wird ein Vektor der Länge m (= Anzahl der verschiedenen Bindungsblöcke von $\mathbf{y}$) erzeugt, der für jeden Bindungsblock den Mittelwert der $\mathbf{z}$-Werte innerhalb des Blocks aufweist; wir nennen diesen Vektor im folgenden $\mathbf{z}^{**}$. Dann wird auf diesen verkürzten Vektor die im vorangegangenen Abschnitt besprochene Projektionsmethode angewendet, also $p[\mathbf{z}^{**}]$ gebildet. Und schließlich wird dieser Vektor wieder expandiert, um die Projektion $p^{**}[\mathbf{z}]$ zu bilden.

4. Unser bisheriges Beispiel kann verwendet werden, um die Vorgehensweise zu illustrieren. Zunächst berechnet man

$$\mathbf{z}^{**} = (1,2.5,4,4,5.5,4.5)'$$

Es gibt 6 Elemente entsprechend der Anzahl der Bindungsblöcke. Wendet man die im vorangegangenen Abschnitt besprochene Projektionsmethode an, findet man $p[\mathbf{z}^{**}] = (1,2.5,4,4,5,5)'$. Expandiert man diesen Vektor, entsteht die gesuchte Projektion

$$p^{**}[\mathbf{z}] = (1,2.5,2.5,4,4,4,4,5,5,5,5,5)'$$

Die gestrichelte Linie in Abbildung 12.2.2 zeigt diesen Vektor. Natürlich ist in vielen Fällen $\| p^{**}[\mathbf{z}] - \mathbf{z} \|^2 > \| p[\mathbf{z}] - \mathbf{z} \|^2$. In unserem Beispiel ist $\| p^{**}[\mathbf{z}] - \mathbf{z} \|^2 = 4.75$ und $\| p[\mathbf{z}] - \mathbf{z} \|^2 = 3.2$.

5. Kruskals *primary approach*, den wir P-Methode nennen, beruht auf einer anderen Überlegung. Es wird nicht gefordert, daß die durch $\mathbf{y}$ gegebenen Bindungen erhalten bleiben, sondern angenommen, daß man die Reihenfolge der Rangzahlen innerhalb jedes Bindungsblocks beliebig wählen darf. Oder anders formuliert: Für die Projektion sind alle Elemente der Menge

$$\mathcal{R}_n^*[\mathbf{y}] := \{(r_1,\ldots,r_n)' \mid \text{wenn } y_j < y_k, \text{ dann } r_j \leq r_k\}$$

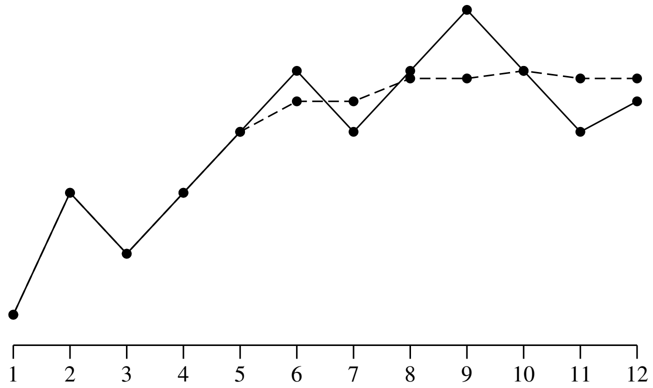


Abb. 12.2.3 Durchgezogene Linie: $\mathbf{z} = (1,3,2,3,4,5,4,5,6,5,4,4.5)'$; gestrichelte Linie: $p^{*}[\mathbf{z}]$.

zugelassen. Die Projektion, die den Abstand zwischen $\mathbf{z}$ und dieser Menge minimiert, nennen wir im folgenden $p^*[\mathbf{z}]$. Offenbar sind in diesem Fall die Restriktionen schwächer, denn im Unterschied zu $\mathcal{R}^{**}[\mathbf{y}]$ ist $\mathcal{R}^{*}[\mathbf{y}]$ keine Teilmenge von $\mathcal{R}_n$.

6. Die praktische Berechnung von $p^*[\mathbf{z}]$ erfolgt wiederum in drei Schritten. In einem ersten Schritt werden die Elemente von $\mathbf{z}$ innerhalb jedes Bindungsblocks (schwach) monoton steigend geordnet; den resultierenden Vektor nennen wir $\mathbf{z}^*$. Dann wird die Standardprojektion $p[\mathbf{z}^*]$ gebildet. Und schließlich werden die Elemente von $p[\mathbf{z}^*]$ wieder in die ursprüngliche Reihenfolge gebracht. Anhand unseres Beispiels sehen diese drei Schritte folgendermaßen aus. Im ersten Schritt entsteht

$$\mathbf{z}^* = (1,2,3,3,4,5,4,5,6,4,4.5,5)'$$

Daraus findet man die Standardprojektion

$$p[\mathbf{z}^*] = (1,2,3,3,4,4.5,4.5,4.875,4.875,4.875,4.875,5)'$$

Und nach dem Rücksortieren entsteht

$$p^*[\mathbf{z}] = (1,3,2,3,4,4.5,4.5,4.875,4.875,5,4.875,4.875)'$$

Abbildung 12.2.3 vergleicht diese Projektion mit $\mathbf{z}$. Der Abstand wird bei dieser Variante natürlich besonders klein, nämlich $\| p^*[\mathbf{z}] - \mathbf{z} \|^2 = 2.69$.

7. Abhängig davon, wie mit Bindungen umgegangen wird, entstehen im allgemeinen auch unterschiedliche Bewertungsfunktionen. Um das zu zeigen, verwenden wir ein weiteres Beispiel aus der Arbeit von Diekmann (1982). Die Befragungsperson Nr. 6 hat folgende Rangordnung r^* für den in Abschnitt 12.1 beschriebenen zweidimensionalen Merkmalsraum $\tilde{\mathcal{X}}_1 \times \tilde{\mathcal{X}}_2$ angegeben:

	$\tilde{x}_{21}$	$\tilde{x}_{22}$	$\tilde{x}_{23}$	$\tilde{x}_{24}$
$\tilde{x}_{11}$	1	3	4	6
$\tilde{x}_{12}$	2	6	8	10
$\tilde{x}_{13}$	6	9	11	12

Verwendet man den ALS-Algorithmus mit Kruskals P-Methode, erhält man folgendes Ergebnis:

	$\phi_2(\tilde{x}_{21})$ 0.00	$\phi_2(\tilde{x}_{22})$ 3.04	$\phi_2(\tilde{x}_{23})$ 4.72	$\phi_2(\tilde{x}_{24})$ 6.41
$\phi_1(\tilde{x}_{11}) = 0.00$	0.00	3.04	4.72	6.41
$\phi_1(\tilde{x}_{12}) = 3.04$	3.04	6.08	7.76	9.45
$\phi_1(\tilde{x}_{13}) = 6.07$	6.07	9.11	10.79	12.48

Man erkennt, daß die in r^* vorhandenen Bindungen nicht erhalten bleiben und außerdem eine neue Bindung erzeugt wird: $\phi(\tilde{x}_{11}, \tilde{x}_{22}) = \phi(\tilde{x}_{12}, \tilde{x}_{21})$. Verwendet man andererseits Kruskals S-Methode, entsteht:

	$\phi_2(\tilde{x}_{21})$ 0.00	$\phi_2(\tilde{x}_{22})$ 3.11	$\phi_2(\tilde{x}_{23})$ 4.69	$\phi_2(\tilde{x}_{24})$ 6.22
$\phi_1(\tilde{x}_{11}) = 0.00$	0.00	3.11	4.69	6.22
$\phi_1(\tilde{x}_{12}) = 3.11$	3.11	6.22	7.80	9.33
$\phi_1(\tilde{x}_{13}) = 6.22$	6.22	9.33	10.91	12.44

In diesem Fall bleiben zwar die in r^* vorhandenen Bindungen erhalten, es werden jedoch zwei neue Bindungen erzeugt.

12.2.5 Strikte monotone Äquivalenz

1. Eine ergänzende Überlegung betrifft Rangordnungen r^*, für die es Bewertungsfunktionen $\phi_1, \ldots, \phi_m$ gibt, so daß ein daraus gebildeter additiver Index $\phi = \Sigma\phi_j$ eine strikt äquivalente Rangordnung liefert. Eine Schwäche des ALS-Algorithmus besteht darin, daß man solche Bewertungsfunktionen nicht unbedingt findet. Als Beispiel betrachten wir folgende Rangordnung, die von Diekmanns Befragungsperson Nr. 5 angegeben worden ist.

	$\tilde{x}_{21}$	$\tilde{x}_{22}$	$\tilde{x}_{23}$	$\tilde{x}_{24}$
$\tilde{x}_{11}$	1	3	6	7
$\tilde{x}_{12}$	2	5	8	9
$\tilde{x}_{13}$	4	10	11	12

Der ALS-Algorithmus liefert folgendes Ergebnis:

	$\phi_2(\tilde{x}_{21})$ 0.00	$\phi_2(\tilde{x}_{22})$ 3.75	$\phi_2(\tilde{x}_{23})$ 6.10	$\phi_2(\tilde{x}_{24})$ 7.09
$\phi_1(\tilde{x}_{11}) = 0.00$	0.00	3.75	6.10	7.09
$\phi_1(\tilde{x}_{12}) = 1.76$	1.76	5.51	7.86	8.85
$\phi_1(\tilde{x}_{13}) = 5.10$	5.10	8.85	11.20	12.19

Der Algorithmus hat also nur eine schwach äquivalente Rangordnung gefunden. Tatsächlich gibt es jedoch in diesem Beispiel auch Bewertungsfunktionen, die zu einer strikt äquivalenten Rangordnung führen, nämlich

	$\phi_2(\tilde{x}_{21})$ 0	$\phi_2(\tilde{x}_{22})$ 6	$\phi_2(\tilde{x}_{23})$ 9	$\phi_2(\tilde{x}_{24})$ 10
$\phi_1(\tilde{x}_{11}) = 0$	0	6	9	10
$\phi_1(\tilde{x}_{12}) = 2$	2	8	11	12
$\phi_1(\tilde{x}_{13}) = 7$	7	13	16	17

Der Grund für diese Eigenschaft des ALS-Algorithmus liegt darin, daß er nur nach schwach äquivalenten Lösungen sucht und das Ergebnis davon abhängt, mit welchen Startwerten die iterative Prozedur begonnen wird. Wir besprechen deshalb im folgenden einen alternativen Algorithmus, mit dem man – wenn es sie gibt – Bewertungsfunktionen finden kann, die eine strikt äquivalente Rangordnung erzeugen.

2. Die Notation knüpft an die Ausführungen in Abschnitt 12.2.2 an. Ausgangspunkt ist ein m-dimensionaler Merkmalsraum $\tilde{\mathcal{X}}$, der aus den Teilmerkmalsräumen $\tilde{\mathcal{X}}_1, \ldots, \tilde{\mathcal{X}}_m$ besteht, also

$$\tilde{\mathcal{X}} = \tilde{\mathcal{X}}_1, \ldots, \tilde{\mathcal{X}}_m$$

Die Anzahl der Merkmalswerte im Teilmerkmalsraum $\tilde{\mathcal{X}}_j$ sei q_j. Somit gibt es im kombinierten Merkmalsraum $\tilde{\mathcal{X}}$ insgesamt

$$q := q_1 \cdots q_m$$

unterschiedliche Merkmalskombinationen. Nun kann man jede dieser Merkmalskombinationen durch einen Zeilenvektor

$$(a_{11}, \ldots, a_{1q_1}, \ldots, a_{m1}, \ldots, a_{mq_m})$$

darstellen, so daß es in jedem der m Blöcke genau eine 1 und ansonsten nur Nullen gibt. Diese Zeilenvektoren kann man zu einer Matrix $\mathbf{A}$ zusammenfassen, die q Zeilen und $p := q_1 + \cdots + q_m$ Spalten hat. Zur Illustration verwenden wir das zu Beginn dieses Abschnitts angegebene Beispiel:

$$
\mathbf{A} \;=\; \begin{pmatrix}
1&0&0&1&0&0&0\\
1&0&0&0&1&0&0\\
1&0&0&0&0&1&0\\
1&0&0&0&0&0&1\\
0&1&0&1&0&0&0\\
0&1&0&0&1&0&0\\
0&1&0&0&0&1&0\\
0&1&0&0&0&0&1\\
0&0&1&1&0&0&0\\
0&0&1&0&1&0&0\\
0&0&1&0&0&1&0\\
0&0&1&0&0&0&1
\end{pmatrix}
\qquad
\begin{pmatrix}
\tilde{x}_{11} & \tilde{x}_{21}\\
\tilde{x}_{11} & \tilde{x}_{22}\\
\tilde{x}_{11} & \tilde{x}_{23}\\
\tilde{x}_{11} & \tilde{x}_{24}\\
\tilde{x}_{12} & \tilde{x}_{21}\\
\tilde{x}_{12} & \tilde{x}_{22}\\
\tilde{x}_{12} & \tilde{x}_{23}\\
\tilde{x}_{12} & \tilde{x}_{24}\\
\tilde{x}_{13} & \tilde{x}_{21}\\
\tilde{x}_{13} & \tilde{x}_{22}\\
\tilde{x}_{13} & \tilde{x}_{23}\\
\tilde{x}_{13} & \tilde{x}_{24}
\end{pmatrix}
$$

Jeder Zeile von **A** entspricht genau eine Merkmalskombination, die wir auf der rechten Seite noch einmal in der bisher verwendeten Form dargestellt haben. Die Matrix **A** wird eine *vollständige Design-Matrix* genannt.

3. Die vorgegebene Rangordnung r^* bezieht sich auf Merkmalskombinationen, so daß man auch sagen kann, daß durch sie eine Rangordnung für die Zeilen der Matrix **A** hergestellt wird. Also kann man die Zeilen von **A** so umsortieren, daß sie der vorgegebenen Rangordnung entsprechen. D.h. man kann einen Vektor **y** der Länge q finden und eine aus **A** durch Umsortieren entstehende Matrix **B**, so daß gilt: Wenn $y_i = y_{i'}$ ist, dann besteht zwischen den Zeilen $\mathbf{b}_i$ und $\mathbf{b}_{i'}$ Indifferenz; und wenn $\mathbf{y}_i < \mathbf{y}_{i'}$ ist, dann wird die Zeile $\mathbf{b}_{i'}$ der Zeile $\mathbf{b}_i$ vorgezogen. Wendet man diese Operationen in unserem Beispiel an, entstehen ein Vektor **y** und eine Matrix **B**, die folgendermaßen aussehen:

$$
\mathbf{y} \;=\; \begin{pmatrix} 1\\2\\3\\4\\5\\6\\7\\8\\9\\10\\11\\12 \end{pmatrix}
\quad \text{und} \quad
\mathbf{B} \;=\; \begin{pmatrix}
1&0&0&1&0&0&0\\
0&1&0&1&0&0&0\\
1&0&0&0&1&0&0\\
0&0&1&1&0&0&0\\
0&1&0&0&1&0&0\\
1&0&0&0&0&1&0\\
1&0&0&0&0&0&1\\
0&1&0&0&0&1&0\\
0&1&0&0&0&0&1\\
0&0&1&0&1&0&0\\
0&0&1&0&0&1&0\\
0&0&1&0&0&0&1
\end{pmatrix}
$$

4. Gesucht sind m Bewertungsfunktionen $\phi_1, \ldots, \phi_m$, so daß aus deren Addition eine zur vorgegebenen Rangordnung strikt äquivalente Rangordnung entsteht. Die Werte dieser Bewertungsfunktionen können in Gestalt

eines Vektors

$$(\phi_1(\tilde{x}_{11}), \ldots, \phi_1(\tilde{x}_{1q_1}), \ldots, \phi_m(\tilde{x}_{m1}), \ldots, \phi_m(\tilde{x}_{mq_m}))'$$

zusammengefaßt werden, der $p := q_1 + \cdots + q_m$ Komponenten hat. Umgekehrt kann man auch sagen, daß jeder Vektor

$$\mathbf{d} \;:=\; (d_1, \ldots, d_p)'$$

als Darstellung einer Menge von m Bewertungsfunktionen aufgefaßt werden kann. Somit kann man sich für das Rechenverfahren darauf konzentrieren, einen Vektor **d** (wir nennen ihn einen *Bewertungsvektor*) zu finden, der eine zu **y** strikt äquivalente Rangordnung erzeugt. Die dafür erforderlichen Bedingungen kann man leicht angeben. Zunächst ist klar, daß man die durch einen Bewertungsvektor **d** erzeugte Rangordnung durch $\mathbf{z} := \mathbf{Bd}$ findet. Offenbar sind **z** und **y** genau dann strikt äquivalent, wenn für $i = 2, \ldots, q$ gilt:

$$z_i - z_{i-1} \begin{cases} = 0 & \text{wenn } y_i = y_{i-1} \\ \geq \epsilon & \text{wenn } y_i > y_{i-1} \end{cases} \tag{12.2.12}$$

Hierbei ist ϵ irgendeine positive Zahl, auf deren genauen Wert es jedoch nicht ankommt; und da Bewertungsfunktionen beliebigen linearen Transformationen unterzogen werden können, kann man insbesondere $\epsilon = 1$ verwenden.

5. Somit kommt man zu folgender Problemformulierung: Gesucht ist ein Bewertungsvektor **d**, so daß $\mathbf{z} := \mathbf{Bd}$ der Bedingung (12.2.12) genügt. Zur Lösung kann man das Verfahren der linearen Programmierung verwenden.[15] Zu diesem Zweck wird eine neue Matrix **U** gebildet, die $q-1$ Zeilen und m Spalten hat und deren Koeffizienten durch $u_{ij} := b_{i+1,j} - b_{ij}$ definiert sind; außerdem ein Spaltenvektor **v** der Länge $q-1$, dessen Komponenten durch

$$v_i \;:=\; \begin{cases} 0 & \text{wenn } y_i = y_{i-1} \\ 1 & \text{wenn } y_i > y_{i-1} \end{cases}$$

definiert sind. Die zu lösende Aufgabe besteht dann darin, einen Bewertungsvektor **d** zu finden, so daß gilt:

$$\mathbf{Ud} \geq \mathbf{v} \quad \text{und} \quad \mathbf{u}_i\mathbf{d} \;=\; 0 \;\text{ wenn } v_i = 0$$

wobei $\mathbf{u}_i$ die i-te Zeile von **U** bezeichnet. Es gibt dann zwei Möglichkeiten. Es gibt keinen Bewertungsvektor **d**, der diesen Anforderungen genügt;

[15] Es handelt sich um ein Standardverfahren, um lineare Funktionen zu minimieren oder maximieren, wenn es Nebenbedingungen gibt, die in Form linearer Gleichungen und/oder Ungleichungen formuliert werden können. Eine gute Darstellung findet man z.B. bei Neumann und Morlock 1993, Kap. 1.

dann gibt es auch keine Bewertungsfunktionen, deren Addition eine strikt äquivalente Rangordnung erzeugen könnte. Oder es gibt beliebig viele Vektoren $\mathbf{d}$, die den Bedingungen genügen. Dann kann man sich einen davon aussuchen und erhält die gesuchten Bewertungsfunktionen. Diese Überlegung zeigt auch, wie sich eine sinnvolle Zielfunktion definieren läßt, die erforderlich ist, um das Verfahren der linearen Programmierung anwendbar zu machen, nämlich

$$\sum_{j=1}^{m} d_j \longrightarrow \min$$

Wenn es einen Bewertungsvektor gibt, der die zuvor genannten Bedingungen erfüllt, dann liefert das Verfahren der linearen Programmierung eine Darstellung dieses Vektors, dessen Komponenten möglichst klein sind. Für das eingangs angeführte Beispiel findet man auf diese Weise den Bewertungsvektor

$$\mathbf{d} = (0, 2, 7, 0, 6, 9, 10)'$$

den wir in diesem Beispiel auch zur Darstellung der Bewertungsfunktionen verwendet haben.

6. Wir nennen das in diesem Abschnitt beschriebene Rechenverfahren den *LP-Algorithmus* (zur Berechnung impliziter Indikatoren). Es kann verwendet werden, um zu prüfen, ob es Bewertungsfunktionen gibt, mit denen sich eine strikt äquivalente Rangordnung erzeugen läßt; und wenn es solche Bewertungsfunktionen gibt, kann man sie mit dem Algorithmus auch finden. Insofern handelt es sich um eine nützliche Ergänzung zum ALS-Algorithmus, der nur nach schwach äquivalenten Lösungen sucht.

12.3 Interpretations- und Anwendungsprobleme

In den folgenden Abschnitten wird besprochen, wie sich Bewertungsfunktionen, die mit der Methode der impliziten Indikatoren konstruiert worden sind, interpretieren lassen.

12.3.1 Unbestimmte Bewertungsfunktionen

1. Zunächst soll überlegt werden, ob und ggf. in welcher Weise Bewertungsfunktionen eindeutig bestimmt sind. Die in den vorangegangenen Abschnitten besprochenen Beispiele zeigen, daß insbesondere dann, wenn die Rangordnung r^* Bindungen aufweist, unterschiedliche Bewertungsfunktionen konstruiert werden können. Aber wie verhält es sich, wenn man Bewertungsfunktionen finden kann, deren Addition eine strikt äquivalente Darstellung der vorgegebenen Rangordnung r^* liefert?

2. Als Beispiel betrachten wir noch einmal die Rangordnung r^*, die von Diekmanns Befragungsperson Nr. 7 angegeben wurde, also

	$\tilde{x}_{21}$	$\tilde{x}_{22}$	$\tilde{x}_{23}$	$\tilde{x}_{24}$
$\tilde{x}_{11}$	1	3	6	9
$\tilde{x}_{12}$	2	5	7	10
$\tilde{x}_{13}$	4	8	11	12

Wie in Abschnitt 12.2.1 gezeigt wurde, kann man Bewertungsfunktionen finden, die eine strikt äquivalente Rangordnung liefern, nämlich:

	$\phi_2(\tilde{x}_{21})$	$\phi_2(\tilde{x}_{22})$	$\phi_2(\tilde{x}_{23})$	$\phi_2(\tilde{x}_{24})$
	0.00	3.00	5.67	8.00
$\phi_1(\tilde{x}_{11}) = 0.00$	0.00	3.00	5.67	8.00
$\phi_1(\tilde{x}_{12}) = 1.25$	1.25	4.25	6.92	9.25
$\phi_1(\tilde{x}_{13}) = 4.00$	4.00	7.00	9.67	12.00

Sie können natürlich in beliebiger Weise linear transformiert werden ($\phi'_j = a + b\,\phi_j$, wobei $a \geq 0$ und $b > 0$ ist), so daß wiederum eine strikt äquivalente Rangordnung entsteht.

3. Diekmann (1982) hat vermutet, daß es keine weiteren Unbestimmtheiten gibt, daß also in diesem Fall die Bewertungsfunktionen bis auf die Möglichkeit einer linearen Transformation eindeutig sind. Man kann sich jedoch leicht davon überzeugen, daß diese Vermutung falsch ist. Z.B. kann man anstelle von ϕ_2 auch die Bewertungsfunktion $\phi'_2(\tilde{x}_{21}) = 0$, $\phi'_2(\tilde{x}_{22}) = 3$, $\phi'_2(\tilde{x}_{23}) = 5.67$, $\phi'_2(\tilde{x}_{24}) = 8.2$ verwenden. Offenbar erhält man ϕ'_2 nicht durch eine lineare Transformation aus ϕ_2, aber dennoch liefert $\phi' = \phi_1 + \phi'_2$ wiederum eine strikt äquivalente Rangordnung.[16] Man gelangt also zu dem Ergebnis, daß die Bewertungsfunktionen, die mit der Methode der impliziten Indikatoren konstruiert werden können, nicht eindeutig (nicht einmal eindeutig bis auf lineare Transformationen) bestimmt sind, und zwar selbst dann nicht, wenn man strikt äquivalente Repräsentationen finden

[16] Man kann darüber spekulieren, wie es zu der falschen Vermutung gekommen ist. Diekmann stützt sich für seine Vermutung auf Überlegungen, die im Kontext der repräsentationalen Meßtheorie entwickelt worden sind. Einige Autoren (zuerst Luce und Tukey 1964) konnten nämlich zeigen, daß man bis auf lineare Transformationen eindeutige Bewertungsfunktionen konstruieren kann, wenn man eine Reihe von Bedingungen voraussetzt. Wie Diekmann zeigt, sind die zunächst wesentlich erscheinenden Bedingungen bei der Rangordnung der Befragungsperson Nr. 7 tatsächlich erfüllt. Allerdings bleibt ein entscheidender Sachverhalt unberücksichtigt. Die Eindeutigkeitsbeweise der repräsentationalen Meßtheorie setzen voraus, daß man sich mit Zahlenmengen beschäftigen kann, in denen unendliche Folgen existieren, deren Eigenschaften man ausnutzen kann. Die „empirischen Relative", die Sozialwissenschaftler durch ihre Daten konstruieren, sind jedoch stets endlich. So hat z.B. der Merkmalsraum im Diekmannschen Beispiel nur 12 Elemente, und die rein mathematischen Überlegungen der repräsentationalen Meßtheorie sind gar nicht übertragbar.

Tabelle 12.3.1 Komponenten eines kombinierten Merkmalsraums.

$\tilde{\mathcal{X}}_1$	Preis	$\tilde{x}_{11}$	2.51 – 3.00 DM
		$\tilde{x}_{12}$	2.01 – 2.50 DM
		$\tilde{x}_{13}$	1.51 – 2.00 DM
$\tilde{\mathcal{X}}_2$	Verwendung	$\tilde{x}_{21}$	als Brotaufstrich geeignet
		$\tilde{x}_{22}$	zum Kochen, Backen, Braten geeignet
		$\tilde{x}_{23}$	universell verwendbar
$\tilde{\mathcal{X}}_3$	Geschmack	$\tilde{x}_{31}$	nach Butter schmeckend
		$\tilde{x}_{32}$	neutral-pflanzlich schmeckend
$\tilde{\mathcal{X}}_4$	Kaloriengehalt	$\tilde{x}_{41}$	kalorienarm (400 kcal/100g)
		$\tilde{x}_{42}$	normaler Kaloriengehalt (700 kcal/100g)

kann. Für die Interpretation folgt daraus, daß man Bewertungsfunktionen bestenfalls als lineare Ordnungen für die Teilmerkmalsräume interpretieren sollte. An diese Überlegung knüpft eine nicht-metrische Variante des Verfahrens an, mit der wir uns in Abschnitt 12.3.3 beschäftigen.

4. Diese Überlegung zeigt auch, daß die Unbestimmtheit von Bewertungsfunktionen nicht daraus resultiert, daß die vorgegebene Rangordnung r^* Bindungen aufweist. Tatsächlich ist es möglich, daß trotz des Vorhandenseins von Bindungen ein *strikt* äquivalenter additiver Index konstruiert werden kann. Als Beispiel kann eine Rangordnung r^* dienen, die von Diekmanns Befragungsperson Nr. 9 angegeben wurde:

	$\tilde{x}_{21}$	$\tilde{x}_{22}$	$\tilde{x}_{23}$	$\tilde{x}_{24}$
$\tilde{x}_{11}$	1	4	6	9
$\tilde{x}_{12}$	2	5	8	11
$\tilde{x}_{13}$	3	6	9	12

Wendet man den LP-Algorithmus an, findet man den Bewertungsvektor $\mathbf{d} = (0, 1, 2, 0, 3, 5, 7)'$, so daß man folgendes Ergebnis erhält:

	$\phi_2(\tilde{x}_{21})$ 0	$\phi_2(\tilde{x}_{22})$ 3	$\phi_2(\tilde{x}_{23})$ 5	$\phi_2(\tilde{x}_{24})$ 7
$\phi_1(\tilde{x}_{11}) = 0$	0	3	5	7
$\phi_1(\tilde{x}_{12}) = 1$	1	4	6	8
$\phi_1(\tilde{x}_{13}) = 2$	2	5	7	9

Offenbar ist die durch die Bewertungsfunktionen erzeugte Rangordnung zur vorgegebenen Rangordnung r^* strikt äquivalent.

Tabelle 12.3.2 Exemplarische Rangordnung für 9 Merkmalskombinationen.

$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$	**y**
$\tilde{x}_{13}$	$\tilde{x}_{22}$	$\tilde{x}_{31}$	$\tilde{x}_{42}$	1
$\tilde{x}_{11}$	$\tilde{x}_{22}$	$\tilde{x}_{32}$	$\tilde{x}_{41}$	2
$\tilde{x}_{11}$	$\tilde{x}_{23}$	$\tilde{x}_{31}$	$\tilde{x}_{42}$	3
$\tilde{x}_{12}$	$\tilde{x}_{22}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	4
$\tilde{x}_{12}$	$\tilde{x}_{21}$	$\tilde{x}_{32}$	$\tilde{x}_{42}$	5
$\tilde{x}_{11}$	$\tilde{x}_{21}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	6
$\tilde{x}_{13}$	$\tilde{x}_{21}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	7
$\tilde{x}_{13}$	$\tilde{x}_{23}$	$\tilde{x}_{32}$	$\tilde{x}_{41}$	8
$\tilde{x}_{12}$	$\tilde{x}_{23}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	9

12.3.2 Ein Beispiel aus der Marktforschung

1. Schwierigkeiten bei der Anwendung des Verfahrens der impliziten Indikatoren entstehen nicht nur daraus, daß es oft nicht gelingt, zumindest schwach äquivalente additive Indizes zu konstruieren. Eine weitere Schwierigkeit resultiert daraus, daß der gemeinsame Merkmalsraum oft eine so große Anzahl von Merkmalskombinationen enthält, daß sich durch Rating-Verfahren keine vollständigen Rangordnungen herstellen lassen. Man denke z.B. an den dreidimensionalen Merkmalsraum, der dem Schichtindex von Scheuch und Daheim zugrunde liegt: Er hat $17 \cdot 12 \cdot 11 = 2244$ Elemente (vgl. Abschnitt 6.3). Auch wenn man davon absieht, daß es in diesem Fall keine vernünftigen Kriterien gibt, um eine Rangordnung herzustellen, erscheint dies bereits wegen der großen Anzahl der zu ordnenden Elemente kaum möglich.

2. Das gleiche Problem tritt bei Anwendungen der Methode der impliziten Indikatoren in der Marktforschung auf. Zur Illustration beziehen wir uns auf ein Beispiel, das von Backhaus, Erichson, Plinke, Schuchard-Ficher und Weiber (1987, S. 363ff) angegeben worden ist. In diesem Beispiel möchte man herausfinden, wie die Präferenzen für unterschiedliche Margarinesorten von deren Eigenschaften abhängen. Die Autoren verwenden einen kombinierten Merkmalsraum, der aus vier Komponenten besteht. Tabelle 12.3.1 zeigt die Teilmerkmalsräume und ihre Elemente. Insgesamt enthält der kombinierte Merkmalsraum 36 Elemente (Produktvarianten), die auf beliebige Weise in eine Rangfolge gebracht werden können. Die Autoren haben Studenten befragt und ihnen 9 der 36 möglichen Merkmalskombinationen vorgegeben mit der Bitte, sie in eine Rangordnung zu bringen. Tabelle 12.3.2 zeigt eine dieser Rangordnungen und zugleich die verwendeten Merkmalskombinationen. Die Person, die diese Rangordnung hergestellt hat, präferiert also am meisten eine Margarinesorte mit mittlerem Preis, die universell verwendbar ist, nach Butter schmeckt und kalorienarm ist.

3. Je nachdem, ob man Rangordnungen für alle oder nur für einen Teil der Elemente des gemeinsamen Merkmalsraums verwendet, spricht man von einem *vollständigen* oder einem *reduzierten Design*. Um festzustellen, ob die Methode der impliziten Indikatoren auch bei einem reduzierten Design angewendet werden kann, repräsentieren wir die insgesamt möglichen Merkmalskombinationen durch Vektoren $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4)$, deren Elemente in den Teilmerkmalsräumen $\tilde{\mathcal{X}}_1, \ldots, \tilde{\mathcal{X}}_4$ variieren können. Wie in Abschnitt 12.2.1 dargestellt worden ist, können dann Indikatorvariablen I_{jk} (= 1, wenn $\tilde{x}_j = \tilde{x}_{jk}$ ist) definiert werden, um einen Regressionsansatz zu bilden, der zur Minimierung des Abstands

$$\| \mathbf{y} - (\alpha + I_{12}\beta_{12} + I_{13}\beta_{13} + I_{22}\beta_{22} + I_{23}\beta_{23} + I_{32}\beta_{32} + I_{42}\beta_{42}) \|^2$$

äquivalent ist. Eine vollständige Design-Matrix hätte 36 Zeilen. Da in diesem Beispiel nur neun Merkmalskombinationen verwendet werden, hat die reduzierte Design-Matrix ebenfalls nur 9 Zeilen. Die dem Regressionsansatz entsprechenden Daten sehen also folgendermaßen aus:

$$\mathbf{y} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{pmatrix} \quad \text{und} \quad \mathbf{X} = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Jetzt kann man überlegen, ob sich das Regressionsproblem mit der OLS-Methode lösen läßt. Das ist genau dann der Fall, wenn die Matrix $\mathbf{X}$ vollen Spaltenrang hat, so daß $\mathbf{X}'\mathbf{X}$ invertierbar ist. Dann erhält man die Lösung durch $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Ob man eine Lösung finden kann, hängt also auch davon ab, welche der insgesamt möglichen Merkmalskombinationen zur Herstellung einer Rangordnung verwendet werden. In unserem Beispiel genügen bereits die 9 ausgewählten Merkmalskombinationen, um der reduzierten Design-Matrix vollen Spaltenrang zu verschaffen und das Regressionsproblem lösbar zu machen. Man erhält die Regressionsparameter

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_{12} \\ \hat{\beta}_{13} \\ \hat{\beta}_{22} \\ \hat{\beta}_{23} \\ \hat{\beta}_{32} \\ \hat{\beta}_{42} \end{pmatrix} = \begin{pmatrix} 5.667 \\ 2.333 \\ 1.667 \\ -3.667 \\ 0.667 \\ 0.000 \\ -3.000 \end{pmatrix}$$

und kann daraus den Vektor

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = (0.67, 2.00, 3.33, 4.33, 5.00, 5.67, 7.33, 8.00, 8.67)'$$

berechnen. Da er ein Element von $\mathcal{R}_9$ ist, liefert er bereits eine perfekte Repräsentation der Rangordnung $\mathbf{y}$, und es ist nicht erforderlich, eine monotone Regression durchzuführen (sie würde in diesem Beispiel zu dem gleichen Ergebnis führen).

4. Aus der Regressionsrechnung gewinnt man die Bewertungsfunktionen

$$\begin{array}{lll} \phi_1(\tilde{x}_{11}) = 0.000 & \phi_1(\tilde{x}_{12}) = 2.333 & \phi_1(\tilde{x}_{13}) = 1.667 \\ \phi_2(\tilde{x}_{21}) = 0.000 & \phi_1(\tilde{x}_{22}) = -3.667 & \phi_1(\tilde{x}_{23}) = 0.667 \\ \phi_3(\tilde{x}_{31}) = 0.000 & \phi_3(\tilde{x}_{32}) = 0.000 & \\ \phi_4(\tilde{x}_{41}) = 0.000 & \phi_4(\tilde{x}_{42}) = -3.000 & \end{array}$$

und kann über die durch das Rechenverfahren erzeugten Rangordnungen für die Teilmerkmalsräume spekulieren. Z.B. findet man für $\tilde{\mathcal{X}}_1$ die Ordnung $(1, 3, 2)$, d.h. die höchste Präferenz besteht für den mittleren, die zweithöchste für den niedrigsten und die niedrigste für den höchsten Preis. Bemerkenswert ist, daß die Bewertungsfunktion für den Teilmerkmalsraum $\tilde{\mathcal{X}}_3$ eine Indifferenz angibt. Das liefert einen ersten Hinweis darauf, daß die Lösung nicht eindeutig ist.

5. Einen weiteren Hinweis gewinnt man daraus, daß bei Backhaus u.a. (1987, S. 368) eine andere Lösung angegeben wird, nämlich

$$\begin{array}{lll} \phi'_1(\tilde{x}_{11}) = 0.701 & \phi'_1(\tilde{x}_{12}) = 1.160 & \phi'_1(\tilde{x}_{13}) = 0.987 \\ \phi'_2(\tilde{x}_{21}) = 0.913 & \phi'_1(\tilde{x}_{22}) = -0.156 & \phi'_1(\tilde{x}_{23}) = 0.776 \\ \phi'_3(\tilde{x}_{31}) = 1.038 & \phi'_3(\tilde{x}_{32}) = 1.201 & \\ \phi'_4(\tilde{x}_{41}) = 1.554 & \phi'_4(\tilde{x}_{42}) = 0.891 & \end{array}$$

Sie liefert zwar in den Teilmerkmalsräumen $\tilde{\mathcal{X}}_1$ und $\tilde{\mathcal{X}}_4$ die gleiche Ordnung wie die zuvor angegebene OLS-Lösung; jedoch unterscheidet sich die Ordnung im Teilmerkmalsraum $\tilde{\mathcal{X}}_2$, und vor allem bemerkenswert ist, daß durch ϕ'_3 eine eindeutige Ordnung in $\tilde{\mathcal{X}}_3$ hergestellt wird. Wir werden im nächsten Abschnitt besprechen, wie diese unterschiedlichen Bewertungsfunktionen entstehen können. (Backhaus u.a. geben an, daß sie das Computerprogramm UNICON verwendet haben, machen aber keine näheren Angaben darüber, wie dieses Programm seine Lösungen berechnet. Schon aus einem Vergleich der β-Koeffizienten erkennt man indessen, daß es sich nicht um bloß lineare Transformationen der Bewertungsfunktionen handelt. Bei der von Backhaus u.a. angegebenen Lösung handelt es sich auch nicht um eine OLS-Lösung.)

6. Hat man Bewertungsfunktionen konstruiert, kann man mit ihrer Hilfe für alle Elemente des gemeinsamen Merkmalsraums eine lineare Ordnung

Tabelle 12.3.3 Varianten der Rangordnung.

$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$	r_a	r_b	r^*
$\tilde{x}_{11}$	$\tilde{x}_{22}$	$\tilde{x}_{31}$	$\tilde{x}_{42}$	0.00	0.00	
$\tilde{x}_{11}$	$\tilde{x}_{22}$	$\tilde{x}_{32}$	$\tilde{x}_{42}$	0.00	0.62	
$\tilde{x}_{13}$	$\tilde{x}_{22}$	$\tilde{x}_{31}$	$\tilde{x}_{42}$	1.67	1.09	1
$\tilde{x}_{13}$	$\tilde{x}_{22}$	$\tilde{x}_{32}$	$\tilde{x}_{42}$	1.67	1.71	
$\tilde{x}_{12}$	$\tilde{x}_{22}$	$\tilde{x}_{31}$	$\tilde{x}_{42}$	2.33	1.75	
$\tilde{x}_{12}$	$\tilde{x}_{22}$	$\tilde{x}_{32}$	$\tilde{x}_{42}$	2.33	2.38	
$\tilde{x}_{11}$	$\tilde{x}_{22}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	3.00	2.53	
$\tilde{x}_{11}$	$\tilde{x}_{22}$	$\tilde{x}_{32}$	$\tilde{x}_{41}$	3.00	3.16	2
$\tilde{x}_{11}$	$\tilde{x}_{23}$	$\tilde{x}_{31}$	$\tilde{x}_{42}$	4.33	3.56	3
$\tilde{x}_{13}$	$\tilde{x}_{22}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	4.67	3.63	
$\tilde{x}_{11}$	$\tilde{x}_{21}$	$\tilde{x}_{31}$	$\tilde{x}_{42}$	3.67	4.09	
$\tilde{x}_{11}$	$\tilde{x}_{23}$	$\tilde{x}_{32}$	$\tilde{x}_{42}$	4.33	4.18	
$\tilde{x}_{13}$	$\tilde{x}_{22}$	$\tilde{x}_{32}$	$\tilde{x}_{41}$	4.67	4.25	
$\tilde{x}_{12}$	$\tilde{x}_{22}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	5.33	4.29	4
$\tilde{x}_{13}$	$\tilde{x}_{23}$	$\tilde{x}_{31}$	$\tilde{x}_{42}$	6.00	4.66	
$\tilde{x}_{11}$	$\tilde{x}_{21}$	$\tilde{x}_{32}$	$\tilde{x}_{42}$	3.67	4.71	
$\tilde{x}_{12}$	$\tilde{x}_{22}$	$\tilde{x}_{32}$	$\tilde{x}_{41}$	5.33	4.91	
$\tilde{x}_{13}$	$\tilde{x}_{21}$	$\tilde{x}_{31}$	$\tilde{x}_{42}$	5.33	5.18	
$\tilde{x}_{13}$	$\tilde{x}_{23}$	$\tilde{x}_{32}$	$\tilde{x}_{42}$	6.00	5.28	
$\tilde{x}_{12}$	$\tilde{x}_{23}$	$\tilde{x}_{31}$	$\tilde{x}_{42}$	6.67	5.32	
$\tilde{x}_{13}$	$\tilde{x}_{21}$	$\tilde{x}_{32}$	$\tilde{x}_{42}$	5.33	5.80	
$\tilde{x}_{12}$	$\tilde{x}_{21}$	$\tilde{x}_{31}$	$\tilde{x}_{42}$	6.00	5.84	
$\tilde{x}_{12}$	$\tilde{x}_{23}$	$\tilde{x}_{32}$	$\tilde{x}_{42}$	6.67	5.94	
$\tilde{x}_{11}$	$\tilde{x}_{23}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	7.33	6.10	
$\tilde{x}_{12}$	$\tilde{x}_{21}$	$\tilde{x}_{32}$	$\tilde{x}_{42}$	6.00	6.46	5
$\tilde{x}_{11}$	$\tilde{x}_{21}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	6.67	6.62	6
$\tilde{x}_{11}$	$\tilde{x}_{23}$	$\tilde{x}_{32}$	$\tilde{x}_{41}$	7.33	6.72	
$\tilde{x}_{13}$	$\tilde{x}_{23}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	9.00	7.19	
$\tilde{x}_{11}$	$\tilde{x}_{21}$	$\tilde{x}_{32}$	$\tilde{x}_{41}$	6.67	7.24	
$\tilde{x}_{13}$	$\tilde{x}_{21}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	8.33	7.72	7
$\tilde{x}_{13}$	$\tilde{x}_{23}$	$\tilde{x}_{32}$	$\tilde{x}_{41}$	9.00	7.81	8
$\tilde{x}_{12}$	$\tilde{x}_{23}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	9.67	7.85	9
$\tilde{x}_{13}$	$\tilde{x}_{21}$	$\tilde{x}_{32}$	$\tilde{x}_{41}$	8.33	8.34	
$\tilde{x}_{12}$	$\tilde{x}_{21}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	9.00	8.38	
$\tilde{x}_{12}$	$\tilde{x}_{23}$	$\tilde{x}_{32}$	$\tilde{x}_{41}$	9.67	8.48	
$\tilde{x}_{12}$	$\tilde{x}_{21}$	$\tilde{x}_{32}$	$\tilde{x}_{41}$	9.00	9.00	

erzeugen. Wenn ein reduziertes Design verwendet wird, entsteht jedoch meistens nur eine schwache lineare Ordnung. Das zeigt sich in der Rangordnung r_a in Tabelle 12.3.3, die mit der OLS-Lösung erzeugt worden ist. Man erkennt, daß sie mit der vorgegebenen Rangordnung r^* vereinbar ist, jedoch zahlreiche Bindungen aufweist. Daneben haben wir eine weitere Rangordnung r_b eingetragen, die aus den von Backhaus et al. angegebenen

Bewertungsfunktionen gebildet worden ist. Bemerkenswert ist, daß diese Rangordnung keine Bindungen aufweist, und die Autoren schließen daraus (S. 369f), daß die Merkmalskombination $(2, 1, 2, 1)$ die „höchstpräferierte Produktalternative“ ist. Da die Bewertungsfunktionen nicht eindeutig bestimmt sind, ist die Schlußfolgerung jedoch fragwürdig.

12.3.3 Implizite partielle Ordnungen

1. Insbesondere bei Anwendungen des Verfahrens der impliziten Indikatoren in der Marktforschung ist es üblich, die konstruierten Bewertungsfunktionen metrisch zu interpretieren. Z.B. schreiben Backhaus u.a.:

Die durch ein Rating-Verfahren erzeugte „Rangreihe bildet die Grundlage zur Ableitung von Teilnutzenwerten für die einzelnen Eigenschaftsausprägungen. Die Auskunftsperson gibt also *ordinale Gesamtnutzenurteile* ab, aus denen durch die Conjoint-Analyse *metrische Teilnutzenwerte* abgeleitet werden. Damit wird es außerdem möglich, durch Addition der Teilnutzenwerte auch metrische Gesamtnutzenwerte zu ermitteln.“ (Backhaus u.a. 1987, S. 346)

Das kann man jedoch schon deshalb bezweifeln, weil sich im allgemeinen keine eindeutigen Bewertungsfunktionen konstruieren lassen. Bestenfalls erhält man lineare Ordnungen für die Teilmerkmalsräume, und auch diese sind, wie das Beispiel im vorangegangenen Abschnitt gezeigt hat, nicht unbedingt eindeutig. Es sollte deshalb überlegt werden, wie man zu einem nicht-metrischen Verständnis des Verfahrens der impliziten Indikatoren gelangen kann.

2. Eine einfache Überlegung verläuft folgendermaßen. Angenommen, es gibt lineare Ordnungen für die Teilmerkmalsräume $\tilde{\mathcal{X}}_1, \ldots, \tilde{\mathcal{X}}_m$. Dann gewinnt man daraus eine partielle Ordnung für den gemeinsamen Merkmalsraum $\tilde{\mathcal{X}}$, die im allgemeinen auf unterschiedliche Weisen zu einer linearen Ordnung fortgesetzt werden kann (vgl. Abschnitt 10.2.1). Hat man nun andererseits eine lineare Ordnung für den gemeinsamen Merkmalsraum, kann man untersuchen, ob sie eine konsistente Fortsetzung einer partiellen Ordnung ist, die durch lineare Ordnungen in den Teilmerkmalsräumen bestimmt wird. Um den Gedankengang formal handhabbar zu machen, verwenden wir folgende Notationen: $r_1, \ldots, r_m$ bezeichnen lineare Ordnungen für die Teilmerkmalsräume, und die (möglicherweise unvollständige) lineare Ordnung für den gemeinsamen Merkmalsraum $\tilde{\mathcal{X}}$ wird wie bisher mit r^* bezeichnet. Diese Ordnung ist eine konsistente Fortsetzung der durch $r_1, \ldots, r_m$ definierten partiellen Ordnung für $\tilde{\mathcal{X}}$, wenn gilt:

$$\begin{aligned} & r_1(\tilde{x}_1) \le r_1(\tilde{x}'_1), \ldots, r_m(\tilde{x}_m) \le r_m(\tilde{x}'_m) \\ & \qquad \Longrightarrow \quad r^*(\tilde{x}_1, \ldots, \tilde{x}_m) \le r^*(\tilde{x}'_1, \ldots, \tilde{x}'_m) \end{aligned} \tag{12.3.1}$$

Die nicht-metrische Variante des Verfahrens der impliziten Indikatoren besteht also darin, nach allen linearen Ordnungen $r_1, \ldots, r_m$ zu suchen, so

Tabelle 12.3.4 Ordnungen in den Teilmerkmalsräumen.

Variante	r_1	r_2	r_3	r_4
(a)	(1,2,3)	(2,1,3)	(2,1)	(2,1)
(b)	(1,3,2)	(2,1,3)	(1,2)	(2,1)
(c)	(1,3,2)	(2,1,3)	(2,1)	(2,1)
(d)	(1,3,2)	(3,1,2)	(1,2)	(2,1)

daß die vorgegebene Rangordnung r^* eine konsistente Fortsetzung bildet.

3. Folgende Möglichkeiten sind in Betracht zu ziehen:

a) Es ist möglich, daß es überhaupt keine partielle Ordnung für $\tilde{\mathcal{X}}$ gibt, so daß r^* eine konsistente Fortsetzung ist.[17]

b) Wenn r^* eine strikte und vollständige lineare Ordnung für $\tilde{\mathcal{X}}$ herstellt, kann es höchstens eine partielle Ordnung $r_1, \ldots, r_m$ geben, so daß r^* eine konsistente Fortsetzung ist.[18]

c) Wenn r^* zwar vollständig, aber nur eine schwache lineare Ordnung ist, kann es mehrere partielle Ordnungen geben, deren konsistente Fortsetzung r^* ist.

d) Unabhängig davon, ob r^* eine strikte oder schwache lineare Ordnung ist, kann es, wenn es sich um eine unvollständige Ordnung handelt (wie bei einem reduzierten Design), ebenfalls mehrere partielle Ordnungen geben, deren konsistente Fortsetzung r^* ist.

4. Zur Illustration verwenden wir das Beispiel aus dem vorangegangenen Abschnitt. In diesem Beispiel gibt es eine zwar strikte, aber nur unvollständige lineare Ordnung r^* für den Merkmalsraum $\tilde{\mathcal{X}}$. Sucht man nach linearen Ordnungen $r_1, \ldots, r_4$ für die vier Teilmerkmalsräume, so daß r^*

[17]Dies ist z.B. in der Untersuchung von Diekmann (1982) bei der Befragungsperson Nr. 2 der Fall, die folgende Rangordnung angegeben hat:

	$\tilde{x}_{21}$	$\tilde{x}_{22}$	$\tilde{x}_{23}$	$\tilde{x}_{24}$
$\tilde{x}_{11}$	2	4	6	11
$\tilde{x}_{12}$	1	5	8	10
$\tilde{x}_{13}$	2	6	8	12

[18]Denn wenn $\tilde{x}_{jk}$ und $\tilde{x}_{jk'}$ zwei Elemente von $\tilde{\mathcal{X}}_j$ sind, gibt es identische Elemente in den anderen Teilmerkmalsräumen, so daß

$$r^*(\ldots, \tilde{x}_{jk}, \ldots) < r^*(\ldots, \tilde{x}_{jk'}, \ldots) \quad \text{oder} \quad r^*(\ldots, \tilde{x}_{jk}, \ldots) > r^*(\ldots, \tilde{x}_{jk'}, \ldots)$$

gilt. Wenn es sich um eine konsistente Fortsetzung handelt, muß dann jedoch $\tilde{x}_{jk} \leq \tilde{x}_{jk'}$ oder $\tilde{x}_{jk} \geq \tilde{x}_{jk'}$ gelten.

