15. Kapitel

Stochastische Modelle

Rolf Steyer

" ... the true logic for this world is the calculus of probabilities ..."

J. Clerk Maxwell

In den Humanwissenschaften wie der Psychologie, Soziologie und Ökonomie verstehen sich viele Forscher als empirische Wissenschaftler. Neben dem Kriterium der ***logischen Widerspruchsfreiheit*** ist für sie die ***Erfahrung*** oder ***Empirie*** das wesentliche Korrektiv für Theorien. Die Theorien einer ***empirischen*** Wissenschaft müssen etwas über unsere ***Erfahrung*** aussagen. Nur dann sind sie an ihr überprüfbar. Die ***deduktivistisch*** oder ***falsifikationistisch*** orientierten Wissenschaftler verlangen von einer empirischen Theorie, daß aus ihr Aussagen über die Empirie ***logisch*** abgeleitet werden können (vgl. hierzu das Abgrenzungsproblem bei Papper, 1984). Aus der Sicht einer deduktivistischen Methodologie sind die ***logische Widerspruchsfreiheit*** und die ***logische Ableitbarkeit*** von Aussagen über beobachtbare Sachverhalte die beiden wichtigsten Kriterien, denen eine empirische Theorie genügen sollte. Alle weiteren Kriterien dienen nur noch dazu, verschiedene empirische Theorien untereinander zu bewerten.

Sowohl um die logische Widerspruchsfreiheit überprüfen zu können, als auch um Aussagen über die Empirie ableiten zu können, muß eine Theorie in einer formalen Sprache formuliert sein oder in diese übersetzt werden können, denn nur dort sind die Regeln des logischen Schließens anwendbar (s. dazu auch Erdfelder & Bredenkamp, Kap. 14, Abschnitt 4.5, in diesem Band). Dabei ist allerdings anzumerken, daß bisher nur wenige empirische Theorien diese beiden Kriterien erfüllen, die m.E. aber dennoch als anzustrebende Ideale unverzichtbar sind, jedenfalls dann, wenn es auf Präzision ankommt. Für viele Alltagszwecke reichen natürlich auch umgangssprachlich formulierte Theorien aus.

Aufgaben stochastischer Modelle: Stochastische Modelle, **um die es in diesem** Artikel geht, genügen den o.g. beiden Kriterien. Sie sind in einer formalen

Sprache - der Sprache der Wahrscheinlichkeitstheorie - formuliert und erlauben daher die Überprüfung ihrer logischen Widerspruchsfreiheit und die Ableitung von Aussagen über die Empirie. Stochastische Modelle sind insbesondere für Anwendungen in der Psychologie und den benachbarten Sozialwissenschaften wegen des ***Meßfehlerproblems*** und allgemein des ***Problems der multiplen Determiniertheit*** geeignet: Beobachtungen und Messungen sind fehlerbehaftet und die zu erklärenden Phänomene haben viele verschiedene Ursachen, die man nur selten alle kennt, so daß deterministische Erklärungen selten möglich sind. In einer empirischen Theorie haben stochastische Modelle im wesentlichen ***zwei Hauptaufgaben*** (zur Beziehung zwischen Theorie und Modell siehe Baizer-, Moulines & Sneed, 1987); sie explizieren die Verknüpfung zwischen:

(a) empirischen und theoretischen Begriffen und damit das ***Meßmodell,***
(b) den theoretischen (bzw. empirischen) Begriffen und damit die ***Abhängigkeitsbegriffe.***

Betrachten wir als Beispiel die Hypothese ,,Frustration führt zu Aggression”! Hier kommen die beiden theoretischen Begriffe ,,Frustration“ und ,,Aggression“ vor, die mit dem Abhängigkeitsbegriff ,,führt zu“ verknüpft sind. Alle drei Begriffe haben zunächst nur umgangssprachliche Bedeutungen, die für den Alltag auch hinreichend präzise sein mögen. Dem Präzisionsanspruch einer empirischen Wissenschaft genügt die umgangssprachliche Formulierung der Frustrations-Aggressions-Hypothese jedoch nicht, da sie allzu viele Fragen offen läßt. Ist mit ihr die deterministische Aussage gemeint:

(a) Für alle Menschen gilt: wenn sie frustriert sind, reagieren sie aggressiv?

Oder ist lediglich die folgende probabilistische Aussage gemeint:

(b) Für alle Menschen gilt: wenn sie frustriert sind, ist die Wahrscheinlichkeit, daß sie aggressiv reagieren, erhöht?

Sind ,,Frustration“ und ,,Aggression“ qualitative, komparative oder gar metrische Begriffe? Wären es nur qualitative Begriffe, die nur die Ausprägungen „vorhanden* und „nicht vorhanden“ haben, dann wären nur die obige deterministische (a) und die probabilistische Präzisierung (b) der Frustrations-Aggressions-Hypothese möglich. Handelt es sich aber um komparative oder gar metrische Begriffe, wären auch weitere Präzisierungen möglich wie z.B.:

(c) Für alle Menschen gilt: je mehr sie frustriert werden, desto stärker reagieren sie aggressiv.
(d) Für alle Menschen gilt: je mehr sie frustriert werden, desto höher ist die Wahrscheinlichkeit, daß sie aggressiv reagieren.
(e) Für alle Menschen gilt: je mehr sie frustriert werden, desto höher ist der Erwartungswert ihrer Aggressivität.

Dieses Beispiel ließe sich leicht über mehrere Seiten fortsetzen. Aber auch mit den obigen Ausführungen dürfte folgendes schon hinreichend klar geworden sein: Damit aus der Frustrations-Aggressions-Hypothese eine Hypothese einer ***deduktiven*** empirischen Wissenschaft werden kann, müssen die theoretischen Begriffe ,,Frustration= und „Aggression“, aber auch der Abhängigkeitsbegriff „führt zu“ präzisiert werden, und zwar so, daß eine Verknüpfung zwischen theoretischen Begriffen und beobachtbaren Sachverhalten hergestellt wird, die logisch-mathematische Deduktionen ermöglicht. Umgangssprachlich formulierte Hypothesen wie die Frustrations-Aggressions-Hypothese als ,,wissenschaftliche Hypothesen” zu bezeichnen (s. z.B. Hager, 1987, **1992),** halte ich für bedenklich. Zwar sind es Hypothesen, die Wissenschaftler zu einem bestimmten Zeitpunkt im Prozeß der Entwicklung ihrer Theorien haben, aber sie sind m.E. nicht das, wohin Theorienentwicklung zielen sollte. Für solche umgangssprachlich formulierte Hypothesen, die im Wissenschaftsprozeß durchaus nützlich sind, gibt es in der Regel nicht nur eine einzige, sondern viele verschiedene Möglichkeiten der Präzisierung. Daher können umgangssprachlich formulierte Hypothesen m. E. nicht das Endziel, sondern nur das Rohmaterial darstellen, aus dem nach entsprechender Bearbeitung Hypothesen konstruiert werden können, aus denen sich empirische Aussagen logisch deduzieren lassen. Sicherlich kommen Wissenschaftler bei ihrer Theorienentwicklung nicht nur mit Deduktion aus, sondern müssen auch an vielen Stellen induktive Schritte tun (s. dazu Westermann & Gerjets, Kap. 10 in diesem Band), aber dennoch sollte man nicht verwischen, wo man induktiv und wo man deduktiv arbeitet. Im letzteren Fall weiß man nämlich, wo man absolute Sicherheit über die Gültigkeit der Schlußfolgerungen hat, im ersteren Fall dagegen, fehlt diese Sicherheit. Die Unterscheidung zwischen deduktiven und induktiven Schlüssen ist also insbesondere bei der Theorienkritik und -revision von Bedeutung, da sie unsere Aufmerksamkeit auf diejenigen Stellen zu richten erlaubt, die möglicherweise falsch sein können, wohingegen andere Teile der Theorie schon aus logischen Gründen nicht falsch sein können. Im Abschnitt 3 werden wir ein Beispiel für den letzten Fall kennenlernen, nämlich die Unkorreliertheit von Fehler- und True-Score-Variablen.