Tabelle 12.3.5 Repräsentation der Produktalternativen.

r^*	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$	$r_1^a(\tilde{x}_1)$	$r_2^a(\tilde{x}_2)$	$r_3^a(\tilde{x}_3)$	$r_4^a(\tilde{x}_4)$
1	$\tilde{x}_{13}$	$\tilde{x}_{22}$	$\tilde{x}_{31}$	$\tilde{x}_{42}$	3	1	2	1
2	$\tilde{x}_{11}$	$\tilde{x}_{22}$	$\tilde{x}_{32}$	$\tilde{x}_{41}$	1	1	1	2
3	$\tilde{x}_{11}$	$\tilde{x}_{23}$	$\tilde{x}_{31}$	$\tilde{x}_{42}$	1	3	2	1
4	$\tilde{x}_{12}$	$\tilde{x}_{22}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	2	1	2	2
5	$\tilde{x}_{12}$	$\tilde{x}_{21}$	$\tilde{x}_{32}$	$\tilde{x}_{42}$	2	2	1	1
6	$\tilde{x}_{11}$	$\tilde{x}_{21}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	1	2	2	2
7	$\tilde{x}_{13}$	$\tilde{x}_{21}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	3	2	2	2
8	$\tilde{x}_{13}$	$\tilde{x}_{23}$	$\tilde{x}_{32}$	$\tilde{x}_{41}$	3	3	1	2
9	$\tilde{x}_{12}$	$\tilde{x}_{23}$	$\tilde{x}_{31}$	$\tilde{x}_{41}$	2	3	2	2

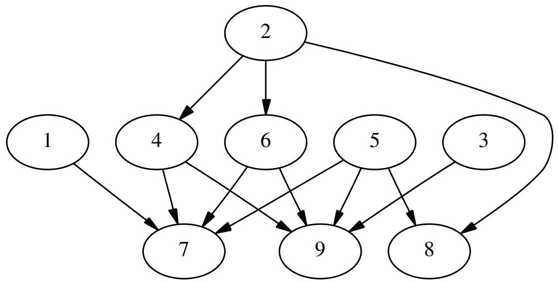


Abb. 12.3.1 Hasse-Diagramm für die durch $r_1^a, \ldots, r_4^a$ definierte partielle Ordnung.

eine konsistente Fortsetzung der durch $r_1, \ldots, r_4$ definierten partiellen Ordnung ist, findet man die in Tabelle 12.3.4 angegebenen vier Varianten. In der Variante (a) wird z.B. im Teilmerkmalsraum $\tilde{\mathcal{X}}_1$ die billigste Produktalternative am meisten präferiert, dann folgt die mittlere und schließlich die teuerste Alternative. Formal kann man die Ordnungen $r_1, \ldots, r_4$ als Bewertungsfunktionen für die Teilmerkmalsräume auffassen und zur numerischen Repräsentation der Produktalternativen verwenden. Als Beispiel verwenden wir die Alternative (a). Tabelle 12.3.5 reproduziert zunächst die Angaben aus Tabelle 12.3.2 und verwendet dann die in Variante (a) ermittelten Ordnungen $r_1^a, \ldots, r_4^a$ zur numerischen Repräsentation. Daß r^* eine konsistente Fortsetzung der durch diese Variante definierten partiellen Ordnung für $\tilde{\mathcal{X}}$ bildet, erkennt man anhand des Hasse-Diagramms in Abbildung 12.3.1. Die Rangordnung r^* ist mit allen Pfeilen verträglich.

5. Tabelle 12.3.4 zeigt, daß es nur für den Teilmerkmalsraum $\tilde{\mathcal{X}}_4$ ein eindeutiges Ergebnis gibt. Bei allen anderen Teilmerkmalsräumen gibt es unterschiedliche Ordnungsmöglichkeiten, die mit der Rangordnung r^* verträglich sind. Dies Ergebnis zeigt auch, warum es in diesem Beispiel zu ganz

Tabelle 12.3.6 10 Rangordnungen (Diekmann 1982, S. 27).

$\tilde{x}_1$	$\tilde{x}_2$	r_1^*	r_2^*	r_3^*	r_4^*	r_5^*	r_6^*	r_7^*	r_8^*	r_9^*	r_{10}^*
$\tilde{x}_{11}$	$\tilde{x}_{21}$	1	2	1	1	1	1	1	1	1	1
$\tilde{x}_{11}$	$\tilde{x}_{22}$	2	4	3	2	3	3	3	4	4	2
$\tilde{x}_{11}$	$\tilde{x}_{23}$	4	6	6	4	6	4	6	6	6	3
$\tilde{x}_{11}$	$\tilde{x}_{24}$	5	11	9	8	7	6	9	9	9	4
$\tilde{x}_{12}$	$\tilde{x}_{21}$	3	1	2	3	2	2	2	2	2	4
$\tilde{x}_{12}$	$\tilde{x}_{22}$	5	5	4	6	5	6	5	5	5	6
$\tilde{x}_{12}$	$\tilde{x}_{23}$	7	8	6	8	8	8	7	8	8	7
$\tilde{x}_{12}$	$\tilde{x}_{24}$	9	10	10	11	9	10	10	9	11	9
$\tilde{x}_{13}$	$\tilde{x}_{21}$	7	2	4	5	4	6	4	3	3	7
$\tilde{x}_{13}$	$\tilde{x}_{22}$	10	6	8	8	10	9	8	6	6	10
$\tilde{x}_{13}$	$\tilde{x}_{23}$	11	8	11	10	11	11	11	11	9	11
$\tilde{x}_{13}$	$\tilde{x}_{24}$	12	12	12	12	12	12	12	12	12	12

unterschiedlichen Bewertungsfunktionen kommen kann. Die OLS-Lösung liefert für den Teilmerkmalsraum $\tilde{\mathcal{X}}_3$ eine Indifferenz, weil in den Varianten (b) und (c) gleiche Ordnungen in den Teilmerkmalsräumen $\tilde{\mathcal{X}}_1$, $\tilde{\mathcal{X}}_2$ und $\tilde{\mathcal{X}}_4$ mit zwei entgegengesetzten Ordnungen in $\tilde{\mathcal{X}}_3$ vereinbar sind. Die von Backhaus u.a. angegebenen Bewertungsfunktionen entsprechen dagegen der Variante (d).

12.3.4 Mehrere Rangordnungen

1. Bisher wurde stets auf eine bestimmte Rangordnung r^* Bezug genommen. Bei praktischen Anwendungen befragt man meistens mehrere Personen und erhält dann von jeder Befragungsperson eine bestimmte Rangordnung r_i^* (wobei $i = 1, \ldots, n$ die Befragungspersonen indiziert). Im wesentlichen gibt es dann zwei Möglichkeiten. Man kann die Rangordnungen zuerst aggregieren, z.B. indem man mittlere oder Median-Rangordnungen berechnet, und dann das Verfahren der impliziten Indikatoren auf die aggregierte Rangordnung anwenden. Oder man kann das Verfahren zunächst separat auf alle individuellen Rangordnungen r_i^* anwenden und dann die Ergebnisse interpretieren. In diesem Abschnitt besprechen wir die zweite Variante. Zur Illustration verwenden wir wiederum die Daten von Diekmann (1982). Tabelle 12.3.6 zeigt die Rangordnungen der insgesamt 10 Personen, die von ihm befragt worden sind. Die Definition der Merkmalskombinationen entspricht den Erläuterungen in Abschnitt 12.1.

2. Der erste Schritt besteht darin, das Verfahren der impliziten Indikatoren für jede der 10 Rangordnungen separat anzuwenden. Tabelle 12.3.7 zeigt die Ergebnisse. Wir haben zuerst den LP-Algorithmus angewendet, um nach Möglichkeit strikt äquivalente Repräsentationen zu finden; sie sind durch eine 1 in der Spalte I gekennzeichnet. Die Werte der Bewertungs-

Tabelle 12.3.7 Bewertungsfunktionen für 10 Rangordnungen.

Person	ϕ_{11}	ϕ_{12}	ϕ_{13}	ϕ_{21}	ϕ_{22}	ϕ_{23}	ϕ_{24}	I
1	0	3	7	0	2	4	5	1
2	0	0	0	0	4.76	4.76	9.87	2
3	0	0	3.35	0	3.35	5.76	8.48	2
4	0	4	6	0	3	5	9	1
5	0	2	7	0	6	9	10	1
6	0	3.11	6.22	0	3.11	4.69	6.22	2
7	0	2	7	0	6	10	14	1
8	0	0	2.87	0	3.35	6.22	8.20	2
9	0	1	2	0	3	5	7	1
10	0	4	7	0	2	3	4	1

funktionen sind durch $\phi_{jk} := \phi_j(\tilde{x}_{jk})$ abgekürzt. Für die übrigen Rangordnungen wurde der ALS-Algorithmus verwendet, und zwar mit Kruskals S-Methode, um Bindungen zu erhalten. In diesen Fällen steht in der Spalte I eine 2 um anzuzeigen, daß nur eine schwache Äquivalenz erreicht werden kann.

3. Die weiteren Überlegungen hängen davon ab, ob man es für sinnvoll hält, die Bewertungsfunktionen metrisch zu interpretieren. In Abschnitt 12.3.3 wurde bereits darauf hingewiesen, daß dies bei Anwendungen des Verfahrens der impliziten Indikatoren in der Marktforschung durchaus üblich ist; und auch Diekmann (1982) folgt dieser Betrachtungsweise. Es ist dann erforderlich, die individuellen Bewertungsfunktionen vergleichbar zu machen. Backhaus u.a. (1987, S. 362) schlagen vor, mit Hilfe der folgenden Formel standardisierte Bewertungsfunktionen zu bilden:

$$\phi_j^*(\tilde{x}_{jk}) = \frac{\phi_j(\tilde{x}_{jk})}{\sum_{l=1}^{m}\sum_{i=1}^{q_l} |\phi_l(\tilde{x}_{li})|} \tag{12.3.2}$$

Als Ergebnis erhält man Bewertungsfunktionen, so daß die Summe ihrer Werte stets gleich 1 ist. Tabelle 12.3.8 zeigt diese standardisierten Bewertungsfunktionen für unser Beispiel. An den Rangordnungen ändert sich natürlich nichts, da es sich um eine lineare Transformation der Bewertungsfunktionen handelt.

4. Wie könnten solche standardisierten Bewertungsfunktionen interpretiert werden? In der Marktforschung ist es üblich, $\phi_j^*(\tilde{x}_{jk})$ als einen „Teilnutzenwert“ zu interpretieren, der dem „Gesamtnutzen“ eines Produkts durch die Merkmalsausprägung $\tilde{x}_{jk}$ hinzugefügt wird. Zugrunde liegt die Vorstellung, daß man durch eine geeignete Wahl von Merkmalsausprägungen Produkte so gestalten kann, daß sie sich in den Rangordnungen vieler potentieller Käufer an einer möglichst hohen Stelle befinden. Bemerkenswert ist, daß auch Diekmann (1982, S. 29f) Überlegungen anstellt, die eine metrische Deutung der Bewertungsfunktionen voraussetzen. Er stellt sich

Tabelle 12.3.8 Standardisierte Bewertungsfunktionen.

Person	ϕ^*_{11}	ϕ^*_{12}	ϕ^*_{13}	ϕ^*_{21}	ϕ^*_{22}	ϕ^*_{23}	ϕ^*_{24}
1	0.00	0.143	0.333	0.00	0.095	0.190	0.238
2	0.00	0.000	0.000	0.00	0.245	0.245	0.509
3	0.00	0.000	0.160	0.00	0.160	0.275	0.405
4	0.00	0.148	0.222	0.00	0.111	0.185	0.333
5	0.00	0.059	0.206	0.00	0.176	0.265	0.294
6	0.00	0.133	0.266	0.00	0.133	0.201	0.266
7	0.00	0.051	0.179	0.00	0.154	0.256	0.359
8	0.00	0.000	0.139	0.00	0.162	0.301	0.397
9	0.00	0.056	0.111	0.00	0.167	0.278	0.389
10	0.00	0.200	0.350	0.00	0.100	0.150	0.200

nämlich die Frage, ob man sagen könne, daß die Befragungspersonen die beiden Aspekte „Bildung“ und „Einkommen“ unterschiedlich gewichten, um ihre individuellen Rangordnungen zu bilden. Er findet drei Gruppen. Eine „bildungsbetonte“ Gruppe (Personen 1 und 10), eine „einkommensbetonte“ Gruppe (Personen 3,7,8,9) und eine „neutrale“ Gruppe (Personen 4,5,6).[19] Zwar verwendet Diekmann, um zu diesem Ergebnis zu gelangen, eine Faktorenanalyse; zu einer ähnlichen Gruppenbildung könnte man indessen unmittelbar durch eine Betrachtung der standardisierten Bewertungsfunktionen gelangen.

5. Eine weitere Frage, die von Diekmann untersucht wird, bezieht sich auf die Unterschiedlichkeit der individuellen Rangordnungen. Diekmann geht dabei von folgender Vorstellung aus:

> „Bei der Konstruktion von Statusskalen wird unterstellt, daß eine Art überindividuelle Statushierarchie existiert, die von den Mitgliedern einer Gesellschaft zumindest in bestimmtem Grade ähnlich wahrgenommen wird. Eine Analyse der individuellen Skalen kann zeigen, wie unterschiedlich die persönlichen Einschätzungen sind.“ (Diekmann 1982, S. 29)

Um die Unterschiede zwischen den individuellen Rangordnungen analysieren zu können, muß man natürlich nicht an die Existenz einer „überindividuellen Statushierarchie“ glauben, sondern man kann sich daran orientieren, wie die Rangordnungen zustande gekommen sind; nämlich durch Rating-Verfahren, bei denen die Befragungspersonen nach ihrem jeweiligen Gutdünken irgendeine Rangordnung angeben können. Die Vorstellung, daß die Befragungspersonen zunächst eine objektiv existierende Rangordnung *wahrnehmen* und dann auf dem Fragebogen angeben, was sie wahrgenommen haben, erscheint sehr unplausibel.[20] Trotzdem kann es natürlich

[19] Die Person mit der Nummer 2 wird von Diekmann aus der Analyse ausgeschlossen.

[20] Diekmann benötigt diese Vorstellung u.a. deshalb, weil er davon sprechen möchte, daß Befragungspersonen bei der Angabe von Rangordnungen „Fehler“ machen können. Tatsächlich modifiziert er die Rangordnungen einiger seiner Befragungspersonen, um sie strikt äquivalent repräsentieren zu können. Er schreibt (S. 26f): „Dieses Verfahren [der Datenmodifikation] ist zulässig, wenn man davon ausgeht, daß die beobachtete Rangfolge mit Meßfehlern behaftet ist.“

sinnvoll sein, Unterschiede zwischen den individuellen Rangordnungen zu analysieren.

6. Dafür ist eine metrische Deutung der Bewertungsfunktionen nicht unbedingt erforderlich. Man kann in einem ersten Schritt dem in Abschnitt 12.3.3 behandelten Gedankengang folgen und untersuchen, ob es lineare Ordnungen für die Teilmerkmalsräume gibt, so daß die von den Befragungspersonen mitgeteilten Rangordnungen konsistente Fortsetzungen sind. Man findet, daß – mit Ausnahme der Person Nr. 2 – alle Rangordnungen nur mit den linearen Ordnungen

$$\tilde{x}_{11} \prec \tilde{x}_{12} \prec \tilde{x}_{13} \quad \text{und} \quad \tilde{x}_{21} \prec \tilde{x}_{22} \prec \tilde{x}_{23} \prec \tilde{x}_{24}$$

verträglich sind, also mit den „natürlich“ erscheinenden Ordnungen für die beiden Merkmalsräume.[21] In einem zweiten Schritt kann man versuchen, mittlere oder Median-Rangordnungen zu berechnen (vgl. Abschnitt 11.4). Wir haben zwei Median-Rangordnungen gefunden. Die erste entspricht der von Person Nr. 7 angegebenen Rangordnung, also

	$\tilde{x}_{21}$	$\tilde{x}_{22}$	$\tilde{x}_{23}$	$\tilde{x}_{24}$
$\tilde{x}_{11}$	1	3	6	9
$\tilde{x}_{12}$	2	5	7	10
$\tilde{x}_{13}$	4	8	11	12

Wir bezeichnen sie mit $\tilde{r}^*$. Die zweite Median-Rangordnung unterscheidet sich nur darin, daß zwischen den Positionen 8 und 9 Indifferenz besteht. Schließlich kann man mit Hilfe der Kemeny-Metrik die Abstände der individuellen von der Median-Rangordnung berechnen:

Person	1	2	3	4	5	6	7	8	9	10
$\mathrm{d}_r(r^*_i, \tilde{r}^*)$	18	15	2	11	6	11	0	6	8	22

Verglichen mit der maximalen Distanz, die in diesem Fall $12 \cdot 11 = 132$ beträgt, gibt es nur geringe individuelle Abweichungen, so daß die Schlußfolgerung mit der von Diekmann (1982, S. 29) übereinstimmt: „daß die individuellen Perzeptionen sehr ähnlich sind.“

[21] Bereits in Abschnitt 12.3.3 wurde darauf hingewiesen, daß sich die von der Person Nr. 2 angegebene Rangordnung überhaupt nicht als konsistente Fortsetzung einer partiellen Ordnung deuten läßt.

Kapitel 13

Varianzmaximierende Indizes

Wenn man Variablen konstruieren möchte, muß man sich in erster Linie Gedanken über ihre Merkmalsräume machen. Ein Teil dieser Aufgabe besteht darin, eine geeignete numerische Repräsentation zu finden, also eine Skala, die in Abschnitt 5.2 als eine Abbildung

$$\phi : \tilde{\mathcal{X}} \longrightarrow \mathbf{R}$$

definiert worden ist. Eine solche Skala ordnet jedem Merkmalswert $\tilde{x} \in \tilde{\mathcal{X}}$ eine Zahl $\phi(\tilde{x}) \in \mathbf{R}$ zu, die dann bei statistischen Berechnungen als numerische Repräsentation des Merkmalswerts dienen soll. Im allgemeinen hängt es von der Art des Merkmalsraums ab, wie geeignete Skalen zu bilden sind und wie sich ihre Werte interpretieren lassen. In der Literatur, die sich mit Skalierungsverfahren beschäftigt, wird dagegen oft ein vollständig anderer Gedankengang verfolgt: Es wird versucht, Skalen zu finden, die im Hinblick auf irgendwelche statistischen Kriterien optimal sind. Hat man z.B. zwei Variablen X und Y, könnte man nach numerischen Repräsentationen für ihre Merkmalsräume suchen, so daß die Variablen maximal korrelieren. Es ist evident, daß dadurch ein Gedankengang entsteht, der mit der ursprünglichen Frage – wie man für einen in seiner inhaltlichen Bedeutung gegebenen Merkmalsraum eine zweckmäßige numerische Repräsentation finden kann – nichts zu tun hat. Ebenfalls ist offensichtlich, daß man dann nicht mehr von einer statistischen Datenanalyse sprechen kann, die sich auf vorab gegebene Daten bezieht. In diesem Kapitel besprechen wir exemplarisch eine Variante der Methode, die sich an der Idee der Varianzmaximierung orientiert.[1] Wir wählen dieses Beispiel auch deshalb, weil sich im nächsten Kapitel zeigen wird, daß der Faktorenanalyse die gleiche Idee zugrunde liegt.

13.1 Der konzeptionelle Ansatz

1. Offenbar gibt es einen Zusammenhang zwischen Variablen- und Skalenkonstruktionen. Hat man eine statistische Variable X mit einem Merkmalsraum $\tilde{\mathcal{X}}$, liefert jede Skala $\phi : \tilde{\mathcal{X}} \longrightarrow \mathbf{R}$ eine neue Variable $\phi(X)$. Dadurch verändert sich allerdings nur die numerische Repräsentation des zunächst gegebenen Merkmalsraums $\tilde{\mathcal{X}}$, so daß die Frage entsteht, welche Gründe es

[1]Es gibt sehr viele Varianten dieses Grundgedankens. Einige von ihnen werden auch durch Statistik-Programme angeboten. Das SPSS-Programm bietet z.B. die Prozeduren ANACOR, HOMALS, PRINCALS und OVERALS an (vgl. SPSS Inc. 1990); das SAS-Programm stellt viele Varianten im Rahmen der Prozedur TRANSREG zur Verfügung (vgl. SAS Institute 1985, S. 1511ff).

überhaupt sinnvoll machen könnten, unterschiedliche Skalen zu verwenden. Eine dieser Überlegungen knüpft an Fragestellungen der Indexkonstruktion an. Ausgangspunkt ist ein Gedanke, der bereits im Zusammenhang mit der Likert-Skalierung (Abschnitt 7.2) angesprochen worden ist: daß Indizes so konstruiert werden sollten, daß sie zwischen Befragungspersonen maximal diskriminieren. Nach Auskunft von Lazarsfeld und Henry wurde dieser Gedanke zuerst von Louis Guttman explizit vorgeschlagen. Den historischen Kontext beschreiben die Autoren so:

„For a long time there had been discussions of how the items of a test should be weighted: by the opinion of competent judges, through bi-serial correlation with a total score, and so on. The full weight of the problem came to the fore only with the appearance of the SSRC monograph on prediction in 1941.[2] It made three major contributions. The main text made clear the distinction between two kinds of 'scales': those for which outside criteria were available and those for which this was not the case. The typical example is concerned with efforts to predict the success of marriage. Studies of this kind require two kinds of classificatory instruments. On the one hand, marriages have to be ordered somehow according to how successful they are. To do this, the investigator has to decide on some criteria of marital success, and has to combine them into one scale which we might call, typically, *intrinsic*. But he also collects information on the state of affairs which existed at the time the marriage was planned: length of engagement, length of time the couple have known each other, similarity of social background, age difference, and so on. These items have to be combined into a predictive instrument to be used to gauge the chances of a successful marriage in the future: this is an *extrinsic* scale because it is derived from its relation to the established scale of success. The main purpose of the study is to get appropriate weights for the items in the predictive scale, based on their correlations with the happiness classification.“ (Lazarsfeld und Henry 1968, S. 7)

2. Die Aufgabenstellung der Studie kann in unserer Terminologie so formuliert werden. Zunächst soll ein Index X^* konstruiert werden, mit dem der „Erfolg von Ehen“ klassifizierend oder quantifizierend erfaßt werden kann. Dann sollen die Werte von X^* durch Rückgriff auf Variablen, die die Umstände der Eheschließung betreffen, voraussagbar gemacht werden; wir nennen diese Variablen $Z_1, \ldots, Z_p$. Wenn X^* ein quantifizierender Index ist, könnte ein einfaches Regressionsmodell

$$\mathrm{M}(X^* \mid Z_1 = z_1, \ldots, Z_p = z_p) \;=\; \alpha + z_1\beta_1 + \cdots + z_p\beta_p$$

ausprobiert werden. Faßt man die rechte Seite formal zu einem Index

$$Z^* \;:=\; Z_1\beta_1 + \ldots + Z_p\beta_p$$

zusammen, kann man den Regressionsansatz auch durch eine Gleichung der Form $\mathrm{M}(X^* \mid Z^* = z^*) = \alpha + z^*$ darstellen. Die Ausführungen von

[2]Paul Horst and others, eds., Prediction of Personal Adjustment. New York, Social Science Research Council, 1941.

Lazarsfeld und Henry beziehen sich auf den unterschiedlichen Status von X^* und Z^*. X^* wird als ein „intrinsischer“ Index bezeichnet, weil es keine „outside criteria“ für seine Bildung gibt.[3] Dagegen wird Z^* ein „extrinsischer“ Index genannt, weil seine Bildung durch die Aufgabe bestimmt werden kann, mit Hilfe der Indikatoren $Z_1, \ldots, Z_p$ möglichst gut die Werte des Index X^* vorauszusagen. Gewichte können in diesem Fall z.B. mit einer OLS-Regression gefunden werden.

3. Offenbar stellen sich in beiden Fällen die Probleme der Indexbildung ganz unterschiedlich. Bei einem „extrinsischen“ Index können optimale Gewichte durch das statistische Modell gefunden werden. Tatsächlich geht es in diesem Fall gar nicht um Indexbildung, denn Z^* hat im Rahmen des Modells nur die Funktion, mit Hilfe der Variablen $Z_1, \ldots, Z_p$ einen „Vorhersagewert“ für X^* zu erzeugen; es ist weder erforderlich noch sinnvoll, Z^* darüber hinaus eine inhaltliche Bedeutung zu geben. Anders verhält es sich jedoch mit X^*. Selbst wenn man sinnvolle Indikatoren für den „Erfolg von Ehen“ finden könnte, ist unklar, ob und ggf. wie sie zu einem Index zusammengefaßt werden könnten, insbesondere wie bei einem additiven Index geeignete Gewichte gefunden werden könnten. An dieser Stelle hat nun Louis Guttman eine neue Möglichkeit aufgezeigt; in den Worten von Lazarsfeld und Henry (1968, S. 7):

„In an important appendix to the SSRC study, Louis Guttman brought to light one basic aspect of intrinsic scales. Because there is no external criterion against which to validate the various items, it is necessary to weight them according to their relations with each other, and this required some mathematical model. Guttman's idea was to weight the items so that the dispersion between the objects to be classified be maximized.“

Die Idee besteht also darin, die Indikatoren so zu einem Index zusammenzufassen, daß der Index zu einer maximalen Diskriminierung zwischen Objekten (in diesem Beispiel Ehepaare) führt. Es ist allerdings offensichtlich, daß diese Idee nichts mit der Frage zu tun hat, ob und ggf. wie einem Index eine inhaltliche Bedeutung gegeben werden kann.

13.2 Das Rechenverfahren

1. In diesem Abschnitt soll überlegt werden, wie ein maximal diskriminierender Index konstruiert werden kann. Das statistische Analogon zur Diskriminierung ist die Varianz. Je größer die Varianz eines Index, desto besser können durch ihn Objekte diskriminiert werden. Es liegt also nahe, einen gewichteten Index zu verwenden und die Gewichte so zu bestimmen, daß sie zu einer Maximierung der Varianz des Index führen. Geht man von

[3] Lazarsfeld und Henry verwenden das Wort 'scale', wo es sich in unserer Terminologie um einen Index handelt; aber diese terminologische Differenz ist unerheblich.

m Indikatoren $X_1, \ldots, X_m$ aus, sieht der Index so aus:

$$X^{**} := \sum_{j=1}^{m} w_j X_j$$

und die Aufgabe besteht darin, die Gewichte $w_1, \ldots, w_m$ so zu wählen, daß die Varianz von X^{**} maximal wird. Das setzt allerdings eine Normierung der Gewichte voraus, denn man kann die Varianz von X^{**} bereits dadurch beliebig groß machen, daß man entsprechend große Gewichte wählt. Wir werden später eine geeignete Normierungsbedingung einführen.

2. Um die Konstruktion eines solchen Index zu besprechen, verwenden wir einige Begriffe der Matrizenrechnung, die im Anhang A näher erläutert werden. Zunächst braucht man natürlich Daten. Wir nehmen an, daß man die Werte der Indikatoren für n Individuen kennt, so daß man sie in einer Datenmatrix

$$\mathbf{X} := (x_{ij}) \qquad i = 1, \ldots, n; \; j = 1, \ldots, m$$

zusammenfassen kann; x_{ij} ist der Wert des Indikators X_j beim Individuum i. Ist eine solche Datenmatrix gegeben, erhält man für jede mögliche Wahl der Gewichte $w_1, \ldots, w_m$ Werte für den Index X^{**}; jeweils einen spezifischen Wert für jedes Individuum. Das kann mit den Mitteln der Matrizenrechnung folgendermaßen ausgedrückt werden. Zunächst werden die Gewichte zu einem Vektor

$$\mathbf{w} := (w_1, \ldots, w_m)'$$

zusammengefaßt. Dann ist das Produkt $\mathbf{x}^{**} := \mathbf{X}\mathbf{w}$ ein Spaltenvektor

$$\mathbf{x}^{**} = (x_1^{**}, \ldots, x_n^{**})'$$

wobei x_i^{**} der Wert des Index für das Individuum i ist, und zwar abhängig von den Gewichten in $\mathbf{w}$.

3. Die weiteren Überlegungen knüpfen an diese Notationen an. Das Interesse gilt der Varianz von X^{**}, also

$$\mathrm{V}(X^{**}) = \frac{1}{n} \sum_{i=1}^{n} (x_i^{**} - \mathrm{M}(X^{**}))^2 \quad \text{wobei} \quad \mathrm{M}(X^{**}) := \frac{1}{n} \sum_{i=1}^{n} x_i^{**}$$

den Mittelwert der Indexwerte bezeichnet. Diese Varianz wird als eine Funktion der Gewichte $w_1, \ldots, w_m$ betrachtet, denn das Ziel besteht darin, diese Gewichte so zu wählen, daß die Varianz maximal wird. Um eine Lösung des Problems zu finden, ist wiederum eine Schreibweise mit Matrizen hilfreich. Zur Vorbereitung drücken wir die Varianz durch die Daten und Gewichte aus. Durch Einsetzen findet man:

$$x_i^{**} - \mathrm{M}(X^{**}) = \sum_{j=1}^{m} x_{ij} w_j - \frac{1}{n} \sum_{k=1}^{n} \sum_{j=1}^{m} x_{kj} w_j$$
$$= \sum_{j=1}^{m} \Big(x_{ij} - \frac{1}{n} \sum_{k=1}^{n} x_{kj}\Big) w_j = \sum_{j=1}^{m} \big(x_{ij} - \mathrm{M}(X_j)\big) w_j$$

Definiert man zur Abkürzung $z_{ij} := x_{ij} - \mathrm{M}(X_j)$, kann man das Ergebnis auch in der Formulierung

$$x_i^{**} - \mathrm{M}(X^{**}) = \sum_{j=1}^{m} z_{ij} w_j$$

schreiben, und man erhält für die Varianz den Ausdruck

$$\mathrm{V}(X^{**}) = \frac{1}{n} \sum_{i=1}^{n} \Big(\sum_{j=1}^{m} z_{ij} w_j \Big)^2$$

Schließlich können die Koeffizienten z_{ij} zu einer neuen Datenmatrix zusammengefaßt werden: $\mathbf{Z} := (z_{ij})$. Sie enthält die Werte der um ihren Mittelwert zentrierten Variablen, d.h. von jedem Wert einer Variablen ist der Mittelwert der Variablen abgezogen worden. Die Varianz des Index kann nun in Matrixnotation durch

$$\mathrm{V}(X^{**}) = \frac{1}{n} \mathbf{w}'\mathbf{Z}'\mathbf{Z}\mathbf{w} \tag{13.2.1}$$

ausgedrückt werden.

4. Wie der Ausdruck (13.2.1) zeigt, kann die Varianz $\mathrm{V}(X^{**})$ schon dadurch beliebig groß gemacht werden, daß man den Vektor $\mathbf{w}$ mit einem beliebig groß gewählten konstanten Wert multipliziert. Um diese Möglichkeit auszuschließen, normieren wir die Varianz $\mathrm{V}(X^{**})$ durch die quadrierte Länge von $\mathbf{w}$, also durch

$$\|\,\mathbf{w}\,\|^2 = \mathbf{w}'\mathbf{w} = \sum_{k=1}^{m} w_k^2 \tag{13.2.2}$$

Das Problem, $\mathrm{V}(X^{**})$ zu maximieren, ist dann äquivalent zu der Aufgabe, die Funktion

$$\lambda(\mathbf{w}) = \frac{\mathbf{w}'\mathbf{Z}'\mathbf{Z}\mathbf{w}}{\mathbf{w}'\mathbf{w}} \tag{13.2.3}$$

für einen Gewichtsvektor $\mathbf{w}$ beliebiger Länge zu maximieren.

5. Eine notwendige Bedingung dafür, daß irgendein bestimmter Vektor ein Maximum der Funktion $\lambda(\mathbf{w})$ liefert, besteht darin, daß die partiellen Ableitungen der Funktion an der Stelle dieses Vektors Null werden müssen. Man muß also die partiellen Ableitungen berechnen. Um die üblichen Regeln für die Bildung von Ableitungen anwenden zu können, ist es zweckmäßig, $\lambda(\mathbf{w})$ ohne Verwendung von Matrizen auszudrücken, also in der Form:

$$\lambda(\mathbf{w}) = \frac{\sum_{i=1}^{n} \big(\sum_{j=1}^{m} z_{ij} w_j \big)^2}{\sum_{j=1}^{m} w_j^2} \tag{13.2.4}$$

Als partielle Ableitungen findet man

$$\frac{\partial \lambda(\mathbf{w})}{\partial w_k} = \frac{2 \sum_{i=1}^{n} \sum_{j=1}^{m} z_{ik} z_{ij} w_j}{\sum_{j=1}^{m} w_j^2} - \frac{2\,\lambda(\mathbf{w}) w_k}{\sum_{j=1}^{m} w_j^2}$$

so daß die Bedingung, daß die partiellen Ableitungen an der Stelle eines Maximums Null werden müssen, in der folgenden Weise ausgedrückt werden kann:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} z_{ik} z_{ij} w_j = \lambda(\mathbf{w})\, w_k \qquad (\text{für } k = 1, \ldots, m) \tag{13.2.5}$$

6. Schließlich ist es zweckmäßig, die Lösungsbedingungen (13.2.5) wiederum mit Hilfe der Matrizenschreibweise darzustellen, also in der Form

$$(\mathbf{Z}'\mathbf{Z})\,\mathbf{w} = \lambda(\mathbf{w})\,\mathbf{w} \tag{13.2.6}$$

Denn in dieser Formulierung sieht man, daß als Lösungen gerade die Eigenvektoren und zugehörigen Eigenwerte der symmetrischen Matrix $\mathbf{Z}'\mathbf{Z}$ in Frage kommen. Jeder Eigenvektor von $\mathbf{Z}'\mathbf{Z}$ liefert einen möglichen Gewichtsvektor $\mathbf{w}$, und der zugehörige Eigenwert liefert $\lambda(\mathbf{w})$. Daraus wird auch unmittelbar klar, wie man ein Maximum findet: Man muß nur nach dem größten Eigenwert von $\mathbf{Z}'\mathbf{Z}$ suchen; der zugehörige Eigenvektor vorgegebener Länge liefert dann die Gewichte, die die Varianz $\mathrm{V}(X^{**})$ maximieren.

13.3 Eine Illustration

1. Um das Rechenverfahren zu illustrieren und über Sinngrenzen nachzudenken, verwenden wir die Klausurdaten aus Abschnitt 6.1. Zunächst folgen wir den einzelnen Rechenschritten.

a) Der erste Schritt besteht darin, die Daten, die bei einer Gesamtheit von Individuen für die Indikatoren ermittelt worden sind, in einer Matrix zusammenzufassen. In unserem Beispiel sind es die in Tabelle 6.1.1

Tabelle 13.1 Eigenwerte und Eigenvektoren der Matrix $\mathbf{Z}'\mathbf{Z}$.

λ_1	λ_2	λ_3	λ_4	λ_5
117.4682	143.5560	264.6131	455.2292	1025.2204
$\mathbf{w}_1$	$\mathbf{w}_2$	$\mathbf{w}_3$	$\mathbf{w}_4$	$\mathbf{w}_5$
0.2847	0.8998	-0.3140	-0.0642	0.0809
0.8473	-0.1869	0.3285	-0.3637	0.0839
0.2217	-0.3917	-0.8867	-0.0636	0.0843
-0.3892	0.0299	-0.0242	-0.8890	0.2379
-0.0210	-0.0325	0.0816	0.2630	0.9606

angegebenen Werte für 46 Personen und 5 Indikatoren (Aufgaben). Es wird also eine $(46, 5)$-Matrix

$$\mathbf{X} := \begin{pmatrix} x_{1,1} & \cdots & x_{1,5} \\ \vdots & & \vdots \\ x_{46,1} & \cdots & x_{46,5} \end{pmatrix}$$

gebildet; x_{ij} ist die Anzahl der Punkte, die die Person i in der Aufgabe j erzielt hat.

b) In einem zweiten Schritt werden die Indikatoren zentriert, d.h. so transformiert, daß sie jeweils den Mittelwert 0 haben. So entsteht eine neue Matrix $\mathbf{Z}$ mit den Elementen

$$z_{ij} := x_{ij} - \mathrm{M}(X_j)$$

Hierbei bezeichnet $\mathrm{M}(X_j)$ den Mittelwert des Indikators X_j.

c) Im nächsten Schritt wird daraus die symmetrische Matrix $\mathbf{A} := \mathbf{Z}'\mathbf{Z}$ gebildet. Sie sieht in unserem Beispiel folgendermaßen aus:

$$\mathbf{A} = \begin{pmatrix} 160.4348 & -5.5217 & 39.3478 & 38.5652 & 60.3043 \\ -5.5217 & 185.3261 & -26.7174 & 126.0217 & 44.9348 \\ 39.3478 & -26.7174 & 244.9783 & 40.1522 & 57.5435 \\ 38.5652 & 126.0217 & 40.1522 & 435.9348 & 128.1957 \\ 60.3043 & 44.9348 & 57.5435 & 128.1957 & 979.4130 \end{pmatrix}$$

d) Schließlich werden die Eigenwerte und zugehörigen Eigenvektoren der Matrix $\mathbf{A}$ berechnet. Tabelle 13.1 zeigt die Ergebnisse. $\lambda_1, \ldots, \lambda_5$ sind die Eigenwerte, $\mathbf{w}_1, \ldots, \mathbf{w}_5$ die zugehörigen Eigenvektoren, die für die Darstellung so normiert wurden, daß sie die euklidische Länge 1 haben.

2. Damit hat man die meisten der erforderlichen Rechnungen durchgeführt. Wie im vorangegangenen Abschnitt gezeigt worden ist, erhält man

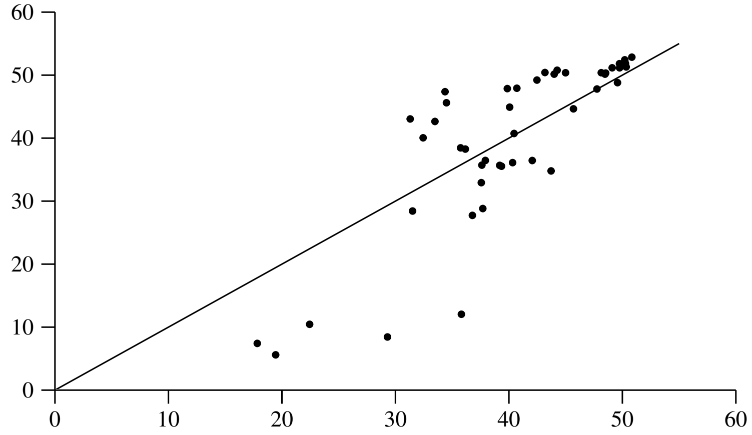


Abb. 13.3.1 Streuungsdiagramm für die Indizes X^* (X-Achse) und X^{***} (Y-Achse).

einen die Varianz maximierenden Index dadurch, daß als Gewichte die Werte desjenigen Eigenvektors verwendet werden, der zum größten Eigenwert gehört. In unserem Beispiel ist der maximale Eigenwert $\lambda_5 = 1025.22$ und der zugehörige Eigenvektor ist $\mathbf{w}_5$. Einen varianzmaximierenden Index erhält man also durch

$$X^{**} := 0.0809\, X_1 + 0.0839\, X_2 + 0.0843\, X_3 + 0.2379\, X_4 + 0.9606\, X_5$$

Dieser Index kann jedoch noch mit einer beliebigen Konstanten multipliziert werden. Das kann man ausnutzen, um den neuen Index möglichst gut mit dem ungewichteten Index X^* (vgl. Tabelle 6.1.1) vergleichbar zu machen. Eine Möglichkeit ist die folgende. Durch den Index X^* werden insgesamt 1854 Punkte vergeben. Der Index X^{**} vergibt jedoch nur 670.5 Punkte. Also multiplizieren wir die Gewichte mit $1854/670.5$ und erhalten so den neuen Index

$$X^{***} := 0.2237\, X_1 + 0.2319\, X_2 + 0.2331\, X_3 + 0.6579\, X_4 + 2.6561\, X_5$$

Jetzt können auch die Varianzen sinnvoll verglichen werden. Die Varianz von X^* beträgt 65.5, die von X^{***} beträgt 170.4.

3. Es ist aufschlußreich, die durch die Indizes X^* und X^{***} vorgenommenen Bewertungen zu vergleichen. Das geschieht in Abb. 13.3.1. Man erkennt, daß sich viele Bewertungen durch die neuen Gewichte erheblich verändern. Zum Beispiel hat der Klausurteilnehmer, der sich im Streuungsdiagramm unten rechts befindet, die Punktzahlen $(10, 10, 5, 10, 0)$, also außer in der letzten Aufgabe stets die höchste Punktzahl erreicht, insgesamt 35 Punkte. Der neue Index gibt jedoch der letzten Aufgabe ein besonders großes Gewicht, so daß die neue Punktzahl nur noch bei 13 liegt.

4. Das Beispiel zeigt, daß das rein statistische Kriterium der Varianzmaximierung keineswegs zu einem besseren Index führt. Der neue Index erlaubt zwar eine größere Diskriminierung der Klausurteilnehmer; aber dies hat offenbar nichts mit der Frage zu tun, was durch einen Index sinnvoll erfaßt werden kann. In unserem Beispiel ist das leicht zu erkennen, denn das Ergebnis des Skalierungsverfahrens kann in diesem Fall mit einer vorgegebenen inhaltlich begründbaren Quantifizierung verglichen werden. In der Regel ist das jedoch nicht der Fall, sondern man hat es ausschließlich mit einer durch das Skalierungsverfahren erzeugten Quantifizierung zu tun. Es sollte auch darauf geachtet werden, daß Indizes, die mit einem Prinzip der Varianzmaximierung erzeugt werden, stets verteilungsabhängige Indizes sind (vgl. Abschnitt 5.3): Der Indexwert für jedes einzelne Individuum hängt dann auch von den Werten ab, die bei allen übrigen Individuen realisiert worden sind. In unserem Beispiel heißt das, daß das durch X^{***} konstruierte Klausurergebnis jedes einzelnen Teilnehmers auch davon abhängt, welche Ergebnisse die Gesamtheit der Klausurteilnehmer erzielt hat. Von einem Meßverfahren zur Begründung eines sozialwissenschaftlichen Größenbegriffs kann also sicherlich nicht gesprochen werden.

Kapitel 14

Faktorenanalyse

Aus zwei Gründen ist es erforderlich, daß wir uns auch mit einigen Aspekten der Faktorenanalyse beschäftigen. Erstens, weil sie entsprechend unserer Unterscheidung nicht zu den statistischen, sondern zu den datenerzeugenden Methoden gehört. Methoden der Faktorenanalyse dienen nicht dazu, Zusammenhänge zwischen empirischen Variablen zu analysieren, sie werden vielmehr verwendet, um aus empirischen Variablen neue „latente“ Variablen, sog. „Faktoren“, zu konstruieren. K. Überla hat das folgendermaßen formuliert:

> „Das Hauptziel der Faktorenanalyse ist die Ableitung hypothetischer Größen oder Faktoren aus einer Menge beobachteter Variablen. Die Faktoren sollen möglichst einfach sein und die Beobachtungen hinreichend genau beschreiben und erklären.“ (Überla 1977, S. 3)

Etwas allgemeiner heißt es bei Comrey und Lee:

> „Factor analytic methods can help scientists to define their variables more precisely and decide what variables they should study and relate to each other in the attempt to develop their science to a higher level.“ (Comrey/Lee 1992, S. 1)

Hier deutet sich auch der zweite Grund an, warum wir uns mit Methoden der Faktorenanalyse beschäftigen: Sie werden oft – z.B. bei der Likert-Skalierung – empfohlen, um Indexkonstruktionen zu überprüfen. In beiden Fällen besteht der Anspruch darin, daß Methoden der Faktorenanalyse der wissenschaftlichen Begriffsbildung dienen können. R. B. Cattell, ein Wegbereiter der Faktorenanalyse, sagt das in folgender Weise:

> „Factor analytic investigation has its function predominantly [...] in *basic* research to provide the measurement foundations for later special problems in pure and applied research.“ (Cattell 1952, S. 17)

Wir werden in diesem Kapitel folgendermaßen vorgehen: In Abschnitt 14.1 beginnen wir mit einigen allgemeinen Überlegungen, um zu verstehen, wie von „latenten Variablen“ und „Faktoren“ gesprochen wird. Es folgen zwei Abschnitte, die sich mit mathematischen Aspekten der Faktorenanalyse beschäftigen. Die erforderliche Mathematik ist tatsächlich weniger kompliziert als es auf den ersten Blick erscheinen könnte.[1] Abschnitt 14.2 beginnt mit der *einfachen Faktorenanalyse*, bei der nur eine latente Variable

[1] Zum Beispiel heißt es bei Comrey und Lee 1992, S. 2: „Factor analysis represents a class of procedures for treating data that are being applied with increasing frequency in the behavioral and health sciences and even in the physical sciences. Despite this popularity, however, factor analysis is not well understood by many of the scientists who use it. There are many reasons for this phenomenon. First of all, factor analysis is based on mathematical foundations that go beyond the level of quantitative sophistication and

konstruiert werden soll. Dabei wird sich zeigen, daß in diesem Fall die Faktorenanalyse und die Konstruktion eines varianzmaximierenden Index äquivalent sind. Abschnitt 14.3 behandelt die *multiple Faktorenanalyse*, die dazu dienen soll, zwei oder mehr latente Variablen zu konstruieren. Schließlich beschäftigt sich Abschnitt 14.4 mit dem Anspruch, daß Methoden der Faktorenanalyse auch der wissenschaftlichen Begriffsbildung, insbesondere der Überprüfung von Indexkonstruktionen, dienen können. Diese Überlegungen werden durch einige Bemerkungen zum Dimensionsbegriff in Abschnitt 14.5 ergänzt.

Es sei schon an dieser Stelle darauf hingewiesen, daß es *die* Faktorenanalyse nicht gibt. Vielmehr umfaßt dieser Name eine Vielzahl unterschiedlicher Verfahren, mit denen mehr oder weniger willkürlich „latente Variablen" bzw. „Faktoren" konstruiert werden. Auf die vielen unterschiedlichen Verfahren, die von Psychologen, Sozialwissenschaftlern und Statistikern ausgedacht worden sind, werden wir indessen nicht näher eingehen, sondern uns auf die Diskussion einiger Grundgedanken beschränken.

14.1 Latente Variablen

1. Durch die Entwicklung von Methoden der Faktorenanalyse hat sich ein Reden von „latenten Variablen" und „Faktoren" verbreitet, das nicht ohne weiteres verständlich ist. In diesem Abschnitt soll einigen Aspekten dieser Rhetorik etwas näher nachgegangen werden. Wir beginnen mit einer Erläuterung, die von Bortz und Döring gegeben worden ist:

„Wir sprechen ferner von einer *manifesten Variablen*, wenn eine Variable direkt beobachtbar ist (z.B. Anzahl gelöster Testaufgaben), und von einer *latenten Variablen*, wenn wir annehmen, daß sie einer manifesten Variablen als hypothetisches Konstrukt zugrunde liegt. Im Beispiel wäre Intelligenz eine latente Variable, deren Ausprägungen wir über die manifeste Variable ‘Anzahl gelöster Testaufgaben’ erschließen." (Bortz und Döring 1995, S. 6)

Hiernach scheint es so, als wäre die Unterscheidung zwischen manifesten und latenten Variablen im wesentlichen mit der Unterscheidung zwischen empirischen und konstruierten Variablen identisch, die in Abschnitt 5.1 eingeführt worden ist. In gewisser Weise ist dies richtig: Latente Variablen sind konstruierte Variablen (Indizes), die aus Indikatoren gebildet werden. Es ist jedoch nicht üblich, jede beliebige konstruierte Variable eine latente Variable zu nennen. Zum Beispiel ist es nicht üblich, Preisindizes latente Variablen zu nennen. Um herauszufinden, welche Vorstellungen mit dem Reden von latenten Variablen verbunden werden, sollte also zunächst der

training possessed by the bulk of behavioral scientists who use it the most." Man wird jedoch sehen, daß die erforderlichen mathematischen Überlegungen relativ leicht nachvollziehbar sind und daß auch kein Grund besteht, von „mathematischen Grundlagen" (= Begründungen?) zu sprechen.

Frage nachgegangen werden, wie der Zusammenhang zwischen Indikatoren und Indizes zu verstehen ist. Denn wenn Indizes als latente Variablen angesprochen werden, verbirgt sich dahinter zumeist eine sehr eigentümliche Ansicht.

2. Unsere Auffassung kann einfach formuliert werden: Indizes sind konstruierte Variablen, die – wenn überhaupt – eine Bedeutung nur aus ihrem Konstruktionsverfahren und den dafür verwendeten Indikatoren gewinnen können. Dasselbe gilt dann natürlich auch für latente Variablen, worin wir insoweit nur eine andere Bezeichnung für konstruierte Variablen sehen können. — Eine ganz andere Auffassung kommt jedoch in der oft verwendeten Rhetorik zum Ausdruck, daß mit Hilfe latenter Variablen die Beschaffenheit der uns empirisch zugänglichen Welt *erklärt* werden könne. Diese Auffassung deutet sich z.B. in folgendem Zitat von Black an:

„Social class can be considered an example of a rather special type of variable, one that is based on several individual characteristics such as occupation, education and income. It is typical of what Kerlinger (1986) calls *latent variables*. These tend to be of greater importance in the social sciences than the single constituent variables because they are seen to be better suited for explaining phenomena and their relationships on a larger scale. Another latent variable whose origins are not found in straight factual information, but in more elaborate instruments, is IQ, tests for which are commonly based on three components: verbal, numerical and spatial ability." (Black 1999, S. 211)

3. Aber wie können latente Variablen irgendetwas erklären? Schnell, Hill und Esser (1999, S. 127) erläutern den aus ihrer Sicht „fruchtbarsten Ansatz", den sie als „kausal-analytischen Ansatz" bezeichnen, mit folgenden Ausführungen:

„Der kausal-analytische Ansatz basiert ebenfalls auf der Annahme latenter Variablen. Indikatoren werden in diesem Ansatz aber als beobachtbare ‘Folgen’ der latenten Variablen aufgefaßt. Theorien werden als Aussagen über die kausalen (ursächlichen) Wirkungen latenter Variablen aufeinander interpretiert. Indikatoren für die latenten Variablen können aus solchen Theorien theoretisch gefolgert werden, da angenommen wird, daß die Phänomene, die durch die Indikatoren erfaßt werden, durch die latenten Variablen tatsächlich verursacht werden."

Ähnliche Formulierungen finden sich auch bei anderen Autoren; hier zur Illustration noch ein Zitat von McCutcheon:

„Many concepts in the social sciences cannot be observed directly. We cannot, for example, directly observe authoritarianism, economic development, racial prejudice, alienation, anomie, or religious commitment. Indeed, there are hundreds of other theoretically interesting concepts for which the available measures are assumed to be imperfect indicators. For example, while we cannot directly observe religious commitment, we are likely to believe that a high level of commitment leads people to attend church more frequently, pray more frequently, conduct themselves with greater behavioral orthodoxy, report that their religious beliefs are of great importance to them, and so forth. Because we believe that each of the observed indicators is caused by an unobserved, or *latent*, variable of interest, we expect covariation among the observed measures, and we study the patterns of interrelationships among the observed indicators to understand and characterize the underlying latent variable.“ (McCutcheon 1987, S. 5)

4. Man sollte sich überlegen, ob und ggf. wie mit einer konstruierten (latenten) Variablen überhaupt irgendetwas erklärt werden kann. Zunächst ist klar, daß Variablen (welcher Art auch immer) nicht *als Ursachen* empirischer Phänomene verstanden werden können. Variablen sind stets Teil einer Sprache, mit der man sich auf empirische Realität bezieht. Weder sind sie noch „bezeichnen“ sie Dinge oder Sachverhalte oder Prozesse, auf die man sich beim Reden über Wirkungszusammenhänge zu beziehen hätte. Somit können sie *als Variablen* auch nichts verursachen. Dennoch kann man Variablen für „statistische Erklärungen“ verwenden. Man kann untersuchen, wie der statistische Zusammenhang zwischen zwei oder mehr Variablen beschaffen ist, indem man ihre gemeinsame Verteilung analysiert; man kann z.B. die Korrelationen zwischen zwei oder mehr Variablen ermitteln.[2] Ob und ggf. in welcher Weise daraus Einsichten in Wirkungszusammenhänge gewonnen werden können, soll an dieser Stelle nicht diskutiert werden. Wichtig ist jedoch der Hinweis, daß man sich bei solchen statistischen Untersuchungen stets auf empirische Variablen beziehen muß, jedenfalls dann, wenn man an der Beschaffenheit der empirischen Realität interessiert ist. Zwar folgt daraus nicht, daß für statistische Analysen ausschließlich empirische Variablen verwendet werden können. Sicherlich kann man aus empirischen Variablen neue Variablen konstruieren und dann die konstruierten Variablen verwenden; das geschieht z.B. bei der Regressionsrechnung. Aber soweit es sich um statistische Analysen handelt, dienen die konstruierten Variablen schließlich nur dazu, in einfacherer Weise über Zusammenhänge zwischen empirischen Variablen sprechen zu können. In den oben angeführten Zitaten kommt jedoch eine andere Auffassung zum Ausdruck. „Latente Variablen“ werden nicht als mehr oder weniger zweckmäßige Konstrukte aufgefaßt, die helfen können, Einsichten in Zusammenhänge zwischen empirischen Variablen zu gewinnen. Sondern sie werden als „Faktoren“ aufgefaßt, die ihren Indikatoren „ursächlich“ zugrundeliegen.

5. Überlegen wir uns das Problem noch einmal anhand unserer Terminologie für Indizes. Den Ausgangspunkt bilden empirische Variablen, und man kann sie natürlich verwenden, um daraus neue Variablen zu konstruieren, z.B. einen Index $X^* := g(X_1, \ldots, X_m)$. Dabei sind $X_1, \ldots, X_m$ empirische Variablen, und X^* ist der daraus gebildete Index, den man jetzt auch als eine „latente Variable“ bezeichnen könnte. Die Frage ist, wie der Zusammenhang zwischen den empirischen Variablen und dem Index aufzufassen ist. Hält man sich daran, wie der Index tatsächlich zustande kommt, gibt es eine einfache Antwort: Indizes werden von Sozialforschern konstruiert, sie erfinden diese Indizes und geben ihnen eine Bedeutung (oder manchmal auch nur einen Namen). Folgendes Bild kann zur Illustration dienen:

$$\begin{array}{ccc} X_1 & \searrow & \\ \vdots & & X^* \\ X_m & \nearrow & \end{array}$$

Die Pfeile gehen von den empirischen Variablen zur konstruierten Variablen, um anzudeuten, daß X^* durch die empirischen Variablen definiert wird und eine Bedeutung – wenn überhaupt – nur aus diesen empirischen Variablen und dem verwendeten Konstruktionsverfahren bekommen kann. Der Rhetorik, die von „latenten Variablen“ als „Faktoren“ oder als „Ursachen“ für empirische Phänomene spricht, liegt dagegen ein anderes Bild zugrunde:

$$\begin{array}{ccc} & \nearrow & X_1 \\ X^* & & \vdots \\ & \searrow & X_m \end{array}$$

Hier sollen die Pfeile andeuten, daß die „latente Variable“ X^* einen Einfluß auf die Phänomene ausübt, die durch die empirischen Variablen erfaßt werden. Aber diese Vorstellung ist nicht nur deshalb fragwürdig, weil Variablen nicht sinnvoll als Ursachen anderer Variablen aufgefaßt werden können; sie ist auch irreführend, weil die Variable X^* durch die empirischen Variablen $X_1, \ldots, X_m$ *definiert* wird, also ihre Bedeutung nur aus diesen Variablen gewinnen kann.

6. Wenn man die Pfeile, um im Bild zu bleiben, von der „latenten Variablen“ zu ihren Indikatoren richtet, ist es im allgemeinen auch kaum möglich, irgendeine sinnvolle Interpretation anzugeben. Man denke an das Klausur-Beispiel, in dem ein Index konstruiert wurde, um die Klausurergebnisse zu bewerten. Es erscheint absurd, in diesem Fall die erzielte Punktzahl X^* als Ursache dafür anzusehen, wie gut oder schlecht die einzelnen Aufgaben gelöst worden sind. Aber was wäre dann von einer Rhetorik zu halten, die X^* als Quantifizierung einer „Fähigkeit zum Lösen von Statistik-Aufgaben“ interpretiert und daraufhin mit einem Verweis auf so ermittelte Fähigkeiten erklärt, wie die Klausurergebnisse ausgefallen sind? Eine scheinbare Analogie zu umgangssprachlichen Erklärungen ist tatsächlich trügerisch. Wenn man sich umgangssprachlich auf Fähigkeiten bezieht, geht es im allgemeinen nicht um Kausalerklärungen. Von jemandem zu sagen, er habe die Fähigkeit, Aufgaben einer gewissen Art

[2] Es sei jedoch angemerkt, daß die Korrelationsrechnung nur sehr beschränkte Einsichten vermitteln kann und keineswegs der einzige oder beste Zugang zur Ermittlung von „Zusammenhängen zwischen Variablen“ ist.

zu lösen, *bedeutet* eben dies: daß er Aufgaben dieser Art meistens lösen kann. Also kann man zum Beispiel in einer bestimmten Situation sagen: Er hat auch *diese* Aufgabe gelöst, weil er gelernt hat, Aufgaben *dieser Art* zu lösen. Aber das Wort 'weil' gewinnt in diesem Zusammenhang seinen Sinn aus einer Subsumtion, nicht aus der Unterstellung eines Kausalzusammenhangs.[3]

7. Diese Überlegung ist wichtig, weil von „latenten Variablen" meistens dann gesprochen wird, wenn man sich, wie im eben angeführten Beispiel, auf *dispositionale* Eigenschaften beziehen möchte. Wenn das nicht der Fall ist, wie zum Beispiel bei einem Preisindex, ist nämlich meistens das Konstruktionsverfahren leicht durchschaubar, und es gibt wenig Ansatzpunkte für eine mystifizierende Rhetorik. Man sieht dann, wie ein Index aus empirischen Variablen konstruiert wird, und es wird dadurch evident, daß der Index infolgedessen den gleichen logischen Status hat wie die empirischen Variablen, aus denen er gebildet wird. Insbesondere wird dann deutlich, daß es ganz sinnlos wäre, von „unbeobachteten Variablen" zu reden. Zum Beispiel ist der Anteil der Mietausgaben am Haushaltseinkommen kein „unbeobachtbarer Sachverhalt", sondern dieser Index bezieht sich ebenso auf einen empirischen Sachverhalt, wie die zu seiner Definition verwendeten empirischen Variablen. Schließlich ist es in diesen Fällen auch evident, daß konstruierte Variablen nicht zur Erklärung derjenigen Sachverhalte verwendet werden können, auf die sich die zu ihrer Bildung verwendeten Indikatoren beziehen. — All diese einfach durchschaubaren Überlegungen werden jedoch obskur, wenn man versucht, sich mit konstruierten Variablen auf dispositionale Eigenschaften zu beziehen.

8. Die für die Rhetorik „latenter Variablen" typische Bezugnahme auf dispositionale Eigenschaften wird verständlich, wenn man daran denkt, daß sie in ihrem historischen Entstehungszusammenhang eng mit der Entwicklung der Faktorenanalyse verbunden ist, die zunächst von Psychologen erfunden wurde, um Fähigkeiten als dispositionale Eigenschaften quantifizierbar zu machen. Als historischer Ausgangspunkt wird meistens eine Arbeit von Charles Spearman (1904) genannt, in der dieser versucht hat, einen als „general intelligence" interpretierbaren „Faktor" zu konstruieren. In der weiteren Entwicklung wurden zwar die Anwendungsfelder der Faktorenanalyse zunehmend ausgeweitet; aber die Vorstellung, mit faktoranalytischen Methoden dispositionale Eigenschaften als „Faktoren" konzipieren zu können, blieb erhalten. Folgendes Zitat von L. Thurstone, einem wichtigen Wegbereiter der Faktorenanalyse, liefert einen Hinweis, wie auf dispositionale Eigenschaften Bezug genommen wird.

[3] Diese Verwechslung durchzieht insbesondere die sozialpsychologische Literatur, die versucht, Verhalten von Menschen durch Dispositionen zu erklären; man vgl. exemplarisch Ajzen 1988.

„The factorial methods have been developed primarily for the purpose of analyzing the relations of human traits. These are defined as follows:

Definition 1. A *trait* is any attribute of an individual.

[. . .] It is useful to distinguish between those traits which are descriptive of the individual as he appears to others and those traits which are exemplified primarily in the things that he can do. This distinction is involved in the definition of 'ability'.

Definition 2. An *ability* is a trait which is defined by what an individual can do.

This definition implies that there are as many abilities as there are enumerable things that individuals can do. Each ability is therefore objectively defined in terms of a specified task and of a specified method of appraising it.

Definition 3. The task, together with the method of appraising it, which defines an ability, is called a *test*.

Definition 4. The numerical evaluation of a test performance is called a *score*.

[. . .] The application of the factorial methods in science rests on a fundamental postulate.

Postulate. The standard scores[4] of all individuals in an unlimited number of abilities can be expressed, in first approximation, as linear functions of their standard scores in a limited number of abilities." (Thurstone 1947, S. 62f)

Hier erkennt man noch sehr deutlich die Intention, mit Hilfe der Faktorenanalyse eine gewisse Anzahl grundlegender „abilities" zu ermitteln, mit denen dann im Prinzip alles, was Menschen tun können, erklärt werden soll. Aber der Begriff 'ability' ist schon so unbestimmt („anything that individuals can do"), daß der konzeptionelle Ansatz auf beliebiges Verhalten anwendbar erscheint. Als einzige Unterscheidung bleibt – wie im eingangs angeführten Zitat von Bortz und Döring – eine Gegenüberstellung von „manifest" und „latent", wobei jedoch das Wort 'latent' seine potentielle Bedeutung nur durch eine Konnotation zum Wort 'ability' – oder ganz abstrakt: 'dispositional' – gewinnt. Dem entspricht eine bis heute gängige Rhetorik, wie z.B. in folgendem Zitat von Roskam:

„Ein 'latent trait' ist eine Fähigkeit oder eine Disposition für ein bestimmtes Verhalten. *Latent traits* sind somit Variablen, nach deren Ausprägung Personen klassifiziert werden können. Beispiele hierfür sind etwa die Dispositionen 'Toleranz gegenüber Fremden' und 'Angst vor Atomkraftwerken' sowie die Fähigkeit 'Sprachgewandtheit'. Klassische Methoden zur Messung derartiger Variablen basieren auf Fragen oder Aussagen, auf welche die Probanden mit 'ja' oder 'nein', 'zustimmend', 'unentschieden' oder 'ablehnend' und dergleichen mehr antworten können." (Roskam 1996, S. 431)

[4] [Gemeint sind "scores", die man als Werte einer Variablen standardisiert hat, so daß die Variable den Mittelwert 0 und die Varianz 1 bekommt; vgl. Abschnitt 5.4.]

„Fähigkeiten“ oder „Dispositionen“ sind in diesen Ausführungen offenbar nur andere Worte für Thurstones „abilities“.[5]

9. Das zuletzt angeführte Zitat deutet auch an, wie sich semantische Assoziationen zu „Einstellungen“ gewinnen lassen. Denn was auch immer man darunter verstehen möchte, bedeutsam erscheint, daß es sich um Dispositionen handelt. In einer berühmten Definition von Allport heißt es:

„An attitude is a mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related.“ (Allport 1935, S. 810)

M. Fishbein gab folgenden Kommentar:

„Im Jahr 1935 verschaffte sich Gordon Allport einen Überblick über den allgemeinen Bereich von Einstellungstheorie und Einstellungsforschung. Nachdem Allport mehr als hundert verschiedene Definitionen von Einstellung untersucht hatte, folgerte er, daß die meisten Forscher die Einstellung übereinstimmend als eine gelernte Prädisposition ansehen, mit der auf ein Objekt oder eine Klasse von Objekten beständig positiv oder negativ reagiert wird. Er stellte weiterhin heraus, daß diese Bipolarität in der Gerichtetheit der Einstellung (d.h. positiv vs. negativ) häufig als die am meisten ausgeprägte Eigenschaft des Konzepts betrachtet wird. Deshalb wurde Einstellung als ein einfaches eindimensionales Konzept begriffen.“ (Fishbein 1967, S. 149)

Fishbein kritisiert dann im weiteren diese „Eindimensionalität“ im bisher dominierenden Verständnis von „Einstellungen“. Wichtig ist an dieser Stelle jedoch nur der Hinweis, daß „Einstellungen“ (auch von Fishbein) als dispositionale Eigenschaften aufgefaßt werden und sich insoweit Thurstones abstraktem Begriff der „abilities“ subsumieren lassen. Das trifft insbesondere für die statistische Einstellungsforschung zu, die davon ausgeht, daß Einstellungen durch Tests erfaßt werden können, die zeigen, wie auf „Einstellungsobjekte“ reagiert wird. Faktorenanalytische Methoden sind infolgedessen nicht nur bei der Erfassung von Fähigkeiten im engeren Sinne des Wortes verbreitet, sondern ebenso im Bereich der Einstellungsforschung.

10. Halten wir fest, daß mit der Rhetorik „latenter Variablen“ meistens auf dispositionale Eigenschaften verwiesen werden soll.[6] Wir möchten zeigen, daß aus dieser Intention einige spezifische Probleme resultieren (im

[5] In diesem Zitat zeigt sich auch die verbreitete Neigung, das Wort 'trait' ganz undifferenziert zum Verweis auf Dispositionen zu verwenden. So auch bei Kerlinger 1964, S. 483: „A *trait* is an enduring characteristic of the individual to respond in a certain manner in all situations.“

[6] Man muß hier einschränkend von „meistens“ sprechen. Denn soweit es sich bei faktoranalytischen Methoden um reine Rechenverfahren hanndelt, kann man mit ihrer Hilfe beliebige „latente Variablen“ konstruieren. Eine Interpretation der Variablen resultiert nicht aus dem Verfahren, sondern entspringt, wie wir noch genauer sehen werden, stets der Phantasie des Variablenkonstrukteurs.

Unterschied zu den generellen Problemen, die mit jeder Art von Indexkonstruktion verbunden sind). Einen Hinweis liefert die Frage, wie dispositionale Eigenschaften erfaßt werden können. Denken wir zum Beispiel daran, daß Zucker in Wasser löslich ist. Man stellt das fest, indem man prüft, wie sich Zucker verhält, wenn er mit Wasser zusammengebracht wird: Er löst sich auf. Soweit bekannt ist, verhält es sich immer so, und deshalb kann man von einer dispositionalen Eigenschaft des Zuckers sprechen.[7] Es ist allerdings klar, daß ein Hinweis auf diese dispositionale Eigenschaft des Zuckers keine Erklärung dafür liefert, *warum* sich Zucker in Wasser auflöst. Der Grund dafür ist unmittelbar einsichtig: diese dispositionale Eigenschaft wurde dadurch *definiert*, daß sich Zucker in Wasser auflöst. Wenn man verstehen will, warum sich Zucker in Wasser auflöst, müßte man nicht nur die Beschaffenheit von Zucker untersuchen, sondern auch die von Wasser und die Interaktionsprozesse, die beim Zusammenbringen von Zucker und Wasser einsetzen.

11. Ein zunächst vollständig vergleichbares Problem tritt auf, wenn Psychologen oder Sozialwissenschaftler versuchen, mit Hilfe von Tests dispositionale Eigenschaften von Menschen zu erfassen. Thurstone hat das in dem oben angeführten Zitat deutlich gesagt: Der Test *definiert* die dispositionale Eigenschaft. Dann aber wird es sinnlos, mit den so definierten dispositionalen Eigenschaften erklären zu wollen, warum Menschen, wenn sie einem Test ausgesetzt werden, jeweils dies oder jenes Verhalten zeigen. Zur Illustration sei auf ein Beispiel verwiesen, das von Thurstone verwendet worden ist.

„The factorial methods have for their object to isolate the primary abilities by objective experimental procedures so that it may be a question of fact how many abilities are represented in a set of tasks, and whether a particular objective performance represents an ability that is in some fundamental sense primary. [...]
The first simplifying assumption of the factorial methods is that, if the objective performance of a task requires a certain number of fundamental or primary abilities, then the performance can be expressed, in first approximation, as a linear function of the primaries. The interpretation of a linear function in this context can be illustrated by a hypothetical example. Consider an arithmetical task which requires two hypothetical abilities that may be called 'number speed' and 'numerical reasoning'. One subject may be superior in setting up the problems, but he may be slow in the numerical calculation. Another subject may be slow in reasoning out the problem, but he may be fast in the calculation. Other subjects may be fast in both, or slow in both. In the factorial analysis an objective performance is stated as a weighted sum of two contributions, one for each of the two abilities that are involved in the task. If the task involves

[7] Zu einer dispositionalen Eigenschaft gehört natürlich immer ein Kontext, in dem sie sich zeigen kann. Schon die Zurechnung dispositionaler Eigenschaften ist deshalb problematisch. In unserem Beispiel könnte man auch von einer dispositionalen Eigenschaft des Wassers sprechen: seiner Fähigkeit, Zucker auflösen zu können.

very little calculating but mostly reasoning, then the weight for the first factor is smaller than for the second factor. The assumption that the performance can be regarded as a sum of the contributions of the two abilities is represented in the expression

$$s = a_1x_1 + a_2x_2$$

in which s represents the standard score of an individual in the arithmetical task, x_1 and x_2 represent the standard scores of the individual in the two fundamental abilities, and a_1 and a_2 are the weights of the two abilities in the arithmetical task. If we know to what extent each of the two abilities are required in the task, the weights a_1 and a_2, and if we know how a person scores in each of the two fundamental abilities that are involved in the task, x_1 and x_2, then the objective performance s can be predicted.“ (Thurstone 1938, S. 1ff)

Der Gedankengang gewinnt eine scheinbare Plausibilität durch die – im umgangssprachlichen Reden oft gemachte – Annahme, daß man unabhängig von den Ergebnissen jeweils bestimmter Tests davon sprechen könne, daß Menschen gewisse Fähigkeiten haben. Ob es sich um eine sinnvolle Annahme handelt, sei dahingestellt. Sie verändert jedoch ihren Charakter, wenn der umgangssprachliche Kontext verlassen und zunächst von „abilities“ und dann von „primary abilities“ gesprochen wird. Denn weder „abilities“ noch „primary abilities“ werden durch Erfahrungen, wie sie dem umgangssprachlichen Reden zugrunde liegen, expliziert, sondern *durch Tests definiert*. Man kann sich deshalb mit den so definierten „primary abilities“ gar nicht mehr auf unabhängig von den Tests explizierbare Fähigkeiten beziehen. In der Terminologie von Thurstone heißt das: x_1 und x_2 werden durch s definiert. Dann aber wird es sinnlos, wie schon bei der Löslichkeit von Zucker in Wasser, durch einen Verweis auf Werte von x_1 und x_2 die jeweils erzielten Werte der Variablen s zu erklären. Die „latenten Variablen“ x_1 und x_2 liefern nicht nur keine Kausalerklärung, wie der „kausal-analytische Ansatz“ behauptet, sie liefern in keinerlei Weise eine Erklärung für die Ergebnisse jeweils bestimmter Tests. Denn die „latenten Variablen“ *entstehen* nur dadurch, daß jeweils durch Tests ermittelte empirische Variablen einer (im übrigen weitgehend beliebigen) mathematischen Transformation unterzogen werden.[8] Insofern kann nicht einmal

[8]Schwierigkeiten beginnen bereits, wenn empirische Variablen *als Indikatoren* interpretiert werden. Da das Wort ‘Indikator’ im Kontext von Indexkonstruktionen ziemlich verbreitet ist, haben wir es in diesem Text übernommen. Man sollte sich jedoch klarmachen, daß die umgangssprachliche Assoziation – daß Indikatoren „etwas anzeigen“ – im Kontext von Indexkonstruktionen fragwürdig ist. Wenn man aus dem Nebenzimmer Geräusche hört, liefern sie vielleicht einen Hinweis auf die Art von Objekten, die sich im Nebenzimmer befinden. Die Geräusche sind dann – im gewöhnlichen Sinn des Wortes – Indikatoren, die helfen können, Vermutungen anzustellen. Das setzt jedoch voraus, daß man über gewisse Vorstellungen von Objekten verfügt, die die Geräusche hervorgebracht haben könnten. Wenn empirische Variablen zur Konstruktion von Indizes verwendet werden, macht jedoch das Reden von Indikatoren in seiner üblichen Bedeutung keinen Sinn. Denn in diesem Fall gewinnt die konstruierte Variable aus den empirischen Variablen und dem verwendeten Konstruktionsverfahren ihren potentiellen Sinn. Es ist vielleicht nützlich auszuprobieren, ob man mit den Geräuschen aus dem Nebenzimmer entsprechend umgehen kann. Man müßte dann versuchen, sie als Indikatoren zur Definition einer dispositionalen Eigenschaft – einer „Fähigkeit (des Nebenzimmers?), Geräusche hervorbringen zu können“ – zu verwenden. Aber schon die Frage, wem dann diese dispositionale Eigenschaft zurechenbar ist, würde offenbar erhebliche Schwierigkeiten erzeugen.

davon gesprochen werden, daß ein statistischer Zusammenhang zwischen Variablen ermittelt wird.

12. Wie kann trotzdem der Eindruck entstehen, daß durch „latente Variablen“ irgendetwas erklärt werden könnte? Denken wir noch einmal an das oben zitierte Beispiel, in dem Thurstone die Ergebnisse eines Rechentests (s) durch „latente Variablen“ x_1 und x_2 erklären möchte. Niemand würde es ernst nehmen, wenn Thurstone stattdessen versuchen würde, die Variable s durch eine Variable

$$x_3 := \sqrt{s}$$

zu erklären. Aber die Konstruktion von x_1 und x_2 unterscheidet sich von der Konstruktion von x_3 nur in den rechentechnischen Details. Ein Unterschied entsteht ausschließlich durch eine Rhetorik, die x_1 und x_2 als „Faktoren“ interpretiert und sie dann in eine „nicht-beobachtbare“ Wirklichkeit hineindichtet. Daß es sich um eine unüberlegte und beliebige Rhetorik handelt, macht Thurstone selbst deutlich:

„Factor analysis assumes that a variety of phenomena within a domain are related and that they are determined, at least in part, by a relatively small number of functional unities or factors. The factors may be called by different names, such as ‘causes’, ‘faculties’, ‘parameters’, ‘functional unities’, ‘abilities’, or ‘independent measurements’. The name for a factor depends on the context, on one’s philosophical preferences and manner of speech, and on how much one already knows about the domain to be investigated.“ (Thurstone 1947, S. 56)

Man kann also ganz nach Belieben sagen, was man möchte. — Aber schon der Gedanke, Dispositionen auch als Ursachen auffassen zu können, erscheint nicht sinnvoll. Zucker löst sich nicht auf, weil er die dispositionale Eigenschaft hat, in Wasser löslich zu sein. Er löst sich nur auf, wenn er mit Wasser in Berührung gebracht wird. Aber auch dann besteht die Ursache nicht in seiner dispositionalen Eigenschaft. Aber beim Zucker ist der Denkfehler noch vergleichsweise harmlos. Bezieht man sich dagegen auf dispositionale Eigenschaften von Menschen, führt der Denkfehler zu einem grundsätzlich falschen Bild menschlicher Akteure, weil dann durch die Begriffsbildung unsichtbar gemacht wird, daß sich Menschen auf ihre Dispositionen reflexiv beziehen können.[9]

[9]Übrig bleibt dann eine Auffassung, die sich bei H. J. Eysenck (1956, S. 12) – selbst ein Vertreter der Faktorenanalyse – unter Berufung auf Thurstone so liest: „‘Das mensch-

13. In den nächsten beiden Abschnitten werden wir von der fragwürdigen Rhetorik, die sich in der Entwicklung faktorenanalytischer Methoden verbreitet hat, weitgehend absehen. Vielmehr soll gezeigt werden, *wie* mit diesen Methoden Variablen *konstruiert* werden können. Versteht man die Verfahren, kann man schließlich auch die sie begleitende Rhetorik besser kritisieren. Mit Fragen der Begriffsbildung werden wir uns zwar in Abschnitt 14.4 noch einmal beschäftigen. Dort geht es allerdings nicht um eine nochmalige Kritik des Redens von „Faktoren", sondern um die Frage, ob faktoranalytische Methoden bei der Konstruktion von Indizes hilfreich sein können.

14.2 Einfache Faktorenanalyse

Wir beginnen mit einer Variante der Faktorenanalyse, bei der nur eine latente Variable konstruiert werden soll. Dementsprechend wird von *einfacher Faktorenanalyse* gesprochen. In dieser Form wurde sie auch zuerst von Charles Spearman (1904) entwickelt, um einen als „general intelligence" interpretierbaren „Faktor" zu konstruieren.

14.2.1 Das Rechenverfahren

1. Ausgangspunkt für das Rechenverfahren sind m empirische Variablen bzw. Indikatoren

$$X_j : \Omega \longrightarrow \tilde{\mathcal{X}}_j$$

für $j = 1, \dots, m$. Gesucht ist eine Variable

$$Y : \Omega \longrightarrow \tilde{\mathcal{Y}}$$

die die Indikatoren in einer noch zu erläuternden Weise „möglichst gut erklärt". Wie wir sehen werden, ist dann Y ein mit Hilfe der Indikatoren $X_1, \dots, X_m$ konstruierter Index. Die in der Faktorenanalyse übliche Rhetorik spricht jedoch nicht von Indizes, sondern von „Faktoren" oder „latenten Variablen", da man mit ihrer Hilfe „etwas erklären" möchte.

2. Der Ansatz der Faktorenanalyse wird verständlich, wenn man zunächst an Regressionsmodelle denkt, etwa in der Form

$$X_j \;=\; a_j + Y b_j + U_j \qquad (\text{für } j = 1, \dots, m) \tag{14.2.1}$$

liche Verhalten', so formuliert es Thurstone, 'ist der eigentliche Gegenstand der Psychologie. Eine ganze Reihe menschlicher Tätigkeiten zeigt individuelle Unterschiede hinsichtlich der erzielten Leistungen. So wie es zweckmäßig ist, bei der Beschreibung von Körperbewegungen physikalische Kräfte anzunehmen, liegt es nahe, beim Menschen vorhandene oder mangelnde Fähigkeiten als tiefere Ursache für die erfolgreiche Bewältigung oder ein Versagen bei ein und derselben Aufgabe zu postulieren.'" Aber vielleicht stimmt es doch; vielleicht wurde die Hand des Autors beim Schreiben dieser Sätze tatsächlich von einer geheimnisvollen Kraft in Bewegung gesetzt?

Wäre Y eine empirische Variable, könnte man für jeden Indikator X_j einen solchen Regressionsansatz nicht nur spekulativ formulieren, sondern die Modellparameter a_j und b_j tatsächlich berechnen. Jetzt ist jedoch Y keine empirische, sondern eine „latente" Variable, die konstruiert werden soll; die Faktorenanalyse ist also ein Verfahren zur Konstruktion von Variablen. Ihre Rhetorik orientiert sich gleichwohl an der Vorstellung eines Regressionsmodells: Man sagt, daß die Indikatoren X_j durch einen *gemeinsamen Faktor* Y und einen jeweils *spezifischen Faktor* U_j „erklärt" werden sollen.

3. Betrachtet man die Regressionsgleichungen, sieht man sofort, daß eine eindeutige Lösung nicht möglich ist. Es gibt vielmehr beliebig viele Möglichkeiten, eine Variable Y und dementsprechend für jeden Indikator X_j auch eine Variable U_j zu bilden. Man kann ja tatsächlich eine beliebige Variable Y annehmen und dann jeweils passende Variablen U_j berechnen. Alle Varianten der Faktorenanalyse beruhen infolgedessen auf zahlreichen mehr oder weniger beliebigen Annahmen, um zu einer Konstruktion der Variablen Y und $U_1, \dots, U_m$ zu gelangen. Es handelt sich allerdings nicht um Annahmen, die empirisch geprüft werden könnten; es sind vielmehr willkürliche Festlegungen.[10]

4. Eine erste Überlegung zeigt, daß sich das Problem vereinfachen läßt, wenn man die Variablen X_j standardisiert, so daß sie den Mittelwert 0 und die Varianz 1 bekommen (vgl. Abschnitt 5.4). Es wird also folgende Transformation vorgenommen:

$$Z_j \;:=\; (X_j - \mathrm{M}(X_j)) \,/\, \sqrt{\mathrm{V}(X_j)} \tag{14.2.2}$$

so daß gilt: $\mathrm{M}(Z_j) = 0$ und $\mathrm{V}(Z_j) = 1$. Im folgenden beziehen wir uns stets auf diese standardisierten Indikatoren $Z_1, \dots, Z_m$.

5. Eine weitere Annahme schließt sich unmittelbar an die Standardisierung der Indikatoren an. Es wird angenommen, daß auch Y als eine standardisierte Variable konstruiert werden soll, so daß $\mathrm{M}(Y) = 0$ und $\mathrm{V}(Y) = 1$ ist. Weiterhin wird gefordert, daß auch bei den Variablen U_j der Mittelwert 0 sein soll, also $\mathrm{M}(U_j) = 0$ für $j = 1, \dots, m$. In den Regressionsansätzen (14.2.1) entfallen dann die Parameter a_j und man kann vereinfacht von

$$Z_j \;=\; Y b_j + U_j \qquad (\text{für } j = 1, \dots, m) \tag{14.2.3}$$

[10] Z.B. heißt es in einer Arbeit zur Faktorenanalyse von Ost 1984, S. 583: Die „latenten Variablen" (in unserer Terminologie Y und U_j) „sind nicht beobachtbar, so daß die Voraussetzungen recht frei festgelegt werden können. Man tut dies in einer modellvereinfachenden und mathematisch zweckmäßigen Art und Weise ... So groß die Freiheit bei der Wahl der Annahmen ... ist, ebensowenig läßt sich allerdings die Gültigkeit dieser Annahmen empirisch überprüfen. Man muß sich im klaren darüber sein, daß der Anwendung und Interpretierbarkeit faktorenanalytischer Ergebnisse dadurch Schranken gesetzt sind, denn mehr als bei anderen statistischen Modellen ist man auf Vermutungen über das Zustandekommen der Daten angewiesen."

ausgehen. In dieser Darstellung werden die Koeffizienten b_j *Faktorladungen* der Variablen (des Faktors) Y für die Indikatoren Z_j genannt.

6. Eine weitere für das Rechenverfahren konstitutive Annahme bezieht sich darauf, daß die Variablen U_j mit der Variablen Y nicht korreliert sein sollen: $\mathrm{Cov}(Y, U_j) = 0$ für $j = 1, \ldots, m$. Daraus folgt

$$\mathrm{V}(Z_j) \;=\; b_j^2\,\mathrm{V}(Y) + \mathrm{V}(U_j)$$

und da es sich sowohl bei Z_j als auch bei Y um standardisierte Variablen handelt, ergibt sich folgende einfache Zerlegung:

$$1 \;=\; b_j^2 + \mathrm{V}(U_j) \tag{14.2.4}$$

In der Literatur zur Faktorenanalyse wird b_j^2 die *Kommunalität* von Z_j und $\mathrm{V}(U_j)$ die *spezifische* Varianz von Z_j genannt. Orientiert man sich am statistischen Verständnis des Wortes ‘erklären’, zeigt die Kommunalität, welcher Anteil an der Varianz der standardisierten Indikatoren Z_j durch die Variable Y erklärt wird.

7. Um die weiteren Konstruktionsschritte zu verstehen, nehmen wir an, daß Daten für die Indikatoren $X_1, \ldots, X_m$ gegeben sind. Ausgangspunkt ist dann eine Datenmatrix

$$\mathbf{X} := \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

für $i = 1, \ldots, n$ Individuen aus Ω. Durch Standardisierung erhält man

$$\mathbf{Z} := \begin{pmatrix} z_{11} & \cdots & z_{1m} \\ \vdots & & \vdots \\ z_{n1} & \cdots & z_{nm} \end{pmatrix}$$

mit den Eigenschaften $\mathrm{M}(Z_j) = 0$ und $\mathrm{V}(Z_j) = 1$.

8. Verwendet man weiterhin die Matrizenschreibweise, ergibt sich eine einfache Darstellung für die Gleichungen (14.2.3). Die Faktorladungen werden zu einem Vektor

$$\mathbf{b} := (b_1, \ldots, b_m)'$$

zusammengefaßt; für die zu konstruierenden Variablen Y und $U_1, \ldots, U_m$ werden hypothetische Werte angenommen, so daß man in Matrixform schreiben kann:

$$\mathbf{y} := \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \text{und} \quad \mathbf{U} := \begin{pmatrix} u_{11} & \cdots & u_{1m} \\ \vdots & & \vdots \\ u_{n1} & \cdots & u_{nm} \end{pmatrix}$$

Dann können die m Gleichungen in (14.2.3) so dargestellt werden:

$$\mathbf{Z} \;=\; \mathbf{y}\mathbf{b}' + \mathbf{U} \tag{14.2.5}$$

9. Betrachten wir jetzt einige Implikationen der bisher getroffenen Annahmen für die Varianzen, Kovarianzen und Korrelationen der Variablen. Da angenommen worden ist, daß die zu konstruierende Variable Y standardisiert sein soll, gilt

$$\mathrm{V}(Y) \;=\; \frac{1}{n}\sum_{i=1}^{n} y_i^2 \;=\; 1$$

Weiterhin gilt für die Kovarianz von jeweils zwei standardisierten Indikatoren

$$\mathrm{Cov}(Z_j, Z_{j'}) \;=\; \frac{1}{n}\sum_{i=1}^{n} z_{ij} z_{ij'}$$

und da es sich um standardisierte Variablen handelt, sind ihre Korrelationen mit ihren Kovarianzen identisch, also $\mathrm{Corr}(Z_j, Z_{j'}) = \mathrm{Cov}(Z_j, Z_{j'})$. Sie können zu einer Kovarianz- bzw. Korrelationsmatrix

$$\mathrm{Corr}(\mathbf{Z}) \;=\; \mathrm{Cov}(\mathbf{Z}) \;=\; \frac{1}{n}\,\mathbf{Z}'\mathbf{Z}$$

zusammengefaßt werden. In der Hauptdiagonalen dieser (m, m)-Matrix stehen die Varianzen von Z_j, also Einsen, da es sich um standardisierte Variablen handelt. Ganz entsprechend kann man eine Kovarianzmatrix für die Variablen $U_1, \ldots, U_m$ bilden. Da ihre Mittelwerte, wie angenommen worden ist, Null sind, findet man zunächst

$$\mathrm{Cov}(U_j, U_{j'}) \;=\; \frac{1}{n}\sum_{i=1}^{n} u_{ij} u_{ij'}$$

und daraus kann die Kovarianzmatrix

$$\mathrm{Cov}(\mathbf{U}) \;=\; \frac{1}{n}\,\mathbf{U}'\mathbf{U}$$

gebildet werden. Da jedoch die Varianzen dieser Variablen unbestimmt sind, ist die Kovarianzmatrix in diesem Fall nicht mit der Korrelationsmatrix identisch.

10. Die Gleichung (14.2.5) kann verwendet werden, um die Korrelationsmatrix der standardisierten Indikatoren zu zerlegen. Man findet:

$$\begin{aligned} \mathrm{Corr}(\mathbf{Z}) \;=\; \frac{1}{n}\,\mathbf{Z}'\mathbf{Z} \;&=\; \frac{1}{n}\,(\mathbf{y}\mathbf{b}' + \mathbf{U})'(\mathbf{y}\mathbf{b}' + \mathbf{U}) \\ &=\; \frac{1}{n}\,(\mathbf{b}\mathbf{y}'\mathbf{y}\mathbf{b}' + \mathbf{U}'\mathbf{U} + \mathbf{b}\mathbf{y}'\mathbf{U} + \mathbf{U}'\mathbf{y}\mathbf{b}') \end{aligned}$$

Nun ist jedoch $\mathbf{y}'\mathbf{y} = n$, und da angenommen worden ist, daß die Variablen U_j nicht mit der Variablen Y korreliert sind, gilt außerdem $\mathbf{y}'\mathbf{U} = 0$. Also resultiert eine einfache Zerlegung der Korrelationsmatrix von $\mathbf{Z}$:

$$\mathrm{Corr}(\mathbf{Z}) = \mathbf{bb}' + \mathrm{Cov}(\mathbf{U}) \quad \text{bzw.} \quad \frac{1}{n}\mathbf{Z}'\mathbf{Z} = \mathbf{bb}' + \frac{1}{n}\mathbf{U}'\mathbf{U} \qquad (14.2.6)$$

Diese Gleichung wird manchmal das „Fundamentaltheorem der Faktorenanalyse“ genannt; sie bildet den Ausgangspunkt, um zunächst den Vektor der Faktorladungen $\mathbf{b}$ und dann die Variablen Y und $U_1, \ldots, U_m$ zu konstruieren. Es sollte allerdings beachtet werden, daß es etwas irreführend ist, hier von einem „Theorem“ zu sprechen, denn die Gleichung ist eine simple Konsequenz der vorher getroffenen Annahmen.

11. Die Gleichung (14.2.6) ist indessen noch nicht hinreichend, um die Faktorladungen $\mathbf{b}$ und die Variablen Y und $U_1, \ldots, U_m$ eindeutig zu bestimmen.[11] An dieser Stelle setzen daher eine Reihe unterschiedlicher Vorschläge an, um die Konstruktionsaufgabe in Angriff zu nehmen. Wir behandeln hier nur eine Variante, die sich daran orientiert, die Varianzen der U-Variablen, also $\mathrm{V}(U_j)$, möglichst klein zu machen. Das Kriterium ist

$$\sum_{j=1}^{m} \mathrm{V}(U_j) = \frac{1}{n} \sum_{j=1}^{m} \sum_{i=1}^{n} u_{ij}^2 \longrightarrow \min \qquad (14.2.7)$$

Wie man sehen wird, gibt es dann einen unmittelbaren Zusammenhang zu der schon in Kap. 13 besprochenen Methode, einen Index (in diesem Fall Y) so zu bilden, daß seine Varianz maximal wird. Zunächst jedoch noch eine Bemerkung zu diesem Kriterium, die an Gleichung (14.2.6) anschließt. Man kann sich überlegen, daß die Summe der Varianzen der U-Variablen gerade die Summe der Hauptdiagonalelemente von $\mathbf{U}'\mathbf{U}/n$ ist. Da $\mathbf{U} = \mathbf{Z} - \mathbf{yb}'$ ist, folgt also aus (14.2.6):

$$\sum_{j=1}^{m} \mathrm{V}(U_j) = \mathrm{tr}(\mathbf{U}'\mathbf{U}/n) = \mathrm{tr}(\mathbf{Z}'\mathbf{Z}/n) - \mathrm{tr}(\mathbf{bb}')$$

(tr bezeichnet die Spur), woraus man folgende Darstellung gewinnt:

$$\sum_{j=1}^{m} \mathrm{V}(U_j) = m - \sum_{j=1}^{m} b_j^2 \qquad (14.2.8)$$

Die Minimierung der Varianzen ist also äquivalent zur Maximierung der Kommunalitäten. Dies liefert eine mögliche Lesart der eingangs verwendeten Formulierung, daß die Variable Y so konstruiert werden soll, daß mit ihr die Indikatoren „möglichst gut erklärt“ werden können.

[11]Man kann sich das schon daran klarmachen, daß ja nur m Indikatoren zur Verfügung stehen, um insgesamt $m+1$ Variablen zu konstruieren.

12. Schließlich muß noch die Frage besprochen werden, wie das Minimierungsproblem (14.2.7) praktisch gelöst werden kann. Dafür gehen wir von Gleichung (14.2.3) aus. Für jedes i gesondert kann sie folgendermaßen geschrieben werden:

$$z_{ij} = y_i\, b_j + u_{ij}$$

Daraus folgt für die zu minimierende Summe der Varianzen:[12]

$$f(b_1, \ldots, b_m, y_1, \ldots, y_n) := \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ij}^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} (z_{ij} - y_i\, b_j)^2$$

Wir betrachten diesen Ausdruck als eine Funktion, deren Argumente aus den angegebenen Parametern bestehen. Eine notwendige Bedingung für ein Minimum besteht darin, daß die partiellen Ableitungen Null werden. Diese Ableitungen lassen sich einfach berechnen:

$$\frac{\partial f(b_1, \ldots, b_m, y_1, \ldots, y_n)}{\partial b_j} = -2 \sum_{i=1}^{n} (z_{ij} - y_i\, b_j)\, y_i \quad (j = 1, \ldots, m)$$

$$\frac{\partial f(b_1, \ldots, b_m, y_1, \ldots, y_n)}{\partial y_i} = -2 \sum_{j=1}^{m} (z_{ij} - y_i\, b_j)\, b_j \quad (i = 1, \ldots, n)$$

Als Bedingungen für die Parameter, die ein Minimum der Funktion f liefern, findet man daraus:

$$\sum_{i=1}^{n} z_{ij}\, y_i = b_j \sum_{i=1}^{n} y_i^2 \quad (j = 1, \ldots, m)$$

$$\sum_{j=1}^{m} z_{ij}\, b_j = y_i \sum_{j=1}^{m} b_j^2 \quad (i = 1, \ldots, n)$$

Die Lösungen dieser Gleichungen liefern das gesuchte Minimum.[13]

13. Aus diesen Bedingungen kann man zunächst ersehen, daß für die Werte der Variablen Y gelten muß:

$$y_i = \frac{1}{\lambda} \sum_{j=1}^{m} z_{ij} b_j \quad \text{wobei} \quad \lambda := \sum_{j=1}^{m} b_j^2 \qquad (14.2.9)$$

Das heißt, wenn man sich an dem Kriterium (14.2.7) orientiert, kann Y als ein additiver Index

$$\mathbf{y} = \frac{1}{\lambda}\mathbf{Zb} \quad \text{bzw.} \quad Y = \frac{1}{\lambda} \sum_{j=1}^{m} Z_j b_j \qquad (14.2.10)$$

[12]Der Faktor $1/n$ wird hier zur Vereinfachung der Schreibweise nicht mitgeführt.

[13]Man kann sich anhand der zweiten Ableitungen davon überzeugen, daß die Funktion f global konvex ist.

dargestellt werden. Die einfache Faktorenanalyse erweist sich insoweit nur als eine Variante der additiven Indexkonstruktion.

14. Weiterhin folgt aus der ersten Bedingung für ein Minimum der Funktion f die Gleichung

$$\frac{1}{n}\sum_{i=1}^{n} z_{ij}\, y_i \;=\; b_j$$

da angenommen wurde, daß Y standardisiert ist ($\mathbf{y}'\mathbf{y} = n$). Dies zeigt auch, daß $b_j = \mathrm{Corr}(X_j, Y)$ gerade die Korrelation zwischen der Variablen Y und dem Indikator X_j angibt. Durch Einsetzen von (14.2.9) erhält man

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m} z_{ij}\, z_{ik}\, b_k \;=\; \lambda\, b_j \qquad (\text{für } j = 1,\ldots,m)$$

Diese m Gleichungen lassen sich in Matrizenschreibweise folgendermaßen zusammenfassen:

$$\frac{1}{n}\,(\mathbf{Z}'\mathbf{Z})\,\mathbf{b} \;=\; \lambda\,\mathbf{b} \tag{14.2.11}$$

Lösungen dieser Gleichung erhält man durch die Eigenwerte und zugehörigen Eigenvektoren der Korrelationsmatrix $\mathbf{Z}'\mathbf{Z}/n$. Da man, wie in (14.2.8) gezeigt worden ist, an einer Lösung interessiert ist, die die Summe der Kommunalitäten (also λ) maximiert, wird der größte Eigenwert und der zugehörige Eigenvektor verwendet.

15. Wie man aus (14.2.11) sieht, können die Eigenvektoren mit einem beliebigen Wert multipliziert werden, ohne daß sich an der Gleichung etwas ändert. Eine geeignete Normierung erhält man durch die (oben schon als Definition verwendete) Forderung

$$\mathbf{b}'\mathbf{b} \;=\; \sum_{j=1}^{m} b_j^2 \;=\; \lambda \tag{14.2.12}$$

Denn dadurch erhält man die Werte der Variablen Y in standardisierter Form. Man sieht das so: Zunächst zeigt Gleichung (14.2.10), wie man die Werte von Y berechnen kann, wenn man zuvor die Faktorladungen $\mathbf{b}$ gefunden hat. Aus dieser Gleichung folgt jedoch unter Zuhilfenahme von (14.2.11) und der Normierungsbedingung (14.2.12):

$$\mathbf{y}'\mathbf{y} \;=\; \frac{1}{\lambda^2}\,\mathbf{b}'\,\mathbf{Z}'\mathbf{Z}\,\mathbf{b} \;=\; n\,\frac{1}{\lambda^2}\,\lambda\,\mathbf{b}'\mathbf{b} \;=\; n$$

16. Fassen wir zusammen. Die einfache Faktorenanalyse ist ein Verfahren, mit dem aus einer Menge standardisierter Variablen $Z_1,\ldots,Z_m$ ein

additiver Index Y konstruiert werden kann. Das Verfahren besteht darin, die Variable Y so zu bilden, daß man eine im statistischen Sinne optimale Erklärung der empirischen Variablen erhält. Es handelt sich um das gleiche Verfahren, das in Kap. 13 verwendet wurde, um die Varianz von Y zu maximieren. — Eine ergänzende Sichtweise liefert folgende Formulierung von K. Überla:

„Die Faktorenanalyse stellt nun die Frage, ob es eine Größe gibt, einen sog. Faktor, der in Verbindung mit einer geeigneten Rechenregel die beobachteten Korrelationen möglichst genau zu reproduzieren gestattet.“ (Überla 1977, S. 44)

Um das zu verstehen, gehen wir von Gleichung (14.2.3) aus. Schreibt man sie in der Form $Z_j \approx Y b_j$, findet man:

$$\mathrm{Corr}(Z_j, Z_{j'}) \;\approx\; \mathrm{Corr}(Y b_j, Y b_{j'}) \;=\; \frac{1}{n}\sum_{i=1}^{n}(y_i b_j)(y_i b_{j'}) \;=\; b_j b_{j'}$$

Bei dieser Betrachtungsweise liefert also $b_j b_{j'}$ eine Näherung für die Korrelation zwischen Z_j und $Z_{j'}$, und man könnte auf die Idee kommen, die Qualität der Indexkonstruktion Y durch den Grad dieser Näherung zu beurteilen.

14.2.2 Eine Illustration

1. Um das Verfahren zu illustrieren, beziehen wir uns auf Spearmans (1904) Versuch, einen als „general intelligence“ interpretierbaren „Faktor“ zu konstruieren. Seine Daten gewann Spearman, indem er Kinder verschiedenen Schulleistungstests unterzog. Wir verwenden für unsere Illustration Daten, die Spearman bei 33 Kindern im Alter von 9 bis 13 Jahren gefunden hat. Dabei beschränken wir uns auf die Ergebnisse von vier Tests: „Classics“ (Z_1), „French“ (Z_2), „English“ (Z_3) und „Mathematics“ (Z_4). Daraus berechnete Spearman (1904, S. 275) folgende Korrelationsmatrix:[14]

$$\mathrm{Corr}(\mathbf{Z}) \;=\; \begin{pmatrix} 1.00 & 0.83 & 0.78 & 0.70 \\ 0.83 & 1.00 & 0.67 & 0.67 \\ 0.78 & 0.67 & 1.00 & 0.64 \\ 0.70 & 0.67 & 0.64 & 1.00 \end{pmatrix}$$

Finden wollte Spearman eine neue Variable Y, die die von ihm ermittelten Korrelationen möglichst gut reproduziert. Der faktoranalytische Ansatz ist also in diesem Fall:

$$Z_j \;=\; Y\,b_j + U_j \qquad \text{für } j = 1,\ldots,4$$

[14]Die von Spearman angegebene Korrelationsmatrix enthält in der Hauptdiagonalen geschätzte Kommunalitäten. Auf das Kommunalitätenproblem werden wir jedoch erst in Abschnitt 14.3.4 eingehen.

Berechnet man die Eigenwerte der Korrelationsmatrix, findet man

$$\lambda_1 = 3.15,\ \lambda_2 = 0.38,\ \lambda_3 = 0.33,\ \lambda_4 = 0.14$$

Aus dem normierten Eigenvektor, der zum größten Eigenwert gehört, gewinnt man die Faktorladungen

$$\mathbf{b} \;=\; (0.94, 0.90, 0.87, 0.84)'$$

Daraus findet man die Korrelationen bzw. Kommunalitäten

$$\mathbf{bb}' \;=\; \begin{pmatrix} 0.88 & 0.84 & 0.82 & 0.79 \\ 0.84 & 0.80 & 0.78 & 0.76 \\ 0.82 & 0.78 & 0.76 & 0.73 \\ 0.79 & 0.76 & 0.73 & 0.71 \end{pmatrix}$$

Die Elemente außerhalb der Hauptdiagonalen können mit den o.a. Korrelationen der Indikatoren verglichen werden. In der Hauptdiagonalen stehen die Kommunalitäten, die angeben, welcher Anteil der Varianz von Z_j durch Y im statistischen Sinne des Wortes erklärt wird.

14.3 Multiple Faktorenanalyse

Die im vorangegangenen Abschnitt behandelte einfache Faktorenanalyse kann verallgemeinert werden, so daß auch zwei oder mehr „latente Variablen“ konstruiert werden können. Diese sog. *multiple Faktorenanalyse* wurde zuerst vor allem von L. Thurstone (1931, 1947) propagiert: „to discover [...] those principal factors that are truly operative in producing the correlation coefficients.“ (1931, S. 421)

14.3.1 Der konzeptionelle Ansatz

1. In der Terminologie folgen wir den im vorangegangenen Abschnitt eingeführten Bezeichnungen. Als Ausgangspunkt nehmen wir sogleich m standardisierte Indikatoren $Z_1, \ldots, Z_m$ an; q sei die Anzahl der Y-Variablen (der gemeinsamen Faktoren), die konstruiert werden sollen. Aus dem Regressionsansatz (14.2.3) wird dann:

$$Z_j \;=\; Y_1\, b_{j1} + \cdots + Y_q\, b_{jq} + U_j \qquad (\text{für } j = 1, \ldots, m) \tag{14.3.1}$$

wobei $Y_1, \ldots, Y_q$ die gemeinsamen Faktoren, $U_1, \ldots, U_m$ die spezifischen Faktoren und $b_{j1,}, \ldots, b_{jq}$ die Faktorladungen für den Indikator Z_j sind.

2. Auch alle Annahmen, die im vorangegangenen Abschnitt gemacht worden sind, werden übernommen; insbesondere also, daß die Y-Variablen standardisiert sein sollen:

$$\mathrm{M}(Y_k) \;=\; 0 \quad \text{und} \quad \mathrm{V}(Y_k) \;=\; 1 \qquad (\text{für } k = 1, \ldots, q)$$

und die U-Variablen den Mittelwert 0 haben sollen:

$$\mathrm{M}(U_j) \;=\; 0 \qquad (\text{für } j = 1, \ldots, m)$$

Übernommen wird auch die Annahme, daß die Y-Variablen und die U-Variablen unkorreliert sein sollen, also

$$\mathrm{Cov}(Y_k, U_j) \;=\; 0 \qquad (\text{für } k = 1, \ldots, q;\ j = 1, \ldots, m)$$

Hinzu kommt jetzt eine weitere Annahme: daß auch die Y-Variablen untereinander unkorreliert sein sollen, also

$$\mathrm{Cov}(Y_k, Y_{k'}) \;=\; 0 \qquad (\text{für } k, k' = 1, \ldots, q;\ k \neq k')$$

Diese Annahme erscheint zwar gerade aus der Sicht der faktoranalytischen Rhetorik problematisch;[15] sie wird jedoch meistens *zunächst* gemacht, um überhaupt zu einer Lösung des Konstruktionsproblems zu gelangen. Wir werden auf diese Annahme später bei der Diskussion des sog. „Rotationsproblems“ noch einmal eingehen.

3. Die bisherigen Annahmen implizieren wiederum eine einfache Varianzzerlegung. Zunächst findet man aus (14.3.1):

$$\mathrm{V}(Z_j) \;=\; \sum_{k=1}^{q} b_{jk}^2\, \mathrm{V}(Y_k) + \mathrm{V}(U_j)$$

und da die Indikatoren und die Y-Variablen standardisiert sind, folgt

$$1 \;=\; \sum_{k=1}^{q} b_{jk}^2 + \mathrm{V}(U_j) \tag{14.3.2}$$

In diesem Fall ist also

$$h_j^2 \;:=\; \sum_{k=1}^{q} b_{jk}^2 \tag{14.3.3}$$

[15] G. Arminger 1979, S. 17, bemerkt dazu: „Allerdings ist es in der Regel inhaltlich nicht sinnvoll, unkorrelierte Faktoren anzunehmen. Gerade in der Testpsychologie oder in der empirischen Sozialforschung wollen wir ja Zusammenhänge zwischen latenten, nicht unmittelbar beobachtbaren Variablen entdecken, sodaß die Annahme der Unkorreliertheit sinnlos ist und auch häufig zu nicht interpretierbaren Ergebnissen führt. Denken wir z.B. an einen psychologischen Test von Persönlichkeitseigenschaften. Wünschenswert ist es, wenn jedes Testitem auf einem Faktor hoch lädt und auf den anderen niedrig, es wäre jedoch unsinnig, zu behaupten, daß Persönlichkeitseigenschaften unkorreliert sind. Zur Erforschung der Persönlichkeitsstruktur suchen wir im Gegenteil nach korrelierten Faktoren.“ Ob unsinnig oder nicht, mit „latenten Variablen“ kann man jedenfalls machen, was man will, und da sie „nicht beobachtbar“ sind, ist schon die Forderung, daß sie „interpretierbar“ sein sollten, fragwürdig.

die Kommunalität (der Variablen Z_j), die angibt, welcher Anteil ihrer Varianz durch die Y-Variablen erklärt wird.