Auch andere Beispiele zeigen, daß unsere Umgangssprache zwar viele Begriffe enthält, mit denen wir nichtdeterministische Abhängigkeiten beschreiben können, aber für viele Zwecke ist sie zu unpräzise. Was bedeutet z.B. die Aussage ,,Rauchen fördert Lungenkrebs“. Ist damit eine korrelative Aussage über eine Population gemeint? Wenn ja, handelt es sich um eine lineare Abhängigkeit oder gibt es bestimmte Schwellen, an denen die Wahrscheinlichkeit an Lungenkrebs zu erkranken, stärker ansteigt? Ist die Abhängigkeit in allen Teilpopulationen gleich, oder gibt es Populationen, die trotz Rauchen weniger gefährdet sind? Ist gar eine kausale Abhängigkeit gemeint und wenn ja in welchem Sinn? (Eine deterministische Abhängigkeit ist wohl auszuschließen.) Handelt es sich vielleicht gar nicht um eine Populationsaussage, sondern bezieht sie sich auf jedes Individuum? Auch hier stellt sich die Frage nach der Art der Abhängigkeit und nach eventuellen interindividuellen Differenzen.

Die Mathematik hat uns in den letzten Jahrhunderten und insbesondere seit dem Erscheinen von Kolmogoroffs (1970) Buch ,,Grundbegriffe der Wahrscheinlichkeitsrechnung” (im Jahre 1933) eine exakte Sprache, die Wahrscheinlichkeitstheorie oder Stochastik, zur Verfügung gestellt, mit der wir die Phänomene der Sozialwissenschaften und ihre Abhängigkeiten präzise beschreiben können. Der Preis dafür ist allerdings, daß wir zunächst diese formale

Sprache lernen müssen, um die Bedeutung verschiedener Arten stochastischer Abhängigkeiten verstehen zu können.

überblick: Der diesem Artikel vorgegebene Umfang verbietet es, exakte Definitionen und mathematische Sätze darzustellen. Stattdessen kann es nur darum gehen, den Leser zum Studium der Wahrscheinlichkeitstheorie und ihrer Anwendungen in der Psychologie zu motivieren. Dies hoffe ich dadurch zu erreichen, daß ich die Relevanz stochastischer Modelle zur Konstruktion empirischer wissenschaftlicher Theorien herauszuarbeiten versuche. Dies ist natürlich nicht möglich, ohne vorher die ***Bestandteile eines stochastischen Modells*** einzuführen und ihre Bedeutung bei ihrer Anwendung in der Psychologie zu erläutern. Dies umreißt den Gegenstand des ***Abschnitts 1.***

Im ***Abschnitt*** 2 geht es um verschiedene ***Arten stochastischer Abhängigkeiten.*** In ***Abschnitt 2.1*** wird die ***stochastische Abhängigkeit*** von Ereignissen und Variablen behandelt und im ***Abschnitt*** 2.2 sind es verschiedene Arten ***regressiver*** und ***korrelativer Abhängigkeit.*** Dabei wird darauf hingewiesen, daß es nicht nur Abhängigkeiten zwischen zwei Variablen oder Ereignissen gibt, sondern daß oft erst die gleichzeitige Betrachtung vieler Variablen oder Ereignisse ein angemessenes Bild der Realität ergibt. Selbst bei einer bivariaten, aber noch mehr bei einer multivariaten Betrachtung, müssen wir zwischen verschiedenen Arten stochastischer Abhängigkeit unterscheiden, die nicht nur von methodischem, sondern auch von inhaltlichem Interesse sind, da sie sozusagen einen abstrahierten Inhalt darstellen, der vielen Anwendungen gemeinsam ist. Jede Art dieser stochastischen Abhängigkeiten ist auch inhaltlich anders zu interpretieren. Im ***Abschnitt 2.3*** geht es um kausale ***regressive Abhängigkeiten.*** Das Kausalitätsproblem stellt sich z.B. dann, wenn zu vermuten ist, daß eine betrachtete stochastische Abhängigkeit einer Variablen Y von einer weiteren Variablen X nicht durch eine Ursache-Wirkungs-Beziehung zwischen diesen beiden Variablen zustande kommt, sondern dadurch, daß beide von einer oder mehreren ,,Drittvariablen“ beeinflußt werden. Eng verknüpft mit dem Kausalitätsproblem sind die verschiedenen Techniken der experimentellen Versuchsplanung, wie z.B. die randomisierte Aufteilung der Beobachtungseinheiten auf die Versuchsbedingungen. Die in vielen umgangssprachlich formulierten Theorien vorkommende ***Ceteris-paribus-Klausel*** (,,unter sonst gleichen Bedingungen gilt: ...“) ist als Versuch anzusehen, Hypothesen über kausale Abhängigkeiten zu formulieren. Als problematisch ist eine derartige Formulierung m.E. deswegen zu bewerten, weil damit meist Unmögliches - die Konstanthaltung ***aller*** Störvariablen - gefordert wird. Dies wird aber selbst im randomisierten Experiment nicht erreicht.

Die Thematik des ***Abschnitts*** 3 knüpft an der ersten der o.g. Aufgaben stochastischer Modelle an, der Verknüpfung zwischen theoretischen und empirischen Begriffen. Es wird exemplarisch gezeigt, wie man in ***stochastischen***

Meßmodellen theoretische Begriffe konstruieren kann. Dabei wird deutlich, daß man über die übliche Praxis hinausgehen kann, nach der Konstrukte nur aus Namen (wie z.B. „Intelligenz“, „Bildhaftigkeit“) bestehen, mit denen wir bestimmte Assoziationen und relativ vage Hypothesen verknüpfen. Stattdessen kann man Konstrukte mathematisch konstruieren, indem man bestimmte Annahmen über die Eigenschaften bekannter Größen einführt und daraus die Existenz des Konstrukts ableitet, dessen Skalenniveau deduziert sowie die explizite mathematische Beziehung zwischen Konstrukt und beobachtbaren Variablen angibt, woraus sich auch die genaue inhaltliche Bedeutung des Konstrukts erschließt. Es liegt auf der Hand, daß es erst mit derartigen Konstrukten möglich wird, eine deduktivistische empirische Wissenschaft zu betreiben, die auf logisch-mathematischer Argumentation und nicht nur auf den weithin üblichen Plausibilitätsüberlegungen basiert.

Damit dürfte klar geworden sein, daß es in diesem Artikel weder um einen Überblick über die verschiedenen stochastischen Modelle gehen soll, noch um einen Überblick über deren Anwendungen in den Sozialwissenschaften. Dazu sei auf einige Arbeiten verwiesen, die wenigsten für bestimmte Teilbereiche einen solchen Überblick versuchen, nämlich Austin und Wolfle (1991), Bentler (1986), Lewis (1986), Mulaik (1986), Rost und Strauß (1992) sowie Torgerson (1986).

1. *Bestandteile und Gegenstand stochastischer Modelle*

Was sind stochastische Modelle? Aus welchen Bestandteilen bestehen sie? Was ist ihr Gegenstand? Allgemein gesprochen handelt es sich dabei um Gebilde, die in der Sprache der Wahrscheinlichkeitstheorie formuliert sind, aus bestimmten Bestandteilen bestehen und bestimmte Phänomene beschreiben und erklären sollen, die in ***Zufallsexperimenten*** auftreten können. Dieser Gegenstand klingt zunächst sehr bescheiden, beinhaltet aber weit mehr, als es auf den ersten Blick erscheinen mag.

1.1 Zufallsexperiment

Alle Aussagen, die im Rahmen eines stochastischen Modells formuliert werden, wie z.B. über die Korrelation zweier Variablen oder irgendeine andere Art des Zusammenhangs zwischen Variablen oder Ereignissen, beziehen sich auf das betrachtete Zufallsexperiment und können daher nicht ohne weiteres auf andere Zufallsexperimente übertragen werden. Bei stochastischen Modellen geht es also nicht um generelle Allaussagen, sondern lediglich um die Gesetzmäßigkeiten in dem jeweils betrachteten Zufallsexperiment. Das zugrundeliegende Zufallsexperiment ist der „Geltungsbereich“ oder das „empirische Phänomen“, von dem in stochastischen Aussagen die Rede ist. Dies schließt natürlich nicht aus, daß in komplexeren Theorien generelle Aussagen über Klassen von stochastischen Modellen gemacht werden, was zu einem entsprechend größeren Geltungsbereich der Theorie führt.

Dies ist in den Sozialwissenschaften nicht anders als in den klassischen Naturwissenschaften, allerdings sind dort die Randbedingungen, die ein Zufallsexperiment definieren, oft überdauernder als in den Sozialwissenschaften. Die Masseverteilung in einem Holzwürfel ist bspw. über viele Jahrzehnte konstant und kann sich eigentlich nur durch die leicht zu diagnostizierenden Aktivitäten eines Holzwurms verändern. Das Zufallsexperiment „Werfen des Würfels“ kann daher nahezu beliebig oft wiederholt werden. In den Sozialwissenschaften dagegen können manche Zufallsexperimente überhaupt nicht wiederholt werden. Das Zufallsexperiment „Lösen einer bestimmten Intelligenztestaufgabe durch Fritz Müller“ kann wegen des zu erwartenden Lerneffekts unter gleichen Bedingungen nicht wiederholt werden. Dennoch ist es auch hier sinnvoll, ein Zufallsexperiment zu unterstellen und Modelle darüber zu formulieren, wie bspw. die Lösungswahrscheinlichkeit von der Fähigkeit der Person und der Schwierigkeit der Aufgabe abhängt (s. dazu das Rasch-Modell z.B. bei de Gruijter & van der Kamp, **1984;** Fischer, **1974; 1981, 1983, 1988;** Kubinger, **1988;** Rasch, **1980;** Rost, **1988;** Steyer & Eid, **1993;** Tutz, **1989).**

Stochastische Modelle haben es zwar oft, aber nicht immer mit Massenphänomenen zu tun, die in großen Stichproben auftreten. Wenn wir bspw. eine Vorhersage der Wahrscheinlichkeit benötigen, mit der Fritz Müller eine bestimmte Aufgabe löst, so können wir dies auch dann tun, wenn diese Aufgabe nur ein einziges Mal zu lösen ist. Jede Entscheidung (z.B. im beruflichen Kontext), eine bestimmte Person für eine bestimmte Aufgabe auszuwählen, beruht letztlich auf der Abschätzung einer solchen Wahrscheinlichkeit. Natürlich benötigen wir dazu Informationen über die Fähigkeit der betreffenden Person und die Aufgabenschwierigkeit, die z.B. aus einem ähnlichen Zufallsexperiment stammen können. Mit stochastischen Modellen erhebt man also keinen Allgemeinheitsanspruch, sondern bezieht sich prinzipiell auf ein bestimmtes, noch durchzuführendes Zufallsexperiment, das auch nicht unbedingt wiederholbar sein muß. Wie bereits erwähnt, spricht jedoch nichts dagegen, stochastische Modelle in umfassendere Theorien einzubetten, für die dann ein allgemeinerer Geltungsanspruch erhoben werden kann.

Beispiel **1:** Ein Zufallsexperiment liegt bspw. vor, wenn wir eine Person aus einer Menge von Personen (einer Grundgesamtheit oder ***Population)*** ziehen, einen psychologischen Test bearbeiten lassen und ihr Antwortmuster registrieren. Diese Art von Zufallsexperimenten wird bei vielen stochastischen Meßmodellen zugrundegelegt (s. ausführlicher Steyer & Eid, 1993). Das Ziehen einer Stichprobe vom Umfang N besteht aus dem N-maligen Wiederholen dieses Zufallsexperiments, wobei allerdings die Grundgesamtheit mit jedem Ziehen um die bereits gezogene Person kleiner wird.

Beispiel 2: Bei anderen Arten von Zufallsexperimenten wird nur das Antwortmuster ***einer*** gegebenen ***Person*** als Ergebnis des Zufallsexperiments betrachtet (s. z.B. Rost, 1988). Bei den einfachsten Meßmodellen wird das Zustandekommen eines Testwerts durch das Zusammenwirken der zu messenden Eigenschaft und eines Meßfehlers erklärt. Bei komplexeren Meßmodellen können die Schwierigkeit des Meßinstruments (des Tests oder eines einzelnen Items; s. z.B. Rasch, 1980), situationale Effekte und zusätzliche Personeigenschaften als weitere Variablen hinzugezogen werden (s. z.B. Steyer, Ferring & Schmitt, 1992), um das Zustandekommen eines Testwerts zu erklären.

Beispiel 3: Ein anderes, in der Psychologie typisches Zufallsexperiment, besteht aus dem Ziehen einer Person aus einer Grundgesamtheit, ihrer Zuweisung zu einer von mehreren experimentellen Bedingungen und der Registrierung ihres Werts auf einer oder mehreren Kriteriumsvariablen. Neben Meßfehlern und anderen nicht bekannten Störvariablen erklären in solchen Modellen die experimentellen Bedingungen das Zustandekommen der Werte der Kriteriumsvariablen (s. z.B. Kap. 10 von Steyer, 1992).

1.2 Wahrscheinlichkeitsraum

Notwendiger Bestandteil jedes stochastischen Modells ist ein Wahrscheinlichkeitsraum, der die (formal)sprachliche Repräsentation des betrachteten Zufallsexperiments und damit des betrachteten empirischen Phänomens darstellt. Ein Wahrscheinlichkeitsraum besteht aus drei Komponenten:

(a) der Menge der möglichen Ergebnisse (des betrachteten Zufallsexperiments),
(b) der Menge der möglichen Ereignisse,
(c) dem Wahrscheinlichkeitsmaß.

Die ***Menge der möglichen Ergebnisse*** beschreibt die formale Struktur des empirischen Phänomens. Die ***Menge der möglichen Ereignisse*** gibt an, von welchen Ereignissen man sprechen können will, und das ***Wahrscheinlichkeitsmaß*** ist eine Funktion, die jedem möglichen Ereignis eine Wahrscheinlichkeit zuordnet. In Anwendungen geschieht diese Zuordnung der Wahrscheinlichkeiten meist nicht explizit, da diese Wahrscheinlichkeiten gar nicht bekannt sind. Wenn man von ***der*** Wahrscheinlichkeit eines Ereignisses spricht, wird aber bereits vorausgesetzt, daß diese Zuordnung existiert.

Ich verwende hier die etwas umständlichen (und auch unüblichen) Bezeichnungen ***„mögliche*** Ergebnisse" und „mögliche Ereignisse", um damit die Unterschiede zwischen einem in einem bereits durchgeführten Zufallsexperiment ***aufgetretenen*** Ergebnis bzw. Ereignis einerseits und einem ***möglichen*** Ergebnis bzw. Ereignis in einem (betrachteten, noch durchzuführenden) Zufallsexperiment andererseits hervorzuheben. Ziel stochastischer Modelle ist nämlich nicht in erster Linie die Beschreibung der Systematik bereits beobachteter Ereignisse, sondern die Angabe der Gesetzmäßigkeiten, die das Zufallsexperiment und die dabei ***möglichen*** Ereignisse charakterisieren. Indirekt werden damit natürlich auch die in einem durchgeführten Experiment ***tatsächlich*** aufgetretenen Ereignisse erklärt.

Menge der (möglichen) Ergebnisse: Zur Illustration betrachten wir das oben bereits angeführte Zufallsexperiment des Ziehens einer Person u aus einer Menge U (Grundgesamtheit) und des Registrierens ihres Testergebnisses, das z.B. aus den möglichen Kombinationen des Lösens (+) oder Nichtlösens (-) von zehn Aufgaben bestehen möge. Ein mögliches Ergebnis des betrachteten Zufallsexperiment wäre dann z.B.: (Fritz Meier, +, +, -, +, -, -, -, -, +, +). Demnach wurde also Fritz Meier gezogen und dieser löste die Aufgaben 1, 2, 4, 9, 10 und die anderen fünf Aufgaben nicht. Die Menge Ω der möglichen Ergebnisse dieses Zufallsexperiments ist dann ein Kreuzprodukt der Form $\Omega = U \times \boldsymbol{M}$, wobei $\boldsymbol{M} := \{+, -\}^{10} = \{+, -\} \times \ldots \times \{+, -\}$ (10 Faktormengen).

Ziel eines stochastischen Modells in einem derartigen Zufallsexperiment könnte z.B. sein zu beschreiben, wie die ***Lösungswahrscheinlichkeit*** einer Aufgabe von ihrer ***Schwierigkeit*** und der ***Fähigkeit*** der gezogenen Person abhängt. Darüber hinaus erlaubt ein solches Modell erst die ***Konstruktion der Schwierigkeits- und Fähigkeitsbegriffe,*** die für die Lösung der betrachteten Aufgaben relevant sind. Beide Begriffe beziehen sich zunächst nur auf das betrachtete Zufallsexperiment.