4. Wiederum vereinfachen sich die Überlegungen, wenn man die m Regressionsgleichungen in (14.3.1) mit Hilfe von Matrizen zusammenfaßt:

$$\mathbf{Y} := \begin{pmatrix} y_{11} & \cdots & y_{1q} \\ \vdots & & \vdots \\ y_{n1} & \cdots & y_{nq} \end{pmatrix} \quad \text{und} \quad \mathbf{B} := \begin{pmatrix} b_{11} & \cdots & b_{1q} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mq} \end{pmatrix}$$

Dann erhält man analog zu (14.2.5) die Gleichung

$$\mathbf{Z} = \mathbf{Y}\mathbf{B}' + \mathbf{U} \tag{14.3.4}$$

Von dieser Gleichung ausgehend kann wiederum eine Zerlegung der Korrelationsmatrix der Indikatoren vorgenommen werden. Aus (14.3.4) folgt zunächst:

$$\begin{aligned} \mathrm{Corr}(\mathbf{Z}) &= \frac{1}{n}\mathbf{Z}'\mathbf{Z} = \frac{1}{n}(\mathbf{Y}\mathbf{B}' + \mathbf{U})'(\mathbf{Y}\mathbf{B}' + \mathbf{U}) \\ &= \frac{1}{n}(\mathbf{B}\mathbf{Y}'\mathbf{Y}\mathbf{B}' + \mathbf{U}'\mathbf{U} + \mathbf{B}\mathbf{Y}'\mathbf{U} + \mathbf{U}'\mathbf{Y}\mathbf{B}') \end{aligned}$$

Da die Y-Variablen weder untereinander noch mit den U-Variablen korreliert sein sollen, gilt außerdem:

$$\mathbf{Y}'\mathbf{Y} = n\,\mathbf{I}_q \quad \text{und} \quad \mathbf{Y}'\mathbf{U} = \mathbf{0}$$

wobei $\mathbf{I}_q$ eine (q,q)-Einheitsmatrix bezeichnet. Das „Fundamentaltheorem der Faktorenanalyse" nimmt dann folgende Form an:

$$\mathrm{Corr}(\mathbf{Z}) = \mathbf{B}\mathbf{B}' + \mathrm{Cov}(\mathbf{U}) \quad \text{bzw.} \quad \frac{1}{n}\mathbf{Z}'\mathbf{Z} = \mathbf{B}\mathbf{B}' + \frac{1}{n}\mathbf{U}'\mathbf{U} \tag{14.3.5}$$

14.3.2 Die Hauptfaktorenmethode

1. Es bleibt zu überlegen, wie man die Y-Variablen konstruieren kann. Dafür verwenden wir wieder das Kriterium (14.2.7), d.h. Minimierung der Varianzen der U-Variablen. Analog zu (14.2.8) findet man jetzt

$$\sum_{j=1}^{m} \mathrm{V}(U_j) = m - \sum_{j=1}^{m}\sum_{k=1}^{q} b_{jk}^2 \tag{14.3.6}$$

Das Kriterium ist identisch mit einer Maximierung der Kommunalitäten, d.h. der durch die Y-Variablen erklärten Varianzanteile.

2. Wie im vorangegangenen Abschnitt betrachten wir das Zielkriterium

als eine Funktion von Parametern, die durch eine Minimierung der Zielfunktion gefunden werden sollen; in diesem Fall also

$$\begin{aligned} &f(b_{11},\ldots,b_{m1},\ldots,b_{1q},\ldots,b_{mq},y_{11},\ldots,y_{n1},\ldots,y_{1q},\ldots,y_{nq}) := \\ &\qquad \sum_{i=1}^{n}\sum_{j=1}^{m} u_{ij}^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\Big(z_{ij} - \sum_{l=1}^{q} y_{il}\,b_{jl}\Big)^2 \end{aligned}$$

Die partiellen Ableitungen sind

$$\begin{aligned} \frac{\partial f}{\partial b_{jk}} &= -2\sum_{i=1}^{n}\Big(z_{ij} - \sum_{l=1}^{q} y_{il}\,b_{jl}\Big)\,y_{ik} \\ \frac{\partial f}{\partial y_{ik}} &= -2\sum_{j=1}^{m}\Big(z_{ij} - \sum_{l=1}^{q} y_{il}\,b_{jl}\Big)\,b_{jk} \end{aligned}$$

Daraus gewinnt man folgende Bedingungen für die Parameter, die ein Minimum der Funktion liefern sollen:

$$\begin{aligned} \sum_{i=1}^{n} z_{ij}\,y_{ik} &= b_{jk}\sum_{i=1}^{n} y_{ik}^2 \\ \sum_{j=1}^{m} z_{ij}\,b_{jk} &= \sum_{l=1}^{q} y_{il}\sum_{j=1}^{m} b_{jl}\,b_{jk} \end{aligned}$$

Um die erste Gleichung zu gewinnen, wird die Annahme verwendet, daß die Y-Variablen wechselseitig unkorreliert sein sollen. Da sie außerdem standardisiert sein sollen, findet man:

$$\mathrm{Corr}(Z_j, Y_k) = \frac{1}{n}\sum_{i=1}^{n} z_{ij}\,y_{ik} = b_{jk} \tag{14.3.7}$$

so daß die Faktorladungen zugleich die Korrelationen der Y-Variablen mit den Indikatoren anzeigen.

3. Für die weiteren Überlegungen ist es hilfreich, diese Bedingungen in Matrixform zu schreiben. Man erhält dann:

$$\mathbf{B} = \frac{1}{n}\mathbf{Z}'\mathbf{Y} \quad \text{und} \quad \mathbf{Z}\mathbf{B} = \mathbf{Y}\mathbf{B}'\mathbf{B} \tag{14.3.8}$$

Die Aufgabe besteht darin, Lösungen $\mathbf{Y}$ und $\mathbf{B}$ dieser Gleichungen zu finden, so daß die Summe der Komunalitäten, also

$$\mathrm{tr}(\mathbf{B}\mathbf{B}') = \mathrm{tr}(\mathbf{B}'\mathbf{B}) = \sum_{j=1}^{m}\sum_{k=1}^{q} b_{jk}^2$$

maximal wird. — Hier tritt nun allerdings eine Schwierigkeit auf. Aus den beiden Matrixgleichungen in (14.3.8) gewinnt man:[16]

$$\frac{1}{n}(\mathbf{Z}'\mathbf{Z})\,\mathbf{B} \;=\; \mathbf{B}\,(\mathbf{B}'\mathbf{B}) \tag{14.3.9}$$

In dieser Form ist das Problem jedoch nicht vollständig bestimmt. Einen möglichen Ansatz, um dennoch zu einer Lösung des Konstruktionsproblems zu gelangen, liefert die Annahme, daß $\mathbf{B}'\mathbf{B}$ eine Diagonalmatrix sein soll, also

$$\mathbf{B}'\mathbf{B} \;=\; \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_q \end{pmatrix} \tag{14.3.10}$$

wobei

$$\lambda_k \;:=\; \sum_{j=1}^{m} b_{jk}^2 \tag{14.3.11}$$

ist. Geht man nämlich von dieser zusätzlichen Bedingung aus, kann man die Gleichung (14.3.9) in Form von q Gleichungen folgendermaßen schreiben ($\mathbf{b}_1,\ldots,\mathbf{b}_q$ sind hier die Spaltenvektoren der (m,q)-Matrix $\mathbf{B}$):

$$\frac{1}{n}(\mathbf{Z}'\mathbf{Z})\,\mathbf{b}_k \;=\; \lambda_k\mathbf{b}_k \qquad (\text{für } k=1,\ldots,q) \tag{14.3.12}$$

Diese Schreibweise zeigt, daß man Lösungen wieder durch die Eigenwerte (λ_k) und die zugehörigen Eigenvektoren ($\mathbf{b}_k$) der Korrelationsmatrix $\mathbf{Z}'\mathbf{Z}/n$ bestimmen kann. Jeder Eigenvektor

$$\mathbf{b}_k \;=\; (b_{1k},\ldots,b_{mk})'$$

liefert die Faktorladungen für eine entsprechende Variable Y_k.

4. Verwendet man die Annahme, daß $\mathbf{B}'\mathbf{B}$ eine Diagonalmatrix ist, wird von der *Hauptfaktorenmethode* (für die Faktorenanalyse) gesprochen. Da die Korrelationsmatrix $\mathbf{Z}'\mathbf{Z}/n$ maximal m Eigenwerte und zugehörige Eigenvektoren besitzt,[17] kann man mit dieser Methode maximal m gemeinsame Faktoren Y_k konstruieren. Es ist üblich, die Eigenwerte zunächst der Größe nach zu ordnen:

$$\lambda_1 \;\geq\; \lambda_2 \;\geq\; \cdots \;\geq\; \lambda_m$$

Dann kann man aus den zugehörigen – durch (14.3.11) normierten – Eigenvektoren der Reihe nach die Variablen $Y_1, Y_2, \ldots$ konstruieren; und man kann mit Y_1 den größten Teil der gemeinsamen Varianz erklären, mit Y_2 den zweitgrößten Teil usw. Schließlich kann man sich überlegen, wieviele solcher Variablen man konstruieren möchte.[18]

[16]Multiplikation der ersten Gleichung mit $\mathbf{B}'\mathbf{B}$, dann Einsetzen der zweiten Gleichung.

[17]Es gibt genau m, wenn alle Indikatoren linear unabhängig sind.

14.3.3 Faktorenanalyse als Indexkonstruktion

1. Verwendet man die Hauptfaktorenmethode, sieht man auf einfache Weise, daß auch die multiple Faktorenanalyse darauf hinausläuft, additive Indizes der Variablen Z_j zu konstruieren. Denn aus (14.3.8) und der Annahme (14.3.10) findet man sofort:

$$\mathbf{y}_k \;:=\; \frac{1}{\lambda_k}\,\mathbf{Z}\mathbf{b}_k$$

Die einzelnen Werte der Variablen Y_k gewinnt man also durch

$$y_{ik} \;:=\; \frac{1}{\lambda_k}\sum_{j=1}^{m} z_{ij}\,b_{jk} \tag{14.3.13}$$

2. Man kann sich auch leicht davon überzeugen, daß diese Berechnung gerade dazu führt, daß die Variablen Y_k standardisiert sind. Denn unter Zuhilfenahme von (14.3.12) findet man:

$$\frac{1}{n}\,\mathbf{y}_k'\mathbf{y}_k \;=\; \frac{1}{n}\,\frac{1}{\lambda_k^2}\,\mathbf{b}_k'\,\mathbf{Z}'\mathbf{Z}\,\mathbf{b}_k \;=\; \frac{\lambda_k}{\lambda_k^2}\,\mathbf{b}_k'\mathbf{b}_k \;=\; 1$$

14.3.4 Das Kommunalitätenproblem

1. In der Literatur zur Faktorenanalyse wird oft die Annahme gemacht, daß auch die spezifischen Faktoren (unsere U-Variablen) nicht miteinander korreliert sein sollen. Diese Annahme führt indessen zu einem Problem, für das es keine vernünftige Lösung gibt (so daß infolgedessen die Annahme selbst fragwürdig wird). Dieses sogenannte „Kommunalitätenproblem" kann man so verstehen: Würde man voraussetzen, daß die Variablen $U_1,\ldots,U_m$ nicht miteinander korreliert sind, wäre ihre Kovarianzmatrix

$$\mathrm{Cov}(\mathbf{U}) \;=\; \frac{1}{n}\,\mathbf{U}'\mathbf{U}$$

eine Diagonalmatrix. Allerdings kennt man die Varianzen nicht, die sich in der Hauptdiagonalen befinden. Aus der Annahme, daß die U-Variablen

[18]Es ist offensichtlich, daß die einfache Faktorenanalyse, so wie wir sie im vorangegangenen Abschnitt besprochen haben, ein Spezialfall dieser Hauptfaktorenmethode ist, bei dem man sich auf die Konstruktion einer Y-Variablen mit Hilfe des größten Eigenwerts beschränkt.

und die Y-Variablen unkorreliert sind, ergibt sich nur die Varianzzerlegung $\mathrm{V}(U_j) = 1 - h_j^2$, wobei h_j^2 die Kommunalität des Indikators Z_j ist, vgl. (14.3.3). Das Kommunalitätenproblem besteht nun darin, daß man diese Kommunalitäten *vorab* bestimmen möchte, d.h. vor und unabhängig von einer Berechnung der Y-Variablen und ihrer Faktorladungen. So schreibt zum Beispiel Überla:

> „Die Entscheidung über die Größe dieser diagonalen Elemente [also der Varianzen der spezifischen Faktoren] muß vor der Faktorenextraktion erfolgen, und ihre Wahl nennt man das Kommunalitätenproblem. Die Notwendigkeit, Kommunalitäten wählen zu müssen, die kleiner als 1 sind, d.h. die Existenz des Kommunalitätenproblems, ergibt sich aus dem faktorenanalytischen Modell. Dieses Modell fordert eine Reihe gemeinsamer Faktoren und für jede Variable einen Einzelrestfaktor.“ (Überla 1977, S. 155)

2. Es gibt jedoch keine objektivierbaren Gesichtspunkte, nach denen eine vorgängige Bestimmung von Kommunalitäten vorgenommen werden könnte. Infolgedessen sind in der Geschichte der Faktorenanalyse eine Vielzahl unterschiedlicher Vorschläge gemacht worden, ohne daß jemals eine Einigung erzielt werden konnte.[19] Hier soll deshalb nur kurz darauf hingewiesen werden, wie sich das Problem stellen würde, *wenn* sich Kommunalitäten vorab bestimmen ließen. Der Gedankengang ist einfach. Hätte man sinnvolle Werte für Kommunalitäten, könnte daraus zunächst die Kovarianzmatrix der U-Variablen berechnet werden, also $\mathbf{U}'\mathbf{U}/n$. Aus dem „Fundamentaltheorem der Faktorenanalyse“, vgl. (14.3.5), erhielte man dann:

$$\mathbf{K} = \mathbf{B}\mathbf{B}'$$

wobei $\mathbf{K} := \mathbf{Z}'\mathbf{Z}/n - \mathbf{U}'\mathbf{U}/n$ die sogenannte „reduzierte Korrelationsmatrix“ der Indikatoren ist. Da dann sowohl $\mathbf{Z}'\mathbf{Z}/n$ als auch $\mathbf{U}'\mathbf{U}/n$ bekannt wären, könnte man die reduzierte Korrelationsmatrix $\mathbf{K}$ berechnen und das Problem würde dann darin bestehen, eine Zerlegung $\mathbf{K} = \mathbf{B}\mathbf{B}'$ zu finden.

3. Das Problem stellt sich nicht, wenn man der Hauptfaktorenmethode folgt, an der wir unsere Diskussion der Faktorenanalyse orientiert haben. Denn dann wird nicht mit einer willkürlich „reduzierten“ Korrelationsmatrix $\mathbf{K}$ begonnen, sondern mit der durch die Daten gegebenen Korrelationsmatrix $\mathbf{Z}'\mathbf{Z}/n$ der empirischen Indikatoren. Die Hauptfaktorenmethode führt allerdings dazu, daß die Unterscheidung zwischen gemeinsamen und spezifischen Faktoren hinfällig wird. Sie liefert vielmehr maximal m unabhängige Variablen $Y_1, \ldots, Y_m$, aus denen sich die Indikatoren $Z_1, \ldots, Z_m$ als lineare Funktionen rekonstruieren lassen. Entscheidet man sich dafür, nur die ersten q Variablen zu verwenden, bilden automatisch Linearkombinationen der restlichen $m - q$ Variablen die spezifischen Faktoren, die jedoch nicht mehr unkorreliert sind. Die Hauptfaktorenmethode liefert deshalb keine Lösung des faktoranalytischen Problems, wenn man dafür die Forderung als konstitutiv ansieht, daß die spezifischen Faktoren unkorreliert sein sollen. Sie hat aber andererseits den Vorteil, daß sie durchschaubar macht, wie in diesem Fall neue Variablen konstruiert werden.

[19] In der Literatur findet man meistens nur eine Aufzählung der verschiedenen Möglichkeiten und den Hinweis, daß man mehr oder weniger beliebig vorgehen kann; man vgl. z.B. Überla 1977, S. 155ff; Arminger 1979, S. 40ff.

14.3.5 Illustration der Rechenschritte

1. Um die Hauptfaktorenmethode mit einem Beispiel zu illustrieren, fassen wir zunächst die erforderlichen Rechenschritte zusammen.

a) Als Ausgangspunkt ist eine (n, m)-Datenmatrix $\mathbf{X} = (x_{ij})$ gegeben, in der sich für m Indikatoren die bei n Individuen erhobenen Werte befinden.

b) Aus der Matrix $\mathbf{X}$ wird die Korrelationsmatrix $\mathbf{Z}'\mathbf{Z}/n$ gebildet.

c) Aus dieser Korrelationsmatrix werden die Eigenwerte $\lambda_1, \ldots, \lambda_m$ und die zugehörigen Eigenvektoren $\mathbf{b}_1, \ldots, \mathbf{b}_m$ berechnet.

d) Die Eigenvektoren werden so normiert, daß $\mathbf{b}_k'\mathbf{b}_k = \lambda_k$ ist.

e) Die normierten Eigenvektoren liefern die Faktorladungen und können nach einer Entscheidung über die Anzahl der Indizes zur Berechnung von Kommunalitäten verwendet werden.

f) Schließlich kann man noch die Werte der Indizes berechnen und sich überlegen, wie gut die Korrelationsmatrix $\mathbf{Z}'\mathbf{Z}/n$ durch sie repräsentiert werden kann.

2. Zur Illustration verwenden wir die in Tabelle 6.1.1 angegebenen Daten aus unserer Statistik-Klausur. Diese Daten liefern die Matrix $\mathbf{X}$, in diesem Fall eine $(46, 5)$-Matrix. Daraus wird die Korrelationsmatrix der Indikatoren berechnet, wie sie bereits in Tabelle 6.5.1 angegeben worden ist. Hier noch einmal die vollständige Matrix, wie wir sie für die weiteren Rechnungen verwendet haben:

$$\frac{1}{n}\,\mathbf{Z}'\mathbf{Z} = \begin{pmatrix} 1.0000 & -0.0320 & 0.1985 & 0.1458 & 0.1521 \\ -0.0320 & 1.0000 & -0.1254 & 0.4434 & 0.1055 \\ 0.1985 & -0.1254 & 1.0000 & 0.1229 & 0.1175 \\ 0.1458 & 0.4434 & 0.1229 & 1.0000 & 0.1962 \\ 0.1521 & 0.1055 & 0.1175 & 0.1962 & 1.0000 \end{pmatrix}$$

Daraus gewinnt man die Eigenwerte und zugehörigen Eigenvektoren. Der Größe nach geordnet gibt es folgende Eigenwerte:

$$\lambda_1 = 1.59,\ \lambda_2 = 1.28,\ \lambda_3 = 0.85,\ \lambda_4 = 0.80,\ \lambda_5 = 0.49$$

Tabelle 14.3.1 Kommunalitäten bei unterschiedlicher Anzahl von Faktoren.

	Anzahl Faktoren				
	1	2	3	4	5
h_1^2	0.16	0.50	0.51	0.99	1.00
h_2^2	0.39	0.77	0.78	0.79	1.00
h_3^2	0.08	0.56	0.68	0.96	1.00
h_4^2	0.66	0.70	0.76	0.77	1.00
h_5^2	0.29	0.35	0.98	1.00	1.00

Die folgende Matrix enthält die normierten Eigenvektoren in entsprechender Reihenfolge:

$$\mathbf{B} = \begin{pmatrix} 0.40 & 0.58 & 0.12 & -0.69 & 0.09 \\ 0.63 & -0.61 & 0.13 & -0.02 & 0.46 \\ 0.28 & 0.69 & 0.35 & 0.54 & 0.19 \\ 0.81 & -0.20 & 0.25 & 0.08 & -0.48 \\ 0.54 & 0.23 & -0.79 & 0.15 & 0.02 \end{pmatrix}$$

Die Elemente dieser Matrix sind die Faktorladungen, die anzeigen, wie die Faktoren mit den Indikatoren korreliert sind.

3. Und wie geht es weiter? Die bisher erzielten Rechenergebnisse sind ersichtlich nicht sehr aufschlußreich. Man könnte versuchen, die Korrelationsmatrix $\mathbf{Z}'\mathbf{Z}/n$ durch die Korrelationen der Faktoren zu approximieren. Das setzt allerdings zunächst eine Entscheidung darüber voraus, wieviele gemeinsame Faktoren man haben möchte. Denn würde man alle konstruierbaren Faktoren verwenden, ergäbe sich aus (14.3.4) zunächst

$$\mathbf{Z} = \mathbf{Y}\mathbf{B}'$$

und daraus würde folgen:

$$\frac{1}{n}\mathbf{Z}'\mathbf{Z} = \frac{1}{n}\mathbf{B}\mathbf{Y}'\mathbf{Y}\mathbf{B}' = \mathbf{B}\mathbf{B}'$$

Die Korrelationsmatrix der Indikatoren könnte dann perfekt reproduziert werden. Das ändert sich erst, wenn man weniger Faktoren verwendet. Aber wieviele sollte man verwenden? Auch für diese Frage gibt es keine objektivierbaren Kriterien. Gewisse Hinweise erhält man aus den Kommunalitäten, die sich in Abhängigkeit von der Anzahl der Faktoren leicht berechnen lassen, vgl. (14.3.3). Tabelle 14.3.1 zeigt die Rechenergebnisse für unser Beispiel. Angenommen, wir beschränken uns auf die ersten drei Faktoren und dementsprechend auf die ersten drei Spaltenvektoren von

$$\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4, \mathbf{b}_5)$$

Die Ausgangsgleichung (14.3.4) wird dann zu $\mathbf{Z} = \mathbf{Y}(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)' + \mathbf{U}$, und daraus findet man:

$$\frac{1}{n}\mathbf{Z}'\mathbf{Z} = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)' + \frac{1}{n}\mathbf{U}'\mathbf{U}$$

$(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)'$ liefert dann die gewünschte Approximation der Korrelationsmatrix $\mathbf{Z}'\mathbf{Z}/n$, bis auf die Hauptdiagonale, in der die Kommunalitäten stehen. Die Matrix sieht folgendermaßen aus:

$$(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)' = \begin{pmatrix} 0.51 & -0.09 & 0.55 & 0.24 & 0.25 \\ -0.09 & 0.78 & -0.20 & 0.67 & 0.10 \\ 0.55 & -0.20 & 0.68 & 0.17 & 0.03 \\ 0.24 & 0.67 & 0.17 & 0.76 & 0.19 \\ 0.25 & 0.10 & 0.03 & 0.19 & 0.98 \end{pmatrix}$$

Man kann sich natürlich darüber streiten, ob es sich um eine akzeptable Approximation handelt oder ob man noch weitere Faktoren hinzunehmen sollte.[20]

14.3.6 Das Rotationsproblem

1. Ein weiteres Problem betrifft die Interpretation der Faktoren. Darauf wird in der faktorenanalytischen Literatur zumeist großer Wert gelegt. Allerdings sollte man sich klarmachen, daß das oft bekundete Interesse an einer Interpretation in einem gewissen Widerspruch zur üblichen Rhetorik steht. Denn wenn man dieser Rhetorik folgte, sollte man sich die Faktoren als „latente Variablen" vorstellen, die in einer empirisch nicht zugänglichen, „nicht beobachtbaren" Realität „wirksam" sind. Dann aber gibt es keinerlei Grund, warum die Faktoren einer Interpretation zugänglich sein sollten.

2. Die übliche Vorgehensweise, um zu einer Interpretation zu gelangen, besteht darin, daß man sich ansieht, wie die Faktoren mit den empirischen Indikatoren korreliert sind. Diese Korrelationen findet man in den Faktorladungen, also den Elementen der Matrix **B**. Darauf bezieht sich Thurstone:

„After the factor loadings have been determined for all of the tests or variables, it is of considerable interest to describe or name the general factors that have been isolated. This can be done best by noting which tests or variables have a relatively high positive loading with the first factor and which have a low or negative loading with this factor. By this inspection it is possible to name the first

[20] Ein gelegentlich vorgeschlagenes „Kriterium" besagt, daß man sich auf die Eigenwerte beschränken sollte, die größer als 1 sind. In unserem Beispiel müßten wir uns dann auf die ersten beiden Faktoren beschränken und hätten dann eine noch wesentlich schlechtere Rekonstruktion der Korrelationsmatrix.

factor although the statistical procedures do not of course concern these matters of describing or naming the factors. In a similar manner one might inspect the second factor loadings to ascertain which tests have positive loadings and which have negative loadings with this factor. A name might then be found for this factor and so on.“ (Thurstone 1931, S. 426)

In unserem Beispiel hilft uns allerdings dieser Ratschlag wenig, denn schon ein flüchtiger Blick auf die Matrix $\mathbf{B}$ zeigt, daß es zunächst keine offensichtlichen Zusammenhänge zwischen den Faktoren und den Aufgaben unserer Statistik-Klausur gibt.

3. Hier setzt indessen eine weitere Überlegung ein, bei der man sich eine Unbestimmtheit des faktorenanalytischen Konstruktionsverfahrens zunutze zu machen versucht, die bisher noch nicht besprochen worden ist. Angenommen, man hat sich für q Faktoren entschieden. (14.3.4) liefert dann die Basisgleichung

$$\mathbf{Z} = \mathbf{Y}\mathbf{B}' + \mathbf{U}$$

wobei $\mathbf{Y}$ eine (n, q)-Matrix ist, die in ihren Spalten die Werte der q konstruierten Variablen (Faktoren) enthält. Als Ergebnis einer Faktorenanalyse erhält man zunächst eine Variante dieser Matrix $\mathbf{Y}$; zum Beispiel mit der Hauptfaktorenmethode, wie wir sie besprochen haben. Der Konstruktionsprozeß kann indessen noch weiter getrieben werden, denn man kann aus den Faktoren

$$\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_q)$$

auf einfache Weise neue Faktoren machen, etwa $\mathbf{Y}^* := \mathbf{Y}\mathbf{T}$, wobei $\mathbf{T}$ eine beliebig gewählte (q, q)-Transformationsmatrix ist. Nun kann man sich überlegen, daß sich an der faktorenanalytischen „Lösung“ überhaupt nichts ändert, wenn man eine bestimmte Sorte von Transformationsmatrizen verwendet, um neue Faktoren zu bilden. Diese besonderen Transformationsmatrizen haben die Eigenschaft

$$\mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}_q$$

und sie werden deshalb *orthogonale* Matrizen genannt.[21] Nimmt man nämlich eine orthogonale Transformationsmatrix, findet man

$$\mathbf{Z} = \mathbf{Y}^*\mathbf{B}^{*'} + \mathbf{U} = \mathbf{Y}\mathbf{T}\mathbf{T}'\mathbf{B}' + \mathbf{U} = \mathbf{Y}\mathbf{B}' + \mathbf{U}$$

Das heißt: Man kann anstelle von $\mathbf{Y}$ genau so gut auch $\mathbf{Y}^* = \mathbf{Y}\mathbf{T}$ verwenden; man muß nur die Matrix der Faktorladungen $\mathbf{B}$ entsprechend transformieren, also

$$\mathbf{B}^* := \mathbf{B}\mathbf{T} \tag{14.3.14}$$

verwenden. Man kann infolgedessen die Faktoren und die Faktorladungen mit Hilfe orthogonaler Transformationsmatrizen weitgehend beliebig verändern, ohne durch die Annahmen des faktoranalytischen Modells eingeschränkt zu werden. Man spricht dann von einer *orthogonalen Rotation* der Faktoren bzw. ihrer Ladungsmatrix. Die Kovarianzmatrix der Faktoren bleibt dabei unverändert, denn

$$\frac{1}{n}\mathbf{Y}^{*'}\mathbf{Y}^* = \frac{1}{n}\mathbf{T}'\mathbf{Y}'\mathbf{Y}\mathbf{T} = \mathbf{T}'\mathbf{T} = \mathbf{I}_q$$

4. Die Idee besteht nun darin, sich diese Unbestimmtheit des faktoranalytischen Modells zunutze zu machen, um die Ladungsmatrix so zu transformieren, daß sich die resultierenden neuen Ladungsmuster möglichst gut – d.h. so, wie man es haben möchte – „interpretieren“ lassen. Dazu schreibt Arminger (1979, S. 79):

„Wir suchen also Ladungsmatrizen, die eine *inhaltlich interpretierbare Struktur* in dem Sinn aufweisen, daß sich die Variablen zu Gruppen zusammenfassen lassen, die jeweils nur auf einem oder möglichst wenigen Faktoren hoch laden und auf allen anderen gering, so daß jeweils ein Faktor mit dieser Variablengruppe identifiziert werden kann.“

Das Zitat geht dann so weiter:

„Wenn wir keine oder nur vage Vermutungen haben, welche Variablen zusammengefaßt werden können und wie die entsprechenden Faktoren korrelieren, betreiben wir *exploratorische* Faktorenanalyse. Wenn wir hingegen Hypothesen über die Ladungen von Variablen auf Faktoren und deren Korrelationen testen, sprechen wir von *konfirmatorischer* Faktorenanalyse [...].“

Der Unterschied zwischen beiden Vorgehensweisen ist allerdings im wesentlichen nur rhetorischer Natur. Bei der exploratorischen Faktorenanalyse wird solange nach transformierten Faktoren gesucht, bis man anhand der korrespondierenden Ladungsmatrix eine Interpretation findet, die einem gefällt. Bei der konfirmatorischen Faktorenanalyse beginnt man mit einer Interpretation – in Gestalt von Annahmen über die Struktur einer Ladungsmatrix – und sucht dann nach einer Transformation der Faktoren, so daß die korrespondierende Ladungsmatrix der vorab angenommenen

---

[21] Man kann sich beliebig viele solcher orthogonalen Matrizen ausdenken. Hier ist ein Beispiel für $q = 2$:

$$\mathbf{T} = \begin{pmatrix} 0.5 & \sqrt{0.75} \\ -\sqrt{0.75} & 0.5 \end{pmatrix}$$

Man kann nachrechnen, daß $\mathbf{T}\mathbf{T}' = \mathbf{I}_2$ ist. Im Fall $q = 2$ kann man auch die Gesamtheit aller orthogonalen Matrizen angeben, nämlich:

$$\mathbf{T} = \begin{pmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{pmatrix}$$

wobei ϕ ein beliebiger Winkel ist. Im vorab gegebenen Beispiel ist $\phi = 60°$.

möglichst nahe kommt. Wenn sich eine akzeptable Übereinstimmung finden läßt, hat sich die vorab vermutete Interpretation bestätigen lassen. Und andernfalls? Dann muß man so oder so nach einer neuen Interpretation suchen oder, was vielleicht eine bessere Idee wäre, den gesamten Ansatz verwerfen.

5. Einige Autoren warnen gelegentlich vor den Beliebigkeiten einer Interpretation. Zum Beispiel schreibt Kerlinger:

„In considering the scientific value of factor analysis, the reader must be cautioned against attributing 'reality' and uniqueness to factors that may not exist. The danger of reification is great. It is easy to name a factor and then to believe there is a reality behind the name. But giving a factor a name does not give it reality." (Kerlinger 1964, S. 683)

Aber die Warnung scheint nicht ernst gemeint; denn derselbe Autor sagt über die konfirmatorische Faktorenanalyse:

„The second major purpose of factor analysis is to test hypotheses. [...] The typical design of such studies has been well outlined by Thurstone, Cattell, Guilford, and others. First, factors are 'discovered'. Their nature is inferred from the tests that are loaded on them. This 'nature' is set up as a hypothesis. New tests are constructed and given to new samples of subjects. The data are factor-analyzed. If the factors emerge *as predicted*, the hypothesis is to this extent confirmed, the factors would seem to have 'reality'. This would certainly not end the matter. One would have to test, among other things, the factors' relation to other factors. One would have to place the factors, as constructs, in a nomological network of constructs." (Kerlinger 1964, S. 682)

Das leitende Motiv richtet sich also gleichwohl darauf, daß man „reale Faktoren" finden möchte; und ihre „Realität" erscheint dadurch nachweisbar, daß sich zeitlich stabile Korrelationen zwischen stets ähnlich interpretierbaren Faktoren finden lassen.

6. Für die praktische Durchführung von Faktorrotationen bei exploratorischen Faktorenanalysen sind eine große Anzahl unterschiedlicher Verfahren vorgeschlagen worden. Sie orientieren sich daran, wie in dem oben angeführten Zitat von Arminger angedeutet wird, eine „möglichst einfache" Struktur für die Ladungsmatrix der transformierten Faktoren zu finden. Was hierbei „möglichst einfach" heißen soll, ist indessen nicht nur unklar, so daß man auch darüber viele unterschiedliche Vorstellungen entwickeln kann. Es kommt auch gar nicht darauf an, an dieser Stelle nach einer begrifflichen Präzisierung zu suchen. Denn wie auch immer man bei der Suche nach einer „möglichst einfachen" Struktur einer Ladungsmatrix vorgeht, das Ziel besteht ja darin, zu einer Interpretation der Faktoren zu gelangen. Die Rotation der Faktoren ist dafür nur als ein Hilfsmittel gedacht und kann infolgedessen beliebig erfolgen.[22] Wichtig erscheint nur, daß man am

[22]So betrachtet, ist es auch nicht erforderlich, sich auf orthogonale Transformationen

Schluß zu *irgendeiner* Interpretation der Faktoren gelangt. Wir werden uns deshalb in diesem Text mit den verschiedenen Verfahren, die zur Rotation von Faktoren vorgeschlagen worden sind, nicht näher beschäftigen.

14.3.7 Prokrustes-Rotation

1. Bei der konfirmatorischen Faktorenanalyse stellt sich das Rotationsproblem in etwas anderer Form dar. Denn dann gibt es einerseits eine Ladungsmatrix $\mathbf{B}$ als Resultat der Faktorenanalyse, z.B. mit der Hauptkomponentenmethode, und andererseits eine gewünschte Ladungsmatrix $\tilde{\mathbf{B}}$, die die gewünschte Struktur der Faktorladungen ausdrückt. Die Aufgabe besteht dann darin, eine orthogonale Transformationsmatrix $\mathbf{T}$ zu finden, so daß sich $\mathbf{BT}$ und $\tilde{\mathbf{B}}$ möglichst ähnlich werden. Um zu präzisieren, was hier „möglichst ähnlich" heißen soll, kann man die Elemente der Matrizen $\tilde{\mathbf{B}}$ und $\mathbf{BT}$ vergleichen, zum Beispiel mit Hilfe der elementweise genommenen quadrierten Differenzen. Die Aufgabe besteht dann darin, eine orthogonale Matrix $\mathbf{T}$ zu finden, so daß

$$\sum_{j=1}^{m} \sum_{k=1}^{q} (\tilde{\mathbf{B}} - \mathbf{BT})_{jk}^2$$

möglichst klein wird. Wenn die Transformationsmatrix $\mathbf{T}$ auf diese Weise berechnet wird, spricht man von einer *Prokrustes-Rotation*.[23]

2. Die mathematischen Überlegungen, wie man eine für das oben angegebene Kriterium optimale Transformationsmatrix finden kann, sind relativ kompliziert und sollen hier nicht besprochen werden.[24] Wir beschränken uns auf eine Illustration, die an die oben beschriebene Hauptfaktorenanalyse unserer Klausurdaten anknüpft. Wenn man sich auf die ersten drei Faktoren beschränkt, findet man die Ladungsmatrix

$$\mathbf{B} = \begin{pmatrix} 0.40 & 0.58 & 0.12 \\ 0.63 & -0.61 & 0.13 \\ 0.28 & 0.69 & 0.35 \\ 0.81 & -0.20 & 0.25 \\ 0.54 & 0.23 & -0.79 \end{pmatrix}$$

zu beschränken. Sie führen ja dazu, wie oben gezeigt worden ist, daß auch die transformierten Faktoren unkorreliert sind. Wenn man dagegen korrelierte Faktoren haben möchte (man erinnere sich an das in Abschnitt 14.3.1 angeführte Zitat von Arminger), kann man auch nicht-orthogonale Transformationsmatrizen verwenden.

[23]Es gibt verschiedene Varianten dieses Verfahrens. Meist werden zusätzlich zur Rotation auch noch Translationen und Skalierungen zugelassen.

[24]Eine Darstellung dieser mathematischen Überlegungen findet sich z.B. bei Mardia, Kent und Bibby 1979, S. 416ff.

Nun könnte man sich zum Beispiel eine Transformation der Faktoren wünschen, so daß der erste neue Faktor vor allem mit der ersten und dritten Klausuraufgabe korreliert, der zweite neue Faktor vor allem mit der zweiten und vierten Klausuraufgabe und der dritte neue Faktor vor allem mit der fünften Klausuraufgabe, so daß die transformierte Ladungsmatrix so aussehen sollte:

$$\tilde{\mathbf{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Gesucht ist dann eine orthogonale Transformationsmatrix $\mathbf{T}$, in unserem Beispiel eine $(3,3)$-Matrix, so daß sich $\mathbf{BT}$ und $\tilde{\mathbf{B}}$ im Sinne des oben genannten Kriteriums möglichst ähnlich werden. Verwendet man das Verfahren der Prokrustes-Rotation, findet man die Transformationsmatrix

$$\mathbf{T} = \begin{pmatrix} 0.40 & 0.81 & 0.42 \\ 0.84 & -0.51 & 0.19 \\ 0.37 & 0.27 & -0.89 \end{pmatrix}$$

und die neue Matrix der Faktorladungen ist

$$\mathbf{BT} = \begin{pmatrix} \underline{0.69} & 0.07 & 0.17 \\ -0.21 & \underline{0.86} & 0.03 \\ \underline{0.82} & -0.03 & -0.06 \\ 0.25 & \underline{0.83} & 0.08 \\ 0.11 & 0.11 & \underline{0.98} \end{pmatrix}$$

Das ist zwar nicht genau das, was man sich gewünscht hat, kommt aber doch der „Hypothese" ziemlich nahe. Wer möchte, könnte jetzt anfangen, die drei neuen Faktoren zu interpretieren.

14.4 Hilfsmittel der Begriffsbildung?

1. In der Methodenliteratur wird oft die Auffassung vertreten, daß faktorenanalytische Methoden auch als ein Hilfsmittel für Indexkonstruktionen verwendet werden können, also zur Überprüfung von Begriffsbildungen im Sinne von Variablenkonstruktionen.[25] Das impliziert zunächst eine Abgrenzung von der in Abschnitt 14.1 kritisierten Rhetorik, bei der faktorenanalytische Methoden dazu dienen sollen, um Faktoren zu finden, mit

[25] Um genau zu sein, sollten Variablen und Begriffe unterschieden werden. Das begriffsbildende Moment bei der Definition von Variablen betrifft die Konzeption des jeweils vorauszusetzenden Merkmalsraums. Mit dem Wort 'Begriffsbildung' wird hier also auf die Konzeption von Merkmalsräumen verwiesen.

denen empirische Sachverhalte erklärt werden können. Statistische Variablen sind keine Faktoren, sondern dienen der Charakterisierung von Gesamtheiten von Objekten. Die im folgenden zu erörternde Frage ist, ob faktorenanalytische Methoden bei der Konstruktion von Indizes hilfreich sein können.

2. Denkt man zunächst an die einfache Faktorenanalyse, die in Abschnitt 14.2 besprochen wurde, kann eine Antwort sofort gegeben werden. Wie man gesehen hat, ist die einfache Faktorenanalyse gleichbedeutend mit einem Verfahren, bei dem ein additiver gewichteter Index

$$Y = \sum_{j=1}^{m} w_j X_j$$

so konstruiert wird, daß seine Varianz maximal wird.[26] Dies Kriterium der Varianzmaximierung ist jedoch, wie bereits in Kap. 13 besprochen worden ist, ganz unerheblich für die Frage, ob und ggf. wie ein sinnvoller Index konstruiert werden kann.

3. Insoweit sich auch die multiple Faktorenanalyse an einem Kriterium der Varianzmaximierung orientiert – nur etwas allgemeiner, da in diesem Fall mehrere Indizes konstruiert werden –, führt eine analoge Überlegung zu der Einsicht, daß auch mit der multiplen Faktorenanalyse keine inhaltlich sinnvollen Indizes konstruiert werden können. Es gibt indessen ein Problem, auf das bereits in Abschnitt 6.5 hingewiesen worden ist: Die Bildung eines additiven Index aus einer Reihe quantitativer Indikatoren erscheint nur dann sinnvoll, wenn sich alle Indikatoren auf die gleiche Dimension beziehen. Wie bereits diskutiert wurde, ist es schwierig, genau zu klären, was „gleiche Dimension" bedeuten soll. Man erinnere sich an die Beispiele, die in Abschnitt 6.1 besprochen worden sind. Bei Preisindizes schien die Annahme, daß sich die Indikatoren auf eine gleiche Dimension (Preisgrößen) beziehen, unproblematisch. Ebenso bei unserem Index zur Bewertung der Ergebnisse einer Statistikklausur. Problematisch erschien andererseits der Schichtindex von Scheuch und Daheim, der auf einer additiven Zusammenfassung von Indikatoren für Einkommen, Bildung und Beruf beruht. Denn das Merkmal „Beruf" kann überhaupt nicht als eine Dimension im Sinne einer quantitativen Größe aufgefaßt werden; und beim Merkmal „Bildung" ist dies zumindest fragwürdig. Aber auch wenn man glaubt, sich einen quantitativen Indikator für „Bildung" verschaffen zu können, hat dieser sicherlich eine andere Dimension als „Einkommen".

4. Hier setzt nun ein in der Methodenliteratur oft vertretener Gedanke ein:

[26] Ein Unterschied liegt nur darin, daß bei der faktorenanalytischen Variante des Verfahrens meistens von standardisierten Indikatoren ausgegangen wird. Für die einfache Faktorenanalyse ist dies jedoch keine wesentliche Voraussetzung.

daß Methoden der multiplen Faktorenanalyse helfen können, um festzustellen, ob sich Indikatoren, aus denen ein additiver Index gebildet werden soll, auf eine gleiche Dimension beziehen. Exemplarisch sei auf folgende Ausführungen von Laatz verwiesen:

„Abschließend wollen wir noch knapp auf die heute gebräuchliche Verwendung der *Faktorenanalyse* für die Prüfung der Skalenqualität eingehen. Der Begriff Faktorenanalyse faßt eine Reihe unterschiedlicher Verfahren zusammen, die es je nach Anwendungsart ermöglichen, aus einer größeren Menge Meßvariablen eine kleinere Menge dahinterstehender *latenter Variablen* zu extrahieren (R-Typ) oder Fälle zu Klassen zu gruppieren (Q-Typ). Man sucht eine lineare Kombination möglichst weniger Faktoren (latenter Variablen), mit der man ohne großen Verlust der Erklärungskraft die Befunde der beobachteten Variablen rekonstruieren kann. In unserem Falle interessiert die erstgenannte Möglichkeit.
Mit ihr kann man einerseits die *Eindimensionalität*, mit anderen Worten die Gültigkeit der Gesamtskala prüfen. Diese wäre gegeben, wenn alle Meßvariablen überwiegend Ausdruck einer einzigen latenten Variablen (eines Faktors) wären. Andererseits kann man sie zur *Itemanalyse* verwenden, denn es ist ja möglich, Eindimensionalität dadurch herzustellen, daß man items aus der Skala ausscheidet, die nicht überwiegend durch die interessierende latente Variable bestimmt sind.“ (Laatz 1993, S. 292f)[27]

Konzentrieren wir uns auf das erste Argument.[28] Kann mit einer Faktorenanalyse geprüft werden, ob sich alle Indikatoren auf die gleiche Dimension beziehen? Was man mit einer Faktorenanalyse herausfinden kann, ist tatsächlich etwas anderes: wieviele Indizes („latente Variablen“) erforderlich sind, um mit ihnen die Korrelationsmatrix der Indikatoren näherungsweise darzustellen. Das hat jedoch mit der Frage, auf welche Dimensionen sich die Indikatoren beziehen, nichts zu tun. Denken wir an unseren Index zur Bewertung der Ergebnisse einer Statistik-Klausur. Die Faktorenanalyse mit unseren Beispieldaten hat gezeigt, daß in diesem Fall ein einziger Indikator sicherlich nicht ausreicht, um die Korrelationsmatrix näherungsweise darzustellen. Dennoch beziehen sich alle Indikatoren auf die gleiche Dimension. Denken wir andererseits an Indikatoren für Einkommen und Bildung. Es gibt eine ziemlich hohe Korrelation, so daß eine Faktorenanalyse zeigen würde, daß man beide Variablen auf eine „latente Variable“ reduzieren kann. Aber dennoch verweisen Einkommen und Bildung auf ganz unterschiedliche Dimensionen. Man kann sich beliebig viele weitere

[27] Ähnliche Ausführungen findet man bei Schnell, Hill und Esser 1999, S. 152ff. Kerlinger 1964, S. 650, schreibt: „Factor analysis serves the cause of scientific parsimony. Generally speaking, if two tests measure the same thing, the scores obtained from them can be added together. If, on the other hand, the two tests do not measure the same thing, their scores cannot be added together. Factor analysis tells us, in effect, what tests or measures can be added and studied together rather than separately.“ Etwas später (S. 680) fügt der Autor hinzu: „Factor analysis can be conceived as a construct validity tool.“

[28] Die zweite Überlegung ist offenbar davon abhängig. Eine kurze Bemerkung dazu erfolgt am Ende des nächsten Abschnitts.

Beispiele ausdenken. Besonders deutlich wird der Denkfehler, wenn man sich auf physikalische Dimensionen bezieht. Man denke z.B. an die Körpergröße und das Gewicht von Menschen. Wenn man entsprechende Variablen bildet, wird man eine hohe Korrelation finden; aber es wäre absurd, mit einem solchen Befund zeigen zu wollen, daß Länge und Gewicht Indikatoren einer gleichen Dimension sind.

14.5 Bemerkungen zum Dimensionsbegriff

1. Um zu verstehen, warum sich dennoch der Gedanke verbreiten konnte, durch Techniken der Korrelationsrechnung Einsichten in Dimensionen gewinnen zu können, ist es nützlich, verschiedene Varianten des Redens von Dimensionen zu unterscheiden. Im Anschluß an den sehr diffusen Sprachgebrauch in der Methodenliteratur haben wir in Abschnitt 6.5 den von uns verwendeten Dimensionsbegriff so fixiert: Mit dem Wort ‘Dimension’ ist ein Größenbegriff gemeint; wird dann durch Bezugnahme auf einen solchen Größenbegriff ein quantitativer Merkmalsraum definiert, kann auch auf ihn mit dem Wort ‘Dimension’ verwiesen werden. Dieses Verständnis des Dimensionsbegriffs orientiert sich an der naturwissenschaftlichen Tradition. Davon kann zunächst unterschieden werden, wie in der Mathematik von Dimensionen gesprochen wird.

2. Hier genügt es, daß wir uns auf den Dimensionsbegriff beziehen, wie er in der analytischen Geometrie verwendet wird. Ausgangspunkt ist der Begriff eines Vektorraums. Im einfachsten Fall kann man sich einen m-dimensionalen Vektorraum als Menge aller Vektoren

$$(x_1, \ldots, x_m)$$

vorstellen, wobei die Elemente dieser Vektoren reelle Zahlen sind. Außerdem sind für diese Vektoren eine Reihe von Operationen definiert. Zum Beispiel kann man Vektoren addieren und subtrahieren, und man kann sie mit einfachen reellen Zahlen (sog. Skalaren) multiplizieren. Eine vollständige Definition ist an dieser Stelle nicht erforderlich. Wichtig ist, daß sich der Dimensionsbegriff in diesem Fall auf einen Raum, im mathematischen Sinne des Wortes, bezieht. Der Vektorraum, der aus den Vektoren $(x_1, \ldots, x_m)$ besteht, wird *m-dimensional* genannt, weil die Komponenten der Vektoren in m Dimensionen unabhängig voneinander variieren können. Man kann das auch so sagen: Um die Gesamtheit der Vektoren eines m-dimensionalen Vektorraums darstellen zu können, werden m Größenbegriffe benötigt, wobei in unserem Beispiel für jeden Größenbegriff eine numerische Repräsentation durch die reellen Zahlen angenommen wird. Den anschaulichen Hintergrund liefern 1-, 2- und 3-dimensionale Koordinatensysteme. Zum Beispiel benötigt man für jeden Punkt in einem 2-dimensionalen Koordinatensystem zwei Zahlangaben, um seine Lage innerhalb des Koordinatensystems anzugeben.

3. Insoweit kann man von einem *raumbezogenen* Dimensionsbegriff sprechen. Er verweist in diesem Zusammenhang auf die Anzahl der Größenbegriffe, die man benötigt, um die innerhalb eines Raums *möglichen* Stellen (Punkte) zu bestimmen. Eine ganz andere Blickweise ergibt sich, wenn man sich nicht auf eine Gesamtheit von Lokalisations*möglichkeiten* in einem Raum bezieht, sondern auf eine jeweils *gegebene* Menge von Daten. Im einfachsten Fall hat man es mit einer einfachen Datenmatrix

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

zu tun, und die Frage betrifft die „Dimensionen" einer solchen Datenmatrix. Offenbar läßt sich ein Zusammenhang zur Idee eines Vektorraums herstellen. Die Datenmatrix kann wahlweise als eine Menge von n Zeilenvektoren oder m Spaltenvektoren aufgefaßt werden. Bei der üblichen Interpretation charakterisieren die Zeilenvektoren jeweils ein Objekt in Ω, und die Spaltenvektoren entsprechen jeweils einer Variablen (wir verwenden die Bezeichnungen $X_1, \ldots, X_m$). Die Idee ist: Die Zeilenvektoren können als n Punkte in einem m-dimensionalen Vektorraum, die Spaltenvektoren als m Punkte in einem n-dimensionalen Vektorraum aufgefaßt werden. Somit erscheint es naheliegend zu sagen: Die Datenmatrix hat die Zeilendimension n und die Spaltendimension m.

4. Hier setzt nun eine weitere Überlegung ein. Um sie zu verfolgen, gehen wir von den n Zeilenvektoren $\mathbf{x}_i := (x_{i1}, \ldots, x_{im})$ aus, die wir als Punkte in einem m-dimensionalen Vektorraum auffassen.[29] Der Vektorraum ist in diesem Fall m-dimensional, denn man benötigt m Dimensionen, um alle in diesem Vektorraum möglichen Vektoren darstellen zu können. Wenn man es dagegen nur mit einer endlichen Menge von Vektoren zu tun hat, bei unserer Datenmatrix $\mathbf{X}$ also nur mit n Vektoren, kann man sich überlegen, ob man wirklich einen m-dimensionalen Vektorraum benötigt, um die n Vektoren darzustellen, oder ob dafür ein Vektorraum mit einer geringeren Anzahl von Dimensionen genügt. Betrachten wir als Beispiel die Datenmatrix

$$\mathbf{X} := \begin{pmatrix} 1 & 2 \\ 3 & 3 \\ 5 & 4 \\ 7 & 5 \end{pmatrix}$$

Wenn man die 4 Zeilenvektoren dieser Matrix als Punkte in ein 2-dimensionales Koordinatensystem einträgt, sieht man, daß sie alle auf einer graden

[29] Alternativ könnte man von den m Spaltenvektoren ausgehen und sich auf einen n-dimensionalen Vektorraum beziehen.

Linie liegen. Es genügt also, so kann man zunächst anschaulich argumentieren, ein 1-dimensionaler Raum (eine Linie), um die 4 Punkte darzustellen.

5. Überlegen wir uns, wie dieser Gedanke präzisiert werden kann. Zu diesem Zweck bleiben wir zunächst bei unserem kleinen Zahlenbeispiel. Man sieht, daß jeweils die eine der beiden Variablen (wir nennen sie X_1 und X_2) als eine lineare Funktion der anderen Variablen dargestellt werden kann, nämlich:

$$\begin{aligned} X_1 &= 2\,X_2 - 3 \\ X_2 &= 0.5\,X_1 + 1.5 \end{aligned}$$

Somit genügt zur Kenntnis der Matrix $\mathbf{X}$, daß man eine der beiden Variablen kennt und weiß, wie man daraus die andere berechnen kann. In diesem Sinne könnte man also die Matrix „eindimensional" nennen. In der analytischen Geometrie ist es allerdings nicht üblich, hier das Wort 'Dimension' zu verwenden, man spricht vielmehr vom *Rang* einer Matrix. In unserem Beispiel hat die Matrix den Rang 1.[30]

6. Fassen wir kurz zusammen. Wir haben gefunden, daß man auf drei unterschiedliche Weisen von Dimensionen sprechen kann.

a) Man kann sich mit dem Wort 'Dimension' erstens auf Größenbegriffe und – davon ausgehend – auf quantitative Merkmalsräume von Variablen beziehen. Dies impliziert, daß zwei Dimensionen unterschiedlich sind, wenn sie sich in ihrer Bedeutung unterscheiden, z.B. Länge und Gewicht.

b) Man kann sich mit dem Wort auch auf die Dimensionen eines Vektorraums beziehen. Hierbei wird zwar von der inhaltlichen Bedeutung der Dimensionen abstrahiert; wichtig ist jedoch, daß sich der Dimensionsbegriff in diesem Fall auf die Vorstellung eines Raums im mathematischen Sinne des Wortes bezieht. Die Anzahl der Dimensionen ist eine Eigenschaft des Raums und wird vorausgesetzt, um Lokalisations*möglichkeiten* innerhalb des Raums zu charakterisieren.

c) Schließlich kann man den Dimensionsbegriff auf Datenmatrizen beziehen, indem man von der Frage ausgeht, wieviele Zeilen- oder Spaltenvektoren erforderlich sind, um daraus die gesamte Matrix berechenbar zu machen. Wenn man für „Berechenbarkeit" lineare Funktionen verlangt, kommt man zum Begriff des Rangs einer Matrix und kann dementsprechend von ihren „Dimensionen" reden.[31]

[30] Diese Begriffsbildung wird im Anhang A.4 genauer erklärt.

[31] Es wird erst kompliziert, wenn man beliebige, nicht nur lineare Funktionen zuläßt. Auf diesen Gedanken ist zuerst Georg Cantor, der Begründer der Mengenlehre, gekommen, indem er sich eine Funktion ausgedacht hat, mit der man die Punkte eines

7. Die beiden ersten Varianten können auf einfache Weise aufeinander bezogen werden; (b) ist gewissermaßen nur eine abstrakte Version von (a). In beiden Fällen kann man sich Dimensionen als Größen vorstellen, die durch reelle Zahlen eine numerische Repräsentation bekommen. Ein Unterschied besteht nur darin, daß in der mathematischen Konzeption eines Vektorraums von der inhaltlichen Bedeutung der Größen (Dimensionen), auf die er sich bezieht, abstrahiert wird. Eine wesentlich andere Perspektive beginnt erst, wenn man den Dimensionsbegriff nicht auf einen Merkmals- oder Vektor*raum* bezieht, sondern auf Mengen von Daten. Dieser Perspektivenwechsel bildet indessen den Ausgangspunkt, wenn bei Datenanalysen – insbesondere bei der Verwendung faktoranalytischer Methoden – von „Dimensionen" gesprochen wird.

8. Versuchen wir, den Unterschied etwas genauer zu verstehen. Wenn im Sinne von (a) oder (b) von Dimensionen gesprochen wird, bezieht man sich auf Größenbegriffe, die *vorab* definiert sein müssen, bevor man Objekte durch jeweils bestimmte Größen charakterisieren kann. Zum Beispiel müssen zunächst die Größenbegriffe 'Länge' und 'Gewicht' definiert werden, bevor man entsprechende Größen bei bestimmten Objekten feststellen kann. Für die Definition solcher Größenbegriffe ist es weiterhin von entscheidender Wichtigkeit, daß sie von den jeweils spezifischen Größen bestimmter Objekte möglichst unabhängig erfolgen kann. Deshalb können auch mit verteilungsabhängigen Indizes keine Größenbegriffe begründet werden. Wenn man zum Beispiel bei Objekten einer bestimmten Art, zum Beispiel bei Menschen, feststellt, daß ihre Körpergröße und ihr Körpergewicht miteinander zusammenhängen (indem man z.B. eine hohe statistische Korrelation ermittelt), ist das für die Definition der Größenbegriffe 'Länge' und 'Gewicht' ganz unerheblich. Sie müssen vielmehr als jeweils selbständige Dimensionen konzipiert werden, damit sie hinterher für empirische Untersuchungen genutzt werden können. Zugrunde liegt die Vorstellung, daß man nur durch empirische Untersuchungen lernen kann, welche Größen die verschiedenen Objekte aufweisen und ob und ggf. wie ihre Größen miteinander zusammenhängen. Solche Zusammenhänge, auch wenn es sie geben sollte, können deshalb nicht in der Definition von Größenbegriffen vorweggenommen werden.

9. Daraus ergibt sich eine wichtige Konsequenz. Dimensionen im Sinne von Größenbegriffen können nicht mit Variablen identifiziert werden, sondern müssen auf ihre Merkmalsräume bezogen werden. Denken wir an m

Quadrats auf die Punkte einer Strecke abbilden kann. Das führte dann zu der irritierenden Vorstellung, daß Flächen und Linien die gleiche „Dimension" haben, und bildete den Ausgangspunkt, um im Rahmen der Mengenlehre und Topologie nach neuen Definitionen für einen mathematischen Dimensionsbegriff zu suchen. Eine gute Einführung findet sich bei Waismann 1996. Eine einfache Darstellung von Cantors Gedankengang findet sich bei Meschkowski 1997, S. 121ff.

quantitative Variablen

$$X_j : \Omega \longrightarrow \tilde{\mathcal{X}}_j \qquad (j = 1, \ldots, m)$$

Eine Variable zeigt, welche Merkmalsausprägungen (Werte in $\tilde{\mathcal{X}}_j$) die Objekte in Ω in einem jeweils gegebenen räumlichen und zeitlichen Kontext aufweisen.[32] Die Werte von Variablen, und mithin die Variablen selbst, können sich also ändern. Andererseits bezieht sich eine Variable, auch wenn sich ihre Werte verändern, immer auf die gleiche Dimension, d.h. auf den gleichen Merkmalsraum. Wenn von der Dimension einer quantitativen Variablen gesprochen wird, ist also ihr Merkmalsraum gemeint bzw. der Größenbegriff, durch den der Merkmalsraum definiert wird.

10. Wie verhält es sich, wenn man sich nicht auf Merkmalsräume bezieht, sondern von einer gegebenen Menge von Daten ausgeht? Beziehen wir uns, wie bisher, auf eine Datenmatrix $\mathbf{X}$, die uns die Werte von m Variablen $X_1, \ldots, X_m$ für n Objekte liefert; und nehmen wir an, daß $\mathbf{X}$ den Rang q hat und daß q kleiner als m ist. Dies bedeutet, daß man maximal q linear unabhängige Spaltenvektoren finden kann. Da die Reihenfolge der Spaltenvektoren in $\mathbf{X}$ keine Rolle spielt, kann man auch annehmen, daß die ersten q Spalten von $\mathbf{X}$ linear unabhängig sind, und die restlichen $m-q$ Spalten durch lineare Funktionen von $X_1, \ldots, X_q$ berechenbar machen:

$$X_j \;=\; f_j(X_1, \ldots, X_m) \quad \text{für} \quad j = q+1, \ldots, m$$

Somit liegt es nahe zu sagen: Es genügen die ersten q Dimensionen, um die gesamte Matrix $\mathbf{X}$ darzustellen. Es ist jedoch wichtig, sich klar zu machen, daß man dann auf eine besondere und eigentlich irreführende Weise von Dimensionen spricht. Folgende Überlegung kann das zeigen. Hätte man sich bei der Erfassung der Daten tatsächlich auf die ersten q Dimensionen, also auf die Variablen $X_1, \ldots, X_q$, beschränkt, gäbe es überhaupt nicht die Möglichkeit, daraus die Werte der Variablen $X_{q+1}, \ldots, X_m$ zu berechnen. Das kann man offenbar nur dann tun, wenn man die Funktionen f_j (für $j = q+1, \ldots, m$) bereits kennt. Diese Funktionen gewinnt man jedoch erst aus einer Kenntnis der gesamten Matrix $\mathbf{X}$. Man muß sich also zunächst auf alle m Dimensionen beziehen, um Werte für die Variablen $X_1, \ldots, X_m$ zu gewinnen. Erst hinterher kann dann gegebenenfalls festgestellt werden, daß der Rang der Datenmatrix kleiner als m ist.

11. Bei dieser Überlegung wird davon ausgegangen, daß die Werte einer Datenmatrix nur durch eine empirische Untersuchung von Objekten (bzw.

[32] Mit dem Begriff einer Variablen wird also keineswegs nur auf einen Begriff Bezug genommen, sondern zugleich auf eine Menge von Objekten. Die Bezugnahme auf eine Begriffsbildung wird durch den Merkmalsraum einer Variablen präzisiert, die Bezugnahme auf Objekte durch eine Objektmenge Ω. Eine solche Bezugnahme auf Objekte ist im übrigen auch Voraussetzung dafür, daß man statistische Variablen als Datenvektoren auffassen kann.

Situationen in denen sie sich befinden) zugänglich werden und daß es sich bei den Daten, die man dann gewinnt, um numerische Repräsentationen *kontingenter* Sachverhalte handelt. Damit ist gemeint, daß sich bei einer Wiederholung der Untersuchung bei den gleichen oder bei anderen Objekten für die gleichen Variablen andere Werte ergeben können. Infolgedessen ist dann auch der Rang einer Datenmatrix ein kontingenter Sachverhalt, der nicht als eine Eigenschaft der Dimensionen $\tilde{\mathcal{X}}_1, \ldots, \tilde{\mathcal{X}}_m$ aufgefaßt werden kann, sondern nur als eine Eigenschaft der Variablen $X_1, \ldots, X_m$, d.h. ihrer Werte. An dieser Stelle wird jedoch eine „weltanschauliche" Frage relevant. Denn man könnte ja *glauben*, daß den Funktionen f_j, die man aus einer empirischen Datenmatrix gewinnen kann, Gesetzmäßigkeiten entsprechen, denen die Objekte, auf die sich die Daten beziehen, unterworfen sind. Aus diesem Glauben würde dann folgen, daß es tatsächlich ausreicht, die Werte der Variablen für die ersten q Dimensionen zu kennen. Denn aus diesen Werten würde sich infolge der als Gesetzmäßigkeiten wirksamen Funktionen f_j zwangsläufig ergeben, welche Werte die übrigen Variablen aufweisen.[33]

12. Die Überlegung führt zu dem Ergebnis, daß zwei Arten von Fragen deutlich unterschieden werden sollten. Einerseits Fragen, die sich auf die Dimension von Variablen beziehen; in diesem Fall geht es darum, ob und ggf. wie ihre Merkmalsräume durch Größenbegriffe explizierbar sind. Andererseits Fragen, die sich auf die Kovariationsmöglichkeiten der Werte von zwei oder mehr Variablen beziehen. Fragen dieser Art sind ganz unabhängig davon, wie die Merkmalsräume (Dimensionen) der Variablen beschaffen sind. Daß auch bei Fragen der zweiten Art oft von „Dimensionsanalysen" gesprochen wird, hat leider viel zur Verwirrung beigetragen. Z.B. heißt es in einem Beitrag von K. Fukunaga (1982, S. 348):

„Whenever we are confronted by large multidimensional data sets, it is usually advantageous for us to discover or impose some structure on the data. Therefore, we might assume that the data is governed by a certain number of underlying parameters. The minimum number, n_0, of parameters required to account for the observed properties of the data is called the *intrinsic dimensionality for representation*, or, equivalently, the data generating process."

Äußerungen dieser Art sollten indessen nicht nur deshalb kritisiert werden, weil sie die Unterscheidung zwischen den beiden vorab genannten Fragestellungen verwirren. Denn die Verwirrung geht ja noch weiter, indem Fragen, die sich auf die *Repräsentation* einer Menge von Daten beziehen, mit Fragen vermengt werden, die sich auf einen „data generating process" beziehen, also darauf, wie die in Form von Daten erfaßten Sachverhalte

[33] Es bleibt dann immer noch die Frage, wie man sich Kenntnisse über die Funktionen f_j verschaffen kann. Aber wenn es sich wirklich um Gesetzmäßigkeiten handeln würde, und nicht nur um ein Mittel zur Darstellung kontingenter Daten, würde es genügen, sie nur einmal festzustellen. Allerdings ist es bisher nicht gelungen, solche Gesetzmäßigkeiten im sozialen Leben zu finden.

in der empirischen Realität entstanden sein könnten. Insofern präsentiert dieses Zitat nur eine besonders unreflektierte Variante der Annahme, daß aus den Repräsentationsmöglichkeiten von Daten durch statistische Modelle Einsichten in Gesetzmäßigkeiten der empirischen Realität gewonnen werden könnten.

13. Kommen wir zum Schluß zurück zur Faktorenanalyse. Nach den bisherigen Überlegungen sollte deutlich geworden sein, warum der Glaube, mit faktorenanalytischen Methoden „Dimensionsanalysen" vornehmen zu können, irreführend ist. Denn der Frage, wie die Merkmalsräume der empirischen Variablen, mit denen eine Faktorenanalyse beginnt, beschaffen sind, wird mit ihren Rechentechniken gar nicht nachgegangen. Insbesondere wird gar nicht überlegt, ob und ggf. wie sich die jeweils verwendeten empirischen Variablen auf Dimensionen beziehen. Stattdessen wird eine vollständig andere Frage verfolgt: wieviele linear unabhängige Variablen (Indizes) ausreichend erscheinen, um die Korrelationsmatrix der empirischen Variablen näherungsweise zu *repräsentieren*. Stellt man fest, daß q Indizes ausreichend erscheinen, könnte man mit Fukunagas Sprachgebrauch daraus eine Aussage über die „intrinsic dimensionality" der Korrelationsmatrix machen, mit der man begonnen hat. Aber nichts rechtfertigt den Glauben, auf diese Weise Dimensionen finden zu können, die zur Beschreibung und Erklärung der empirischen Realität relevant sein könnten. Im Gegenteil, man findet bestenfalls alternative Repräsentationsmöglichkeiten für statistische Zusammenhänge zwischen empirischen Variablen, deren Merkmalsräume durch das Verfahren nicht reflektiert werden können. Infolgedessen erweist sich auch der zweite Gedanke, den Laatz in dem in Abschnitt 14.4 angeführten Zitat andeutet, als eine Sackgasse, die auf einer Verwechslung von Fragestellungen beruht.

Kapitel 15

Skalierungsmodelle

Skalierungsverfahren dienen zunächst dem Zweck, Indizes zu konstruieren und zu prüfen. Als Beispiel kann man an die Methode der Likert-Skalierung denken, die in Abschnitt 7.2 besprochen wurde. Daneben gibt es ein breites Feld von Ansätzen zur Datenkonstruktion, in denen sich Gesichtspunkte der Skalierung mit spekulativen Modellbildungen vermischen. In diesem Kapitel besprechen wir zunächst einige Grundgedanken, dann werden exemplarisch drei Skalierungsmodelle dargestellt.

15.1 Skalierung und Modellbildung

1. Fassen wir die bisherigen Überlegungen zu Skalierungsverfahren kurz zusammen. Ausgangspunkt sind empirische Variablen

$$(X_1, \ldots, X_m) : \Omega \longrightarrow \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$$

Ein Skalierungsverfahren, so wie dies Wort in Abschnitt 4.3 erläutert worden ist, soll ein Verfahren liefern und begründen, mit dem aus diesen Variablen eine neue Variable

$$X^* : \Omega \longrightarrow \tilde{\mathcal{X}}^*$$

konstruiert werden kann; und zwar mit der Intention, daß mit Hilfe der neuen Variablen X^* sinnvolle Aussagen über die Mitglieder von Ω gemacht werden können. Aus diesem Anspruch ergeben sich die Fragen, die an ein Skalierungsverfahren zu stellen sind. Liefert es Variablenkonstrukte, die sinnvolle Aussagen ermöglichen? Liefert es eine Begründung für den neuen Merkmalsraum $\tilde{\mathcal{X}}^*$ und dafür, wie den Mitgliedern von Ω durch X^* konstruierte Eigenschaften zugewiesen werden? Kann man zeigen, daß das Verfahren einen quantitativen Merkmalsraum liefert, so daß $\tilde{\mathcal{X}}^*$ durch einen sinnvollen Größenbegriff expliziert werden kann?

2. In der Literatur, die sich mit Skalierungsverfahren beschäftigt, werden diese Fragestellungen oft mit einer wesentlich anderen Fragestellung vermischt: Wie kann *erklärt* werden, daß die Variablen $X_1, \ldots, X_m$ bei den Mitgliedern von Ω ihre jeweils realisierten Werte annehmen? Wenn im folgenden von „Modellbildung" gesprochen wird, beziehen wir uns auf diese Fragestellung. In diesem Kontext soll ein Modell Einsichten in einen Prozeß erlauben, durch den Werte statistischer Variablen realisiert werden. Es handelt sich ersichtlich um eine Fragestellung, die mit den Intentionen eines Skalierungsverfahrens zunächst überhaupt nichts zu tun hat. Denn bei Skalierungsverfahren (im engeren Sinne dieses Wortes) geht es um die

Konstruktion von Variablen, also um Fragen der Begriffsbildung; bei der Modellbildung geht es dagegen darum, wie empirische Sachverhalte erklärt werden können. Es ist auch nicht ohne weiteres klar, wie Modelle dieser Art überhaupt konstruiert werden können. Aber mit dieser Frage werden wir uns in diesem Text nicht näher beschäftigen. Der Hinweis auf Modellbildung ist hier nur deshalb erforderlich, weil Skalierungsverfahren oft in der Form von Modellen dargestellt und vorgeschlagen werden.

3. Wir sprechen dann von *Skalierungsmodellen.* Ihr Grundgedanke besteht darin, ein Skalierungsverfahren mit einer spezifischen Rhetorik auszustatten. Das Skalierungsverfahren liefert zunächst ein Verfahren zur Konstruktion eines oder mehrerer Indizes aus einer gegebenen Menge von Indikatoren. Die zusätzliche Rhetorik besteht darin, die Indizes als „latente Variablen" anzusprechen, durch die die jeweils beobachteten Werte der Indikatoren „erklärt" werden können. Prototypische Beispiele liefern Verfahren der Indexkonstruktion, die im Umkreis der Faktorenanalyse entwickelt worden sind. Die Vorstellung, daß man Skalierungsverfahren als Skalierungsmodelle konzipieren könne, durchzieht jedoch die gesamte Skalierungsliteratur; und als Leitgedanke tritt dann stets die merkwürdige Vorstellung auf, daß Indizes auch zur Erklärung von Werten der Indikatoren, aus denen sie konstruiert worden sind, verwendet werden können. C. H. Coombs, ein Protagonist solcher Skalierungsmodelle, hat das so formuliert:

„A measurement or scaling model is actually a theory about behavior, admittedly on a miniature level, but nevertheless theory." (Coombs 1964, S. 5)

Ein Skalierungsmodell hat dementsprechend für Coombs zwei Aspekte: Es besteht aus einem „scaling criterion" und einer „scaling method". Coombs fügt zur Erläuterung hinzu:

„This distinction between a scaling criterion and a scaling method is one of relative emphasis. On the one hand the technique may be used to test the hypothesis that a unidimensional latent attribute can account for the observed behavior, in which case it is a scaling criterion. On the other hand it may be used to construct such a unidimensional scale, in which case it is a scaling method." (S. 81)

4. Es ist zwar fragwürdig, ob und ggf. wie man davon sprechen könnte, daß mit einem Skalierungsmodell überhaupt irgendetwas erklärt werden kann; denn sicherlich geht es in diesem Kontext nicht um eine Erklärung im üblichen Sinn des Wortes, was ja voraussetzen würde, daß man sich auf einen empirisch explizierbaren Prozeß beziehen kann, der die zu erklärenden Phänomene hervorgebracht haben könnte. Ein gewisses Verständnis für Skalierungsmodelle kann man aber vielleicht anhand des folgenden Zitats von L. Guttman (1950, S. 64) finden:

"*The definition of scale.* For a given population of objects, the multivariate frequency distribution of a universe of attributes will be called a *scale* if it is possible to derive from the distribution a quantitative variable with which to characterize the objects such that each attribute is a simple function of that quantitative variable. Such a quantitative variable is called a scale variable."

Dies entspricht zwar nicht der Definition des Worts 'Skala', die in Abschnitt 4.3 besprochen worden ist. Guttmans Konzept einer „scale variable" kann aber verwendet werden, um eine Intention von Skalierungsmodellen zu erläutern: Es soll nicht genügen, ausgehend von einer Menge empirischer Indikatoren $X_1, \ldots, X_m$ einen Index X^* zu konstruieren und mit inhaltlichen Überlegungen zu begründen; sondern es soll außerdem möglich sein, die individuellen Werte der empirischen Variablen aus den Werten des Index berechenbar zu machen.[1] Das heißt, es sollen Funktionen

$$g_j : \tilde{\mathcal{X}}^* \longrightarrow \tilde{\mathcal{X}}_j$$

konstruierbar sein, so daß für alle $\omega \in \Omega$ gilt: $X_j(\omega) = g_j(X^*(\omega))$. So betrachtet soll ein Skalierungsmodell nicht nur ein Verfahren liefern, mit dem ein Index X^* konstruiert werden kann; sondern es soll zugleich die Funktionen g_j liefern, mit deren Hilfe die Werte der empirischen Indikatoren aus den Werten des Index berechnet werden können.[2] Wir nennen dies im folgenden die *Reproduzierbarkeitsbedingung*.[3]

5. Es ist offensichtlich, daß diese Reproduzierbarkeitsbedingung höhere Anforderungen stellt als die Verfahren zur Konstruktion von Indizes, die wir bisher besprochen haben. Bei diesen Indizes können die Werte der empirischen Indikatoren, aus denen sie gebildet werden, im allgemeinen nicht aus den Indexwerten zurückgewonnen werden. Andererseits ist jedoch klar, daß mit Guttmans Reproduzierbarkeitsforderung ein Gesichtspunkt formuliert wird, der für die Intentionen eines Skalierungsverfahrens ganz irrelevant ist. Sie liefert insbesondere kein Kriterium für sinnvolle Indizes. Man denke z.B. an den Index für den Mietanteil am Haushaltseinkommen, der

[1] Das angeführte Zitat und auch Guttmans weitere Ausführungen in dem Aufsatz, dem das Zitat entstammt, sind nicht vollständig klar. Gelegentlich erscheint es, als ob er etwas anderes fordern möchte: daß man aus der statistischen Verteilung des Index X^* die gemeinsame Verteilung der Indikatoren berechnen könne. Unsere Interpretation entspricht jedoch dem von ihm vorgeschlagenen Skalierungsmodell, mit dem wir uns in Abschnitt 15.2 beschäftigen werden. Sie entspricht auch einem verbreiteten Verständnis seiner Intentionen. Zum Beispiel gibt es in dem gleichen Sammelband (Stouffer u.a. 1950) einen Beitrag von Suchman, der mit folgenden Worten beginnt: „Scale analysis provides an answer to the problem of how to rank people from high to low with respect to the particular area being studied. From the rank order of a person it is possible to reproduce his answers to all the questions." (Suchman 1950, S. 213)

[2] In der Testpsychologie und gelegentlich auch in der Methodenliteratur werden die Funktionen g_j „Itemcharakteristik" oder „item characteristic curve" (für das „Item" X_j) genannt.

[3] Guttman 1950, S. 77, spricht von „reproducibility" und hat auch vorgeschlagen, einen „coefficient of reproducibility" zu berechnen, der bei seinem Skalierungsmodell das Ausmaß erfassen soll, in dem die Reproduzierbarkeitsbedingung erfüllt werden kann.

in Abschnitt 6.1 besprochen wurde und bei dem die Reproduzierbarkeitsbedingung nicht erfüllt ist. Weiterhin ist auch ersichtlich, daß diese Bedingung ganz irrelevant für die Frage ist, ob mit einem Skalierungsmodell die Werte empirischer Indikatoren erklärt werden können. Wenn ein Skalierungsmodell die Reproduzierbarkeitsbedingung erfüllt, kann man zwar die Werte der empirischen Indikatoren aus den Werten des konstruierten Index berechnen. Aber dies bedeutet insoweit nur, daß der Index eine äquivalente Repräsentation des durch die empirischen Indikatoren gegebenen Informationsgehalts liefert, eine mathematisch äquivalente Reformulierung der empirischen Indikatoren. Das sollte natürlich nicht mit einer Erklärung der Werte der empirischen Indikatoren verwechselt werden.

6. Soweit wir Guttman verstehen können, hat er selbst auch nicht den Anspruch vertreten, daß mit Skalierungsmodellen Werte empirischer Indikatoren *erklärt* werden können. Seine Reproduzierbarkeitsforderung soll vielmehr dazu dienen, die „Eindimensionalität" eines Index zu prüfen:

„The need for scale analysis arises out of the fundamental problem of attitude scaling and opinion polling of how to determine the dimensions of meaning which the questions asked have for the respondents. Scale analysis affords a rigorous test for the existence of single-meaning for an area and provides a rank order of individuals for such areas as are found scalable." (Guttman 1950, S. 88)

Die Problemstellung entspricht derjenigen, die bereits in Abschnitt 7.2 im Zusammenhang mit der Likert-Skalierung besprochen wurde: Man möchte herausfinden, ob sich die Indikatoren auf eine „gleiche Dimension" beziehen. Bei der Likert-Skalierung versucht man, eine Antwort aus den Korrelationskoeffizienten der Indikatoren zu gewinnen. Guttman schlägt vor, als Kriterium seine Reproduzierbarkeitsforderung zu verwenden. Aber es ist evident, daß auch dieser Vorschlag darauf hinausläuft, inhaltliche Überlegungen, die sich auf die Bedeutung der Indikatoren beziehen müßten, durch formale Kriterien zu substituieren. Es ist deshalb nicht erforderlich, die Kritik, die in Abschnitt 7.2 an der „Item-Analyse" im Kontext der Likert-Skalierung vorgetragen wurde, hier am Beispiel von Guttmans Reproduzierbarkeitskriterium zu wiederholen.

7. Es sollte beachtet werden, daß die oben verwendete Formulierung für Guttmans Reproduzierbarkeitsbedingung schon bei einer rein formalen Betrachtung nicht hinreicht, um eine sinnvoll interpretierbare „Eindimensionalität" zu garantieren. Denn nehmen wir zum Beispiel an, daß die Indikatoren nur die Werte $0, \ldots, 9$ annehmen können, dann kann man z.B. folgenden Index definieren:

$$X^*(\omega) := \sum_{j=1}^{m} X_j(\omega)\, 10^{j-1}$$

Dieser Index erfüllt offensichtlich die Reproduzierbarkeitsbedingung, aber liefert tatsächlich nur eine Rekodierung der Indikatorenwerte. Es muß also noch eine weitergehende Idee hinzukommen, um der Vorstellung eines „eindimensionalen“ Index einen Sinn zu geben. In den folgenden Abschnitten beschäftigen wir uns exemplarisch mit Skalierungsmodellen, bei denen diese weitergehende Idee – Coombs' „scaling criterion“ – durch die Vorstellung eingeführt wird, daß sowohl die Indikatoren als auch die Mitglieder von Ω in eine lineare Ordnung gebracht werden können. Wir beginnen mit Guttmans Skalogramm-Analyse, dann besprechen wir ein Modell für Rangordnungsdaten, das von Coombs vorgeschlagen worden ist, und zum Abschluß werden am Beispiel eines einfachen Rasch-Modells Grundzüge probabilistischer Skalierungsmodelle behandelt.

15.2 Guttmans Skalogramm-Analyse

1. In diesem Abschnitt besprechen wir ein Skalierungsmodell, das von L. Guttman (1944, 1950) vorgeschlagen worden ist. In der Methodenliteratur wird von *Skalogramm-Analyse* oder *Guttman-Skalierung* gesprochen. Ausgangspunkt ist eine Menge binärer Indikatoren:

$$X_j : \Omega \longrightarrow \tilde{\mathcal{X}}_j := \{0, 1\} \qquad (\text{für } j = 1, \ldots, m)$$

In einer oft verwendeten Rhetorik repräsentieren diese Indikatoren die Ergebnisse einer Bearbeitung von Aufgaben bei einem Test. Es gibt dann m Aufgaben $A_1, \ldots, A_m$, und X_j bekommt den Wert 1, wenn die Aufgabe A_j erfolgreich gelöst worden ist, andernfalls den Wert 0. Um das Skalierungsverfahren zu erläutern, werden wir zunächst diese Rhetorik übernehmen.[4] Zur Bezugnahme auf die Individuen in der Gesamtheit Ω verwenden wir wie bisher $\omega_1, \ldots, \omega_n$ (bzw. $i = 1, \ldots, n$).

2. Die Grundidee besteht darin, für jede Aufgabe einen „Grad ihrer Schwierigkeit“ und für jede Person einen „Grad ihrer Fähigkeit“ (zur Lösung der Aufgaben) anzunehmen. Die Frage ist, ob und ggf. wie solche „Größen“ gefunden werden können. Guttman hat vorgeschlagen, die Fragestellung folgendermaßen zu konzipieren.

a) Es wird zunächst hypothetisch angenommen, daß man von Schwierigkeiten (der Aufgaben) und Fähigkeiten (der Personen) als Größen reden kann und daß man sie in einer gleichen Dimension als Zahlen, die auf der reellen Zahlenachse definiert sind, vergleichen kann.

b) Die Annahme, daß es sich bei der Schwierigkeit von Aufgaben um Größen handelt, impliziert, daß es möglich sein muß, jede (im Rahmen eines spezifischen Tests definierte) Menge von Aufgaben entsprechend ihrem Schwierigkeitsgrad zu ordnen; insbesondere also die Aufgaben $A_1, \ldots, A_m$, von denen wir ausgegangen sind und deren Schwierigkeitsgrade im folgenden mit $s_1, \ldots, s_m$ bezeichnet werden.

c) Analog impliziert die Annahme, daß es sich bei den Fähigkeiten um Größen handelt, daß man die Personen entsprechend dem Grad ihrer Fähigkeit (zur Lösung der durch den Test definierten Aufgaben) in eine lineare Ordnung bringen kann; insbesondere also auch die Personen $\omega_1, \ldots, \omega_n$, auf die wir uns oben bezogen haben. Ihre hypothetisch angenommenen Fähigkeiten bezeichnen wir im folgenden mit $x_1^*, \ldots, x_n^*$. Diese Notation orientiert sich an der Vorstellung, daß schließlich ein Index X^* konstruiert werden soll, mit dem Fähigkeiten quantifiziert werden können.

d) Schließlich wird angenommen, daß eine Person mit der Fähigkeit x^* eine Aufgabe mit dem Schwierigkeitsgrad s genau dann lösen kann, wenn gilt: $s \leq x^*$.

Wichtig ist vor allem die zuletzt genannte Annahme, die – wie im vorangegangenen Abschnitt besprochen worden ist – unter zwei Aspekten betrachtet werden kann. Einerseits als ein Erklärungsprinzip, mit dem (vorgeblich) Testergebnisse erklärt werden können; andererseits als ein Skalierungsprinzip, mit dem (möglicherweise) aus einer Menge vorgegebener Daten Schwierigkeitsgrade und Fähigkeitsgrade konstruiert werden können.

3. Während es sich bei den Annahmen (a), (b) und (c) um nicht-empirische Voraussetzungen des Verfahrens handelt, kann bei der zuletzt genannten Annahme (d) geprüft werden, ob sie mit Daten, die aus der Durchführung eines Tests gewonnen worden sind, vereinbar ist. Gehen wir von der oben eingeführten Notation aus und nehmen wir an, daß mit einem Test für jede Person ermittelt worden ist, welche der Aufgaben sie gelöst und welche sie nicht gelöst hat. Es sei

$$x_{ij} := \begin{cases} 1 & \text{wenn Person } \omega_i \text{ die Aufgabe } A_j \text{ gelöst hat} \\ 0 & \text{andernfalls} \end{cases}$$

Dann kann für jede Person eine Größe $x_{i\cdot} := \sum_{j=1}^m x_{ij}$ definiert werden, die angibt, wieviele der Aufgaben von ihr gelöst worden sind; und analog kann für jede Aufgabe eine Größe $x_{\cdot j} := \sum_{i=1}^n x_{ij}$ definiert werden, die angibt, wieviele Personen die Aufgabe gelöst haben. Man kann sich leicht klarmachen, daß die oben genannte Annahme (d) genau dann mit den Daten vereinbar ist, wenn simultan eine lineare Ordnung der Personen und eine lineare Ordnung der Aufgaben gefunden werden kann, so daß nach geeigneter Umbenennung der Indizes gilt: $x_{1\cdot} \leq x_{2\cdot} \leq \cdots \leq x_{n\cdot}$ und

[4] Guttman selbst hat sein Modell primär für die Skalierung von Meinungen und Einstellungen entwickelt. Sein Hauptgedanke, die Reproduzierbarkeitsforderung, ist jedoch rein formaler Natur, und die Skalogramm-Analyse ist infolgedessen von inhaltlichen Bezugnahmen ganz unabhängig. Wie man sehen wird, kann man sie auf beliebige Mengen binärer Indikatoren anwenden.

$x_{\cdot 1} \le x_{\cdot 2} \le \cdots \le x_{\cdot m}$. Gelingt dies, kann man den Schwierigkeitsgrad der Aufgaben durch

$$s_j := n + 1 - x_{\cdot j}$$

und den Fähigkeitsgrad der Personen durch

$$x_i^* := \max\{\, s_j \mid \text{Person } i \text{ hat Aufgabe } j \text{ gelöst} \,\}$$

definieren und sich dann davon überzeugen, daß jede Person ω_i genau die Aufgaben A_j gelöst hat, für die $s_j \le x_i^*$ gilt.

4. Ein Beispiel soll den Gedankengang verdeutlichen. Es sei angenommen, daß es 4 Aufgaben gibt, die von 5 Personen bearbeitet worden sind, und folgende Ergebnisse entstanden sind:

	1	2	3	4
ω_1	1	1	1	0
ω_2	1	1	1	1
ω_3	0	0	1	0
ω_4	0	1	1	0
ω_5	1	1	1	0

Daraus findet man: $x_{\cdot 1} = 3, x_{\cdot 2} = 4, x_{\cdot 3} = 5, x_{\cdot 4} = 1$ und folgende Schwierigkeitsgrade der Aufgaben: $s_1 = 3,\ s_2 = 2,\ s_3 = 1,\ s_4 = 5$. Weiterhin findet man für die Personen

$$x_{1\cdot} = 3,\ x_{2\cdot} = 4,\ x_{3\cdot} = 1,\ x_{4\cdot} = 2,\ x_{5\cdot} = 3$$

und für ihre Fähigkeiten

$$x_1^* = 3,\ x_2^* = 5,\ x_3^* = 1,\ x_4^* = 2,\ x_5^* = 3$$

Ordnet man schließlich die Aufgaben nach ihrer Schwierigkeit und die Personen nach ihrer Fähigkeit, erhält man folgende geordnete Datenmatrix:

	3	2	1	4
ω_3	1	0	0	0
ω_4	1	1	0	0
ω_1	1	1	1	0
ω_5	1	1	1	0
ω_2	1	1	1	1

Ersichtlich gilt: Wenn die vorgegebenen Daten so geordnet werden können, daß die Bedingung (d) erfüllbar wird, wird auch die in Abschnitt 15.1 besprochene Reproduzierbarkeitsforderung erfüllt. Für jede Person kann dann aufgrund ihrer durch das Verfahren definierten Fähigkeit berechnet werden, welche der Aufgaben sie gelöst hat.

5. Es hängt allerdings von den jeweils vorliegenden Daten ab, ob das Verfahren zum Erfolg führt, d.h. in diesem Fall, ob die Aufgaben und Personen simultan so geordnet werden können, daß die Bedingung (d) erfüllt wird. In dem eben zur Illustration verwendeten Beispiel haben wir die Daten gerade so gewählt, daß das Verfahren funktioniert; in den meisten Fällen funktioniert es nicht. Aber auch wenn es funktioniert, ist fragwürdig, welche Schlußfolgerungen gezogen werden können. Wenn man bereit ist, das Verfahren als eine sinnvolle Explikation der Begriffe 'Schwierigkeit von Aufgaben' und 'Fähigkeit von Personen' anzuerkennen, kann man es – wenn es funktioniert – dazu verwenden, den Aufgaben eines Tests Schwierigkeitsgrade und Personen, die den Test durchgeführt haben, Grade ihrer Fähigkeit zuzuweisen. Man sollte sich jedoch klarmachen, daß mit dem Verfahren keine Größenbegriffe begründet werden können. Denn nicht nur hängt es von den jeweils verwendeten Daten ab, ob das Verfahren funktioniert; noch wichtiger ist, daß die Bedeutung der Begriffe selbst von den jeweils verwendeten Daten abhängt. Würde man den Test noch einmal durchführen, sei es mit den gleichen oder mit anderen Personen, bekäme man andere Daten. Nicht nur gibt es keinerlei Garantie, daß das Verfahren dann immer noch funktioniert; auch wenn es funktioniert, könnten andere Schwierigkeitsgrade für die Aufgaben resultieren. Indizes, die mit der Guttman-Skalierung erzeugt werden, sind also typische Beispiele für die in Abschnitt 5.3 besprochenen verteilungsabhängigen Indizes.

6. Man sollte sich klarmachen, wie die Argumentation verläuft. Zunächst werden hypothetisch Größen unterstellt, und es wird eine Annahme über ihren Zusammenhang formuliert. Dann wird anhand einer Menge von Daten untersucht, ob die Annahme mit den Daten vereinbar ist. So kann jedoch nicht geprüft werden, ob die entscheidende Voraussetzung – daß überhaupt Größenbegriffe unterstellt werden können – sinnvoll ist. Tatsächlich verhält es sich umgekehrt: *Wenn* auf einen Größenbegriff Bezug genommen werden kann, kann daraus eine Ordnung gewonnen werden. Exemplarisch kann man an einen Sehtest denken, bei dem die Aufgaben darin bestehen anzugeben, ob man gewisse Schriftzeichen in unterschiedlichen Entfernungen noch erkennen kann. Dann können die Aufgaben entsprechend der jeweils verwendeten Entfernung geordnet werden, und es sollte möglich sein, für die Personen einen Grad ihrer Sehfähigkeit zu ermitteln. Es wäre offenbar unsinnig, in diesem Sehtest zugleich ein Verfahren zu sehen, mit dem geprüft werden kann, ob die Aufgaben entsprechend der Entfernung der Schriftzeichen geordnet werden können. Vielmehr wäre dies die Voraussetzung, um den Begriff einer Sehfähigkeit zu definieren. Im übrigen gibt es in diesem Fall auch keinen Grund, von der oben unter (d) genannten Annahme auszugehen; denn es ist ja durchaus vorstellbar, daß man Objekte nur in einer mittleren Entfernung gut erkennen kann, nicht jedoch, wenn sie sehr nah oder sehr weit entfernt sind. — Es sei auch erwähnt, daß bereits Guttman selbst sein Verfahren gar nicht ernsthaft im

Hinblick darauf konzipiert hat, daß mit ihm Meßbarkeitsannahmen überprüft werden können. Angesichts der Tatsache, daß das Verfahren fast nie funktioniert (es sei denn, daß man sich von vornherein auf einen Größenbegriff beziehen kann), hat er vorgeschlagen, daß man ein gewisses Ausmaß von Abweichungen von der als ideal gedachten Ordnungsstruktur durchaus akzeptieren könne, ohne die Hypothese der Skalierbarkeit verwerfen zu müssen.[5]

7. Soweit es um Fähigkeiten geht, kann man sich zumindest an einigen Beispielen – etwa einem Sehtest – verdeutlichen, was mit einer Quantifizierung gemeint sein könnte. In der Sozialforschung wird das Verfahren jedoch – wenn überhaupt – nicht für Fähigkeiten, sondern zur Skalierung von „Einstellungen“ verwendet. Schnell, Hill und Esser (1999, S. 185f) geben folgendes Beispiel an:

„Das Prinzip der Guttman-Skala läßt sich an einem einfachen Beispiel verdeutlichen. Angenommen, die unterschiedliche Stärke der Ablehnung von Atomkraftwerken sollte mit einer Guttman-Skala gemessen werden. Die drei Items könnten z.B. folgendermaßen aussehen:

1. Es sollten keine weiteren Atomkraftwerke geplant werden.
2. Es sollten keine weiteren Atomkraftwerke in Betrieb genommen werden.
3. Alle Atomkraftwerke sollten sofort abgeschaltet werden.

Wenn die Ablehnung von Atomkraftwerken mit diesen drei Items eindimensional gemessen werden könnte, so müßte die Zustimmung zum Item 2 die Zustimmung zu Item 1 bedingen. Folglich müßte die Zustimmung zu Item 3 die Zustimmung zu Item 2 und zu Item 1 bedingen; andererseits müßte ein entschiedener Befürworter der Kernenergie alle Items ablehnen.“

Aber angenommen, man könnte sich auf eine Menge von Daten beziehen, die mit diesen Fragen erzeugt worden sind; und angenommen, man könnte die Items und die Befragungspersonen mit dem Guttman-Verfahren ordnen (ggf., der Empfehlung von Guttman folgend, „näherungsweise“). Könnte man daraus die Schlußfolgerung ziehen, daß es sich bei der „Stärke der Ablehnung von Atomkraftwerken“ um eine quantifizierbare Eigenschaft von Menschen handelt?

15.3 Entfaltung von Rangordnungen

In den folgenden zwei Abschnitten besprechen wir ein Skalierungsmodell für Rangordnungsdaten, das von C. H. Coombs (1950) vorgeschlagen wurde. Es wird in der englischsprachigen Literatur als *Unfolding* bezeichnet; wir sprechen von einer *Entfaltung von Rangordnungen*.

[5] Da es dann keine eindeutig bestimmten linearen Ordnungen für die Aufgaben und Personen gibt, gibt es unterschiedliche Vorschläge, wie das Verfahren näherungsweise angewendet werden kann. Eine ausführliche Diskussion findet man bei Staufenbiel 1992.

15.3.1 Coombs' Modell für Präferenzen

1. Erinnern wir uns, wie in Abschnitt 11.1 Rangordnungsdaten durch statistische Variablen repräsentiert worden sind. Gegeben ist eine Menge von Befragungspersonen Ω und außerdem eine Menge Ω^*, deren Elemente als Alternativen für Präferenzen oder als Objekte für Bewertungen angesehen werden können. In beiden Fällen wird angenommen, daß die Befragungspersonen Rangordnungen für die Elemente aus Ω^* angeben können. Also kann man eine Variable

$$R : \Omega \longrightarrow \tilde{\mathcal{R}}$$

definieren, so daß $R(\omega)$ die Rangordnung ausdrückt, die von der Befragungsperson ω für die Elemente von Ω^* angegeben wird. Nehmen wir an, daß Ω^* aus m Elementen besteht, und indizieren wir die Befragungspersonen durch $\omega_1, \ldots, \omega_n$. Dann ist $\tilde{\mathcal{R}} \subset \mathbf{R}^m$, und man kann die individuellen Rangordnungen durch

$$R(\omega_i) = (r_{i1}, \ldots, r_{im})$$

repräsentieren; r_{ij} liefert den Rangplatz, der von der Befragungsperson ω_i für das Element $j \in \Omega^*$ vergeben wird. Es wird folgende Interpretationsmöglichkeit angenommen: Wenn $r_{ij} = r_{ij'}$, ist ω_i indifferent zwischen den Alternativen ω_j^* und $\omega_{j'}^*$; wenn $r_{ij} < r_{ij'}$, zieht ω_i die Alternative $\omega_{j'}^*$ der Alternative ω_j^* vor.

2. Das Skalierungsmodell, das im folgenden besprochen werden soll, wurde zuerst von C. H. Coombs (1950) vorgeschlagen. Es gibt ein- und mehrdimensionale Varianten. Die eindimensionale Variante wurde von Coombs so erläutert:

„The basic assumptions of the theory of preferential choice on which the unfolding technique in one dimension is based are as follows: Each individual [Mitglieder von Ω] and each stimulus [Elemente von Ω^*] may be represented by a point on a common dimension called a *J scale*, and each individual's preference ordering of the stimuli from most to least preferred corresponds to the rank order of the absolute distances of the stimulus points from the ideal point, the nearest being the most preferred. The individual's preference ordering is called an *I scale* and may be thought of as the J scale folded at the ideal point with only the rank order of the stimuli given in order of increasing distance from the ideal point [...]. The data consist of a set of I scales from a number of individuals, and the analytical problem is how to unfold these I scales to recover the J scale.

The problems the psychologist may be interested in include those of discovering whether a common latent attribute may be underlying the preferences of individuals among the alternatives, measuring such an attribute, determining the degree to which such a common attribute exists, or determining for what subpopulations such a common attribute exists.“ (Coombs 1964, S. 80)

Es ist bemerkenswert, daß Coombs hier von einer „theory of preferential choice“ spricht.[6] Die Absicht scheint darin zu bestehen, Überlegungen zu entwickeln, mit denen verständlich gemacht und erklärt werden kann, wie Menschen ihre Präferenzen bilden.[7] Aber was hat diese Absicht mit Skalierung zu tun? — Diese Frage kann man immer stellen, wenn Skalierungsmodelle mit einem Modellbildungsanspruch vertreten werden. Da indessen ein solcher Anspruch meistens nicht ernsthaft verfolgt wird, erscheint es angemessener, von einem anderen Verständnis auszugehen: daß die Modellbildung gar nicht dem Zweck dienen soll, irgendetwas zu erklären, vielmehr – in der Form von Skalierungsmodellen – dazu dienen soll, Skalierungsverfahren von empirisch ausweisbaren Ansprüchen der Begriffsbildung unabhängig zu machen und sie gegen mögliche Einwände zu immunisieren.[8]

3. Um das von Coombs vorgeschlagene Skalierungsmodell zu verstehen, kann ein Beispiel helfen, das von Crano und Brewer verwendet worden ist.

„Das Problem, das Coombs' Ansatz zu lösen versucht, ist die Bestimmung der

[6] Er fügt hinzu: „Unfolding theory is an assumption about a reality, the nature of the judgment process in preferential choice behavior.“ (S. 81)

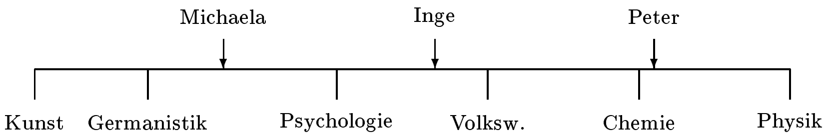
[7] Auch Borg und Staufenbiel 1993, S. 161, formulieren diesen Anspruch: „Gewisse Präferenzaussagen erscheinen merkwürdig. Nehmen wir an, eine Vp [„Versuchsperson“] hätte uns mitgeteilt, daß ihre Präferenzen für einige politische Parteien und Gruppierungen wie folgt geordnet seien: 1. Nationalisten, 2. Kommunisten, 3. CDU, 4. Grüne, 5. SPD und 6. CSU. Irgendwie erscheint es schwierig, sich auf diese Rangordnung einen Reim zu machen. Fragen wir anders herum: Welche Rangordnung würde nicht merkwürdig erscheinen? Nach welchen plausiblen Kriterien oder nach welcher Theorie sollten derartige Präferenzordnungen erzeugt werden?“ Das Wort ‘sollten’ verweist allerdings auf eine bemerkenswerte Zweideutigkeit.

[8] Dazu heißt es bei Rost 1988, S. 10f: „Wie lassen sich Relevanzkriterien an formale Theorien anlegen, die per definitionem keinen inhaltlich abgrenzbaren Realitätsbereich zum Gegenstand haben? Das hinter dieser Frage stehende Verständnis von ‘Testtheorie’ ist das einer Klasse von Modellen, eben ‘Testmodellen’, die, jeweils angewendet auf einen bestimmten Test und eine bestimmte Personenpopulation, zu einer inhaltlichen psychologischen Theorie über das Verhalten dieser Personen in dieser Testsituation werden. ‘Sinn’ einer solchen inhaltlichen Theorienbildung ist es nicht, mit ihrer ‘Verifizierung’ (dieser Begriff steht hier und im folgenden für ‘nicht gelungene Falsifizierung’ im Sinne von Popper, 1972) das psychologische Wissen über das menschliche Verhalten anzureichern. Vielmehr dient eine solche Miniaturtheorie, nachdem sie sich als gültig für einen Test und eine Personenpopulation erwiesen hat, lediglich als Prämisse, unter der diagnostische Aussagen über die getesteten Personen möglich werden. Diese Rolle von Testmodellen als Mittel zu dem Zweck, mit einem Test überhaupt arbeiten zu können, erklärt auch die oft große Trivialität der durch die Modelle formalisierten psychologischen Annahmen, die das Prädikat ‘psychologische Theorie’ als übertrieben erscheinen lassen. Die Frage nach der Relevanz formaler Testmodelle für die Weiterentwicklung der Psychologie stellt sich daher als die Frage nach dem, wofür man die Testmodelle benutzen *kann* oder wofür man sie *benötigt*.“ Skalierungsmodelle haben also vielleicht einen Sinn in der Entwicklung psychologischer Testverfahren. Es kann allerdings bezweifelt werden, daß solche Modelle für die Sozialforschung nützlich sein könnten; es sei denn – wie Methodenlehrbücher gelegentlich nahezulegen scheinen –, daß man Sozialforschung als eine Variante psychologischer Testtheorie konzipieren möchte.