Menge der (möglichen) Ereignisse: Ereignisse in einem solchen Zufallsexperiment sind Teilmengen von Ω. Beim oben beschriebenen Zufallsexperiment ist das Ereignis, Fritz Meier zu ziehen, (Fritz Meier) $\times$ ***M.*** Das Ereignis, Fritz Meier ***oder*** Franz Müller zu ziehen, ist {Fritz Meier] $\times$ ***M*** U [Franz Müller} $\times$ ***M,*** und das Ereignis, daß die gezogene Person (wer auch immer das ist) die erste Aufgabe löst, ist $U \times \{+\} \times \{+, -\}^9$. In jedem Fall ist das aufgeführte Ereignis eine Teilmenge von Ω.

Die möglichen Ereignisse kann man wieder zu einer Menge zusammenfassen, z.B. der Menge aller Teilmengen von Ω, der ***Potenzmenge*** von Ω. Man muß jedoch nicht immer alle Teilmengen von Ω als mögliche Ereignisse betrachten. Wahrscheinlichkeiten können auch dann schon sinnvoll definiert werden, wenn man nur eine Teilmenge der Potenzmenge von Ω betrachtet, die man als σ-*Algebra* bezeichnet und mit Zeichen wie z.B. $\mathcal{A}$ notiert (zur Definition einer σ-Algebra s. Bauer, 1978, S. 16). Eine σ-Algebra - und daher jede Ereignismenge - ist abgeschlossen gegenüber abzählbaren Vereinigung und Schnittmengenbildungen, d.h. die Vereinigungs- und die Schnittmenge von Elementen aus $\mathcal{A}$ sind ebenfalls Elemente aus $\mathcal{A}$.

Wahrscheinlichkeitsmaß: Jedem (möglichen) Ereignis ***A*** ist durch das Wahrscheinlichkeitsmaß (W-Maß) ***P*** eine Wahrscheinlichkeit ***P(A)*** zugeordnet. In den meisten Anwendungen sind diese Wahrscheinlichkeiten unbekannt. Dennoch muß natürlich vorausgesetzt werden, daß diese Wahrscheinlichkeiten existieren. Empirische Untersuchungen dienen in der Regel dazu, diese Wahrscheinlichkeiten zu schätzen oder allgemeiner formuliert, einige Aussagen über diese Wahrscheinlichkeiten machen zu können.

Die Wahrscheinlichkeit eines Ereignisses soll eine Zahl zwischen 0 und 1 (einschließlich) sein. Eine weitere wichtige definierende Eigenschaft ist die Additivität eines Wahrscheinlichkeitsmaßes, d.h. die Eigenschaft

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots , \qquad (1.1)$$

falls diese Ereignisse ***paarweise disjunkt*** sind, falls also für jedes Paar dieser Ereignisse gilt: $A_i \cap A_j = 0$, falls $i \neq j$. Wenn die Ereignisse paarweise disjunkt sind, dann addieren sich also ihre Einzelwahrscheinlichkeiten zur Wahrscheinlichkeit dafür, daß mindestens eines dieser Ereignisse eintritt. Man beachte auch, daß es erst mit der Einführung eines Wahrscheinlichkeitsmaßes sinnvoll wird, von Ereignissen etc. zu sprechen. Vorher handelt es sich nur um Teilmengen der zugrundegelegten Menge Ω.

Durch die beiden Mengen Ω, $\mathcal{A}$ und durch das W-Maß P, d.h. durch den Wahrscheinlichkeitsraum (W-Raum) <~2, $\mathcal{A}$, P>, ist ein Zufallsexperiment beschreibbar. Damit stellt der W-Raum die formalsprachliche Repräsentation des in einem stochastischen Modell betrachteten empirischen Phänomens dar. In einem solchen W-Raum stecken prinzipiell alle Informationen und alle Aussagen, die man über ein betrachtetes Zufallsexperiment formulieren kann. In Ω sind alle möglichen Ergebnisse aufgeführt, die bei diesem Zufallsexperiment auftreten können, in $\mathcal{A}$ sind alle Ereignisse angegeben, von denen man in diesem Kontext sprechen kann und mit dem W-Maß P liegen die Wahrscheinlichkeiten aller (möglichen) Ereignisse fest, auch wenn sie in der Regel unbekannt sind. Damit liegt auch fest, wie diese Ereignisse voneinander abhängen, da die Schnittmengen von Ereignissen auch Ereignisse sind (s. dazu untendie Definition der stochastischen Abhängigkeit von Ereignissen).

1.3 Zufallsvariablen

Zufallsvariablen (Zvn) oder, synonym, ***stochastische Variablen,*** ordnen jedem Ergebnis der Ergebnismenge Ω einen Wert zu (z.B. dem Ergebnis „Augenzahl eins" beim Werfen eines Würfels den Wert 1). Zufallsvariablen sind also Abbildungen mit dem Definitionsbereich Ω. Die zugeordneten Werte können Zahlen, aber auch Elemente beliebiger anderer Mengen sein (s. dazu Bauer, 1978, S. 136). Sind die Werte der Zufallsvariablen reelle Zahlen, dann sind unter bestimmten Voraussetzungen ***Erwartungswerte*** (theoretische Mittelwerte), ***Varianzen, Kovarianzen*** und ***Korrelationen*** definiert. Diese Größen kennzeichnen bestimmte Eigenschaften der ***Verteilung*** bzw. der gemeinsamen Verteilung von Zufallsvariablen. (Zur exakten Definition dieser Begriffe sei auf Bauer, 1978, oder Steyer & Eid, 1993, verwiesen.) Mit solchen numerischen Zufallsvariablen kann man quantitative Gesetze formulieren. Dadurch lassen sich empirische Phänomene sehr viel einfacher beschreiben, als dies durch ausschließliche Verwendung von Ereignissen möglich wäre.

Zufallsvariablen bilden die Ergebnisse $\omega \in \Omega$ eines Zufallsexperimentes nach einer ***festen*** Zuordnungsvorschrift ab. Die Ergebnisse $\omega \in \Omega$ sind zufällig und damit indirekt auch die Werte einer Zufallsvariablen. Zufallsvariablen können sehr systematisch voneinander abhängen, auch wenn diese Systematik in der Regel nicht deterministisch ist. So kann man bspw. die Körpergröße (X) und das Geschlecht (Y) als Zufallsvariablen in einem Zufallsexperiment einführen, das aus dem Ziehen einer Person aus einer Population und dem Registrieren des X- und Y-Wertes besteht. Die beiden Variablen X und Y werden sicherlich nicht unabhängig sein, sondern in einer bestimmten Stärke miteinander in Zusammenhang stehen.

Bei Ereignissen haben wir zwischen einem ***möglichen*** Ereignis und einem ***tatsächlich aufgetretenen*** Ereignis unterschieden. Die analoge Unterscheidung bei Zufallsvariablen ist die zwischen der Zufallsvariablen selbst und ihrem tatsächlich aufgetretenen Wert bzw. ihrer Realisierung.

1.4 Zusammenfassende Bemerkungen

Natürlich konnten hier nur einige grundlegende stochastische Begriffe eingeführt werden. Vollständigere Einführungen, die auch für empirische Wissenschaftler geeignet sind, findet man z.B. bei Bosch (1986), Härtter (1987), Hinderer (1980), Krickeberg und Ziezold (1979), Rényi (1977) und Viertl (1990). An den mathematisch vorgebildeten Leser wenden sich Ash (1972), Bauer (1978), Breiman (1968), Gänssler und Stute (1977), Loeve (1977, 1978) und Plachky (1981).

Mit den oben behandelten drei Bestandteilen des W-Raums und den Zufallsvariablen sind die elementaren Bausteine stochastischer Modelle angegeben. Zugleich ist damit der Rahmen abgesteckt, in dem sich stochastische Modelle bewegen. Allgemein geht es also bei einem stochastischen Modell um die Beschreibung der Gesetzmäßigkeiten des Auftretens von Ereignissen und der Realisierungen von Zufallsvariablen sowie um die Beschreibung der Abhängigkeiten zwischen Ereignissen und zwischen Zufallsvariablen in einem Zufallsexperiment. Aus bereits vorliegenden Daten, die aus einem bereits durchgeführten Zufallsexperiment stammen, versucht man, bestimmte Parameter zu schätzen und Hypothesen über deren Größe zu testen. Diese Parameter sollen aber nicht primär die vorliegenden ***Daten*** oder die tatsächlich aufgetretenen Ereignisse beschreiben, sondern die Gesetzmäßigkeiten des betrachteten Zufallsexperiments charakterisieren. ***Das empirische Phänomen, um das es in erster Linie bei stochastischen Modellen geht, sind also nicht die erhobenen Da-***

ten und deren Systematisierung, sondern das ***(durch den W-Raum repräsentierte) betrachtete Zufallsexperiment und die darin herrschenden Gesetzmäßigkeiten.*** Die Daten sind - bildlich gesprochen - lediglich die Spuren, aus denen wir auf die Gesetze schließen wollen, die das Zufallsexperiment steuern. Hat man diese Gesetze erkannt, werden damit natürlich auch die Daten und deren Abhängigkeitsmuster erklärt. Stochastische Modelle unterscheiden sich daher in ihrer Zielsetzung von anderen Modellen und Verfahren (z.B. Multidimensionale Skalierung, Clusteranalyse), die oft nur der Beschreibung und Strukturierung vorliegender Daten dienen.