Ordnung oder Struktur, die einer Anzahl von Items oder Objekten zugrunde liegt, und die verläßliche Anordnung von Respondenten an verschiedenen Punkten innerhalb dieser Struktur in einer solchen Art und Weise, daß die Rangordnung der Wahlen eines jeden Respondenten bestimmbar wird. Folgendes Beispiel dient zur Illustration: Peter, Michaela und Inge stehen kurz vor dem Abitur und müssen ein Studienfach wählen. Nehmen wir an, daß 6 verschiedene Studienplatzmöglichkeiten offenstehen und daß folgende Rangordnungen der möglichen Studiengänge von den drei Abiturienten vorgenommen werden:

Peter	Michaela	Inge
1. Chemie	1. Germanistik	1. Volkswirtschaft
2. Physik	2. Psychologie	2. Psychologie
3. Volkswirtschaft	3. Kunst	3. Chemie
4. Psychologie	4. Volkswirtschaft	4. Germanistik
5. Germanistik	5. Chemie	5. Physik
6. Kunst	6. Physik	6. Kunst

Obwohl die Wahlen der Abiturienten recht eindeutig und unterschiedlich erscheinen, ist es möglich, eine Präferenzstruktur zu bestimmen, die die Einstellungen aller drei Personen zu den möglichen Studienfächern darstellt. Dies zeigt folgende Abbildung:



Um zu entscheiden, ob die mit der Hilfe der Faltungstechnik von Coombs erstellte Skala in Abb. 10-1 [s.o.] die Präferenzmuster unserer Abiturienten angemessen darstellt, schauen wir uns als Beispiel den Punkt an, an dem sich Inge auf der Präferenzskala befindet. Gemessen in Entfernungen auf der Skala ist der nächste Studiengang für Inge Volkswirtschaft, dann Psychologie, dann Chemie usw. Eine graphische Illustration der Präferenzen von Inge wird dadurch möglich, daß man die Skala an ihrem Idealpunkt (d.h. an dem Punkt, an dem sie sich auf der Skala befindet) ‘faltet’, daher der Name ‘Faltungstechnik’. Wenn man die Entfernung vom Idealpunkt als Indikator der relativen Präferenz nimmt, ist es möglich, das Präferenzmuster dieser Abiturientin, das in der obigen Tabelle zu sehen ist, nachzuvollziehen. Eine Untersuchung der Idealpunkte der beiden anderen Abiturienten und der Entfernungen zwischen diesen Punkten und den verschiedenen Studiengängen auf der Skala zeigt, daß auch ihre Präferenzmuster in der Skala dargestellt sind. Wie Galanter (1966) feststellt: ‘Durch die Benutzung dieser (Faltungs-) Technik ist es möglich, trotz der unterschiedlichen Präferenzen verschiedener Personen ... die Möglichkeit zu untersuchen, daß die Unterschiede in Wirklichkeit eine versteckte Konsistenz der Präferenzstruktur verdecken’. Die Nützlichkeit der Faltungstechnik wird dadurch erhöht, daß die Dimension, entlang derer die Stimuli geordnet sind, identifizierbar wird. Im vorliegenden Beispiel könnte man annehmen, daß die Dimension, aufgrund derer die Präferenzen der Abiturienten variieren, durch den Grad der mathematischen

Fähigkeit, die von den verschiedenen Studienfächern vorausgesetzt wird, determiniert wird. Physik stellt unter den gegebenen Studienfächern den höchsten Anspruch an mathematische Fähigkeit, Chemie den nächsthöchsten und Kunst den geringsten.“ (Crano und Brewer 1975, S. 238f)

Die Interpretation, die Crano und Brewer geben, ist natürlich fragwürdig. Denn warum sollte man zum Beispiel annehmen, daß Michaela deshalb am liebsten Germanistik studieren möchte, *weil* ihre mathematischen Fähigkeiten gering sind? Tatsächlich sind ja mathematische Fähigkeiten in diesem Beispiel gar nicht untersucht worden; und selbst wenn dies der Fall gewesen wäre, würde es sich um eine unbegründete Spekulation handeln. Im folgenden konzentrieren wir uns jedoch darauf, das Skalierungsverfahren genauer zu verstehen.

4. Übersetzen wir das Beispiel in unsere Notation. Die Menge

$$\Omega := \{\omega_1, \omega_2, \omega_3\}$$

besteht aus den drei Abiturienten: Peter (ω_1), Michaela (ω_2) und Inge (ω_3). Die Menge

$$\Omega^* := \{1, 2, 3, 4, 5, 6\}$$

bezieht sich auf die Studienfächer: Physik (1), Chemie (2), Volkswirtschaft (3), Psychologie (4), Germanistik (5) und Kunst (6). Also kann man die Rangordnungen durch eine Rangordnungsvariable

$$R : \Omega \longrightarrow \tilde{\mathcal{R}} \subset \mathbf{R}^6$$

repräsentieren, die folgende Werte hat:

$$\begin{aligned} R(\omega_1) &= (5, 6, 4, 3, 3, 1) \\ R(\omega_2) &= (1, 2, 3, 5, 6, 4) \\ R(\omega_3) &= (2, 4, 6, 5, 3, 1) \end{aligned}$$

In der Sprechweise von Coombs ist $R(\omega_i)$ die *I-Skala* für die Person $\omega_i \in \Omega$, also einfach eine numerische Repräsentation der von ihr angegebenen Rangordnung.

5. Der Vorschlag von Coombs besteht nun darin, eine Art prototypischer Rangordnung zu konstruieren, die von ihm *J-Skala* genannt wird. Wir verweisen auf sie durch

$$\mathbf{r}^* := (r_1^*, \ldots, r_6^*)$$

Da diese Rangordnung dazu dienen soll, die individuellen Rangordnungen $R(\omega_i)$ berechenbar zu machen, muß irgendwie ein Zusammenhang zwischen $\mathbf{r}^*$ und R hergestellt werden. Coombs schlägt vor, eine neue Variable

$$S : \Omega \longrightarrow \mathbf{R}$$

zu konstruieren, die für jede Person $\omega_i \in \Omega$ einen *Idealpunkt* $S(\omega_i)$ liefert. Die beiden Konstrukte, die Rangordnung $\mathbf{r}^*$ und die Idealpunkt-Variable S, werden dann durch Coombs folgendermaßen in einen Zusammenhang gebracht. Zunächst werden Abstände konzipiert:[9]

$$d_{ij} := |\, S(\omega_i) - r_j^* \,|$$

Dann wird ein „Erklärungsprinzip“ eingeführt:[10]

$$\text{Für alle } \omega_i \in \Omega,\ j, j' \in \Omega^*\text{:}\ \ d_{ij} < d_{ij'} \implies r_{ij} < r_{ij'} \qquad (15.3.1)$$

Die Annahme ist also: die prototypische Rangordnung $\mathbf{r}^*$ und die Werte der Idealpunkt-Variablen S liefern die von den Personen geäußerten Rangordnungen.

6. Insoweit handelt es sich offensichtlich nur um eine recht merkwürdige Idee zur „Erklärung“ von Präferenzen. Sowohl die prototypische Rangordnung $\mathbf{r}^*$ als auch die Idealpunkt-Variable S sind rein spekulativ gebildete Konstrukte, für die es weder in unserer Erfahrung noch in unserer Sprache irgendwelche empirischen Anhaltspunkte gibt. Hier setzt indessen ein zweiter Gedankengang ein: daß man diesem Modell durch ein Skalierungsverfahren eine „empirische Bedeutung“ geben könne. Der Gedanke besteht darin, das Erklärungsprinzip (15.3.1) einfach umzukehren, also:

$$\text{Für alle } \omega_i \in \Omega,\ j, j' \in \Omega^*\text{:}\ \ r_{ij} < r_{ij'} \implies d_{ij} < d_{ij'} \qquad (15.3.2)$$

Denn in dieser Formulierung kann man offenbar versuchen, es als Ausgangspunkt für ein Skalierungsverfahren zu verwenden; das heißt: eine prototypische Rangordnung $\mathbf{r}^*$ und eine Idealpunkt-Variable S so zu konstruieren, daß das Zielkriterium (15.3.2) erfüllbar wird. So sind Crano und Brewer vorgegangen, um die in ihrem Schaubild dargestellte prototypische Rangordnung der Studienfächer und die Idealpunkte der drei Abiturienten zu erzeugen.

7. Es handelt sich offenbar um ein Skalierungsmodell, so wie dies Wort in Abschnitt 15.1 erläutert worden ist. Einerseits besteht es aus einem spekulativen Modell, das mit theoretischen Konstrukten und einem mit ihrer Hilfe formulierten Erklärungsprinzip operiert. Andererseits wird dieses Modell durch eine einfache Umkehrung seines Erklärungsprinzips in

[9] Da es sich um reelle Zahlen handelt, sind die Rechenoperationen sicherlich zulässig. Ob und ggf. wie den hier definierten Abständen ein Sinn gegeben werden kann, ist allerdings fragwürdig. In der psychologischen Testtheorie wird dies Problem meistens ignoriert, oder es wird die Vorstellung eines „psychologischen Raums“ gebildet, um wenigstens metaphorisch von „Abständen“ sprechen zu können.

[10] In dieser Formulierung wird angenommen, daß es – wie in unserem Beispiel – keine Indifferenzen gibt.

ein Skalierungsverfahren verwandelt, mit dem prototypische Rangordnungen und Idealpunkt-Variablen konstruiert werden können.[11] Wenn es gelingt, eine Menge von Rangordnungsdaten mit Hilfe des von Coombs vorgeschlagenen Skalierungsmodells zu repräsentieren, wird ersichtlich auch die Guttmansche Reproduzierbarkeitsforderung erfüllt, die in Abschnitt 15.1 besprochen wurde. Es ist weiterhin ersichtlich, daß diese Forderung durch Coombs' Skalierungsmodell in einer spezifischen Weise zu erfüllen versucht wird: nämlich durch eine simultane lineare Ordnung der Personen in Ω und der Alternativen in Ω^*.

8. Es ist allerdings keineswegs sicher, daß eine solche Repräsentation von Rangordnungsdaten stets gefunden werden kann. Denn angenommen, es gäbe in Ω noch einen vierten Abiturienten, dessen Rangordnung für die möglichen Studienfächer so aussieht: $R(\omega_4) = (1, 6, 2, 5, 3, 4)$. Dann wäre es nicht möglich, eine prototypische Rangordnung und vier Idealpunkte zu finden, so daß das Coombs'sche Erklärungsprinzip erfüllt ist. Es hängt also von den jeweils vorliegenden Daten ab, ob das Skalierungsverfahren anwendbar ist oder nicht. Einige Autoren, zum Beispiel Borg und Staufenbiel (1993, S. 166), sehen hierin einen besonderen Vorzug von Skalierungsmodellen: daß solche Modelle „falsifizierbar" sind. Der Gedankengang scheint etwa folgendermaßen zu verlaufen. Wenn eine Menge von Rangordnungsdaten gegeben ist, kann man zunächst hypothetisch annehmen, daß man für sie eine Darstellung entsprechend dem Coombs'schen Skalierungsmodell finden kann. Dann kann mit Hilfe der Daten festgestellt werden, ob dies tatsächlich der Fall ist. Wenn dies der Fall ist, hat man das Modell „bestätigt", andernfalls hat man es „falsifiziert". Aber der Gedankengang ist ziemlich unsinnig;[12] denn es gibt beliebig viele Beispiele, mit denen das Modell „falsifiziert" werden kann. Im Zweifelsfall braucht man sich nur eine Rangordnung auszudenken, die mit einer prototypischen Rangordnung nicht repräsentiert werden kann.

9. Der Gedankengang zeigt indessen, daß es bei Skalierungsmodellen gar nicht um Modellbildung im üblichen Verständnis des Wortes geht. Ginge es wirklich darum, ein Modell zu konstruieren, mit dem die Herausbildung von Präferenzen verständlich gemacht und erklärt werden kann, müßte man sich explizit auf Bildungsprozesse von Präferenzen beziehen. Die Frage, ob sich irgendeine gegebene Menge von Rangordnungsdaten durch eine prototypische Rangordnung und eine Menge von Idealpunkten *repräsentieren* läßt, ist dafür ganz irrelevant. Denn eine Antwort auf diese Frage sagt bestenfalls etwas über die Rangordnungsdaten, die als kontingente Daten gegeben sind. Der entscheidende Punkt ist, daß das Coombs'sche Skalierungsmodell gar nicht unabhängig von einer Bezugnahme auf jeweils kontingente Daten formuliert werden kann. Denn die Konstruktion der für die Modellformulierung verwendeten „theoretischen Konstrukte" ist bereits *durch ihre Definition* von den jeweils gegebenen Daten abhängig.[13]

[11] Ob es sich um ein plausibles Modell für die Bildung von Präferenzen handelt, ist sicherlich fragwürdig, soll hier jedoch nicht weiter diskutiert werden. Man kann sich in dieser Frage wohl der oben zitierten Ansicht von Rost anschließen, daß gar kein ernsthafter Modellbildungsanspruch vertreten wird. Unser Interesse gilt dem Skalierungsverfahren, also dem Verfahren der Datenkonstruktion. Bemerkenswert ist aber, daß in diesem Fall die Datenkonstruktion von einem fragwürdigen spekulativen Modell *abhängig* gemacht wird. Insofern kann man auch hier von Pseudo-Daten sprechen. Die Konstruktion einer prototypischen Rangordnung $\mathbf{r}^*$ und einer Idealpunkt-Variablen S liefert keine Daten, die als Hinweise auf empirische Sachverhalte explizierbar sind, sondern sie liefert Werte für theoretische Konstrukte, die ihrerseits nicht empirisch explizierbar sind.

[12] Tatsächlich wird er auch in der Literatur, die sich mit Unfolding-Modellen beschäftigt, nicht ernst genommen. Zum Beispiel sagt van Blokland-Vogelesang 1993, S. 243: „For a variety of reasons, however, individuals generally do not all produce rankings consistent with one underlying qualitative or quantitative J scale, and a variety of methods have been developed for the unfolding of imperfect data."

15.3.2 Berechnung und Illustration

1. Die im vorangegangenen Abschnitt skizzierte Kritik richtete sich gegen den theoretischen Anspruch des Coombs'schen Modells für Präferenzen. Davon unabhängig kann überlegt werden, ob man die Entfaltungstechnik als eine Methode zur Analyse von Rangordnungsdaten verwenden kann. In diesem Abschnitt besprechen wir ein Berechnungsverfahren und illustrieren es mit Daten für Parteipräferenzen. Ausgangspunkt sind n Rangordnungen $\mathbf{r}_i = (r_{i1}, \ldots, r_{im})$ (für $i = 1, \ldots, n$), die sich auf m Alternativen beziehen. Zu überlegen ist, wie man eine prototypische Rangordnung $\mathbf{r}^* = (r_1^*, \ldots, r_m^*)$ finden kann, die die vorgegebenen Rangordnungen im Sinne der Entfaltungsidee am besten repräsentiert. Wir nehmen an, daß die vorgegebenen Rangordnungen keine Bindungen aufweisen. Zur Illustration des Rechenverfahrens verwenden wir folgende vier Rangordnungen aus dem vorangegangenen Abschnitt:

$$\mathbf{R} = \begin{pmatrix} 5 & 6 & 4 & 3 & 2 & 1 \\ 1 & 2 & 3 & 5 & 6 & 4 \\ 2 & 4 & 6 & 5 & 3 & 1 \\ 1 & 6 & 2 & 5 & 3 & 4 \end{pmatrix}$$

Die ersten drei Rangordnungen sind von Peter, Michaela und Inge; die vierte haben wir hinzugefügt, weil sie nicht in das Schema von Crano und Brewer paßt.

2. Um eine im Sinn der Entfaltungstechnik optimale Rangordnung $\mathbf{r}^*$ zu finden, ist es erforderlich, alle möglichen Rangordnungen zu untersuchen. Das geschieht am übersichtlichsten dadurch, daß man der Reihe nach alle

[13] Deshalb ist es irreführend, wenn E. H. Witte 1989, S. 368, schreibt: „Die Methode des 'unfolding' setzt voraus, daß es eine allgemeingültige Rangordnung der Items gibt, die je nach individueller Einstellung verändert wird."

Permutationen der m Alternativen betrachtet. Permutationen fassen wir als Funktionen

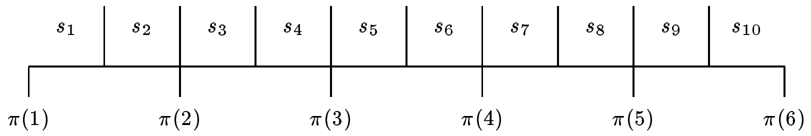
$$\pi : \{1, \ldots, m\} \longrightarrow \{1, \ldots, m\}$$

auf, wobei die Menge $\{1, \ldots, m\}$ die Indizes für die m Alternativen enthält. Z.B. könnte eine der Permutationen folgendermaßen aussehen.

j	1	2	3	4	5	6
$\pi(j)$	6	5	4	3	2	1

In diesem Beispiel wird durch die Permutation π die Reihenfolge invertiert. Jede eineindeutige Abbildung liefert natürlich eine unterschiedliche Permutation. Insgesamt gibt es $m!$ mögliche Permutationen, in unserem Beispiel $6! = 720$.

3. Jeder Permutation entsprechen $2\,(m-1)$ Möglichkeiten zur Plazierung von Idealpunkten. Folgende Abbildung illustriert diese Möglichkeiten anhand unseres Beispiels.



Jeder möglichen Plazierung eines Idealpunktes entspricht dann eine bestimmte Rangordnung. In diesem Beispiel:

s	$\pi(1)$	$\pi(2)$	$\pi(3)$	$\pi(4)$	$\pi(5)$	$\pi(6)$
s_1	6	5	4	3	2	1
s_2	5	6	4	3	2	1
s_3	4	6	5	3	2	1
s_4	3	5	6	4	2	1
s_5	2	4	6	5	3	1
s_6	1	3	5	6	4	2
s_7	1	2	4	6	5	3
s_8	1	2	3	5	6	4
s_9	1	2	3	4	6	5
s_{10}	1	2	3	4	5	6

Die Menge der Rangordnungen, die mit Hilfe einer bestimmten Permutation π erzeugt werden können, bezeichnen wir im folgenden mit $R(\pi)$.

4. Am besten wäre es natürlich, wenn man eine Permutation π finden könnte, so daß alle vorgegebenen Rangordnungen $\mathbf{r}_i$ Elemente von $R(\pi)$ sind. Das ist jedoch im allgemeinen nicht möglich, denn jede Permutation

erlaubt nur $2\,(m-1)$ unterschiedliche Rangordnungen. Z.B. gibt es keine Permutation, mit deren Hilfe die eingangs genannten vier Rangordnungen darstellbar sind. Wenn man also nicht bei einer solchen negativen Aussage stehen bleiben möchte, muß man sich irgendein Kriterium ausdenken, um eine optimale Permutation zu definieren. Wir verwenden Kendalls Rangkorrelation (vgl. Abschnitt 11.3). Ist eine Permutation π gegeben, kann man für jede vorgegebene Rangordnung $\mathbf{r}_i$ diejenige (nicht unbedingt eindeutig bestimmte) Rangordnung aus $R(\pi)$ finden, deren Rangkorrelation mit $\mathbf{r}_i$ einen maximalen Wert annimmt. Diesen maximalen Wert bezeichnen wir mit $\tau_i(\pi)$. Offenbar liefert

$$\tau(\pi) := \sum_{i=1}^{n} \tau_i(\pi)$$

ein mögliches Kriterium: Suche nach derjenigen Permutation π, die den Wert von $\tau(\pi)$ maximiert. In unserem Beispiel findet man dann die Permutation $\pi = (1,2,3,4,5,6)$ und folgende Rangkorrelationen:

$\mathbf{r}_i$	bestes Element aus $R(\pi)$	$\tau_i(\pi)$
$\mathbf{r}_1$	$(5,6,4,3,2,1)$	1.0
$\mathbf{r}_2$	$(1,2,3,5,6,4)$	1.0
$\mathbf{r}_3$	$(2,4,6,5,3,1)$	1.0
$\mathbf{r}_4$	$(1,3,5,6,4,2)$	0.2

Man erkennt, daß zwar die ersten drei Rangordnungen mit Hilfe der Permutation π erzeugt werden können, nicht jedoch die vierte, für die die beste Rangordnung aus $R(\pi)$ nur eine Rangkorrelation von 0.2 liefert.

5. Zur Illustration verwenden wir Daten über Parteipräferenzen, die von H. Norpoth (1979) analysiert worden sind. Die Daten wurden 1972 im Rahmen der Wahlstudie 1972 erhoben.[14] Es wurden Präferenzen für fünf Parteien erfragt, und zwar: (1) DKP, (2) SPD, (3) FDP, (4) CDU/CSU, (5) NPD. Zunächst wurden Paarvergleichsdaten erhoben, dann die inkonsistenten Angaben ausgeschlossen. Indifferenzen wurden nicht zugelassen. Tabelle 15.3.1 zeigt die Rangordnungen von 1785 Personen, so wie sie von Norpoth (1979, S. 355) veröffentlicht wurden. Z.B. haben 315 Personen die Rangordnung

$$(1,5,4,3,2)$$

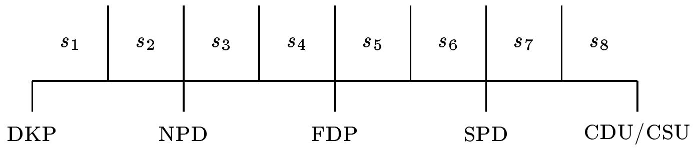
angegeben, d.h. die SPD wird am meisten präferiert, dann folgen FDP, CDU/CSU, NPD und DKP.

[14]Die Ausgangsdaten sind im Zentralarchiv für empirische Sozialforschung (Köln) unter der Nummer ZA-Nr. 635-637 verfügbar. Wir verwenden hier jedoch die von Norpoth aufbereiteten Daten in der von ihm 1979, S. 355, veröffentlichten Form.

Tabelle 15.3.1 Parteipräferenzen von 1785 Personen im Jahr 1972 (Norpoth 1979, S. 355).

1	2	3	4	5	h	1	2	3	4	5	h
1	5	4	3	2	315	3	4	2	5	1	2
2	5	4	3	1	210	1	4	2	5	3	6
1	5	4	2	3	13	4	3	2	5	1	1
3	5	4	2	1	45	1	3	2	5	4	7
2	5	4	1	3	25	3	1	2	5	4	2
3	5	4	1	2	9	3	4	1	5	2	2
1	5	3	4	2	201	3	2	1	5	4	1
2	5	3	4	1	94	1	4	5	3	2	35
4	5	3	2	1	13	2	4	5	3	1	33
2	5	3	1	4	1	3	4	5	2	1	3
4	5	3	1	2	1	3	4	5	1	2	2
1	5	2	4	3	7	1	3	5	4	2	31
1	5	2	3	4	1	2	3	5	4	1	10
3	5	1	4	2	1	1	3	5	2	4	2
2	5	1	3	4	1	1	2	5	4	3	1
3	5	1	2	4	1	4	1	2	3	5	1
2	3	4	5	1	40	1	3	4	2	5	1
1	3	4	5	2	178	1	4	3	2	5	2
1	2	4	5	3	17	1	2	3	4	5	1
3	1	4	5	2	2	1	4	2	3	5	1
2	4	3	5	1	97	1	3	2	4	5	1
1	4	3	5	2	351	5	1	3	2	4	1
1	2	3	5	4	9	5	4	3	1	2	1
2	1	3	5	4	2	5	4	3	2	1	4

6. Man erkennt sofort, daß eine perfekte Entfaltung dieser Rangordnungsdaten nicht möglich ist. Denn es gibt 48 unterschiedliche Rangordnungen; bei jeder Permutation können jedoch nur $2\,(m-1) = 8$ Rangordnungen perfekt dargestellt werden. Wendet man nun das zuvor besprochene Rechenverfahren an, findet man folgende J-Skala: $\mathbf{r}^* = (1, 5, 3, 2, 4)$. Als Bild dargestellt:



Folgte man der Coombs'schen Idee, sollte man z.B. sagen können, daß die Präferenzen der in Tabelle 15.3.1 zuerst angeführten 315 Personen dadurch zustande gekommen sind, daß sie sich den Idealpunkt s_6 ausgewählt haben. Aber was könnte man über die dann folgenden 210 Personen sagen, deren Rangordnung sich durch dieses Bild nicht perfekt entfalten läßt? Das Beispiel ist auch deshalb interessant, weil es zeigt, daß eine mit der Entfaltungstechnik gewonnene J-Skala keineswegs immer zu einer sinnvoll interpretierbaren Ordnung der Alternativen führt.[15]

15.4 Probabilistische Skalierung

Etwa seit den 50er Jahren haben einige Autoren begonnen, auch probabilistische Skalierungsmodelle zu entwickeln. In diesem Abschnitt besprechen wir einige Grundgedanken und behandeln dann exemplarisch ein Modell, das von Georg Rasch vorgeschlagen worden ist.

15.4.1 Der konzeptionelle Ansatz

1. Um zu verstehen, wie probabilistische Skalierungsmodelle konstruiert werden, beginnen wir mit einer kurzen Erinnerung an einfache Logit-Modelle, wie sie in Einführungen in die Statistik erklärt werden. Es gibt eine Variable Y mit zwei Ausprägungen, die durch 0 und 1 repräsentiert werden, und – der Einfachheit halber – nur eine weitere Variable X mit dem Merkmalsraum $\tilde{\mathcal{X}}$. Der Ansatz für das Logit-Modell ist dann

$$\mathrm{P}(Y = 1 \,|\, X = \tilde{x}) \approx \frac{\exp(\alpha + \tilde{x}\beta)}{1 + \exp(\alpha + \tilde{x}\beta)}$$

wobei $\tilde{x}$ Werte im Merkmalsraum $\tilde{\mathcal{X}}$ annehmen kann und α und β Modellparameter sind. Sind die Modellparameter gegeben, erhält man durch das Modell für jeden Wert $\tilde{x} \in \tilde{\mathcal{X}}$ einen Wert

$$\mathrm{P}(Y = 1 \,|\, X = \tilde{x})$$

der sich interpretieren läßt als die Wahrscheinlichkeit, daß Y den Wert 1 annimmt, wenn X den Wert $\tilde{x}$ annimmt.[16] Insoweit handelt es sich um ein einfaches statistisches Modell, mit dem man versuchen kann, Einsichten in den Zusammenhang der Variablen X und Y zu gewinnen.

2. Mit der probabilistischen Skalierung wird zwar ein anderes Ziel verfolgt: Es sollen Indizes (latente Variablen) konstruiert werden, durch deren Werte Individuen charakterisiert werden können. Ein Grundgedanke kann jedoch schon anhand des einfachen Logit-Modells illustriert werden. Man kann nämlich eine neue Variable

$$X^* := \alpha + X\beta$$

[15] Es sei angemerkt, daß Norpoth (1979) eine konzeptionell unterschiedliche Variante der Entfaltungstechnik verwendet und zu anderen Ergebnissen gelangt.

[16] Wir folgen hier dem in der Skalierungsliteratur üblichen Reden von „Wahrscheinlichkeiten“. Da stets eine endliche Gesamtheit Ω zugrunde liegt, kann man stattdessen auch von relativen Häufigkeiten sprechen.

definieren und sie zur Charakterisierung der Mitglieder der Gesamtheit Ω, für die die Variablen X und Y definiert sind, verwenden. Allerdings muß man aufpassen, wie das praktisch erreicht werden kann, denn α und β sind ja zunächst ganz unbestimmt. Erst durch eine Berechnung des Modells mit einer Menge von Daten (d.h. jeweils bestimmten Werten für die Variablen X und Y) erhält man bestimmte Parameterwerte $\hat{\alpha}$ und $\hat{\beta}$, aus denen dann eine Variable

$$\hat{X}^* := \hat{\alpha} + X\hat{\beta}$$

tatsächlich gebildet werden kann. Dann ist jedoch offensichtlich, daß die Werte dieser Variablen nicht allein durch die Werte von X bestimmt werden, sondern ebenfalls durch die Werte von Y *und* durch das verwendete Modell *und* durch die Daten, die zur Modellberechnung verwendet worden sind. $\hat{X}^*$ ist also insbesondere eine verteilungsabhängige Variable, so wie dieser Ausdruck in Abschnitt 5.3 eingeführt worden ist.[17]

3. Das Beispiel illustriert, wie latente Variablen mit einem probabilistischen Verfahren konstruiert werden können. Bei der probabilistischen Skalierung gibt es jedoch eine wichtige Besonderheit. Es gibt dann nicht zwei oder mehr empirische Variablen, deren statistische Zusammenhänge untersucht werden sollen, sondern nur eine Menge von Indikatoren $X_1, \ldots, X_m$, aus denen mit Hilfe eines probabilistischen Modells ein Index X^* konstruiert werden soll. An dieser Stelle wird nun die eingangs erwähnte Vorstellung wichtig: Der Index X^* wird so behandelt, *als ob* mit ihm statistisch erklärt werden könnte, mit welchen Wahrscheinlichkeiten die Indikatoren $X_1, \ldots, X_m$ ihre möglichen Werte annehmen.[18] Den Ausgangspunkt bildet

[17] Dies gilt ganz unabhängig davon, ob für die latenten Variablen selbst irgendwelche spezifischen Verteilungsannahmen gemacht werden. Auch dafür finden sich Beispiele in der Skalierungsliteratur. Es ist indessen kaum möglich, darüber sinnvoll zu reden; bestenfalls kann man sagen: Da latenten Variablen keine empirischen Sachverhalte entsprechen, kann man mit ihnen machen, was man will. So heißt es zum Beispiel bei Schnell, Hill und Esser 1999, S. 142: „Die weitaus meisten theoretischen Begriffe in den Sozialwissenschaften beziehen sich auf kontinuierliche 'latente' (also nicht direkt meßbare) Variablen, wie z.B. 'Intelligenz', 'Prestige', 'Dogmatismus', 'Anomie', 'Legitimität' usw. Theorien, in denen solche theoretischen Begriffe als latente Variablen enthalten sind, machen meist jedoch keine Aussage darüber, daß die entsprechenden Eigenschaften nur immer als ganzzahliges Vielfaches einer Einheit vorkommen können. Somit können latente Variablen in der Regel als kontinuierlich aufgefaßt werden, d.h., zwischen zwei Punkten auf der entsprechenden Dimension liegen immer unendlich viele Punkte. Weiterhin können latente Variablen meist als näherungsweise normalverteilt angesehen werden, d.h. (stark vereinfacht), die meisten Objekte liegen nahe am Mittelwert, und die Zahl der Objekte, die vom Mittelwert abweichen, wird um so geringer, je stärker sie vom Mittelwert abweichen." Der Spekulation werden also keine Grenzen gesetzt.

[18] Zu diesem „als ob" heißt es bei Lord und Novick 1968, S. 537: „Though much of psychological theory is based on a trait orientation, it is not necessary that these traits exist in any physical or physiological sense. It is sufficient for most purposes of psychology that a person behaves as if he were in possession of a certain amount of each of a number of relevant traits and that he behaves as if his value on these traits substantially determined his behavior."

also ein Wahrscheinlichkeitsausdruck

$$\mathrm{P}(X_1 = \tilde{x}_1, \ldots, X_m = \tilde{x}_m \,|\, X^* = x^*)$$

der sich in scheinbarer Analogie zu $\mathrm{P}(Y = \tilde{y} \,|\, X = \tilde{x})$ so interpretieren läßt: Es ist die bedingte Wahrscheinlichkeit dafür, daß die Indikatoren die Werte $\tilde{x}_1, \ldots, \tilde{x}_m$ annehmen, wenn der Index den Wert x^* annimmt. Aber die Analogie hat einen Haken. Denn der Index X^* steht zunächst noch gar nicht zur Verfügung, um als Bedingung in einem Wahrscheinlichkeitsausdruck verwendet zu werden; er kann vielmehr erst als Ergebnis der Modellkonstruktion (und anschließender Modellberechnung) gebildet werden. So zeigt sich auch an dieser Stelle die Zirkularität der Vorstellung, daß Indikatoren durch einen aus ihnen gebildeten Index erklärt werden könnten.

15.4.2 Einfache Rasch-Modelle

1. Diese Zirkularität bezieht sich auf die Begriffsbildung, schließt es jedoch nicht aus, probabilistische Skalierungsmodelle zu konstruieren. Wie man das machen kann, besprechen wir in diesem Abschnitt anhand eines Skalierungsmodells, das Ende der 50er Jahre von Georg Rasch entwickelt worden ist.[19] In der Methodenliteratur wird es deshalb oft als *Rasch-Modell* bezeichnet.

2. Das Modell geht davon aus, daß die Indikatoren $X_1, \ldots, X_m$ binäre Variablen sind, die jeweils nur zwei mögliche Werte annehmen können, 0 und 1.[20] Die Aufgabe besteht darin, aus diesen binären Indikatoren mit Hilfe eines probabilistischen Modells einen Index

$$X^* := g(X_1, \ldots, X_m)$$

zu konstruieren. Eine für den Modellansatz grundlegende Annahme besteht darin: daß unter der Bedingung, daß ein Wert für den Index X^*

[19] Einige historische Bemerkungen zum Kontext finden sich in der Einleitung von Andersen 1973; außerdem im Vorwort von B. D. Wright zu Rasch 1980.

[20] Dieses einfache Modell kann also nicht verwendet werden, um Indizes zu konstruieren, die auf Rating-Skalen mit mehr als zwei Wahlmöglichkeiten beruhen. Allerdings ist es leicht möglich, Rasch-Modelle so zu erweitern, daß sie bei beliebigen Merkmalsräumen für die Indikatoren verwendet werden können. Zu ihrer Erweiterung für Indikatoren, die auf Rating-Skalen beruhen, vgl. man z.B. Andrich 1978; weitere Entwicklungen werden von Rost und Georg (1991) und Rost (1996) diskutiert. Um die Grundgedanken der probabilistischen Skalierung zu verdeutlichen und zu kritisieren, genügt jedoch schon das einfache Rasch-Modell.

gegeben ist, die einzelnen Indikatoren ihre Werte unabhängig voneinander annehmen. Man spricht bei dieser Annahme von *konditionaler Unabhängigkeit* der Indikatoren, wobei in diesem Fall ein aus ihnen zu bildender Index als Bedingung dienen soll. Es folgt dann:

$$\mathrm{P}(X_1 = \tilde{x}_1, \ldots, X_m = \tilde{x}_m \,|\, X^* = x^*) = \prod_{j=1}^{m} \mathrm{P}(X_j = \tilde{x}_j \,|\, X^* = x^*)$$

und das Problem reduziert sich darauf, geeignete Modelle für die einzelnen Wahrscheinlichkeiten $\mathrm{P}(X_j = \tilde{x}_j \,|\, X^* = x^*)$ zu finden.[21] Die Annahme konditionaler Unabhängigkeit ist allerdings nicht nur grundlegend, sondern auch problematisch.[22] Sie widerspricht nicht nur – wie man noch sehen wird – den meisten Anwendungen des Modells, sie kann auch nicht geprüft werden. Denn um sie zu prüfen, müßte man die Werte der latenten Variablen X^*, die als Bedingung der konditionalen Unabhängigkeit dienen soll, kennen; aber andererseits erhält man ihre Werte erst durch das Modell.[23]

3. Eine zweite wichtige Annahme bezieht sich auf den Zusammenhang zwischen der latenten Variablen (dem Index) X^* und den Wahrscheinlichkeiten $\mathrm{P}(X_j = \tilde{x}_j \,|\, X^* = x^*)$. Im einfachen Rasch-Modell wird der gleiche Ansatz verwendet wie bei einfachen Logit-Modellen:

$$\mathrm{P}(X_j = 1 \,|\, X^* = x^*) \approx \frac{\exp(x^* - \alpha_j)}{1 + \exp(x^* - \alpha_j)} \tag{15.4.1}$$

Die besondere Form der Parametrisierung verdankt sich dem intendierten Verwendungszweck des Modells. Man kann das am besten verstehen, wenn

[21]Wir verwenden die allgemeine Schreibweise $X_j = \tilde{x}_j$. Zwar kann jeder Indikator nur die Werte 0 oder 1 annehmen; aber bei m Indikatoren können natürlich viele verschiedene Kombinationen von Nullen und Einsen auftreten.

[22]Lord und Novick 1968, S. 538, haben diese Annahme so erläutert: „At this point, however, it will be convenient to describe and then make explicit the concept of a single underlying trait, and to define and characterize the conditions under which this concept can be properly applied. We might give one possible interpretation of this term by saying that an individual's performance depends on a single underlying trait if, given his value on that trait, nothing further can be learned from him that can contribute to the explanation of his performance. The proposition is that the latent trait is the only important factor and, once a person's value on the trait is determined, the behaviour is random, in the sense of statistical independence (Anderson, 1959)."

[23]Probleme dieser Art werden in der Literatur selten diskutiert; stattdessen wird schlagwortartig auf „Theorie" verwiesen. Z.B. schreiben Michel und Conrad 1982, S. 27, in einem Abschnitt über die „Grundlagen probabilistischer Testmodelle": „Die Entwicklung eines Tests, mit Hilfe dessen die individuellen Positionen auf dem latenten Merkmalskontinuum begründet festgelegt werden sollen, setzt eine *psychologische Theorie* voraus, die sich auf ein allgemeines Gesetz über die Relation zwischen manifestem Verhalten im Test und der in Frage stehenden latenten Dimension bezieht und den Zufallsprozeß beschreibt, der die Vertrauensintervalle der zu schätzenden unbekannten Parameter verursacht."

man an den ursprünglich von Rasch intendierten Verwendungszusammenhang denkt: daß das Modell einer Analyse von Fähigkeitstests dienen soll. Die Situation entspricht dann derjenigen, die bereits in Abschnitt 15.2 zur Erläuterung der Guttman-Skalierung zugrunde gelegt wurde. Man bezieht sich auf einen Test, bei dem es Aufgaben $A_1, \ldots, A_m$ gibt, deren Bearbeitung durch die Indikatoren $X_1, \ldots, X_m$ erfaßt wird. Das Modell soll dazu dienen, für jede Aufgabe einen „Grad ihrer Schwierigkeit" und für jede Person, die an dem Test teilnimmt, einen „Grad ihrer Fähigkeit" (zur Lösung der Testaufgaben) zu erzeugen. Dem entsprechen dann die Modellparameter α_j und x^*. — Ausgehend von dieser Interpretation kann man Gleichung (15.4.1) als eine Funktion interpretieren, die zeigt, wie die Wahrscheinlichkeit einer erfolgreichen Aufgabenlösung sowohl vom Schwierigkeitsgrad der Aufgabe (α_j) als auch von dem Fähigkeitsgrad der die Aufgabe bearbeitenden Person (x^*) abhängt. Man erkennt, daß es sich um probabilistische Varianten der in Abschnitt 15.1 besprochenen „Itemcharakteristiken" g_j handelt. Bei der in (15.4.1) gewählten Formulierung steigt die Wahrscheinlichkeit einer erfolgreichen Aufgabenlösung mit steigenden Fähigkeiten und sinkt mit steigendem Schwierigkeitsgrad der Aufgaben.[24]

4. Fassen wir den Modellansatz zusammen. Für jeden einzelnen Indikator X_j hat man eine Itemcharakteristik entsprechend der Gleichung (15.4.1). Infolge der Annahme konditionaler Unabhängigkeit findet man daraus für die Gesamtheit der Indikatoren den Modellansatz

$$\mathrm{P}(X_1 = \tilde{x}_1, \ldots, X_m = \tilde{x}_m \,|\, X^* = x^*) \approx \prod_{j=1}^{m} \frac{\exp(x^* - \alpha_j)^{\tilde{x}_j}}{1 + \exp(x^* - \alpha_j)}$$

Es bleibt die Frage, wie man mit diesem Modellansatz rechnen und tatsächlich eine latente Variable X^* konstruieren kann.

5. Man benötigt dafür zunächst Daten, also Werte für die Indikatoren $X_1, \ldots, X_m$. Wir nehmen an, daß solche Werte für n Personen zur Verfügung stehen; x_{ij} sei der Wert des Indikators X_j bei der Person i, wobei es sich um Werte in $\{0, 1\}$ handelt. Wenn man jedoch mit diesen Daten das

[24]Es sei erwähnt, daß das einfache Rasch-Modell, das wir in diesem Abschnitt besprechen, ein Spezialfall einer größeren Klasse von Modellen ist, die in der Literatur als „latent trait"-Modelle oder auch als Modelle der „Item Response Theory" bezeichnet werden. Hambleton, Swaminathan und Rogers 1991, S. 7, geben folgende Erläuterungen: „Item response theory (IRT) rests on two basic postulates: (a) The performance of an examinee on a test item can be predicted (or explained) by a set of factors called traits, latent traits, or abilities; and (b) the relationship between examinees' item performance and the set of traits underlying item performance can be described by a monotonically increasing function called an *item characteristic function* or *item characteristic curve* (ICC). This function specifies that as the level of the trait increases, the probability of a correct response to an item increases. [...] Many possible item response models exist, differing in the mathematical form of the item characteristic function and/or the number of parameters specified in the model. All IRT models contain one or more parameters describing the item and one or more parameters describing the examinee."

Modell berechnen möchte, tritt ein Problem auf: Wie soll man sich zu den möglichen Werten der latenten Variablen X^* verhalten? Denn diese Werte sind ja zunächst nicht bekannt, sondern sollen erst mit Hilfe des Modells konstruiert werden. Das Verfahren besteht darin, den Personen zunächst hypothetische Werte für den Index X^* zu geben; der Person i wird der hypothetische Indexwert x_i^* gegeben. Das Modell hat dann die Parameter

$$\alpha_1, \ldots, \alpha_m \quad \text{und} \quad x_1^*, \ldots, x_n^*$$

Man wird jedoch sehen, daß – abhängig von den jeweils verwendeten Daten – nicht für jede Person ein individueller Wert der latenten Variablen berechnet werden kann.

6. Das zur Modellberechnung verwendete Verfahren ist die Maximum-Likelihood-Methode. Die Grundidee kann folgendermaßen erläutert werden. Man betrachtet zunächst eine Person i mit den Indikatorwerten

$$(x_{i1}, \ldots, x_{im})$$

und dem hypothetischen Indexwert x_i^*. Eingesetzt in die Modellgleichung erhält man

$$\mathrm{P}(X_1 = x_{i1}, \ldots, X_m = x_{im} \mid X^* = x_i^*) \approx \prod_{j=1}^{m} \frac{\exp(x_i^* - \alpha_j)^{x_{ij}}}{1 + \exp(x_i^* - \alpha_j)}$$

Dies ist die durch das Modell angenommene bedingte Wahrscheinlichkeit, daß sich bei der Person i die Indikatorwerte $(x_{i1}, \ldots, x_{im})$ zeigen, wenn bei ihr die latente Variable X^* den (unbekannten) Wert x_i^* hat. Durch den Modellansatz ist dies außerdem eine Funktion der Modellparameter $\alpha_1, \ldots, \alpha_m$. Diese Funktion, also

$$\mathcal{L}_i(\alpha_1, \ldots, \alpha_m, x_i^*) := \prod_{j=1}^{m} \frac{\exp(x_i^* - \alpha_j)^{x_{ij}}}{1 + \exp(x_i^* - \alpha_j)}$$

wird als Likelihood für die Daten der Person i bezeichnet. Die Likelihood der Daten für alle Personen ergibt sich dann – vergleichbar mit der Konjunktion von Wahrscheinlichkeiten – aus dem Produkt aller individuellen Beiträge, also

$$\mathcal{L}(\alpha_1, \ldots, \alpha_m, x_1^*, \ldots, x_n^*) := \prod_{i=1}^{n} \mathcal{L}_i(\alpha_1, \ldots, \alpha_m, x_i^*) = \prod_{i=1}^{n} \prod_{j=1}^{m} \frac{\exp(x_i^* - \alpha_j)^{x_{ij}}}{1 + \exp(x_i^* - \alpha_j)}$$

Das Rechenverfahren besteht nun darin, Werte für die Modellparameter aus einer Maximierung dieser Likelihoodfunktion zu gewinnen. Wir bezeichnen diese Werte, wenn sie die Likelihoodfunktion maximieren, mit

$$\hat{\alpha}_1, \ldots, \hat{\alpha}_m \quad \text{und} \quad \hat{x}_1^*, \ldots, \hat{x}_n^*$$

7. Die verschiedenen Rechentechniken, die man zur Maximierung der Likelihoodfunktion verwenden kann, brauchen hier nicht besprochen zu werden. Die Überlegung muß jedoch noch einen Schritt weiter verfolgt werden, damit man versteht, welche Parameter überhaupt berechnet werden können. Zunächst ist klar, daß man anstelle der Likelihoodfunktion ihren Logarithmus betrachten kann, da sich für die Lösung der Maximierungsaufgabe daraus kein Unterschied ergibt. In unserem Fall sieht die sog. Log-Likelihoodfunktion folgendermaßen aus:

$$\ell(\alpha_1, \ldots, \alpha_m, x_1^*, \ldots, x_n^*) := \log(\mathcal{L}(\alpha_1, \ldots, \alpha_m, x_1^*, \ldots, x_n^*)) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}(x_i^* - \alpha_j) - \log(1 + \exp(x_i^* - \alpha_j))$$

Eine notwendige (nicht hinreichende) Bedingung dafür, daß Parameterwerte ein Maximum dieser Funktion liefern, besteht darin, daß an der Stelle des Maximums alle partiellen Ableitungen der Funktion Null werden. Diese partiellen Ableitungen können leicht berechnet werden:

$$\frac{\partial \ell}{\partial \alpha_k} = -\Big(\sum_{i=1}^{n} x_{ik} - \frac{\exp(x_i^* - \alpha_k)}{1 + \exp(x_i^* - \alpha_k)} \Big)$$

$$\frac{\partial \ell}{\partial x_k^*} = \sum_{j=1}^{m} x_{kj} - \frac{\exp(x_k^* - \alpha_j)}{1 + \exp(x_k^* - \alpha_j)}$$

Es müssen also folgende Gleichungen für die zu berechnenden Werte der Itemparameter $\hat{\alpha}_j$ erfüllt sein:

$$\sum_{i=1}^{n} x_{ij} = \sum_{i=1}^{n} \frac{\exp(\hat{x}_i^* - \hat{\alpha}_j)}{1 + \exp(\hat{x}_i^* - \hat{\alpha}_j)} \quad (\text{für } j = 1, \ldots, m)$$

und folgende Gleichungen für die Werte der Personenparameter $\hat{x}_i^*$:

$$\sum_{j=1}^{m} x_{ij} = \sum_{j=1}^{m} \frac{\exp(\hat{x}_i^* - \hat{\alpha}_j)}{1 + \exp(\hat{x}_i^* - \hat{\alpha}_j)} \quad (\text{für } i = 1, \ldots, n)$$

8. Uns interessieren hier die Gleichungen für die Personenparameter, denn sie zeigen eine bemerkenswerte Implikation des Modellansatzes: Wenn man zunächst Werte für die Itemparameter berechnet hat, kann man daraus Werte für die Personenparameter berechnen; und zwar hängen diese dann nur noch von den Größen $x_{i\cdot} := \sum_{j=1}^{m} x_{ij}$ ab. Diese Größen erfassen jedoch nur, bei *wievielen* der m Items eine Person eine 1 erzielt hat, und sind ganz unabhängig davon, bei *welchen* Aufgaben eine 1 und bei welchen eine 0 herausgekommen ist. Man kann das auch so formulieren: Das Modell impliziert, daß man die Personen entsprechend den möglichen Größen von

Tabelle 15.4.1 Daten für das Klausurbeispiel.

Muster	A_1	A_2	A_3	A_4	A_5	Anzahl	Score-Gruppe
1	0	1	0	0	0	1	1
2	0	1	0	1	1	1	3
3	0	1	1	0	1	1	3
4	1	0	0	0	0	1	1
5	1	0	1	0	0	1	2
6	1	0	1	0	1	3	3
7	1	1	0	0	0	1	2
8	1	1	0	0	1	8	3
9	1	1	0	1	0	4	3
10	1	1	0	1	1	5	4
11	1	1	1	0	0	2	3
12	1	1	1	0	1	1	4
13	1	1	1	1	0	3	4
14	1	1	1	1	1	14	5

$x_{i\cdot}$ in $m+1$ Gruppen – gelegentlich „Score-Gruppen" genannt – einteilen kann. Für jede Gruppe kann nur ein einziger Personenparameter berechnet werden. Wenn man also das Rasch-Modell für Skalierungszwecke verwendet, erhalten alle Personen, die sich in der gleichen Score-Gruppe befinden, auch den gleichen Wert der latenten Variablen.

9. Zur Illustration verwenden wir die Klausurdaten aus Abschnitt 6.1. Da das einfache Rasch-Modell binäre Indikatoren voraussetzt, nehmen wir eine Dichotomisierung vor. Das geschieht in zwei Schritten. In einem ersten Schritt werden die in jeder Aufgabe erzielten Ergebnisse durch eine Note von 1 bis 5 bewertet.[25] In einem zweiten Schritt werden binäre Indikatoren definiert:

$$x_{ij} := \begin{cases} 1 & \text{wenn Person } i \text{ in Aufgabe } j \text{ die Note 1 oder} \\ & \text{die Note 2 erzielt hat} \\ 0 & \text{andernfalls} \end{cases}$$

Insgesamt kann es bei den resultierenden Indikatoren $(x_{i1}, \ldots, x_{i5})$ 32 unterschiedliche Muster geben, davon treten 14 in unserem Datensatz auf (vgl. Tabelle 15.4.1). Natürlich kann mit dem Test zwischen *diesen* Mustern nicht unterschieden werden. Das Rasch-Modell bezieht sich jedoch

[25] Es wird folgende Kodierung verwendet. Die tatsächlich erzielten Punktzahlen werden in den Aufgaben A_1, A_2 und A_4 mit dem Faktor 3, in Aufgabe A_3 mit dem Faktor 6 und in Aufgabe A_5 mit dem Faktor 2 multpliziert, so daß in jeder Aufgabe maximal 30 Punkte erzielt werden können. Dann wird festgelegt: Note 1 (25 – 30 Punkte), Note 2 (19 – 24 Punkte), Note 3 (13 – 18 Punkte), Note 4 (7 – 12 Punkte), Note 5 (0 – 6 Punkte).

nicht auf Muster dieser Art, sondern auf Score-Gruppen, die nur durch die Anzahl der Aufgaben (Indikatoren) bestimmt werden. In unserem Beispiel gibt es 5 verschiedene Score-Gruppen, und das Modell kann nur zwischen diesen Gruppen unterscheiden. Alle Personen, die der gleichen Score-Gruppe angehören, erhalten den gleichen Wert der latenten Variablen, in diesem Beispiel den gleichen Wert des mit dem Modell konstruierten Fähigkeitsindex.

10. Hierin zeigt sich auch ein bemerkenswerter Unterschied zwischen dem Rasch-Modell und Guttmans Skalierungsmodell. In gewisser Weise kann man das Rasch-Modell als eine probabilistische Version von Guttmans Modell auffassen. Aber durch die probabilistische Modellformulierung geht der entscheidende Gedanke verloren: die Reproduzierbarkeitsbedingung. Das zeigt sich am deutlichsten darin, daß mit dem Rasch-Modell nicht mehr zwischen Konstellationen von Indikatorenwerten unterschieden werden kann, die der gleichen Score-Gruppe angehören. Das Rasch-Modell abstrahiert also davon, *welche* Aufgaben bei einem Test gelöst werden.[26]

11. Diese Eigenschaft des Rasch-Modells hängt mit einer anderen Besonderheit zusammen, die von Rasch „spezifische Objektivität" genannt worden ist. Darin wird oft ein besonderer Vorzug des Rasch-Modells gesehen. So heißt es z.B. bei Schnell, Hill und Esser (1999, S. 194):

„Das Rasch-Modell bietet gegenüber anderen Skalierungsverfahren eine Reihe von Vorteilen. Das Rasch-Modell gestattet eine klare Trennung zwischen den Eigenschaften der Personen (in Form der Personenparameter) und den Eigenschaften der Items (in Form der Itemparameter). Daraus resultiert ein großer Vorteil des Rasch-Modells: Messungen mit einer Rasch-Skala sind ‘spezifisch objektiv’. Spezifisch objektive Messungen erlauben den Vergleich von Items unabhängig von Personen und den Vergleich von Personen unabhängig von Items."

Versuchen wir zu verstehen, was gemeint ist. Die „spezifische Objektivität" des Rasch-Modells beruht zunächst darauf, daß nur zwischen Score-Gruppen unterschieden werden kann. Der Beitrag jeder Person zur Likelihood des Modells kann *deshalb* auf ihre Mitgliedschaft in einer Score-Gruppe konditioniert werden; und die spezifische Itemcharakteristik, die für das Modell angenommen wird, ermöglicht es schließlich, für die Mo-

[26] Daß davon abstrahiert wird, *welche* Aussagen Befragungs- oder Testpersonen machen, ist indessen nicht nur beim Rasch-Modell, sondern bei den meisten psychologischen „Meßverfahren" der Fall; zum Beispiel auch bei der Likert-Skalierung. Dazu bemerkte Loevinger 1948, S. 507: „Psychological tests, whether measuring abilities, attitudes, or personality traits, commonly consist of a series of items each scored plus or minus. Plus may mean to do correctly, to agree with, to exhibit the characteristic, and so on. [...] In contemporary psychometric practice, it is the rule rather than the exception that two people having the same score on a test will have scored plus on different items. When each score stands for a variety of patterns of pluses, it is difficult to conceive of these scores as measuring anything. Such scores are crude empirical devices known to have some predictive efficiency, but they cannot be called measurements in any strict sense."

dellberechnung die Abhängigkeit von der latenten Variablen X^* zu eliminieren. Man sieht das folgendermaßen. Sei G eine Variable, die für jede Person angibt, welcher Score-Gruppe sie angehört; die möglichen Werte sind $g = 0, 1, \ldots, m$. Sei weiterhin

$$D_g := \{ d = (d_1, \ldots, d_m) \,|\, d_j \in \{0, 1\}, \Sigma_j d_j = g \}$$

die Menge aller möglichen Muster in der Score-Gruppe g. Dann kann der Beitrag einer Person i zur konditionalen Likelihood folgendermaßen geschrieben werden:

$$\begin{aligned}
\ell_i^c &= \mathrm{P}(X_1 = x_{i1}, \ldots, X_m = x_{im} \,|\, X^* = x_i^*, G = x_{i\cdot}) \\
&= \frac{\mathrm{P}(X_1 = x_{i1}, \ldots, X_m = x_{im}, G = x_{i\cdot} \,|\, X^* = x_i^*)}{\mathrm{P}(G = x_{i\cdot} \,|\, X^* = x_i^*)} \\
&= \frac{\mathrm{P}(X_1 = x_{i1}, \ldots, X_m = x_{im} \,|\, X^* = x_i^*)}{\sum_{d \in D_{x_{i\cdot}}} \mathrm{P}(X_1 = d_1, \ldots, X_m = d_m \,|\, X^* = x_i^*)}
\end{aligned}$$

Für das Rasch-Modell findet man dann die spezielle Form

$$\begin{aligned}
\ell_i^c &= \frac{\prod_{j=1}^m \exp(x_i^* - \alpha_j)^{x_{ij}} / (1 + \exp(x_i^* - \alpha_j))}{\sum_{d \in D_{x_{i\cdot}}} \prod_{j=1}^m \exp(x_i^* - \alpha_j)^{d_j} / (1 + \exp(x_i^* - \alpha_j))} \\
&= \frac{\prod_{j=1}^m \exp(x_i^* - \alpha_j)^{x_{ij}}}{\sum_{d \in D_{x_{i\cdot}}} \prod_{j=1}^m \exp(x_i^* - \alpha_j)^{d_j}} \\
&= \frac{\exp(x_i^*)^{x_{i\cdot}} \prod_{j=1}^m \exp(-\alpha_j)^{x_{ij}}}{\sum_{d \in D_{x_{i\cdot}}} \exp(x_i^*)^{x_{i\cdot}} \prod_{j=1}^m \exp(-\alpha_j)^{d_j}} \\
&= \frac{\prod_{j=1}^m \exp(-\alpha_j)^{x_{ij}}}{\sum_{d \in D_{x_{i\cdot}}} \prod_{j=1}^m \exp(-\alpha_j)^{d_j}}
\end{aligned}$$

Man sieht, daß in dieser konditionalen Likelihood die Abhängigkeit von den Personenparametern x_i^* verschwunden ist. Das wiederum bedeutet, daß man die Itemparameter $\alpha_1, \ldots, \alpha_m$ berechnen kann, ohne auf die Personenparameter Bezug zu nehmen. Hat man jedoch Werte für die Itemparameter berechnet, können daraus, wie oben schon gezeigt worden ist, direkt auch Werte für die Personenparameter gewonnen werden. „Spezifische Objektivität" des Rasch-Modells bedeutet also, daß sich die Berechnung der Item- und Personenparameter separieren läßt.

12. „Spezifische Objektivität" erscheint vielen Autoren deshalb als eine wünschenswerte Eigenschaft, weil dadurch der Anschein erweckt werden kann, daß Fähigkeiten oder beliebige „latent traits" analog zu physikalischen Eigenschaften als „meßbar" aufgefaßt werden können. Z.B. sagen J. Rost und W. Georg (1991, S. 54):

„Die Skalierung mittels eines Fragebogens stellt – wie auch viele Meßvorgänge in der Physik – einen Meßvorgang dar, an dem zwei unterschiedliche Mengen von Meßobjekten beteiligt sind (Personen und Items) und bei dem sich die beobachteten Daten stets nur aus der Interaktion von je einem Element jeder Menge ergeben (die Reaktion einer Person auf ein Item). Die Idee spezifisch objektiver Messungen besagt nun, daß die Meßwerte der Elemente der einen Menge unabhängig davon sein müssen, welche Elemente der anderen Menge gerade für den Meßvorgang ausgewählt wurden. Das bedeutet, daß die Skalenwerte der Personen unabhängig davon sein müssen, ob eher leichte oder schwere Items zu einem Konstrukt im Test enthalten sind, und daß die Skalenwerte der Items unabhängig davon sein müssen, ob Personen mit eher niedriger oder eher hoher Eigenschaftsausprägung in der Stichprobe enthalten sind. Die Forderung nach spezifisch objektiven Messungen ist eigentlich eine pure Selbstverständlichkeit, nicht nur in der Wissenschaft, sondern auch im Alltagsleben (denn wer würde schon akzeptieren, daß ein Päckchen auf der Waage im Postamt mehr wiegt als zu Hause auf der Küchenwaage?)."

Der Vergleich zwischen Messungen physikalischer Eigenschaften und der Erfassung von Fähigkeiten mit Hilfe von Tests ist jedoch irreführend. Ein Päckchen kann keinen wesentlichen Einfluß darauf nehmen, wie sein Gewicht gemessen wird; und die heute institutionalisierten Meßverfahren für Gewichte erfassen im wesentlichen auch nur das Gewicht des jeweiligen Objekts. Bei der Erfassung von Fähigkeiten mittels Tests verhält es sich jedoch anders. Wie die durch den Test jeweils vorgegebenen Aufgaben gelöst werden – und ggf. auch, welche der Aufgaben überhaupt zur Bearbeitung ausgesucht werden –, hängt nicht allein von einer den Aufgaben zurechenbaren Schwierigkeit und einer den Personen zurechenbaren Fähigkeit zur Lösung der Aufgaben ab, sondern die Bearbeitung der Aufgaben erfolgt in einer jeweils konkreten Situation von einer jeweils konkreten Person und ist somit – wenn man sich überhaupt solcher Redeweisen bedienen will – ein Interaktionsprozeß einer Vielzahl von Situations- und Persönlichkeitsmerkmalen. Die Frage ist nicht, ob hiervon bei einem Test abstrahiert werden *darf*; jedenfalls bei der Durchführung standardisierter Tests kann auf die Begleitumstände ihrer Anwendung im allgemeinen nicht viel Rücksicht genommen werden, und das kann vielleicht pragmatisch gerechtfertigt werden. Hier geht es jedoch nicht um die Frage, ob und unter welchen Umständen Tests pragmatisch gerechtfertigt werden können, sondern um die Vorstellung, daß es sich um Meßverfahren handelt, mit denen – in Analogie zum Gewicht von Päckchen – eine begrifflich isolierbare dispositionale Eigenschaft von Menschen ermittelt werden kann. Schon in Abschnitt 2.2 wurde darauf hingewiesen, daß dispositionale Eigenschaften nur durch Bezugnahmen auf einen Kontext definiert werden können. Das Rasch-Modell nimmt jedoch auf den Kontext ausdrücklich keine Rücksicht. Im Gegenteil, die Idee „spezifischer Objektivität" besagt gerade, daß der Kontext (und das umfaßt auch die Art, den Schwierigkeitsgrad, die Präsentation und die Reihenfolge der jeweils gestellten Aufgaben) für die *Definition* von Fähigkeiten *keine Rolle spielen soll*. Die Frage, ob und ggf. wie Fähigkeiten sinnvoll bestimmt und erfaßt werden können, wird gar nicht thematisiert, sondern es wird einfach angenommen, daß man Menschen Fähigkeiten auf die gleiche Weise zurechnen kann, wie man z.B. Päckchen Gewichte zurechnen kann.

13. Tatsächlich stimmt auch die oben zitierte Aussage von Rost und Georg nicht, „daß die Skalenwerte der Personen unabhängig davon sein müssen, ob eher leichte oder schwere Items zu einem Konstrukt im Test enthalten sind". Oder genauer gesagt: Dies ist beim Rasch-Modell nicht der Fall (insofern ist auch die oben zitierte Aussage von Schnell, Hill und Esser falsch). Denn angenommen, man hätte Daten aus einem Test, bei dem es m Aufgaben gibt, und würde diese Daten verwenden, um mit einem Rasch-Modell Fähigkeitswerte der Testteilnehmer zu berechnen. Was würde geschehen, wenn man dann die Ergebnisse für eine oder mehrere Aufgaben aus dem Datensatz entfernt und das Modell noch einmal berechnet? Im allgemeinen würden andere Werte für die Fähigkeiten der Testteilnehmer resultieren.

14. Hier zeigt sich noch einmal ein bemerkenswerter Unterschied zwischen der Guttman-Skalierung und dem Rasch-Modell. *Wenn* die aus einem Test gewonnenen Daten der Guttman'schen Skalierbarkeitsbedingung genügen, könnte man tatsächlich die Daten für eine oder mehrere Aufgaben ignorieren, ohne daß sich an den Rangordnungen der Testteilnehmer etwas ändert (abgesehen davon, daß die Rangordnung gröber wird und ggf. Bindungen entstehen). Der Grund dafür liegt darin, daß die Guttman'sche Skalierbarkeitsbedingung impliziert, daß jedes unterschiedliche Muster in den Aufgabenlösungen genau einer Score-Gruppe entspricht. Diese Entsprechung geht jedoch beim probabilistischen Rasch-Modell verloren, wie oben anhand der Klausurdaten gezeigt worden ist. Infolgedessen kann aber auch die Annahme einer unabhängig von der Auswahl von Aufgaben existierenden Rangordnung der Personen nicht aufrechterhalten werden. Dies wäre nur dann der Fall, wenn die zur Berechnung eines Rasch-Modells verwendeten Daten der Guttman'schen Skalierbarkeitsbedingung genügen würden. Aber wenn dies der Fall wäre, gäbe es überhaupt keinen Grund, ein probabilistisches Modell zu verwenden. Es ist gerade erfunden worden, um Fähigkeitswerte mit Daten konstruieren zu können, die sich nicht durch eine Skalogramm-Analyse ordnen lassen.[27]

[27]Wenn bei einem Datensatz Guttmans Skalierbarkeitsbedingung erfüllt ist, gibt es nicht nur keinen Grund, ein Rasch-Modell zu berechnen, eine Modellberechnung kann dann gar nicht durchgeführt werden. Man erkennt das an der oben dargestellten konditionalen Likelihood-Funktion. Sie nimmt dann für alle möglichen Parameterwerte den gleichen Wert an und kann infolgedessen nicht sinnvoll maximiert werden.

15. Mit einer analogen Überlegung kann die zweite Hälfte der Aussage von Rost und Georg kritisiert werden, „daß die Skalenwerte der Items unabhängig davon sein müssen, ob Personen mit eher niedriger oder eher hoher Eigenschaftsausprägung in der Stichprobe enthalten sind." Denn ein Rasch-Modell würde im allgemeinen durchaus unterschiedliche Parameterwerte liefern, wenn man aus einem gegebenen Datensatz unterschiedliche Teilgruppen für die Modellberechnung auswählt. Offenbar liegt eine Verwechslung vor. Die Annahme der konditionalen Unabhängigkeit und die spezifische Parametrisierung des Rasch-Modells erlauben es, Schwierigkeitswerte für die Aufgaben unabhängig von *den* Fähigkeitswerten zu berechnen, die das Modell schließlich für die Personen erzeugen soll. Aber dies bedeutet keineswegs, daß die Parameterwerte, die das Modell liefert, von den Daten unabhängig sind, die von den Testteilnehmern erhoben werden. Verändert man die Daten, erhält man im allgemeinen auch andere Ergebnisse. Die Ergebnisse, die das Modell für die Schwierigkeitsgrade der Aufgaben liefert, sind also keineswegs von den Personen unabhängig, die – als Testteilnehmer – die Daten liefern; und sie sind infolgedessen auch nicht von deren Fähigkeiten unabhängig.[28] Das Rasch-Modell liefert also auch keine verteilungsunabhängigen Indizes (für Personen und Aufgaben), wie gelegentlich behauptet wird.

[28]Die Verwechslung, die wir hier kritisieren, kommt gut in folgender Äußerung von Mokken 1971, S. 110, zum Ausdruk: „We may therefore conclude that in this model [das Rasch-Modell], due to the separability of subject and item parameters, subjects θ [unser x^*] may be estimated 'item-selection free', i.e. independently of the items sampled or selected for their measurement, and item δ's [unsere α_j] may be determined 'population free', i.e. independently of the particular subjects sampled or selected for the measurement."

Anhang A

Rechnen mit Matrizen

Die Darstellung einiger der in diesem Text behandelten Verfahren wird durch die Verwendung der Matrizenrechnung wesentlich erleichtert. In diesem Anhang erklären wir die erforderlichen Grundbegriffe.

A.1 Grundbegriffe

1. Unter einer *Matrix* verstehen wir eine rechteckig angeordnete Menge von Zahlen, wobei es sich im allgemeinen um beliebige reelle Zahlen handeln kann. Wenn es n Zeilen und m Spalten gibt, spricht man von einer (n, m)-*Matrix* oder von einer Matrix der *Ordnung* (n, m). Zum Beispiel ist

$$\mathbf{A} := \begin{pmatrix} 1 & 3 & 9 \\ 4 & 1 & 7 \\ 2 & 7 & 8 \\ 2 & 1 & 5 \end{pmatrix}$$

eine $(4, 3)$-Matrix. Die Schreibweise $\mathbf{A} := \ldots$ sagt, daß man dieser Matrix den Namen $\mathbf{A}$ geben möchte. Beliebige andere Namen sind möglich; in diesem Text verwenden wir (meistens) fettgedruckte Buchstaben.[1]

2. Um allgemein auf eine (n, m)-Matrix Bezug nehmen zu können, verwenden wir die Schreibweise

$$\mathbf{A} = (a_{ij})$$

wobei sich der Index i auf die Zeilen, der Index j auf die Spalten der Matrix bezieht. In ausführlicher Schreibweise hat man also

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$

Das Matrixelement a_{ij} steht in der i-ten Zeile und j-ten Spalte. Wenn $\mathbf{A}$ eine Matrix ist, verwenden wir gelegentlich auch die Schreibweise $(\mathbf{A})_{ij}$, um auf das Matrixelement Bezug zu nehmen, das sich in der i-ten Zeile und j-ten Spalte von $\mathbf{A}$ befindet.

[1] Eine Ausnahme ist das Symbol $\mathbf{R}$, mit dem stets auf die Menge der reellen Zahlen verwiesen wird.

3. *Vektoren* sind Matrizen, die nur aus einer Zeile oder einer Spalte bestehen. $(1, m)$-Matrizen werden als *Zeilenvektoren*, $(n, 1)$-Matrizen werden als *Spaltenvektoren* bezeichnet. Im Kontext der Matrizenrechnung werden $(1, 1)$-Matrizen als *Skalare* bezeichnet und mit dem entsprechenden Matrixelement identifiziert. Zur Bezeichnung von Vektoren verwenden wir (meistens) fettgedruckte Kleinbuchstaben. Um uns auf die Spaltenvektoren einer (n, m)-Matrix $\mathbf{A}$ zu beziehen, schreiben wir oft

$$\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_m)$$

Gemeint ist, daß $\mathbf{a}_1$ die erste Spalte von $\mathbf{A}$ ist, $\mathbf{a}_2$ die zweite Spalte usw.

4. Es seien $\mathbf{A}$ und $\mathbf{B}$ zwei Matrizen der gleichen Ordnung. Wir schreiben

$$\mathbf{A} = \mathbf{B} \quad \text{wenn für alle } i, j \text{ gilt: } a_{ij} = b_{ij}$$

Analog werden die Relationen $<$ und $\leq$ sowie $>$ und $\geq$ definiert, z.B.:

$$\mathbf{A} \leq \mathbf{B} \quad \text{wenn für alle } i, j \text{ gilt: } a_{ij} \leq b_{ij}$$

Man beachte, daß diese Relationen nur eine partielle Ordnung der Matrizen liefern.

5. Um mit Matrizen rechnen zu können, werden einige Operationen vereinbart. Die einfachste Operation ist die Multiplikation einer Matrix mit einem Skalar. Sei $\mathbf{A} = (a_{ij})$ eine Matrix und c ein Skalar; die Definition lautet: $c\,\mathbf{A} := (c\,a_{ij})$. Es wird also jedes Element von $\mathbf{A}$ mit der Zahl c multipliziert. Wenn $c = 0$ ist, entsteht eine *Nullmatrix*, bei der alle Elemente gleich Null sind. Wir werden für Nullmatrizen das Symbol $\mathbf{0}$ verwenden und annehmen, daß sich ihre Ordnung aus dem Kontext ergibt.

6. Matrizen, die die gleiche Ordnung haben, können addiert und subtrahiert werden. Seien $\mathbf{A}$ und $\mathbf{B}$ zwei (n, m)-Matrizen, dann ist

$$\mathbf{A} + \mathbf{B} := (a_{ij} + b_{ij}) \quad \text{und} \quad \mathbf{A} - \mathbf{B} := (a_{ij} - b_{ij})$$

Man wendet also die Operationen einzeln auf alle sich entsprechenden Matrixelemente an.

7. Eine Matrix $\mathbf{A}$ kann mit einer Matrix $\mathbf{B}$ multipliziert werden, wenn die Anzahl der Spalten von $\mathbf{A}$ und die Anzahl der Zeilen von $\mathbf{B}$ gleich sind. Sei also $\mathbf{A} = (a_{ij})$ eine (n, m)-Matrix und $\mathbf{B} = (b_{ij})$ eine (m, p)-Matrix. Dann ist das Produkt $\mathbf{C} := \mathbf{AB}$ eine (n, p)-Matrix mit den Elementen

$$c_{ij} := \sum_{k=1}^{m} a_{ik} b_{kj}$$

Zum Beispiel kann die zu Beginn definierte $(4, 3)$-Matrix $\mathbf{A}$ mit einem

$(3,1)$-Spaltenvektor $\mathbf{b} := \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ multipliziert werden. Man erhält als Ergebnis den $(4,1)$-Spaltenvektor $\mathbf{Ab} = \begin{pmatrix} 13 \\ 12 \\ 17 \\ 8 \end{pmatrix}$, der die Zeilensummen der Matrix $\mathbf{A}$ enthält. Man beachte, daß die Matrixmultiplikation nicht kommutativ ist, d.h. im allgemeinen ist $\mathbf{AB} \neq \mathbf{BA}$.

8. Unter Verwendung der bisher gegebenen Definitionen kann man sich davon überzeugen, daß folgende Regeln für das Rechnen mit Matrizen gültig sind. Dabei sind $\mathbf{A}$, $\mathbf{B}$ und $\mathbf{C}$ jeweils Matrizen passender Ordnung, und a und b sind Skalare.

$$\begin{array}{rclrcl}
\mathbf{A}+\mathbf{B} &=& \mathbf{B}+\mathbf{A} & a\,(\mathbf{B}+\mathbf{C}) &=& a\,\mathbf{B}+a\,\mathbf{C} \\
\mathbf{A}+(\mathbf{B}+\mathbf{C}) &=& (\mathbf{A}+\mathbf{B})+\mathbf{C} & a\,(\mathbf{B}-\mathbf{C}) &=& a\,\mathbf{B}-a\,\mathbf{C} \\
\mathbf{A}\,(\mathbf{BC}) &=& (\mathbf{AB})\,\mathbf{C} & (a+b)\,\mathbf{C} &=& a\,\mathbf{C}+b\,\mathbf{C} \\
\mathbf{A}\,(\mathbf{B}+\mathbf{C}) &=& \mathbf{AB}+\mathbf{AC} & (a-b)\,\mathbf{C} &=& a\,\mathbf{C}-b\,\mathbf{C} \\
(\mathbf{B}+\mathbf{C})\,\mathbf{A} &=& \mathbf{BA}+\mathbf{CA} & a\,(b\,\mathbf{C}) &=& (a\,b)\,\mathbf{C} \\
\mathbf{A}\,(\mathbf{B}-\mathbf{C}) &=& \mathbf{AB}-\mathbf{AC} & a\,(\mathbf{BC}) &=& (a\,\mathbf{B})\,\mathbf{C} \\
(\mathbf{B}-\mathbf{C})\,\mathbf{A} &=& \mathbf{BA}-\mathbf{CA} & a\,\mathbf{B} &=& \mathbf{B}\,a
\end{array}$$

9. Eine oft verwendete Operation ist die *Transposition*, die darin besteht, daß die Zeilen und Spalten einer Matrix vertauscht werden. Sei $\mathbf{A} = (a_{ij})$ eine (n,m)-Matrix. Durch Transposition entsteht eine (m,n)-Matrix

$$\mathbf{A}' := (a_{ji})$$

Wir verwenden stets ein einfaches Hochkomma, um die Transponierte einer Matrix zu bezeichnen. Es gilt folgende wichtige Regel:

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$$

Die Transposition ist auch für das Rechnen mit Vektoren wichtig. Seien $\mathbf{x}$ und $\mathbf{y}$ zwei $(n,1)$-Spaltenvektoren. Dann ist $\mathbf{x}'\mathbf{y}$ ein Skalar, jedoch $\mathbf{xy}'$ eine (n,n)-Matrix.

10. Wenn $\mathbf{x} = (x_1, \ldots, x_n)'$ ein Spaltenvektor ist, kann man insbesondere das Produkt $\mathbf{x}'\mathbf{x} = \sum_{i=1}^{n} x_i^2$ bilden. Das Ergebnis ist eine stets nichtnegative Zahl, deren Wurzel

$$\| \mathbf{x} \| := \sqrt{\mathbf{x}'\mathbf{x}}$$

als *euklidische Länge* des Vektors $\mathbf{x}$ bezeichnet wird.

11. Eine Matrix heißt *quadratisch*, wenn sie gleich viele Zeilen und Spalten hat. Ist $\mathbf{A} = (a_{ij})$ eine quadratische (n,n)-Matrix, ist ihre *Spur* (*trace*) durch

$$\mathrm{tr}(\mathbf{A}) := \sum_{i=1}^{n} a_{ii}$$

definiert. Es gilt folgender Zusammenhang:

$$\mathrm{tr}(\mathbf{A}'\mathbf{A}) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2 = \mathrm{tr}(\mathbf{AA}')$$

Diese Gleichung gilt auch für beliebige (m,n)-Matrizen.

12. Eine quadratische Matrix $\mathbf{A}$ heißt *symmetrisch*, wenn sie mit ihrer Transponierten identisch ist, wenn also $\mathbf{A} = \mathbf{A}'$ gilt. Man überlege sich, daß $\mathbf{A}'\mathbf{A}$ stets eine symmetrische Matrix ist.

13. Eine quadratische (n,n)-Matrix $\mathbf{A} = (a_{ij})$ wird eine *Diagonalmatrix* genannt, wenn alle Elemente außerhalb der Hauptdiagonalen Null sind. Zur Definition von Diagonalmatrizen verwenden wir oft die Schreibweise:

$$\mathrm{diag}(a_1, \ldots, a_n) := \begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_n \end{pmatrix}$$

Offenbar ist jede Diagonalmatrix auch symmetrisch.

14. Ein wichtiges Beispiel für eine Diagonalmatrix ist die Einheitsmatrix. Eine (n,n)-*Einheitsmatrix* ist durch

$$\mathbf{I}_n := \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

definiert; in der Hauptdiagonalen stehen Einsen und außerhalb stehen Nullen. Wenn $\mathbf{A}$ eine beliebige (n,m)-Matrix ist, gilt offenbar: $\mathbf{AI}_m = \mathbf{A}$ und $\mathbf{I}_n\mathbf{A} = \mathbf{A}$. Sei nun außerdem $\mathbf{B}$ eine (m,p)-Matrix. Dann gilt auch: $\mathbf{AI}_m\mathbf{B} = \mathbf{AB}$.

A.2 Datenmatrizen

1. Die Matrizenrechnung dient uns in diesem Text hauptsächlich dafür, um auf einfache Weise mit Datenmatrizen umgehen zu können. In diesem Abschnitt definieren wir das Konzept einer Datenmatrix und erinnern kurz an einige Begriffe der Statistik. — Ausgangspunkt für die Konzeption von

Datenmatrizen sind statistische Variablen. Nehmen wir an, daß man sich auf m statistische Variablen $X_1, \ldots, X_m$ beziehen möchte und daß es Werte dieser Variablen für n Objekte (Individuen, Situationen) gibt. Dann kann man mit x_{ij} den Wert der Variablen X_j beim Objekt i bezeichnen und diese Werte in einer Datenmatrix

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

zusammenfassen. Die j-te Spalte dieser Matrix enthält die Werte der Variablen X_j, die i-te Zeile enthält die Werte der Variablen $X_1, \ldots, X_m$ für das Objekt i.

2. Der Mittelwert einer Variablen X_j ist durch

$$\mathrm{M}(X_j) := \frac{1}{n} \sum_{i=1}^{n} x_{ij}$$

definiert. Sei jetzt $\mathbf{1}_n$ ein $(n,1)$-Spaltenvektor, bei dem alle Elemente gleich 1 sind. Dann kann man in folgender Weise einen Zeilenvektor bilden, der die Mittelwerte aller Variablen enthält:

$$\frac{1}{n}\,\mathbf{1}_n'\mathbf{X} = (\mathrm{M}(X_1), \ldots, \mathrm{M}(X_m))$$

3. Wichtig für die Überlegungen des Haupttextes sind auch die Begriffe Varianz, Kovarianz und Korrelation. Wir erinnern deshalb kurz an die Definitionen. Die *Varianz* einer Variablen X_j ist durch

$$\mathrm{V}(X_j) := \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \mathrm{M}(X_j))^2$$

definiert. Die *Kovarianz* von zwei Variablen X_j und $X_{j'}$ ist durch

$$\mathrm{Cov}(X_j, X_{j'}) := \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \mathrm{M}(X_j))(x_{ij'} - \mathrm{M}(X_{j'}))$$

definiert; und schließlich die *Korrelation* von X_j und $X_{j'}$ durch

$$\mathrm{Corr}(X_j, X_{j'}) := \frac{\mathrm{Cov}(X_j, X_{j'})}{\sqrt{\mathrm{V}(X_j)\mathrm{V}(X_{j'})}}$$

4. Es ist oft praktisch, die Kovarianzen einer Menge von Variablen $X_1, \ldots, X_m$ in einer *Kovarianzmatrix*

$$\mathrm{Cov}(\mathbf{X}) := \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \cdots & \mathrm{Cov}(X_1, X_m) \\ \vdots & & \vdots \\ \mathrm{Cov}(X_m, X_1) & \cdots & \mathrm{Cov}(X_m, X_m) \end{pmatrix}$$

zusammenzufassen. Es ist eine symmetrische (m,m)-Matrix. Die Hauptdiagonale enthält die Varianzen. (Wir verwenden wahlweise $\mathrm{V}(X_j)$ oder $\mathrm{Cov}(X_j, X_j)$, um die Varianz von X_j zu bezeichnen.) Ganz analog kann man eine *Korrelationsmatrix*

$$\mathrm{Corr}(\mathbf{X}) := \begin{pmatrix} \mathrm{Corr}(X_1, X_1) & \cdots & \mathrm{Corr}(X_1, X_m) \\ \vdots & & \vdots \\ \mathrm{Corr}(X_m, X_1) & \cdots & \mathrm{Corr}(X_m, X_m) \end{pmatrix}$$

definieren.

5. Die Darstellung von Kovarianzmatrizen wird besonders einfach, wenn man um ihre Mittelwerte *zentrierte Variablen* verwendet, also

$$Z_j := X_j - \mathrm{M}(X_j)$$

anstelle von X_j. Für die Kovarianzen findet man dann nämlich

$$\mathrm{Cov}(Z_j, Z_{j'}) = \frac{1}{n} \sum_{i=1}^{n} z_{ij} z_{ij'}$$

und es gibt die einfache Darstellung

$$\mathrm{Cov}(\mathbf{X}) = \mathrm{Cov}(\mathbf{Z}) = \frac{1}{n}\,\mathbf{Z}'\mathbf{Z}$$

für die Kovarianzmatrix der Variablen $Z_1, \ldots, Z_m$.

6. Wenn man nicht nur zentrierte, sondern *standardisierte Variablen* verwendet (vgl. Abschnitt 5.4), also

$$Z_j := \frac{X_j - \mathrm{M}(X_j)}{\sqrt{\mathrm{V}(X_j)}}$$

ergibt sich eine analoge Vereinfachung für die Korrelationsmatrix. Denn die standardisierten Variablen haben die Varianz 1 und die Korrelationen zwischen den standardisierten Variablen sind mit ihren Kovarianzen identisch. Also gilt bei standardisierten Variablen:

$$\mathrm{Corr}(\mathbf{X}) = \mathrm{Corr}(\mathbf{Z}) = \mathrm{Cov}(\mathbf{Z}) = \frac{1}{n}\,\mathbf{Z}'\mathbf{Z}$$

A.3 Invertierbare Matrizen

1. Sei $\mathbf{A}$ eine quadratische (n,n)-Matrix. $\mathbf{A}$ heißt *invertierbar*, wenn es eine (n,n)-Matrix $\mathbf{B}$ gibt, so daß gilt:

$$\mathbf{BA} = \mathbf{AB} = \mathbf{I}_n \tag{A.3.1}$$

Man beachte, daß nur bei quadratischen Matrizen von Invertierbarkeit gesprochen wird.

2. Nicht alle quadratischen Matrizen sind invertierbar. Die folgende Matrix

$$\mathbf{A} := \begin{pmatrix} 1 & 2 \\ 4 & 2 \end{pmatrix}$$

ist jedoch invertierbar. Denn wenn man

$$\mathbf{B} := \begin{pmatrix} -1/3 & 1/3 \\ 2/3 & -1/6 \end{pmatrix}$$

verwendet, kann man sich von der Gültigkeit der Beziehung (A.3.1) überzeugen.

3. Wenn $\mathbf{A}$ invertierbar ist, gibt es genau eine Matrix $\mathbf{B}$, die die Gleichung (A.3.1) erfüllt. Denn angenommen, es gäbe noch eine zweite Matrix $\mathbf{C}$, so daß $\mathbf{CA} = \mathbf{AC} = \mathbf{I}_n$ ist. Dann folgt: $\mathbf{B} = \mathbf{BAC} = \mathbf{C}$. Die bei einer invertierbaren Matrix $\mathbf{A}$ durch (A.3.1) eindeutig bestimmte Matrix $\mathbf{B}$ wird die *zu* $\mathbf{A}$ *inverse Matrix* (kurz *Inverse*) genannt und mit $\mathbf{A}^{-1}$ bezeichnet. Also: wenn $\mathbf{A}$ invertierbar ist, gibt es genau eine inverse Matrix $\mathbf{A}^{-1}$, und es gilt:

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$$

4. Wenn $\mathbf{A}$ eine invertierbare Matrix ist, ist auch $\mathbf{A}^{-1}$ invertierbar, und es gilt:

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}$$

Wenn $\mathbf{A}$ und $\mathbf{B}$ invertierbare (n,n)-Matrizen sind, ist auch $(\mathbf{AB})$ invertierbar, und es gilt:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

denn: $(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}_n$. Wenn $\mathbf{A}$ eine invertierbare (n,n)-Matrix ist, dann ist auch $\mathbf{A}'$ eine invertierbare Matrix, und es gilt:

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$$

denn: $\mathbf{A}'(\mathbf{A}^{-1})' = (\mathbf{A}^{-1}\mathbf{A})' = \mathbf{I}_n' = \mathbf{I}_n$.

5. Bei kleinen invertierbaren Matrizen kann man ihre Inversen per Hand ausrechnen; die Verfahren sind jedoch mühselig und sollen hier nicht besprochen werden. Bei größeren Matrizen wird man stets einen Computer

verwenden, um ihre Inversen zu berechnen. Eine Ausnahme bilden Diagonalmatrizen, deren Inverse (wenn es sie gibt) direkt hingeschrieben werden können. Sei nämlich

$$\mathbf{A} := \operatorname{diag}(a_1, \ldots, a_n)$$

eine Diagonalmatrix. Dann ist $\mathbf{A}$ genau dann invertierbar, wenn alle Koeffizienten $a_1, \ldots, a_n$ ungleich Null sind; und die Inverse ist

$$\mathbf{A}^{-1} = \operatorname{diag}(1/a_1, \ldots, 1/a_n)$$

A.4 Linearkombinationen

1. Sei $\mathbf{a}_1, \ldots, \mathbf{a}_m$ eine Menge von $(n,1)$-Spaltenvektoren. Eine *Linearkombination* dieser Vektoren ist ein Ausdruck der Form

$$k_1\mathbf{a}_1 + \cdots + k_m\mathbf{a}_m$$

wobei $k_1, \ldots, k_m$ reelle Zahlen (Skalare) sind; sie werden die *Koeffizienten* der Linearkombination genannt. Es ist klar, daß eine solche Linearkombination wiederum einen $(n,1)$-Spaltenvektor liefert. Eine Linearkombination wird *nicht-trivial* genannt, wenn mindestens einer der Koeffizienten ungleich Null ist.

2. Die Vektoren $\mathbf{a}_1, \ldots, \mathbf{a}_m$ heißen *linear abhängig*, wenn es eine Linearkombination

$$k_1\mathbf{a}_1 + \cdots + k_m\mathbf{a}_m = \mathbf{0}$$

gibt, wobei mindestens einer der Koeffizienten ungleich Null ist. Die Vektoren heißen *linear unabhängig*, wenn sie nicht linear abhängig sind. Es folgt aus der Definition: Wenn einer der Vektoren $\mathbf{a}_1, \ldots, \mathbf{a}_m$ ein Nullvektor ist, sind die Vektoren linear abhängig. Daraus folgt umgekehrt: Wenn die Vektoren linear unabhängig sind, kann keiner der Vektoren ein Nullvektor sein. Weiterhin folgt unmittelbar aus der Definition: Wenn die Vektoren $\mathbf{a}_1, \ldots, \mathbf{a}_m$ linear unabhängig sind, ist auch jede Teilmenge dieser Vektoren linear unabhängig.

3. Wenn die Vektoren $\mathbf{a}_1, \ldots, \mathbf{a}_m$ linear abhängig sind, kann mindestens einer der Vektoren als eine Linearkombination der übrigen dargestellt werden. Denn es gibt dann eine Linearkombination

$$k_1\mathbf{a}_1 + \cdots + k_m\mathbf{a}_m = \mathbf{0}$$

wobei mindestens einer der Koeffizienten ungleich Null ist. Sei etwa $k_j \neq 0$. Dann ist

$$\mathbf{a}_j = \sum_{l \neq j} \left(-\frac{k_l}{k_j}\right) \mathbf{a}_l$$

Man kann sich weiterhin überlegen: Die Vektoren $\mathbf{a}_1, \ldots, \mathbf{a}_m$ sind genau dann linear unabhängig, wenn keiner der Vektoren als eine Linearkombination der übrigen Vektoren dargestellt werden kann.

4. Ein Vektor $\mathbf{b}$ wird ein *skalares Vielfaches* eines Vektors $\mathbf{a}$ genannt, wenn es eine Zahl c gibt, so daß $\mathbf{b} = c\,\mathbf{a}$ ist. Offenbar gilt, daß zwei Vektoren $\mathbf{a}$ und $\mathbf{b}$ genau dann linear unabhängig sind, wenn keiner der beiden ein skalares Vielfaches des jeweils anderen ist.

5. Wenn $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$ eine invertierbare (n,n)-Matrix ist, dann sind ihre Spaltenvektoren $\mathbf{a}_1, \ldots, \mathbf{a}_n$ linear unabhängig. Denn angenommen, daß $\mathbf{Ak} = \mathbf{0}$ ist, wobei $\mathbf{k} = (k_1, \ldots, k_n)'$ irgendein Vektor ist. Dann folgt aus der Invertierbarkeit von $\mathbf{A}$:

$$\mathbf{k} = (\mathbf{A}^{-1}\mathbf{A})\,\mathbf{k} = \mathbf{A}^{-1}(\mathbf{Ak}) = \mathbf{0}$$

Es gibt also keine nicht-triviale Linearkombination der Spalten von $\mathbf{A}$, mit der man einen Nullvektor erzeugen könnte. Es gilt auch umgekehrt: Wenn die $(n,1)$-Spaltenvektoren $\mathbf{a}_1, \ldots, \mathbf{a}_n$ linear unabhängig sind, dann ist die Matrix

$$\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$$

invertierbar. Ein Beweis ist jedoch etwas komplizierter und soll hier ausgelassen werden.

6. Sei $\mathbf{a}_1, \ldots, \mathbf{a}_m$ eine Menge von $(n,1)$-Spaltenvektoren. Wenn $m > n$ ist, sind diese Vektoren linear abhängig. Denn angenommen, sie wären linear unabhängig. Dann wären auch die Vektoren $\mathbf{a}_1, \ldots, \mathbf{a}_{n+1}$ linear unabhängig, und keiner von ihnen wäre ein Nullvektor. Also bestünde dann auch die Matrix $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$ aus linear unabhängigen Spaltenvektoren und wäre invertierbar. Dann könnte man das Gleichungssystem

$$\mathbf{Ak} = \mathbf{a}_1 k_1 + \ldots + \mathbf{a}_n k_n = \mathbf{a}_{n+1}$$

betrachten. Da $\mathbf{a}_{n+1} \neq \mathbf{0}$ ist, muß auch $\mathbf{k} \neq \mathbf{0}$ sein. Das aber bedeutet, daß einer der Vektoren, nämlich $\mathbf{a}_{n+1}$, als eine nicht-triviale Linearkombination von $\mathbf{a}_1, \ldots, \mathbf{a}_n$ dargestellt werden kann, was ein Widerspruch zur Annahme der linearen Unabhängigkeit ist.

7. Aus der vorstehenden Überlegung folgt insbesondere:

a) Wenn $\mathbf{A}$ eine (n,m)-Matrix und $m > n$ ist, dann sind die Spaltenvektoren von $\mathbf{A}$ linear abhängig.

b) Wenn $\mathbf{A}$ eine (n,m)-Matrix ist, dann ist sowohl die Anzahl ihrer linear unabhängigen Spaltenvektoren als auch die Anzahl ihrer linear unabhängigen Zeilenvektoren höchstens gleich $\min\{n,m\}$.

8. Sei $\mathbf{A}$ eine (n,m)-Matrix. Die maximale Anzahl linear unabhängiger Spaltenvektoren von $\mathbf{A}$ wird ihr *Spaltenrang* genannt. Ganz analog wird die maximale Anzahl linear unabhängiger Zeilenvektoren von $\mathbf{A}$ ihr *Zeilenrang* genannt. Oder anders formuliert: Der Zeilenrang von $\mathbf{A}$ ist der Spaltenrang von $\mathbf{A}'$. Man kann zeigen, daß bei jeder Matrix ihr Zeilenrang und ihr Spaltenrang identisch sind. Man spricht deshalb kurz vom *Rang einer Matrix* und bezeichnet ihn mit $\mathrm{rg}(\mathbf{A})$. Offenbar gilt

$$\mathrm{rg}(\mathbf{A}) \leq \min\{n,m\}$$

wie die zuvor angestellte Überlegung gezeigt hat.

9. Daß bei jeder Matrix Zeilen- und Spaltenrang gleich sind, kann man sich folgendermaßen überlegen (diese Überlegung folgt Searle 1982, S. 169f). Sei $\mathbf{A}$ eine (n,m)-Matrix mit dem Zeilenrang r und dem Spaltenrang s. Zunächst ist klar, daß der Zeilenrang unabhängig von der Anordnung der Spalten und der Spaltenrang unabhängig von der Anordnung der Zeilen in der Matrix $\mathbf{A}$ ist. Man kann auch sicherlich durch Vertauschen von Zeilen und Spalten $\mathbf{A}$ in eine solche Form bringen, daß die ersten r Zeilen und die ersten s Spalten linear unabhängig sind. Diese neue Matrix, die wir $\mathbf{A}^*$ nennen, hat den gleichen Zeilen- und Spaltenrang wie $\mathbf{A}$ und kann also folgendermaßen dargestellt werden:

$$\mathbf{A}^* = \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{W} \end{pmatrix}$$

wobei $\mathbf{X}$ eine (r,s)-Matrix, $\mathbf{Y}$ eine $(r, m-s)$-Matrix, $\mathbf{Z}$ eine $(n-r, s)$-Matrix und $\mathbf{W}$ eine $(n-r, m-s)$-Matrix ist. Die Zeilen von $\mathbf{Z}$ sind nach Voraussetzung linear abhängig von den Zeilen von $\mathbf{X}$. Also gibt es eine Matrix $\mathbf{T}$, so daß

$$\mathbf{Z}' = \mathbf{X}'\mathbf{T}' \quad \text{bzw.} \quad \mathbf{Z} = \mathbf{TX}$$

ist. Jetzt kann man zeigen, daß die Spalten von $\mathbf{X}$ nicht linear abhängig sind. Denn angenommen, sie wären linear abhängig; dann gäbe es einen Vektor $\mathbf{a} \neq \mathbf{0}$, so daß $\mathbf{Xa} = \mathbf{0}$ ist. Dann aber wäre auch $\mathbf{Za} = \mathbf{TXa} = \mathbf{0}$, also auch:

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Z} \end{pmatrix} \mathbf{a} = \mathbf{0}$$

Das aber wäre ein Widerspruch zur Voraussetzung, daß die ersten s Spalten von $\mathbf{A}^*$ linear unabhängig sind. Also hat $\mathbf{X}$ s linear unabhängige Spalten. Und da, wie bereits gezeigt wurde, die Anzahl der linear unabhängigen Spalten einer Matrix nicht größer als die Anzahl ihrer Zeilen sein kann, folgt $s \leq r$. Mit einer ganz analogen Überlegung kann man zeigen, daß $r \leq s$ gelten muß. Also müssen Zeilen- und Spaltenrang gleich sein.

10. Zwei $(n,1)$-Vektoren $\mathbf{x}$ und $\mathbf{y}$ heißen *orthogonal*, wenn $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x} = 0$ ist. Es gilt: wenn $\mathbf{x} \neq \mathbf{0}$ und $\mathbf{y} \neq \mathbf{0}$ und $\mathbf{x}$ und $\mathbf{y}$ orthogonal sind, dann sind $\mathbf{x}$ und $\mathbf{y}$ auch linear unabhängig. Denn angenommen, sie wären linear abhängig; dann müßte es eine Zahl $a \neq 0$ geben, so daß $\mathbf{x} = a\mathbf{y}$ ist. Da $\mathbf{x}$ und $\mathbf{y}$ orthogonal sind, würde daraus indessen folgen: $\mathbf{x}'\mathbf{x} = a\mathbf{x}'\mathbf{y} = 0$, also auch $\mathbf{x} = \mathbf{0}$, was ein Widerspruch zur Voraussetzung ist.

11. Eine (n,n)-Matrix $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$ heißt *orthogonal*, wenn gilt:

$$\mathbf{a}_j'\mathbf{a}_k = \begin{cases} 0 & \text{wenn } j \neq k \\ 1 & \text{wenn } j = k \end{cases} \qquad (\text{für alle } j,k = 1,\ldots,n)$$

Wenn $\mathbf{A}$ eine orthogonale Matrix ist, sind offenbar alle ihre Spaltenvektoren ungleich Null; und aus der Orthogonalitätsbedingung folgt auch, daß ihre Spaltenvektoren linear unabhängig sind. Eine orthogonale Matrix ist also invertierbar. Weiterhin gilt:

$$\mathbf{A}'\mathbf{A} = \begin{pmatrix} \mathbf{a}_1'\mathbf{a}_1 & \cdots & \mathbf{a}_1'\mathbf{a}_n \\ \vdots & & \vdots \\ \mathbf{a}_n'\mathbf{a}_1 & \cdots & \mathbf{a}_n'\mathbf{a}_n \end{pmatrix} = \mathbf{I}_n$$

Da $\mathbf{A}$ invertierbar ist, folgt:

$$\mathbf{A}' = \mathbf{A}'\mathbf{I}_n = \mathbf{A}'\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n\mathbf{A}^{-1} = \mathbf{A}^{-1}$$

Bei einer orthogonalen Matrix $\mathbf{A}$ ist also ihre Inverse mit $\mathbf{A}'$ identisch.

A.5 Eigenwerte und Eigenvektoren

1. Sei $\mathbf{A}$ eine quadratische (n,n)-Matrix. Ein $(n,1)$-Vektor $\mathbf{v}$ heißt ein *Eigenvektor* von $\mathbf{A}$, wenn $\mathbf{v} \neq \mathbf{0}$ ist und es eine Zahl λ gibt, so daß gilt:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

λ wird dann ein *Eigenwert* von $\mathbf{A}$ genannt. Man sagt auch: λ ist der zum Eigenvektor $\mathbf{v}$ korrespondierende Eigenwert.

2. In diesem Text benötigen wir Eigenwerte und Eigenvektoren nur für symmetrische Matrizen. Dann gibt es zwei grundlegende Feststellungen.[2] Für jede symmetrische (n,n)-Matrix $\mathbf{A}$ gilt:

a) Die Eigenwerte und Eigenvektoren von $\mathbf{A}$ sind stets reell.

b) $\mathbf{A}$ hat genau n linear unabhängige Eigenvektoren $\mathbf{v}_1, \ldots, \mathbf{v}_n$ mit zugehörigen Eigenwerten $\lambda_1, \ldots, \lambda_n$.

[2]Die Beweise sind ziemlich kompliziert und sollen hier deshalb nicht nachvollzogen werden.

Betrachten wir als Beispiel eine Diagonalmatrix

$$\mathbf{A} = \operatorname{diag}(a_1, \ldots, a_n)$$

Sie hat offenbar die Eigenwerte $a_1, \ldots, a_n$ und die zugehörigen Eigenvektoren sind gerade die Einheitsvektoren $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Dies Beispiel zeigt auch, daß Eigenwerte Null sein können und daß mehrere Eigenwerte einer Matrix gleich sein können.

3. Im weiteren beziehen wir uns auf eine beliebige symmetrische (n,n)-Matrix $\mathbf{A}$. Es seien $\mathbf{v}_1, \ldots, \mathbf{v}_n$ ihre Eigenvektoren und $\lambda_1, \ldots, \lambda_n$ die zugehörigen Eigenwerte, so daß man schreiben kann:

$$\mathbf{A}\mathbf{v}_j = \lambda_j\mathbf{v}_j \qquad (\text{für } j = 1,\ldots,n) \tag{A.5.1}$$

Man erkennt: Wenn $\mathbf{v}_j$ ein Eigenvektor von $\mathbf{A}$ ist, dann ist auch $k\mathbf{v}_j$ ein Eigenvektor, wobei k eine beliebige Zahl sein kann. Eigenvektoren können also beliebig normiert werden. Insbesondere können sie so normiert werden, daß gilt: $\mathbf{v}_j'\mathbf{v}_j = 1$. Wir werden im folgenden stets diese Normierung voraussetzen.

4. Die Eigenvektoren können zu einer Matrix

$$\mathbf{V} := (\mathbf{v}_1, \ldots, \mathbf{v}_n)$$

zusammengefaßt werden. Da die Spalten dieser Matrix linear unabhängig sind, ist es eine invertierbare Matrix. Weiterhin können die Eigenwerte zu einer Diagonalmatrix

$$\mathbf{\Lambda} := \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$$

zusammengefaßt werden. Also kann (A.5.1) auch folgendermaßen geschrieben werden (man beachte, daß die Matrizenmultiplikation auf beiden Seiten nicht kommutativ ist):

$$\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{\Lambda} \tag{A.5.2}$$

Diese Gleichung zeigt auch, daß man die Reihenfolge der Eigenvektoren (Spalten) von $\mathbf{V}$ beliebig verändern kann, wenn man nur die Reihenfolge der Eigenwerte in $\mathbf{\Lambda}$ in entsprechender Weise verändert.

5. Man kann auf einfache Weise zeigen, daß die Eigenvektoren zu zwei verschiedenen Eigenwerten orthogonal sind. Beziehen wir uns auf zwei Eigenwerte λ_j und λ_k. Dann kann man zunächst schreiben:

$$\mathbf{A}\mathbf{v}_j = \lambda_j\mathbf{v}_j \quad \text{und} \quad \mathbf{A}\mathbf{v}_k = \lambda_k\mathbf{v}_k$$

Aus der ersten Gleichung folgt: $\mathbf{v}_k'\mathbf{A}\mathbf{v}_j = \lambda_j\mathbf{v}_k'\mathbf{v}_j$. Aus der zweiten Gleichung folgt zunächst, da $\mathbf{A}$ symmetrisch ist: $\mathbf{v}_k'\mathbf{A} = \lambda_k\mathbf{v}_k'$; und daraus folgt durch Multiplikation mit $\mathbf{v}_j$: $\mathbf{v}_k'\mathbf{A}\mathbf{v}_j = \lambda_k \mathbf{v}_k'\mathbf{v}_j$. Beide Überlegungen zusammen ergeben also

$$\lambda_j \mathbf{v}_k'\mathbf{v}_j = \lambda_k \mathbf{v}_k'\mathbf{v}_j$$

Wenn $\lambda_j \neq \lambda_k$ ist, folgt daraus $\mathbf{v}_k'\mathbf{v}_j = 0$, d.h., daß die beiden Eigenvektoren orthogonal sind.

6. Aus dieser Überlegung ergibt sich: Wenn alle Eigenwerte einer symmetrischen Matrix $\mathbf{A}$ verschieden sind, besteht die Matrix ihrer Eigenvektoren aus paarweise orthogonalen Spaltenvektoren. Da wir vereinbart hatten, daß die Eigenvektoren normiert sind, d.h. $\mathbf{v}_j'\mathbf{v}_j = 1$, handelt es sich um eine orthogonale Matrix, also

$$\mathbf{V}'\mathbf{V} = \mathbf{V}\mathbf{V}' = \mathbf{I}_n \tag{A.5.3}$$

Die Überlegung, um die Orthogonalität von zwei Eigenvektoren zu zeigen, beruhte allerdings darauf, daß man sich auf zwei verschiedene Eigenwerte beziehen kann. Wenn zwei oder mehr Eigenwerte den gleichen Wert haben, versagt die Überlegung. Man kann aber zeigen, daß auch dann die Eigenvektoren so gewählt werden können, daß sie zueinander orthogonal sind.[3] Das heißt, daß die Eigenvektoren einer symmetrischen Matrix immer derart bestimmt werden können, daß die Matrix der Eigenvektoren orthogonal ist, also (A.5.3) gilt.

7. Eine symmetrische (n, n)-Matrix $\mathbf{A}$ hat genau dann den Rang n (und ist also invertierbar), wenn alle ihre Eigenwerte ungleich Null sind. Einen Beweis kann man sich leicht überlegen. (a) Nehmen wir zunächst an, daß $\mathrm{rg}(\mathbf{A}) = n$ ist. Dann müssen alle Eigenwerte ungleich Null sein. Denn wäre z.B. $\lambda_k = 0$, wäre auch $\mathbf{A}\mathbf{v}_k = \mathbf{0}$; und dies würde zeigen, daß die Spalten von $\mathbf{A}$ linear abhängig sind. (b) Nehmen wir umgekehrt an, daß alle Eigenwerte von $\mathbf{A}$ ungleich Null sind. Dann müssen die Spalten von $\mathbf{A}$ linear unabhängig sein. Denn angenommen, es gäbe einen Vektor $\mathbf{k} \neq \mathbf{0}$, so daß $\mathbf{A}\mathbf{k} = \mathbf{0}$ ist. Dann würde aus (A.5.2) folgen:

$$\mathbf{V}'\mathbf{A}\mathbf{k} = \mathbf{\Lambda}\mathbf{V}'\mathbf{k} = 0$$

Dies wäre aber ein Widerspruch, da $\mathbf{V}'$, wie $\mathbf{V}$, vollen Rang und damit linear unabhängige Spalten hat.