Zur Beschreibung der Gesetzmäßigkeiten eines Zufallsexperiments werden die (einfachen und bedingten) Wahrscheinlichkeiten von Ereignissen bzw. die Verteilungen von Zufallsvariablen mittels Aussagen über ihre Erwartungswerte, Varianzen u.ä. charakterisiert. Bei der Beschreibung der ***Abhängigkeiten*** zwischen Zufallsvariablen macht man Aussagen über ihre gemeinsamen Verteilungen, z.B. durch Aussagen über ihre Korrelation, bedingten Erwartungswerte, bedingten Varianzen etc. Verschiedene Arten von in stochastischen Modellen verwendeten Abhängigkeiten und Gesetzen sollen im folgenden Abschnitt behandelt werden.

2. Stochastische Abhängigkeitsbegriffe

Abhängigkeiten zwischen Ereignissen oder Variablen sind von zentralem Interesse in jeder empirischen Wissenschaft. Dabei ist oft nicht die Frage, ***ob*** eine Abhängigkeit zwischen zwei oder mehr Variablen besteht, sondern ***wie*** diese Abhängigkeit aussieht. Selbst auf der Ebene von Ereignissen können dabei recht komplexe Phänomene auftreten. So kann es z.B. sein, daß drei Ereignisse paarweise unabhängig sind, aber dennoch aus jeweils zwei Ereignissen eine perfekte Vorhersage des dritten möglich ist. In komplexen Systemen, die in der Psychologie und den anderen Sozialwissenschaften betrachtet werden, haben wir es mit dem ***Problem der multiplen Determiniertheit zu*** tun. Damit ist klar, daß man sich nicht auf die Betrachtung bivariater Abhängigkeiten beschränken, sondern multivariate Zusammenhänge betrachten sollte. Die Behaltensleistung in einer Gedächtnisaufgabe bspw. hängt nicht nur von der Gedächtnisstärke der betrachteten Person ab, sondern auch von der Bildhaftigkeit des Lernmaterials, seiner Bedeutungshaltigkeit, dem Grad der Aktiviertheit der Person, der Stärke ihrer Motiviertheit, bei dieser Aufgabe eine gute Leistung zu erbringen, dem Grad ihrer Müdigkeit, ihrem Trainingsstand, der Art der vorherigen Tätigkeit (Transfereffekte) etc. Die verschiedenen Arten stochastischer Abhängigkeiten zwischen Ereignissen und zwischen Zufallsvariablen, um die es in diesem Abschnitt geht, sind daher von zentraler

Bedeutung für stochastische Modelle, die empirische Phänomene in den Sozialwissenschaften beschreiben und erklären sollen.

2.1 Stochastische Abhängigkeit

Der Begriff der stochastischen Abhängigkeit ist sowohl für Ereignisse als auch für Zufallsvariablen definiert. Wir betrachten zunächst die stochastische Abhängigkeit zwischen Ereignissen.

2.1.1 Stochastische Abhängigkeit zwischen Ereignissen

Bei der Einführung der stochastischen Abhängigkeit bzw. Unabhängigkeit zwischen Ereignissen kann man vom Begriff der bedingten Wahrscheinlichkeit $P(A|B)$ ausgehen. Ist diese verschieden von $P(A)$, dann sind die beiden Ereignisse A und B ***abhängig.***

Bedingte Wahrscheinlichkeit: Zunächst soll daher der Begriff der bedingten Wahrscheinlichkeit eingeführt werden. Sei $P(B) > 0$. Dann heißt die Zahl $P(A \mid B)$ ***bedingte Wahrscheinlichkeit des Ereignisses A gegeben das Ereignis B*** (kurz: B-bedingte Wahrscheinlichkeit von A) genau dann, wenn

$$P(A|B) := \frac{P(A \cap B)}{P(B)} . \tag{2.1}$$

Die linke Seite der obigen Gleichung ist ***nicht*** etwa so zu lesen, daß $A|B$ eine Wahrscheinlichkeit P zugeordnet wird. Da A und B Mengen sind, ist der Ausdruck $A|B$ gar nicht definiert. Stattdessen ist der Ausdruck $P(A|B)$ *so zu* verstehen, daß dem Ereignis A eine bedingte Wahrscheinlichkeit gegeben B zugewiesen wird. Um dies auch in der Schreibweise auszudrücken, kann man statt $P(A|B)$ auch $P_B(A)$ schreiben. Genau wie P ist auch P_B ein W-Maß auf $\mathcal{A}$ und verfügt daher über alle Eigenschaften eines W-Maßes (s. insbesondere G1.1.1). Für verschiedene Ereignisse B und C ist im allgemeinen natürlich auch die B-bedingte Wahrscheinlichkeit $P(A|B) = P_B(A)$ von der C-bedingten Wahrscheinlichkeit $P(A|C) = P_C(A)$ verschieden.

Stochastische Unabhängigkeit zweier Ereignisse: Im Falle der Unabhängigkeit sollten die bedingten und unbedingten Wahrscheinlichkeiten gleich sein, d. h. $P(A|B) = P(A)$ und $P(B|A) = P(B)$. In diesem Fall gelten also:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A) \text{ und } P(B|A) = \frac{P(A \cap B)}{P(A)} = P(B) . \tag{2.2}$$

Um auch den Fall $P(A) = 0$ und/oder $P(B) = 0$ zuzulassen, multipliziert man beide Seiten dieser Gleichungen mit $P(B)$ bzw. $P(A)$ und definiert Unabhängigkeit über die resultierende Gleichung. Zwei Ereignisse ***A*** und ***B*** heißen ***unabhängig*** (bezüglich des Wahrscheinlichkeitsmaßes ***P***, kurz: ***P-unabhängig)*** genau dann, wenn

$$P(A \cap B) = P(A) \cdot P(B). \tag{2.3}$$

Der oben definierte Begriff der Unabhängigkeit zweier Ereignisse ist also ein symmetrischer Begriff. Sprechweisen, die eine Asymmetrie nahelegen, wie „das Ereignis ***A*** ist unabhängig vom Ereignis ***B***“, sollte man daher möglichst vermeiden. Man beachte auch, daß hier ein spezieller Begriff der Unabhängigkeit definiert wurde, der sich auf eine bestimmte Eigenschaft der ***Wahrscheinlichkeiten*** der beiden Ereignisse und ihrer Schnittmenge bezieht. Daher spricht man auch von ***stochastischer*** Unabhängigkeit oder Unabhängigkeit ***bzgl. des W-Maßes P.*** Die ***stochastische Abhängigkeit*** zweier Ereignisse ist als Negation ihrer stochastischen Unabhängigkeit definiert.

Stochastische Unabhängigkeit mehrerer Ereignisse: Wenn wir z.B. drei Ereignisse ***A, B*** und C betrachten, dann kann es durchaus sein, daß für alle drei möglichen Paare stochastische Unabhängigkeit besteht, man aber dennoch aus zwei von ihnen das dritte perfekt vorhersagen kann ***(Meehlsches Paradoxon;*** vgl. Krauth & Lienert, 1973). Würde man sich also nur auf die Betrachtung ***paarweiser*** Abhängigkeit oder Unabhängigkeit beschränken, wurde man in einem solchen Fall eine perfekte Abhängigkeit nicht entdecken. Bei der Definition der stochastischen Unabhängigkeit dreier Ereignisse postuliert man also nicht nur für jedes Paar die Gleichung 2.3, sondern auch

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C). \tag{2.4}$$

Gilt diese Gleichung nicht, dann heißen die drei Ereignisse ***stochastisch abhängig.*** Aus Gleichung 2.4 folgt 2.3 für jedes Paar dieser drei Ereignisse.

Komplexe stochastische Abhängigkeitsstrukturen zwischen Ereignissen und zwischen kategoriellen Variablen können auf verschiedene Arten beschrieben werden (s. z.B. Goodman & Kruskal, 1979). Man kann solche Abhängigkeitsstrukturen sehr gut mit ***log-linearen Modellen*** beschreiben, deren Darstellung jedoch hier nicht möglich ist. Stattdessen sei auf Andersen (1990), Bishop, Fienberg und Holland (1977), Langeheine (1986), Hartung (1989) und Knoke und Burke (1980) verwiesen.

2.1.2 Stochastische Abhängigkeit zwischen Zufallsvariablen

Die Definition der stochastischen Unabhängigkeit von Zufallsvariablen ist nicht ganz so elementar wie die zwischen Ereignissen. Verbal und notwendigerweise etwas unpräzise kann man formulieren, daß zwei stochastische Variablen X und Y stochastisch unabhängig sind, genau dann, wenn alle mit ihnen verknüpften Ereignisse im oben definierten Sinn unabhängig sind; genauer, wenn für alle Ereignisse ***A***, daß X einen Wert in einer bestimmten

Teilmenge ihrer Wertemenge annimmt und ***B***, daß Y einen Wert in einer bestimmten Teilmenge ihrer Wertemenge annimmt, die Gleichung 2.3 gilt. Eine präzise Definition der Unabhängigkeit von Zufallsvariablen findet man z.B. bei Bauer (1978) oder bei Steyer und Eid (1993).

Eine andere Umschreibung der stochastischen Unabhängigkeit zwischen zwei Zufallsvariablen ist, daß kein Aspekt der Verteilung der einen Variablen von der Ausprägung (dem Wert) der anderen Variablen abhängt. Insbesondere hängen weder der Erwartungswert noch die Varianz oder irgendein anderer Kennwert der Verteilung der einen Variablen ab vom Wert, den die andere Variable annimmt. Die stochastische Abhängigkeit ist natürlich wieder durch die Negation der Unabhängigkeit definiert. Da die Definition der stochastischen Unabhängigkeit von Zufallsvariablen auf die Unabhängigkeit von Ereignissen zurückgeführt wird, kann man die Definition der Unabhängigkeit mehrerer Zufallsvariablen ebenfalls auf die entsprechende Definition für Ereignisse zurückführen.

Die stochastische Unabhängigkeit von Zufallsvariablen spielt z.B. bei randomisierten Experimenten eine Rolle. Bei einem randomisierten Experiment wird eine Beobachtungseinheit (z. B. eine Versuchsperson) definitionsgemäß nach Zufall einer der experimentellen Bedingungen zugeordnet. Damit wird die stochastische Unabhängigkeit zwischen der Treatmentvariablen und allen Variablen hergestellt, die eine Eigenschaft der Beobachtungseinheit ***vor*** der Zuweisung repräsentieren. Dies hat wichtige Konsequenzen für die kausale Interpretierbarkeit der Abhängigkeit der Kriteriums- von der Treatmentvariablen (näheres dazu in Abschnitt 2.3).

2.2 Regressive und korrelative Abhängigkeiten

Während der Begriff der stochastischen ***Unabhängigkeit*** bei vielen stochastisehen und statistischen Modellen eine wichtige Rolle spielt, sind Aussagen über eine stochastische ***Abhängigkeit*** zu unspezifisch. Es gibt nämlich sehr viele verschiedene Arten stochastischer Abhängigkeit. Die wichtigsten Arten sind die regressiven und korrelativen Abhängigkeiten.

Aussagen über regressive und korrelative Abhängigkeiten können dann getroffen werden, wenn neben einem W-Raum mindestens zwei Zufallsvariablen X und Y vorliegen. Bei ***korrelativen*** Abhängigkeiten müssen beide Variablen numerisch sein (d.h. Zahlen als Werte annehmen). Aussagen über ***regressive*** Abhängigkeiten dagegen sind auch dann möglich, wenn nur der Regressand Y numerisch ist. Der Regressor X dagegen kann qualitativ sein, also z.B. die Werte ,,Experimentalgruppe“ oder ,,Kontrollgruppe“ annehmen.

Bei einer Aussage über eine regressive Abhängigkeit einer numerischen Zufallsvariablen Y von einer Zufallsvariablen X macht man eine Aussage über die ***Regression*** oder ***bedingte Erwartung*** $E(Y|X)$. Bei der Regression $E(Y|X)$

des ***Regressanden*** Y auf den ***Regressor*** X handelt es sich um eine weitere Zufallsvariable, deren Werte die ***bedingten Erwartungswerte*** $E(Y|X = x)$ sind (zur genaueren Definition s. z.B. Bauer, 1978, Steyer, 1988, 1992 oder Steyer & Eid, 1993, Anhang G). Aussagen über eine solche Regression sind nicht nur der Kern der einfachen oder multiplen Regressionsanalyse, sondern auch der Varianzanalyse, der Faktorenanalyse und der Strukturgleichungsmodelle.

Aussagen über Regressionen können in verschiedener Weise formuliert werden, z.B.
(a) als Regressionskurve in einem Kartesischen Koordinatensystem,
(b) als Säulendiagramm, mit dem man bedingte Wahrscheinlichkeiten oder Erwartungswerte angibt,
(c) als Tabelle, in der man bedingte Erwartungswerte in Gruppen angibt,
(d) als Pfaddiagramm oder auch
(e) als Gleichung.

Die Darstellungsform hat jedoch nichts mit der logischen Struktur zu tun, die gemeint ist, wenn von Regressionen die Rede ist. In allen genannten Fällen geht es um Aussagen darum, wie die bedingten Erwartungswerte $E(Y|X = x)$ einer Variablen Y von den Werten einer (bzw. mehrerer) Variablen X (bzw. $X_1, ..., X_m$) abhängen, oder um globale Aussagen darüber, wie stark diese regressive Abhängigkeit ist, z.B. durch Angabe des ***Determinationskoeffizienten***

$$R^2_{Y|X} := Var[E(Y|X)] \,/\, Var(Y). \tag{2.5}$$

Diese Kenngröße gibt den Anteil der Variation der bedingten Erwartungswerte $E(Y|X = x)$ um den theoretischen Erwartungswert $E(Y)$ an der Gesamtvarianz der betrachteten Y-Variablen an, die sich ihrerseits additiv aus der Varianz $Var[E(Y|X)]$ der Regression $E(Y|X)$ und der Varianz $Var(\varepsilon)$ der Fehlervariablen $\varepsilon := Y - E(Y|X)$ zusammensetzt.

Der Begriff der Regression eignet sich dazu, verschiedene Arten von Abhängigkeiten und Unabhängigkeiten des Regressanden Y vom Regressor X zu beschreiben. ***Regressive Unabhängigkeit*** der Variablen Y von X liegt vor genau dann, wenn gilt:

$$E(Y|X) = E(Y). \tag{2.6}$$

Andernfalls heißt ***Y*** von ***X regressiv abhängig.*** Bei regressiver Unabhängigkeit nimmt also die Regression $E(Y|X)$ für jede Ausprägung x von X den gleichen Wert, nämlich $E(Y)$ an. Ist X eine kontinuierliche numerische Variable, kann man den Fall der regressiven Unabhängigkeit auch durch eine Regressionsgerade darstellen, die parallel zur X-Achse verläuft.

2.2.1 Einfache regressive Abhängigkeit

Besteht eine regressive Abhängigkeit eines Regressanden Y von einem Regressor X, so kann sie verschiedener Art sein. Die einfachste Art der regressiven Abhängigkeit ist die ***lineare,*** die dann vorliegen kann, wenn es sich bei X um einen ***numerischen*** Regressor handelt. ***Einfache lineare regressive Abhängigkeit*** ist dadurch definiert, daß die Regression $E(Y|X)$ eine lineare Funktion von X ist:

$$E(Y|X) = \alpha_0 + \alpha_1 \cdot X, \qquad \text{wobei } \alpha_0, \alpha_1 \in \mathbb{R}. \tag{2.7}$$

Kann X nur zwei verschiedene Werte annehmen, dann muß sich die Regression E(YIX) sogar als eine solche lineare Funktion von X darstellen lassen. Um dies zu erkennen, trage man die beiden Punkte [x_1, $E(Y|X = x_1)$] und [x_2, $E(Y|X = x_2)$] in ein 2-dimensionales Koordinatensystem ein. Dabei sieht man, daß zwei Punkte ***immer*** auf einer Geraden liegen, die sich durch den Achsenabschnitt cr_0 und die Steigung a_1 angeben läßt. Nimmt X mehr als zwei verschiedene Werte an, muß die regressive Abhängigkeit möglicherweise auch durch eine andere Funktion dargestellt **werden** (z.B. als Polynom).