8. Eine symmetrische Matrix $\mathbf{A}$ heißt *positiv semi-definit*, wenn für alle Vektoren $\mathbf{x} \neq \mathbf{0}$ gilt:[4]

$$\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0 \tag{A.5.4}$$

[3]Die Überlegung ist etwas komplizierter und soll hier nicht nachvollzogen werden.

[4]Entsprechend heißt die Matrix *positiv definit*, wenn in der folgenden Bedingung ein strenges $>$ gefordert wird.

Die meisten symmetrischen Matrizen, mit denen wir uns in diesem Text beschäftigen, haben diese Eigenschaft. Insbesondere ist jede Matrix, die als ein *Kreuzprodukt*

$$\mathbf{A} = \mathbf{X}'\mathbf{X}$$

dargestellt werden kann, positiv semi-definit; also insbesondere auch Kovarianz- und Korrelationsmatrizen. Denn gilt diese Darstellung, kann man für jeden beliebigen Vektor $\mathbf{x}$ auch einen Vektor $\mathbf{y} = \mathbf{X}\mathbf{x}$ bilden und findet dann:

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{X}'\mathbf{X}\mathbf{x} = \mathbf{y}'\mathbf{y} \geq 0$$

9. Wenn eine symmetrische Matrix $\mathbf{A}$ positiv semi-definit ist, folgt daraus, daß alle ihre Eigenwerte größer oder gleich Null sind. Denn geht man von der Darstellung (A.5.1) aus, folgt

$$\mathbf{v}_j'\mathbf{A}\mathbf{v}_j = \lambda_j \mathbf{v}_j'\mathbf{v}_j \geq 0$$

Also muß gelten, daß $\lambda_j \geq 0$ ist. Entsprechend kann man sich überlegen, daß alle Eigenwerte positiv sind, wenn die Matrix positiv definit ist.

10. Wenn $\mathbf{A}$ positiv semi-definit ist, kann man mit ihren Eigenwerten und Eigenvektoren eine sehr einfache Darstellung finden. Es gibt dann nämlich keine negativen Eigenwerte, und man kann eine Diagonalmatrix

$$\mathbf{\Lambda}^{1/2} := \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_n} \end{pmatrix}$$

definieren. Dann findet man ausgehend von (A.5.2) und (A.5.3):

$$\mathbf{A} = \mathbf{A}\mathbf{V}\mathbf{V}' = \mathbf{V}\mathbf{\Lambda}\mathbf{V}' = \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{V}' = (\mathbf{V}\mathbf{\Lambda}^{1/2})(\mathbf{V}\mathbf{\Lambda}^{1/2})'$$

also $\mathbf{A} = \mathbf{W}\mathbf{W}'$, wobei $\mathbf{W} = \mathbf{V}\mathbf{\Lambda}^{1/2}$ ist, also auch $\mathbf{W}'\mathbf{W} = \mathbf{\Lambda}$.

11. Wenn $\mathbf{A}$ vollen Rang hat, also invertierbar ist, kann mit Hilfe dieser Darstellung die Inverse von $\mathbf{A}$ konstruiert werden. Es kann dann nämlich, wie oben gezeigt worden ist, kein Eigenwert gleich Null, sondern alle Eigenwerte müssen größer als Null sein. Also kann man eine Matrix

$$\mathbf{\Lambda}^{-1} := \begin{pmatrix} 1/\lambda_1 & & 0 \\ & \ddots & \\ 0 & & 1/\lambda_n \end{pmatrix}$$

definieren, und mit ihrer Hilfe auch $\mathbf{A}^{-1} := \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}'$. Einfaches Ausrechnen zeigt, daß es sich um die Inverse von $\mathbf{A}$ handelt.

A.6 Lineare Regression

1. Gelegentlich benötigen wir in diesem Text Verfahren der linearen Regression, wie sie aus Einführungen in die Statistik bekannt sind. Hier soll deshalb gezeigt werden, wie mit Hilfe der Matrizenschreibweise eine einfache Darstellung erreicht werden kann. Als Ausgangspunkt nehmen wir an, daß eine abhängige Variable Y und unabhängige Variablen $X_2, \ldots, X_m$ gegeben sind. Ein linearer Regressionsansatz sieht dann folgendermaßen aus:

$$Y = \beta_1 + X_2\beta_2 + \cdots + X_m\beta_m + U(\beta_1, \ldots, \beta_m)$$

wobei die Variable $U(\beta_1, \ldots, \beta_m)$ durch den Regressionsansatz definiert wird, so daß ihre Werte auch von den Regressionsparametern $\beta_1, \ldots, \beta_m$ abhängen. Definiert man eine Hilfsvariable X_1, die stets den Wert 1 annimmt, kann man auch schreiben:

$$Y = X_1\beta_1 + X_2\beta_2 + \cdots + X_m\beta_m + U(\beta_1, \ldots, \beta_m) \qquad (A.6.1)$$

Die Idee ist, Werte für die Parameter $\beta_1, \ldots, \beta_m$ so zu bestimmen, daß die Werte der durch den Regressionsansatz definierten Variablen $U(\beta_1, \ldots, \beta_m)$ insgesamt „möglichst klein“ werden. Das meistens verwendete Kriterium bezieht sich auf die Summe der quadrierten Werte von $U(\beta_1, \ldots, \beta_m)$. Man spricht dann von einer *OLS-Regression* („ordinary least squares“).

2. Um diese Idee praktisch zu verfolgen, benötigt man Daten. Nehmen wir also an, daß uns Werte für die Variablen Y und $X_1, \ldots, X_m$ zur Verfügung stehen. Gibt es n Werte, können sie mit Hilfe der Matrizenschreibweise so dargestellt werden:

$$\mathbf{y} := \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \text{und} \quad \mathbf{X} := \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

Definiert man außerdem einen Vektor $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_m)'$ für die Parameter und einen Vektor $\mathbf{u}(\boldsymbol{\beta}) := (u_1(\boldsymbol{\beta}), \ldots, u_n(\boldsymbol{\beta}))'$ für die Werte der Variablen $U(\beta_1, \ldots, \beta_m)$, kann der Regressionsansatz (A.6.1) in Matrizenschreibweise so formuliert werden:

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \cdots + \mathbf{x}_m\beta_m + \mathbf{u}(\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}(\boldsymbol{\beta}) \qquad (A.6.2)$$

wobei $\mathbf{x}_1, \ldots, \mathbf{x}_m$ die Spalten der Matrix $\mathbf{X}$ sind. $\mathbf{X}\boldsymbol{\beta}$ ist also eine Linearkombination der Spalten von $\mathbf{X}$.

3. Um Werte für die Parameter des Regressionsansatzes zu bestimmen, muß der folgende Ausdruck minimiert werden:

$$\sum_{i=1}^{n} u_i(\boldsymbol{\beta})^2 = \mathbf{u}(\boldsymbol{\beta})'\mathbf{u}(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \|^2 \qquad (A.6.3)$$

Es muß also das Minimum der Funktion

$$f(\boldsymbol{\beta}) := (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

bestimmt werden. Eine notwendige Bedingung ist, daß an der Stelle des Minimums die ersten Ableitungen der Funktion Null werden. In der Matrixschreibweise können diese Ableitungen so ausgedrückt werden:[5]

$$\frac{\partial f(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = -2\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = -2\mathbf{X}'\mathbf{u}(\boldsymbol{\beta})$$

Parameterwerte, die ein Minimum der Funktion f liefern, müssen also die Bedingung

$$\mathbf{X}'\mathbf{u}(\boldsymbol{\beta}) = \mathbf{0} \quad \text{bzw.} \quad \mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y} \qquad (A.6.4)$$

erfüllen. Offenbar gibt es genau dann eine Lösung, wenn $\mathbf{X}'\mathbf{X}$ eine invertierbare Matrix ist, und die eindeutige Lösung ist dann

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

4. Definiert man $\hat{\mathbf{y}} := \mathbf{X}\hat{\boldsymbol{\beta}}$ und $\mathbf{u}(\hat{\boldsymbol{\beta}}) := \mathbf{y} - \hat{\mathbf{y}}$, erhält man die Darstellung

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{u}(\hat{\boldsymbol{\beta}})$$

Unmittelbar aus (A.6.4) folgt außerdem $\mathbf{X}'\mathbf{u}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$, so daß der Vektor $\mathbf{u}(\hat{\boldsymbol{\beta}})$ zu allen Spalten von $\mathbf{X}$, und somit auch zu jeder Linearkombination dieser Spalten, orthogonal ist.

A.7 Pyrrhons Lemma

1. Um mit der Regressionsrechnung etwas vertrauter zu werden, besprechen wir *Pyrrhons Lemma*, das zeigt, wie man durch Hinzufügen von Variablen zu einem Regressionsansatz die Modellparameter verändern kann. Wir folgen einer Darstellung durch T. K. Dijkstra (1995), der auch die Bezeichnung *Pyrrhons Lemma* vorgeschlagen hat, wohl um zum Ausdruck zu bringen, daß man der Regressionsrechnung mit Skepsis begegnen sollte.

[5] Faßt man $\mathbf{Ax}$ als eine vektor-wertige Funktion des Vektors $\mathbf{x}$ auf, gilt folgende Differentiationsregel:

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \mathbf{A}'$$

Entsprechend gilt, wenn man $\mathbf{x}'\mathbf{Ax}$ als eine skalar-wertige Funktion des Vektors $\mathbf{x}$ auffaßt, die Regel

$$\frac{\partial \mathbf{x}'\mathbf{Ax}}{\partial \mathbf{x}} = 2\mathbf{Ax}$$

2. Als Ausgangspunkt dient eine lineare Regression der Form

$$\mathbf{y} \;=\; \mathbf{x}_1\hat\beta_1 + \cdots + \mathbf{x}_m\hat\beta_m + \mathbf{u}(\hat{\boldsymbol\beta}) \tag{A.7.1}$$

wobei vorausgesetzt wird, daß die Länge der Vektoren gleich $n > m$ ist und die Regressionsparameter $\hat\beta_1,\dots,\hat\beta_m$ mit der OLS-Methode berechnet worden sind. Sei jetzt angenommen, daß man sich anstelle der zunächst gefundenen Modellparameter $\hat\beta_1,\dots,\hat\beta_m$ andere Werte wünscht, etwa $\hat\beta_1^*,\dots,\hat\beta_m^*$. Pyrrhons Lemma besagt, daß man dies dadurch erreichen kann, daß man eine weitere Variable, wir nennen sie $\mathbf{z}$, konstruiert und in den Regressionsansatz mit aufnimmt:

$$\mathbf{y} \;=\; \mathbf{x}_1\hat\beta_1^* + \cdots + \mathbf{x}_m\hat\beta_m^* + \mathbf{z}\,\hat\gamma^* + \mathbf{u}(\hat{\boldsymbol\beta}^*,\hat\gamma^*)$$

Um das zu erreichen, geht man von folgender Darstellung für die neue Variable $\mathbf{z}$ aus:

$$\mathbf{z} \;=\; \mathbf{x}_1\alpha_1 + \cdots + \mathbf{x}_m\alpha_m + \mathbf{u}(\hat{\boldsymbol\beta})\,\delta + \mathbf{d} \tag{A.7.2}$$

wobei $\delta \neq 0$ irgendeine Zahl ist und $\mathbf{d}$ irgendein Vektor der Länge n, der sowohl zu $\mathbf{x}_1,\dots,\mathbf{x}_m$ als auch zu $\mathbf{u}(\hat{\boldsymbol\beta})$ orthogonal ist, so daß gilt:

$$\mathbf{d}'\mathbf{u}(\hat{\boldsymbol\beta}) = 0,\; \mathbf{d}'\mathbf{x}_1 = 0,\dots,\mathbf{d}'\mathbf{x}_m = 0$$

Wenn $n > m+1$ ist, kann ein solcher Vektor stets gefunden werden.

3. Verwendet man in dem erweiterten Regressionsansatz diese Darstellung für $\mathbf{z}$, findet man:

$$\begin{aligned}\mathbf{y} \;&=\; \mathbf{x}_1\hat\beta_1^* + \cdots + \mathbf{x}_m\hat\beta_m^* + \mathbf{z}\,\hat\gamma^* + \mathbf{u}(\hat{\boldsymbol\beta}^*,\hat\gamma^*)\\ &=\; \mathbf{x}_1(\hat\beta_1^* + \alpha_1\hat\gamma^*) + \cdots + \mathbf{x}_m(\hat\beta_m^* + \alpha_m\hat\gamma^*) + \mathbf{u}(\alpha,\hat{\boldsymbol\beta}^*,\hat\gamma^*)\end{aligned} \tag{A.7.3}$$

wobei $\mathbf{u}(\alpha,\hat{\boldsymbol\beta}^*,\hat\gamma^*) := (\mathbf{u}(\hat{\boldsymbol\beta})\delta+\mathbf{d})\hat\gamma^* + \mathbf{u}(\hat{\boldsymbol\beta}^*,\hat\gamma^*)$ ist. Faßt man die Vektoren $\mathbf{x}_1,\dots,\mathbf{x}_n$ zu einer Matrix $\mathbf{X}$ zusammen und bildet man den Vektor

$$\mathbf{w} \;:=\; (\hat\beta_1 - (\hat\beta_1^* + \alpha_1\hat\gamma^*),\dots,\hat\beta_m - (\hat\beta_m^* + \alpha_m\hat\gamma^*))'$$

kann man die Gleichungen (A.7.1) und (A.7.3) folgendermaßen zusammenfassen:

$$\mathbf{X}\mathbf{w} \;=\; \mathbf{u}(\hat{\boldsymbol\beta}) - \mathbf{u}(\alpha,\hat{\boldsymbol\beta}^*,\hat\gamma^*)$$

Also gilt auch

$$\mathbf{X}'\mathbf{X}\mathbf{w} \;=\; \mathbf{X}'(\mathbf{u}(\hat{\boldsymbol\beta}) - \mathbf{u}(\alpha,\hat{\boldsymbol\beta}^*,\hat\gamma^*))$$

Nun sind jedoch nicht nur $\mathbf{u}(\hat{\boldsymbol\beta})$ und $\mathbf{u}(\hat{\boldsymbol\beta}^*,\hat\gamma^*)$ orthogonal zu den Regressorvariablen $\mathbf{x}_1,\dots,\mathbf{x}_m$, sondern dies gilt nach Voraussetzung ebenfalls für den Vektor $\mathbf{d}$. Das gleiche gilt somit auch für $\mathbf{u}(\alpha,\hat{\boldsymbol\beta}^*,\hat\gamma^*)$, und es folgt

$$\mathbf{X}'(\mathbf{u}(\hat{\boldsymbol\beta}) - \mathbf{u}(\alpha,\hat{\boldsymbol\beta}^*,\hat\gamma^*)) \;=\; \mathbf{0}$$

Also ist auch

$$\mathbf{w} \;=\; (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{u}(\hat{\boldsymbol\beta}) - \mathbf{u}(\alpha,\hat{\boldsymbol\beta}^*,\hat\gamma^*)) = \mathbf{0}$$

und man findet:

$$\mathbf{u}(\hat{\boldsymbol\beta}) \;=\; \mathbf{u}(\alpha,\hat{\boldsymbol\beta}^*,\hat\gamma^*) \quad\text{und}\quad \hat\beta_j \;=\; \hat\beta_j^* + \alpha_j\hat\gamma^* \tag{A.7.4}$$

für $j = 1,\dots,m$. Hieraus läßt sich zunächst ein beliebiger Wert für $\hat\gamma^*$ bestimmen. Denn aus $\mathbf{u}(\hat{\boldsymbol\beta}) = \mathbf{u}(\alpha,\hat{\boldsymbol\beta}^*,\hat\gamma^*)$ folgt zunächst

$$\mathbf{u}(\hat{\boldsymbol\beta}) \;=\; (\mathbf{u}(\hat{\boldsymbol\beta})\,\delta + \mathbf{d})\,\hat\gamma^* + \mathbf{u}(\hat{\boldsymbol\beta}^*,\hat\gamma^*)$$

Multipliziert man diese Gleichung von links mit $(\mathbf{u}(\hat{\boldsymbol\beta})\,\delta + \mathbf{d})'$, und berücksichtigt man, daß aus den Orthogonalitätsbeziehungen insbesondere auch $(\mathbf{u}(\hat{\boldsymbol\beta})\,\delta + \mathbf{d})'\mathbf{u}(\hat{\boldsymbol\beta}^*,\hat\gamma^*) = \mathbf{0}$ folgt, findet man:

$$\mathbf{u}(\hat{\boldsymbol\beta})'\mathbf{u}(\hat{\boldsymbol\beta})\,\delta \;=\; (\mathbf{u}(\hat{\boldsymbol\beta})\,\delta + \mathbf{d})'(\mathbf{u}(\hat{\boldsymbol\beta})\,\delta + \mathbf{d})\,\hat\gamma^* \;=\; (\mathbf{u}(\hat{\boldsymbol\beta})'\mathbf{u}(\hat{\boldsymbol\beta})\,\delta^2 + \mathbf{d}'\mathbf{d})\,\hat\gamma^*$$

Und daraus gewinnt man schließlich die Darstellung

$$\hat\gamma^* \;=\; \frac{\mathbf{u}(\hat{\boldsymbol\beta})'\mathbf{u}(\hat{\boldsymbol\beta})\,\delta}{\mathbf{u}(\hat{\boldsymbol\beta})'\mathbf{u}(\hat{\boldsymbol\beta})\,\delta^2 + \mathbf{d}'\mathbf{d}} \tag{A.7.5}$$

Beginnt man also mit den Residuen $\mathbf{u}(\hat{\boldsymbol\beta})$ des ursprünglichen Regressionsansatzes und einem in weiten Grenzen frei wählbaren Vektor $\mathbf{d}$, kann man sich durch die Wahl der Zahl δ zunächst einen beliebigen Koeffizienten $\hat\gamma^*$ verschaffen. Dann aber kann man ausgehend von beliebig vorgegebenen neuen Modellparametern $\hat\beta_1^*,\dots,\hat\beta_m^*$ aufgrund von (A.7.4) die erforderlichen Koeffizienten $\alpha_1,\dots,\alpha_m$ bestimmen, um schließlich mit Hilfe von (A.7.2) die Werte der zusätzlichen Regressorvariablen $\mathbf{z}$ zu berechnen.

4. Um den Rechengang zu illustrieren, verwenden wir folgende (willkürlich ausgedachten) Daten:

$\mathbf{y}$	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{y}$	$\mathbf{x}_1$	$\mathbf{x}_2$
1600	1	1	2900	1	0
2000	1	1	3500	1	0
1800	1	0	2600	1	1
2500	1	1	4000	1	0

Man kann sich vorstellen, daß es sich um Daten für 8 Personen handelt, wobei $\mathbf{y}$ das monatliche Erwerbseinkommen und $\mathbf{x}_2$ das Geschlecht (0 männlich, 1 weiblich) erfaßt. Der Regressionsansatz ist dann:

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + \mathbf{u}(\boldsymbol{\beta}) \tag{A.7.6}$$

und man erhält die Modellparameter $\hat{\beta}_1 = 3050$ und $\hat{\beta}_2 = -875$, woraus erkenntlich wird, daß Frauen im Durchschnitt weniger verdienen als Männer. Aber angenommen, jemand hätte lieber andere Modellparameter, etwa

$$\hat{\beta}_1^* = 3050 \quad \text{und} \quad \hat{\beta}_2^* = 0$$

so daß die Geschlechtsvariable keinen Beitrag zum bedingten Mittelwert liefert. Pyrrhons Lemma zeigt dann, wie man eine Variable $\mathbf{z}$ konstruieren kann, deren Aufnahme in den Regressionsansatz zu diesem Ergebnis führt. Zunächst muß man irgendeinen Vektor $\mathbf{d}$ finden, der sowohl zu $\mathbf{x}_1$ und $\mathbf{x}_2$ als auch zu $\mathbf{u}(\hat{\boldsymbol{\beta}})$ orthogonal ist. Dafür gibt es viele verschiedene Möglichkeiten. Die folgende Tabelle zeigt zunächst in den ersten beiden Spalten noch einmal $\mathbf{x}_1$ und $\mathbf{x}_2$ und in der dritten Spalte den Residualvektor $\mathbf{u}(\hat{\boldsymbol{\beta}})$ aus der Regression (A.7.6). Dann folgen in der vierten Spalte Werte eines Vektors $\mathbf{d}$, der zu $\mathbf{x}_1$, $\mathbf{x}_2$ und $\mathbf{u}(\hat{\boldsymbol{\beta}})$ orthogonal ist.

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{u}(\hat{\boldsymbol{\beta}})$	$\mathbf{d}$	$\mathbf{z}$
1	1	−575	−15.25	−161.677
1	1	−175	17.25	−89.177
1	0	−1250	3.00	−122.000
1	1	325	−1.00	−57.427
1	0	−150	−1.00	−16.000
1	0	450	−1.00	44.000
1	1	425	−1.00	−47.426
1	0	950	−1.00	94.000

Somit findet man $\mathbf{u}(\hat{\boldsymbol{\beta}})'\mathbf{u}(\hat{\boldsymbol{\beta}}) = 3337500$ und $\mathbf{d}'\mathbf{d} = 544.125$. Wählt man z.B. $\delta = 0.1$, erhält man aus (A.7.5) den Wert $\hat{\gamma}^* \approx 9.8396$. Also gewinnt man aus den gewünschten neuen Modellparametern und aus (A.7.4) die Werte $\alpha_1 = 0$ und $\alpha_2 = -88.9265$. Und so kann man schließlich die Werte der neuen Variablen $\mathbf{z}$ berechnen, die in der letzten Spalte der obigen Tabelle angegeben sind. Der Regressionsansatz

$$\mathbf{y} = \mathbf{x}_1\beta_1^* + \mathbf{x}_2\beta_2^* + \mathbf{z}\,\gamma^* + \mathbf{u}(\hat{\boldsymbol{\beta}}^*, \gamma^*)$$

liefert dann die gewünschten neuen Modellparameter, nämlich $\hat{\beta}_1^* \approx 3050$, $\hat{\beta}_2^* \approx 0$ und $\hat{\gamma}^* \approx 9.84$.

Literatur

Adams, E. W. 1966. On the Nature and Purpose of Measurement. Synthese 16, 125–169.

Adorno, T. W. 1957. Soziologie und empirische Forschung. In: E. Topitsch (Hg.), Logik der Sozialwissenschaften, 511–525. Köln: Kiepenheuer & Witsch 1972.

Ajzen, I. 1988. Attitudes, Personality, and Behavior. Stony Stratford: Open University Press.

Alber, J. 1988. Die Gesundheitssysteme der OECD-Länder im Vergleich. In: M. G. Schmidt (Hg.), Staatstätigkeit. Internationale und historisch vergleichende Analysen, 116–150. Opladen: Westdeutscher Verlag.

Albrecht, G. 1975. Nicht-reaktive Messung und Anwendung historischer Methoden. In: Techniken der empirischen Sozialforschung, Band 2, 9–81. München: Oldenbourg.

Allmendinger, J., Schmidt, P., Wegener, B. 1983. ZUMA-Handbuch sozialwissenschaftlicher Skalen. Dokumentarische Bearbeitung: H. P. Ohly, T. Eikelmann. Bonn: Informationszentrum Sozialwissenschaften.

Allport, G. W. 1935. Attitudes. In: C. Murchison (ed.), A Handbook of Social Psychology. Worcester: Clark University Press.

Anastasi, A. 1982. Psychological Testing. 5th ed. New York: Macmillan.

Andersen, E. B. 1973. Conditional Inference and Models for Measuring. Copenhagen: Mentalhygiejnisk Forlag.

Anderson, N. H. 1982. Cognitive Algebra and Social Psychophysics. In: B. Wegener (ed.), Social Attitudes and Psychophysical Measurement, 123–148. Hillsdale: Lawrence Earlbaum.

Andrich, D. 1978. A Rating Formulation for Ordered Response Categories. Psychometrika 43, 561–573.

Arminger, G. 1979. Faktorenanalyse. Stuttgart: Teubner.

Atteslander, P. 1975. Methoden der empirischen Sozialforschung. Berlin: de Gruyter.

Atteslander, P., Kopp, M. 1984. Befragung. In: E. Roth, K. Heidenreich (Hg.), Sozialwissenschaftliche Methoden. Lehr- und Handbuch für Forschung und Praxis, 144–172. München: Oldenbourg.

Backhaus, K., Erichson, B., Plinke, W., Schuchard-Ficher, C., Weiber, R. 1987. Multivariate Analysemethoden. Berlin: Springer-Verlag.

Bailey, K. D. 1982. Methods of Social Research. 2nd ed. New York: Free Press.

Bateson, N. 1984. Data Construction in Social Surveys. London: Allen & Unwin.

Bender, S., Hilzendegen, J., Rohwer, G., Rudolph, H. 1996. Die IAB-Beschäftigtenstichprobe 1975–1990. BeitrAB 197. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung.

Besozzi, C., Zehnpfennig, H. 1976. Methodologische Probleme der Index-Bildung. In: Techniken der empirischen Sozialforschung, Band 5, 9–55. München: Oldenbourg.

Birchfield, V., Crepaz, M. M. L. 1998. The Impact of Constitutional Structures and Collective and Competitive Veto Points on Income Inequality in Industrialized Democracies. European Journal of Political Research 34, 175–200.

Black, T. R. 1999. Doing Quantitative Research in the Social Sciences. An Integrated Approach to Research Design, Measurement and Statistics. London: Sage.

Blokland-Vogelesang, R. van 1993. A Nonparametric Distance Model for Unidimensional Unfolding. In: M. A. Fligner, J. S. Verducci (eds.), Probability Models and Statistical Analyses for Ranking Data, 241–277. New York: Springer-Verlag.

Blumer, H. 1956. Sociological Analysis and the „Variable“. American Sociological Review 21, 683–690. Dt. Übers. in: K. Acham (Hg.), Methodologische Probleme der Sozialwissenschaften, 386–402. Darmstadt: Wissenschaftliche Buchgesellschaft 1978.

Borg, I., Staufenbiel, T. 1993. Theorien und Methoden der Skalierung. Bern: Huber.

Bortz, J. 1984. Lehrbuch der empirischen Forschung für Sozialwissenschaftler. Berlin: Springer-Verlag.

Bortz, J., Döring, N. 1995. Forschungsmethoden und Evaluation. Berlin: Springer-Verlag.

Braun, M., Mohler, P. P. 1991. Die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS): Rückblick und Ausblick in die neunziger Jahre. ZUMA-Nachrichten 29, 7–25.

Buhmann, B., Rainwater, L., Schmaus, G., Smeeding, T. M. 1988. Equivalence Scales, Well-Being, Inequality, and Poverty: Sensitivity Estimates Across Ten Countries Using the Luxembourg Income Study (LIS) Database. Review of Income and Wealth 34, 115–142.

Bulmer, M. 1979. Concepts in the Analysis of Qualitative Data. Sociological Review 27, 651–677.

Campbell, N. 1921. Measurement. In: J. R. Newman (ed.), The World of Mathematics, Vol. 3, 1797–1813. New York: Simon and Schuster 1956.

Campenhausen, C. von 1981. Die Sinne der Menschen. Band I: Einführung in die Psychophysik der Wahrnehmung. Stuttgart: Thieme.

Carnap, R. 1956. The Methodological Character of Theoretical Concepts. In: H. Feigl, M. Scriven (eds.), The Foundations of Science and the Concepts of Psychology and Psychoanalysis, 38–76. Minneapolis: University of Minnesota Press.

Cattell, R. B. 1952. Factor Analysis. An Introduction and Manual for the Psychologist and Social Scientist. Westport: Greenwood.

Cattell, R. B. 1967. The Scientific Analysis of Personality. Harmondsworth: Penguin.

Cattell, R. B. 1973. Die empirische Erforschung der Persönlichkeit. Weinheim: Beltz.

Coleman, J. S. 1990. Foundations of Social Theory. Cambridge: Harvard University Press.

Comrey, A. L., Lee, H. B. 1992. A First Course in Factor Analysis. Hillsdale: Lawrence Erlbaum.

Coombs, C. H. 1950. Psychological Scaling without a Unit of Measurement. Psychological Review 57, 145–158.

Coombs, C. H. 1964. A Theory of Data. New York: Wiley.

Coombs, C. H., Coombs, L. C., Lingoes, J. C. 1978. Stochastic Cumulative Scales. In: S. Shye (ed.), Theory Construction and Data Analysis in the Behavioral Sciences, 280–298. San Francisco: Jossey-Bass.

Coombs, C. H., Dawes, R. M., Tversky, A. 1975. Mathematische Psychologie. Weinheim: Beltz.

Crano, W. D., Brewer, M. B. 1975. Einführung in die sozialpsychologische Forschung. Methoden und Prinzipien. Dt. Ausg. hrsg. von K. R. Scherer. Köln: Kiepenheuer & Witsch.

Cronbach, L. J. 1951. Coefficient Alpha and the Internal Structure of Tests. Psychometrika 16, 297–334.

Danziger, K. 1990. Constructing the Subject. Historical Origins of Psychological Research. Cambridge: Cambridge University Press.

Denz, H. 1989. Einführung in die empirische Sozialforschung. Wien: Springer-Verlag.

Diekmann, A. 1982. Eine additiv verbundene Messung des Sozialprestiges. Zeitschrift für Sozialpsychologie 13, 22–31.

Diekmann, A. 1995. Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek: Rowohlt.

Diez, J. A. 1997. A Hundred Years of Numbers. A Historical Introduction to Measurement Theory 1887 – 1990. Part I: The Formation Period. Studies in History and Philosophy of Science 28, 167–185. Part II: Suppes and the Mature Theory. Studies in History and Philosophy of Science 28, 237–265.

Dijkstra, T. K. 1995. Pyrrho's Lemma, or Have it Your Way. Metrika 42, 119–125.

Dixon, B. R., Bouma, G. D., Atkinson, G. B. J. 1987. A Handbook of Social Science Research. Oxford: Oxford University Press.

Duhem, P. 1908. Ziel und Struktur der physikalischen Theorien. Hamburg: Felix Meiner 1978.

Duncan, O. D. 1984. Notes on Social Measurement. Historical and Critical. New York: Russell Sage Foundation.

Dunn-Rankin, P. 1983. Scaling Methods. Hillsdale: Lawrence Erlbaum.

Durkheim, E. 1984. Die Regeln der soziologischen Methode (1895). Hrsg. von R. König. Frankfurt: Suhrkamp.

Esser, H. 1993. Soziologie. Allgemeine Grundlagen. Frankfurt: Campus.

Eysenck, H. J. 1956. Wege und Abwege der Psychologie. Reinbek: Rowohlt.

Fahrmeir, L., Künstler, R., Pigeot, I., Tutz, G. 1999. Statistik. Der Weg zur Datenanalyse. Berlin: Springer-Verlag.

Feger, H. 1982. Multidimensional Scaling of Attitudes: Intra- and Interindividual Variations in Preferences and Cognitions. In: B. Wegener (ed.), Social Attitudes and Psychophysical Measurement, 219–235. Hillsdale: Lawrence Erlbaum.

Feger, H., Graumann, C. F. 1983. Beobachtung und Beschreibung von Erleben und Verhalten. In: H. Feger, J. Bredenkamp (Hg.), Datenerhebung, 76–134, Göttingen: Verlag für Psychologie Hogrefe.

Fischer, G. H. 1974. Einführung in die Theorie psychologischer Tests. Bern: Huber.

Fishbein, M. 1967. Einstellung und die Vorhersage des Verhaltens. Dt. Übers. in: S. E. Hormuth (Hg.), Sozialpsychologie der Einstellungsänderung, 148–173. Königstein: Athenäum-Hain-Scriptor-Hanstein 1979.

Fisher, R. A. 1922. On the Mathematical Foundations of Theoretical Statistics. Philosophical Transactions of the Royal Society of London. Series A, Vol. 222, 309–368.

Fowler, F. J. 1988. Survey Research Methods. Newbury Park: Sage.

Friedrichs, J. 1974. Situation als soziologische Erhebungseinheit. Zeitschrift für Soziologie 3, 44–53.

Friedrichs, J. 1990. Methoden empirischer Sozialforschung. 14. Auflage. Opladen: Westdeutscher Verlag.

Fukunaga, K. 1982. Intrinsic Dimensionality Extraction. In: P. R. Krishnaiah, L. N. Kanal (eds.), Handbook of Statistics 2, 347–360. Amsterdam: North-Holland.

Fürstenberg, F. 1995. Soziale Handlungsfelder. Strukturen und Orientierungen. Opladen: Leske und Budrich.

Galton, F. 1889. Natural Inheritance. London: Macmillan.

Gigerenzer, G. 1981. Messung und Modellbildung in der Psychologie. München: Reinhardt.

Gosztonyi, A. 1976. Der Raum. Geschichte seiner Probleme in Philosophie und Wissenschaften. Freiburg: Alber.

Groeben, N., Westmeyer, H. 1975. Kriterien psychologischer Forschung. München: Juventa.

Guilford, J. P. 1954. Psychometric Methods. New York: McGraw-Hill.

Gulliksen, H. 1946. Paired Comparisons and the Logic of Measurement. Psychological Review 53, 199–213.

Gustafsson, A., Herrmann, A., Huber, F. 2000. Conjoint Analysis as an Instrument of Market Research Practice. In: dies. (eds.), Conjoint Measurement. Methods and Applications, 5–45. Berlin: Springer-Verlag.

Guttman, L. 1944. A Basis for Scaling Qualitative Data. American Sociological Review 9, 139–150.

Guttman, L. 1950. The Basis for Scalogram Analysis. In: S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, J. A. Clausen, Measurement and Prediction, 60–90. Princeton: Princeton University Press.

Habermehl, W. 1992. Angewandte Sozialforschung. München: Oldenbourg.

Hacking, I. 1983. Representing and Intervening. Cambridge: Cambridge University Press.

Halmos, P. R. 1950. Measure Theory. London: van Nostrand.

Hambleton, R. K., Swaminathan, H., Rogers, H. J. 1991. Fundamentals of Item Response Theory. Newbury Park: Sage.

Hamblin, R. L. 1974. Social Attitudes: Magnitude Measurement and Theory. In: H. M. Blalock (ed.), Measurement in the Social Sciences, 61–120. London: Macmillan.

Heidenreich, K. 1984. Die Verwendung standardisierter Tests. In: E. Roth, K. Heidenreich (Hg.), Sozialwissenschaftliche Methoden. Lehr- und Handbuch für Forschung und Praxis, 399–416. München: Oldenbourg.

Helmholtz, H. von 1887. Zählen und Messen erkenntnistheoretisch betrachtet. Reprint: Darmstadt: Wissenschaftliche Buchgesellschaft 1959.

Herkner, W. 1991. Lehrbuch Sozialpsychologie. Bern: Huber.

Holm, K. 1970. Zuverlässigkeit von Skalen und Indizes. Kölner Zeitschrift für Soziologie und Sozialpsychologie 22, 356–385.

Holm, K. 1974. Theorie der Fragebatterie. Kölner Zeitschrift für Soziologie und Sozialpsychologie 26, 316–341.

Holm, K. 1975. Zweck und Verlauf einer Befragung. In: K. Holm (Hg.), Die Befragung, Band 1, 9–13. München: Francke.

International Statistical Institute 1986. Declaration on Professional Ethics. International Statistical Review 54, 227–242.

Jacob, R., Eirmbter, W. H. 2000. Allgemeine Bevölkerungsumfragen. München: Oldenbourg.

Jakoby, N., Jacob, R. 1999. Messung von internen und externen Kontrollüberzeugungen in allgemeinen Bevölkerungsumfragen. ZUMA-Nachrichten, Jg. 23, Nr. 45, 61–71.

Jevons, W. S. 1958. The Principles of Science. Reprint of 2nd ed. 1877. New York: Dover.

Jöckel, K.-H., Babitsch, B., Bellach, B.-M., Bloomfield, K., Hoffmeyer-Zlotnik, J., Winkler, J., Wolf, C. 1998. Messung und Quantifizierung soziodemographischer Merkmale in epidemiologischen Studien. In: W. Ahrens, B.-M Bellach, K.-H Jöckel (Hg.), Messung soziodemographischer Merkmale in der Epidemiologie, 7–38. München: MMV Medizin Verlag.

Johnson, H. M. 1936. Pseudo-Mathematics in the Mental and Social Sciences. American Journal of Psychology 48, 342–351.

Kaplan, A. 1964. The Conduct of Inquiry. Methodology for Behavioral Science. Scranton: Chandler Publ.

Kelly, J. S. 1988. Social Choice Theory. New York: Springer-Verlag.

Kemeny, J. G. 1959. Mathematics without Numbers. Daedalus 88, 577–591. Reprint in: D. Lerner (ed.), Quantity and Quality, 35–52. New York: Free Press of Glencoe 1961.

Kemeny, J. G., Snell, J. L. 1962. Mathematical Models in the Social Sciences. London: Blaisdell.

Kempski, J. v. 1952. Zur Logik der Ordnungsbegriffe, besonders in den Sozialwissenschaften. Studium Generale, Bd. 5. Reprint in H. Albert (Hg.), Theorie und Realität, 115–137. Tübingen: Mohr 1972.

Kendall, M. G. 1938. A New Measure of Rank Correlation. Biometrika 30, 81–93.

Kendall, M. G. 1962. Rank Correlation Methods (3rd ed.). London: Griffin.

Kerlinger, F. N. 1964. Foundations of Behavioral Research. Educational and Psychological Inquiry. New York: Holt, Rinehart and Winston.

Kline, M. 1980. Mathematics. The Loss of Certainty. Oxford: Oxford University Press.

Költringer, R. 1993. Gültigkeit von Umfragedaten. Köln: Böhlau.

Kromrey, H. 1994. Empirische Sozialforschung. 6. Aufl. Opladen: Leske und Budrich.

Krug, W., Nourney, M., Schmidt, J. 1999. Wirtschafts- und Sozialstatistik. Gewinnung von Daten. 5. Aufl. München: Oldenbourg.

Kruskal, J. B. 1964a. Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. Psychometrika 29, 1–27.

Kruskal, J. B. 1964b. Nonmetric Multidimensional Scaling: A Numerical Method. Psychometrika 29, 115–129.

Kyburg, H. E. 1992. Measuring Errors of Measurement. In: C. W. Savage, P. Ehrlich (eds.), Philosophical and Foundational Issues in Measurement Theory, 75–91. Hillsdale: Lawrence Erlbaum.

Laatz, W. 1993. Empirische Methoden. Thun/Frankfurt: Harri Deutsch.

Lazarsfeld, P. F. 1993. On Social Research and Its Language. Ed. by R. Boudon. Chicago: University of Chicago Press.

Lazarsfeld, P. F, Henry, N. W. 1968. Latent Structure Analysis. Boston: Houghton Mifflin Comp.

Leeuw, J. de 1977. Correctness of Kruskal's Algorithms for Monotone Regression with Ties. Psychometrika 42, 141–144.

Lewis-Beck, M. S. 1995. Data Analysis. Thousand Oaks: Sage.

Lienert, G. A., Raatz, U. 1998. Testaufbau und Testanalyse. 6. Aufl. Weinheim: Psychologie Verlags Union.

Likert, R. 1932. A Technique for the Measurement of Attitudes. Archives of Psychology 140. New York.

Lipset, S. M. 1958. A Sociologist Looks at History. Dt. Übers.: Bemerkungen zum Verhältnis von Soziologie und Geschichtswissenschaft. In: E. Topitsch (Hg.), Logik der Sozialwissenschaften, 477–486. Köln: Kiepenheuer & Witsch 1972.

Lodge, M. 1982. The Social-Psychophysical Scaling of Political Opinion. In: B. Wegener (ed.), Social Attitudes and Psychophysical Measurement, 177–198. Hillsdale: Lawrence Erlbaum.

Loevinger, J. 1948. The Technique of Homogeneous Tests Compared with Some Aspects of 'Scale Analysis' and Factor Analysis. Psychological Bulletin 45, 507–529.

Lord, F. M., Novick, M. R. 1968. Statistical Theories of Mental Test Scores. Reading: Addison-Wesley.

Luce, R. D., Tukey, J. W. 1964. Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. Journal of Mathematical Psychology 1, 1–27.

Mardia, K. V., Kent, J. T., Bibby, J. M. 1979. Multivariate Analysis. New York: Academic Press.

Marquardt, M., Döring, A., Stieber, J., Filipiak, B., Löwel, H., Holle, R. 1998. Inkonsistenzen von Probandenangaben bei Wiederholungsbefragungen in Gesundheitssurveys. In: W. Ahrens, B.-M Bellach, K.-H Jöckel (Hg.), Messung soziodemographischer Merkmale in der Epidemiologie, 87–94. München: MMV Medizin Verlag.

Matthes, J. 1993. Über das Erfahren von Erfahrung. In: H. J. Schneider, R. Inhetveen (Hg.), Enteignen uns die Wissenschaften? Zum Verhältnis zwischen Erfahrung und Empirie, 101–123. München: Fink-Verlag.

Mayntz, R., Holm, K., Hübner, P. 1969. Einführung in die Methoden der empirischen Soziologie. Opladen: Westdeutscher Verlag.

McCutcheon, A. L. 1987. Latent Class Analysis. London: Sage.

McGaw, D., Watson, G. 1976. Political and Social Inquiry. New York: Wiley.

McIver, J. P., Carmines, E. G. 1994. Unidimensional Scaling. In: M. S. Lewis-Beck (ed.), Basic Measurement, 139–228. London: Sage.

McKennell, A. C. 1977. Attitude Scale Construction. In: C. A. O'Muircheartaigh, C. Payne (eds.), The Analysis of Survey Data, Vol. 1, 183–220. New York: Wiley.

Mehrens, W. A., Lehmann, I. J. 1975. Measurement and Evaluation in Education and Psychology. 2nd ed. New York: Holt, Rinehart and Winston.

Meschkowski, H. 1997. Mathematik verständlich dargestellt. Eine Entwicklungsgeschichte von der Antike bis zur Gegenwart. Wiesbaden: VMA.

Michel, L., Conrad, W. 1982. Theoretische Grundlagen psychometrischer Tests. In: K.-J. Groffmann, L. Michel (Hg.), Grundlagen psychologischer Diagnostik, 1–129. Göttingen: Verlag für Psychologie Hogrefe.

Micko, H. C. 1970. Eine Verallgemeinerung des Meßmodells von Rasch mit einer Anwendung auf die Psychophysik der Reaktionen. Psychologische Beiträge 12, 4–22.

Mokken, R. J. 1971. A Theory and Procedure of Scale Analysis. The Hague: Mouton.

Mueller, J. H., Schuessler, K. F., Costner, H. L. 1970. Statistical Reasoning in Sociology. 2nd ed. New York: Houghton Mifflin.

Müller, P. J. (Hg.) 1977. Die Analyse prozeß-produzierter Daten. Stuttgart: Klett-Cotta.

Müller, W. 1979. Schulbildung und Weiterbildung als soziologische Hintergrundvariablen. In: F. A. Pappi (Hg.), Sozialstrukturanalysen mit Umfragedaten, 169–206. Königstein: Athenäum.

Nachmias, C., Nachmias D. 1981. Research Methods in the Social Sciences. 2nd ed. London: Edward Arnold Publ.

Narens, L., Luce, R. D. 1986. Measurement: The Theory of Numerical Assignments. Psychological Bulletin 99, 166–180.

Neubauer, W. 1996. Preisstatistik. München: Vahlen.

Neumann, K., Morlock, M. 1993. Operations Research. München: Hanser.

Noelle-Neumann, E., Petersen, T. 2000. Alle, nicht jeder. Einführung in die Methoden der Demoskopie. 3. Aufl. Berlin: Springer-Verlag.

Norpoth, H. 1979. Dimensionen des Parteienkonflikts und Präferenzordnungen der deutschen Wählerschaft: Eine Unfoldinganalyse. Zeitschrift für Sozialpsychologie 10, 350–362.

Orth, B. 1974. Einführung in die Theorie des Messens. Stuttgart: Kohlhammer.

Ost, F. 1984. Faktorenanalyse. In: L. Fahrmeir, A. Hamerle (Hg.), Multivariate statistische Verfahren, 575–632. Berlin: de Gruyter.

Rasch, G. 1980. Probabilistic Models for Some Intelligence and Attainment Tests. Reprint of 1961 edition. Chicago: University of Chicago Press.

Rinne, H. 1996. Wirtschafts- und Bevölkerungsstatistik. 2. Aufl. München: Oldenbourg.

Roche, J. J. 1998. The Mathematics of Measurement. A Critical History. London: Athlone Press.

Rohwer, G., Pötter, U. 2001. Grundzüge der sozialwissenschaftlichen Statistik. Weinheim: Juventa.

Roskam, E. E. 1996. Latent-Trait-Modelle. In: E. Erdfelder, R. Mausfeld, T. Meiser, G. Rudinger (Hg.), Handbuch Quantitative Methoden, 431–458. Weinheim: Beltz.

Rost, J. 1988. Quantitative und qualitative probabilistische Testtheorie. Bern: Huber.

Rost, J. 1996. Testtheorie und Testkonstruktion. Bern: Huber.

Rost, J., Georg, W. 1991. Alternative Skalierungsmöglichkeiten zur klassischen Testtheorie am Beispiel der Skala „Jugendzentrismus". ZA-Information 28, 52–74.

Sacks, P. 1999. Standardized Minds. The High Price of America's Testing Culture and What We Can Do to Change It. Cambridge: Perseus Publishing.

Sagoroff, S. 1973. Wissenschaft und Statistik. Das statistische Denken in den empirischen Wissenschaften. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

SAS Institute 1985. SAS/STAT User's Guide. Version 6. Cary, NC: SAS Institute Inc.

Savage, C. W., Ehrlich, P. (eds.) 1992. Philosophical and Foundational Issues in Measurement Theory. Hillsdale: Lawrence Erlbaum.

Scheuch, E. K. 1973. Das Interview in der Sozialforschung. In: R. König (Hg.), Handbuch der empirischen Sozialforschung, Band 2. Stuttgart: Enke.

Scheuch, E. K. 1999. Die Entwicklung der Umfrageforschung in der Bundesrepublik Deutschland in den siebziger und achtziger Jahren. ZUMA-Nachrichten Jg. 23, Nr. 45, 7–22.

Scheuch, E. K., Daheim, H. 1961. Sozialprestige und soziale Schichtung. In: D. V. Glass, R. König (ed.), Soziale Schichtung und soziale Mobilität, 65–103 (= Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft 5). Opladen: Westdeutscher Verlag.

Scheuch, E. K., Rüschemeyer, D. 1956. Soziologie und Statistik. Kölner Zeitschrift für Soziologie und Sozialpsychologie 8, 272–291.

Schneider, H. J. 1993. Zur Einführung: Der Begriff der Erfahrung und die Wissenschaften vom Menschen. In: H. J. Schneider, R. Inhetveen (Hg.), Enteignen uns die Wissenschaften? Zum Verhältnis zwischen Erfahrung und Empirie, 7–27. München: Fink-Verlag.

Schneider, H. J. 1997. „Den Zustand meiner Seele beschreiben" – Bericht oder Diskurs? In: W. R. Köhler (Hg.), Davidsons Philosophie des Mentalen, 33–52. Paderborn: Schöningh.

Schnell, R., Hill, P. B., Esser, E. 1992. Methoden der empirischen Sozialforschung. 3. Aufl. München: Oldenbourg.

Schnell, R., Hill, P. B., Esser, E. 1999. Methoden der empirischen Sozialforschung. 6. Aufl. München: Oldenbourg.

Schönemann, P. H. 1994. Measurement: The Reasonable Ineffectiveness of Mathematics in the Social Sciences. In: I. Borg, P. P. Mohler (eds.), Trends and Perspectives in Empirical Social Research, S. 149–160. Berlin: de Gruyter.

Schumann, S. 1999. Repräsentative Umfrage. Praxisorientierte Einführung in empirische Methoden und statistische Analyseverfahren. München: Oldenbourg.

Searle, S. R. 1982. Matrix Algebra Useful for Statistics. New York: Wiley.

Sirkin, R. M. 1995. Statistics for the Social Sciences. Thousand Oaks: Sage.

Sixtl, F. 1976. Skalierungsverfahren: Grundzüge und ausgewählte Methoden sozialwissenschaftlichen Messens. In: K. Holm (Hg.), Die Befragung, Band 4, 9–95. München: Francke.

Sjöberg, L. 1982. Beliefs and Values as Attitude Components. In: B. Wegener (ed.), Social Attitudes and Psychophysical Measurement, 199–217. Hillsdale: Lawrence Erlbaum.

Spearman, C. 1904. 'General Intelligence', Objectively Determined and Measured. American Journal of Psychology 15, 201–293.

Spector, P. E. 1992. Summated Rating Scale Construction. London: Sage.

SPSS Inc. 1990. SPSS Categories. Chicago: SPSS Inc.

Staufenbiel, T. 1992. Eindimensionale Skalogramm-Analyse. Frankfurt: Peter Lang.

Stevens, S. S. 1946. On the Theory of Scales of Measurement. Science 103 (No. 2684), 677–680.

Stouffer, S. A., Guttman, L., Suchman, E. A., Lazarsfeld, P. F., Star, S. A., Clausen, J. A. 1950. Measurement and Prediction. Princeton: Princeton University Press.

Suchman, E. A. 1950. The Intensity Component in Attitude and Opinion Research. In: S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, J. A. Clausen, Measurement and Prediction, 213–276. Princeton: Princeton University Press.

Thurstone, L. L. 1928. Attitudes Can be Measured. American Journal of Sociology 33, 529–554.

Thurstone, L. L. 1931. Multiple Factor Analysis. Psychological Review 38, 406–427.

Thurstone, L. L. 1932. The Measurement of Social Attitudes. Journal of Abnormal and Social Psychology 26, 249–269.

Thurstone, L. L. 1938. Primary Mental Abilities. Chicago: University of Chicago Press.

Thurstone, L. L. 1947. Multiple-Factor Analysis. A Development and Expansion of *The Vectors of Mind*. Chicago: University of Chicago Press.

Thurstone, L. L., Chave, E. J. 1929. The Measurement of Attitude. Chicago: University of Chicago Press.

Überla, K. 1977. Faktorenanalyse. Berlin: Springer-Verlag.

Upshaw, H. S. 1971. Attitude Measurement. In: H. M. Blalock, A. B. Blalock (eds.), Methodology in Social Research, 60–111. London: McGraw-Hill.

Ven, A. van der 1977. Einführung in die Skalierung. Bern: Huber.

Waismann, F. 1996. Einführung in das mathematische Denken. Darmstadt: Wissenschaftliche Buchgesellschaft.

Weber, M. 1976. Wirtschaft und Gesellschaft. 5. Aufl., hrsg. von J. Winckelmann. Tübingen: Mohr.

Weis, C., Steinmetz, P. 1998. Marktforschung. 3. Aufl. Ludwigshafen: Kiehl.

Winch, P. 1966. Die Idee der Sozialwissenschaft und ihr Verhältnis zur Philosophie. Frankfurt: Suhrkamp.

Wise, M. N. (ed.) 1995. The Values of Precision. Princeton: Princeton University Press.

Witte, E.H. 1989. Sozialpsychologie. Ein Lehrbuch. München: Psychologie Verlags Union.

Wittgenstein, L. 1967. Bemerkungen über die Grundlagen der Mathematik. Dt.-engl. Ausgabe, hrsg. von G.H. von Wright, R. Rhees, G.E.M. Anscombe. Oxford: Basil Blackwell.

Wolf, J. 1988. Marktforschung. Landsberg: Verlag Moderne Industrie.

Young, G., Householder, A.S. 1938. Discussion of a Set of Points in Terms of Their Mutual Distances. Psychometrika 3, 19–22.

Zajonc, R.B. 1980. Feeling and Thinking. American Psychologist 35, 151–175.

Stichwortverzeichnis