Interpretation der Regressionskoeffizienten: In Gleichung 2.7 lassen sich a_0 als Achsenabschnitt und a_1 als Steigung einer Geraden in einem 2-dimensionalen Koordinatensystem interpretieren. Demnach gibt CC_1 die Differenz ***E(YIX*** = x_1) - ***E(YIX*** = x_2) der bedingten Erwartungswerte an, wenn die Differenz $x_1 - x_2$ zweier Werte von X gleich eins ist.

In vielen Büchern findet man eine falsche Interpretation des Regressionskoeffizienten CZ_1, derzufolge dieser Koeffizient angäbe, wie sich der bedingte Erwartungswert von Y verändert, wenn man X um eine Einheit verändert. Bei einer solchen Interpretation wird jedoch übersehen, daß mit einer ***Veränderung*** einer X-Variablen von einem ganz anderen Zufallsexperiment gesprochen wird, als wenn man X und Y lediglich ***beobachtet.*** Bei einer Intervention zur Veränderung von X können bspw. auch indirekte Effekte auf Y auftreten, die den direkten Effekt einer Veränderung von X auf Y überlagern. In den Sozialwissenschaften müssen wir immer damit rechnen, daß wir es mit einem ganzen System vieler interdependenter Variablen zu tun haben, bei dem wir zwar die Kovariation zweier Variablen beobachten und mit einer Regression beschreiben können; aber aus diesen Beobachtungsergebnissen können wir nicht ***logisch -*** allenfalls heuristisch - schließen, was passiert, wenn wir in das System eingreifen, da damit ein neues System geschaffen wird. In Termini der Wahrscheinlichkeitstheorie formuliert, liegt mit einer Intervention ein anderes Zufallsexperiment vor, das durch einen anderen W-Raum repräsentiert wird.

2.2.2 Einfache lineare quasi-regressive Abhängigkeit

Ist die Regression ***nicht*** linear, kann man dennoch betrachten, wie gut die Abhängigkeit einer numerischen Zufallsvariablen Y von einer zweiten numerischen Zufallsvariablen X durch eine ***lineare*** Funktion beschrieben werden kann, wobei diese Funktion dann aber nicht als Regression im oben behandelten Sinn interpretiert werden kann. Eine solche lineare Funktion, die die nach dem Kleinst-Quadrat-Kriterium optimale lineare Beziehung zwischen X und Y angibt, bezeichne ich als ***lineare Quasi-Regression.*** (Müller, 1975, spricht stattdessen von der „Regression 2. Art".) Diese darf nicht mit dem oben behandelten Begriff der einfachen linearen Regression verwechselt werden. Unter der linearen Quasi-Regression ist diejenige lineare Funktion f(X) = α + ß $\cdot X$ von X zu verstehen, die die folgende Funktion von a und ***b*** - das sogenannte Kleinst-Quadrat-Kriterium - minimiert:

$$LS(a, b) := E\,[[Y - (a + b \cdot X)]^2]. \tag{2.8}$$

Weist LS(a, ***b)*** für die Zahlen α und ß ihr Minimum auf, bezeichnet man die Funktion *f(X)* = α + ß . X auch mit $\hat{Y}$. ***Die so definierte lineare Quasi-Regression $\hat{Y}$ ist nur dann mit der (echten) Regression $E(Y|X)$*** identisch, wenn ***E(YIX)*** tatsächlich eine lineare Funktion von X ist. Läßt sich ***E(YIX)*** dagegen z.B. nur durch eine Parabel beschreiben, dann unterscheiden sich die Regression und die lineare Quasi-Regression voneinander. Im allgemeinen gibt also die lineare Quasi-Regression Y ***nicht*** an, wie die bedingten Erwartungswerte ***E(YIX*** = x) von den Werten x von X abhängen. Stattdessen gibt $\hat{Y}$ ***nur einen Wert an, der dann im Sinn des Kleinst-Quadrat-Kriteriums (s. Gl.2.8)*** eine optimale Vorhersage für Y ist, falls man sich für die Vorhersage auf die Klasse der ***linearen*** Funktionen von X beschränkt. Der einzige Grund, lineare Quasi-Regressionen und nicht die echte Regression zu verwenden, könnte in der größeren Einfachheit liegen. Dies wird aber mit einem Verzicht an Information bezahlt. Natürlich kann man auch die Begriffe einer quadratischen, kubischen Quasi-Regression etc. einführen. Dabei ist lediglich a + ***b*** . X durch die entsprechende quadratische, kubische Funktion etc. von X zu ersetzen.

2.2.3 Korrelative Abhängigkeit

Die wohl am häufigsten verwendete Kenngröße zur Beschreibung einer Abhängigkeit zwischen zwei numerischen Zufallsvariablen ist der Korrelationskoeffizient

$$Kor(X, Y) := Cov(X, Y)\ /\ [Std(X) \cdot Std(Y)], \tag{2.9}$$

der als Kovarianz geteilt durch das Produkt der beiden Standardabweichungen definiert ist, falls die beiden Standardabweichungen größer 0 sind; andernfalls ist ***Kor(X,*** Y) := 0. (Zur Definition der Kovarianz siehe z.B. Bauer, 1978, oder Steyer & Eid, 1993.) Mit dieser Kenngröße wird die Stärke der durch Gleichung 2.7 beschriebenen ***linearen*** regressiven Abhängigkeit des Regressanden Y vom Regressor X angegeben. Liegt eine ***lineare*** regressive Abhängigkeit vor, dann ist der quadrierte Korrelationskoeffizient mit dem in Gleichung 2.5 definierten Determinationskoeffizienten identisch. Ist die Regression dagegen ***nicht*** linear, dann gibt der Korrelationskoeffizient an, wie stark der durch die ***lineare Quasi-Regression*** beschreibbare Zusammenhang zwischen X und Y ist.

2.2.4 Partielle lineare regressive Abhängigkeit

Betrachtet man ***zwei*** numerische Regressoren X und Z, ergeben sich wiederum verschiedene Abhängigkeitsarten des Regressanden Y von den beiden Regressoren X und Z. Die einfachste Art ist dabei beschreibbar durch die multiple Regressionsgleichung

$$E(Y|X, Z) = \beta_0 + \beta_1 X + \beta_2 Z, \qquad \beta_0, \beta_1, \beta_2 \in \mathbb{R}. \tag{2.10}$$

Dabei kann es durchaus sein, daß der ***partielle Regressionskoeffizient*** β_1 nicht mit dem entsprechenden einfachen Regressionskoeffizienten $\boldsymbol{a_1}$ aus der einfachen linearen Regression (s. Gl. 2.7) identisch ist. In diesem Fall heißen X und ***Z konfundiert.*** Setzt man die Gültigkeit der Gleichungen 2.7 und 2.10 voraus, dann tritt der Fall $\alpha_1 = \beta_1$ dann und nur dann ein, wenn (a) Z von X regressiv unabhängig ist (s. G1.2.6) oder (b) wenn β_2 gleich 0 ist. Im allgemeinen kann es durchaus sein, daß a_1 positiv und β_1 negativ sind und umgekehrt, oder, daß einer der beiden Parameter gleich null ist und der andere stark von null verschieden. (Ein Beispiel findet man in Steyer, 1992, Kap.3). Man beachte, daß hier von den wahren Parametern die Rede ist und nicht von ihren Schätzungen anhand einer Stichprobe.

Spätestens hier wird klar, daß es verschiedene Arten regressiver Abhängigkeit gibt, die in Anwendungen auch inhaltlich völlig verschiedenes besagen und die auf den ersten Blick sogar widersprüchlich erscheinen können, bspw. dann, wenn $\boldsymbol{a_1} < 0$ und $\beta_1 > 0$. Es gibt also nicht ***die*** Abhängigkeit einer Variablen Y von einer Variablen X, nicht einmal ***die regressive*** Abhängigkeit der Variablen Y von X und selbst nicht ***die partielle regressive*** Abhängigkeit der Variablen Y von X, da auch die partiellen regressiven Abhängigkeiten der Variablen Y von X, die in multiplen Regressionen mit zwei, drei, vier Regressoren beschrieben werden, völlig verschieden aussehen können.

2.2.5 Bedingte regressive Abhängigkeit

Zum Verständnis der Bedeutung der regressiven Abhängigkeit eines Regressanden Y bei der Berücksichtigung von zwei (oder mehr) Regressoren X und Z ist die Betrachtung der ***bedingten Regression*** E_Z von Nutzen, d.h.

der Regression von Y auf X bzgl. des bedingten Wahrscheinlichkeitsmaßes $P_{Z=z}$. Bezeichnet Z bspw. die Geschlechtsvariable mit den beiden Werten 1 (für Frauen) und 0 (für Männer), dann ist $E_{Z=0}(Y|X)$ die Regression von Y auf X in der Subpopulation der Männer.

Die wichtigste Eigenschaft der durch die Gleichung 2.10 beschriebenen Art regressiver Abhängigkeit ist, daß die bedingte Regression $E_{Z=z}(Y|X)$ bei festem Wert z von Z wiederum eine ***lineare*** Funktion von X ist, und zwar mit dem ***gleichen Steigungskoeffizienten*** β_1 für alle möglichen Werte z von Z. Die Achsenabschnitte dieser bedingten Regressionen für verschiedene Werte z von Z unterscheiden sich dagegen, falls β_2 ungleich null ist:

$$E_{Z=z}(Y|X) = (\beta_0 + \beta_2 z) + \beta_1 X, \quad \beta_0, \beta_1, \beta_2 \in \mathbb{R}. \tag{2.11}$$

Die durch β_1 beschriebene Abhängigkeit des Regressanden Y von X ist damit über alle Werte von Z generalisierbar. Dies ist keineswegs selbstverständlich, wie in den nachfolgenden Abschnitten deutlich wird. Darüber hinaus ist damit gezeigt, daß man den ***partiellen Regressionskoeffizienten*** β_1 als ***einfachen Regressionskoeffizienten in der bedingten Regression*** $E_{Z=z}(Y|X)$ verstehen kann. (Durch Vertauschung von X und Z erhält man die analoge Aussage für β_2.)

Eine weitere Art regressiver Abhängigkeit resultiert, wenn man zwar bei jedem gegebenen Wert z von Z eine lineare Regression von Y auf X hat, dabei jedoch der Steigungskoeffizient vom jeweiligen Wert von Z abhängt. Etwas präziser läßt sich diese Art der Abhängigkeit durch die Gleichung

$$E(Y|X, Z) = g_0(Z) + g_1(Z) \cdot X \tag{2.12}$$

charakterisieren, wobei $g_0(Z)$ und $g_1(Z)$ beliebige reellwertige Funktionen von Z bezeichnen. Für feste Werte z von Z resultiert hier die ***bedingte lineare Regression***

$$E_{Z=z}(Y|X) = g_0(z) + g_1(z) \cdot X, \tag{2.13}$$

deren Steigungskoeffizienten vom Wert z von Z abhängen und durch die Funktion $g_1(Z)$ berechnet werden können, falls diese Funktion bekannt ist.

Ist nun $g_1(Z)$ keine konstante Funktion, die für alle Werte z von Z den gleichen Wert β_1 annimmt, sagt man auch, daß Z den Effekt von X ***moderiert*** (besser wäre eigentlich: ***modifiziert***). Man spricht dann daher auch von einem ***Moderatormodell*** und nennt Z einen ***Moderator.*** Anwendungen findet man z.B. bei Bredenkamp (1984a, b) oder Erdfelder und Steyer (1984).

Ein recht häufig vorkommendes spezielles Moderatormodell liegt vor, wenn es sich bei den Funktionen g_0 und g_1 um ***lineare*** Funktionen von Z handelt:

$$E(Y|X, Z) = (\beta_0 + \beta_2\, Z) + (\beta_1 + \beta_3 \cdot Z)\, X$$
$$= \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 X \cdot Z. \quad (2.14)$$

Das Moderatormodell wird gelegentlich mit dem oben bereits erwähnten Phänomen der Konfundierung verwechselt (s. z.B. Wermuth, 1989), was von Schmitt (1990) sowie von Schmitt und Götz-Baltes (1992) kritisiert wurde. Die wesentliche Eigenschaft eines ***Moderatormodells*** ist, daß sich die bedingten Regressionskoeffizenten $g_1(z)$ für verschiedene Werte z von Z voneinander unterscheiden. Bei der ***Konfundierung*** dagegen wird vorausgesetzt, daß die bedingten Regressionskoeffizienten für alle Werte z von Z gleich β_1 sind. Dieser Koeffizient β_1 ist nur nicht mit dem einfachen Regressionskoeffizienten α_1 identisch.

Bei allen bisherigen Betrachtungen waren die bedingten Regressionen ***lineare*** Funktionen von X. Dieser Fall wird am allgemeinsten durch die Gleichung 2.12 beschrieben, die sich leicht verallgemeinern läßt, z.B. zu

$$E(Y|X, Z) = g_0(Z) + g_1(Z) \cdot X + g_2(Z) \cdot X^2, \quad (2.15)$$

falls X mindestens drei verschiedene Werte annehmen kann.

Das ***allgemeine Prinzip der bedingten Regression*** ist die Beschreibung der regressiven Abhängigkeit des Regressanden Y vom Regressor X bei konstanten Ausprägungen einer oder mehrerer anderer Variablen Z. Dabei beachte man, daß Z in den Gleichungen 2.12 und 2.15 durchaus auch für eine vektorielle Variable stehen kann. Ist z.B. $Z = (Z_1, Z_2)$ ein 2-dimensionaler Vektor numerischer Zufallsvariablen und handelt es sich bei $g_0(Z)$ und $g_1(Z)$ jeweils um lineare Funktionen von Z_1 und Z_2, dann ist ein Spezialfall von Gleichung 2.12

$$E(Y|X, Z) = (\beta_0 + \beta_1\, Z_1 + \beta_2\, Z_2) + (\gamma_0 + \gamma_1\, Z_1 + \gamma_2\, Z_2) \cdot X \quad (2.16)$$

Auch wenn diese Formeln recht kompliziert erscheinen, sollte man dennoch bedenken, daß damit eine große Vereinfachung erzielt werden kann, wenn man beschreiben will, wie die regressive Abhängigkeit des Regressanden Y vom Regressor X von den Ausprägungskombinationen der Variablen Z_1 und Z_2 modifiziert wird.

2.2.6 Zusammenfassende Bemerkungen

Bei Regressions- und Korrelationsmodellen äußert sich das Problem der multiplen Determiniertheit darin, daß im allgemeinen ***mit jedem*** neu ***hinzugenommenen Regressor eine andere regressive (bzw. korrelative) Abhängigkeit*** betrachtet wird. Auf der Modellebene schlägt sich dies darin nieder, daß sich der einfache bzw. partielle Regressionskoeffizient und damit die jeweils betrachtete einfache bzw. partielle regressive Abhängigkeit mit jedem neu hinzugenommenen Regressor ändern kann, und zwar ***nicht*** nur graduell und ***nicht*** nur wegen stichprobenbedingter Schätzfehlerprobleme. So können aus ***positi-***

ven Abhängigkeiten ***negative*** oder auch Unabhängigkeiten werden und umgekehrt (s. auch die Erläuterungen zu G1.2.10).

Demnach gibt es nicht ***den Effekt*** eines Regressors X auf einen Regressanden Y, sondern prinzipiell mindestens soviele Effekte von X auf Y wie es weitere Variablen gibt, die man zu dem gerade betrachteten Regressor X als zusätzliche Regressoren hinzuziehen kann. Man siehe hierzu auch die Literatur zu ,,Suppressorvariablen" und ,,Moderatorvariablen" (z. B. Bartussek, 1970; Cheek, 1982; Darlington, 1968; Ghiselli, 1963; Kenny & Judd, 1984; Moosbrugger, 1981; Penner & Wymer, 1983; Saunders, 1956; Schmitt, 1990).

2.3 Kausale regressive Abhängigkeiten

Nach den obigen Ausführungen ist leicht einzusehen, daß Regressionsmodelle (und folglich auch Spezialfalle wie Varianz- und faktoranalytische Modelle) nicht zur Theorienformulierung oder -prüfung beitragen können, wenn die mit ihnen definierten Abhängigkeitsbegriffe so vielfältig und zunächst auch beliebig sind. Regressive Abhängigkeiten sind damit ***für sich genommen*** zur Theorienformulierung ungeeignet. Tatsächlich kommt in psychologischen Hypothesen oft eine ***kausale*** Terminologie wie ,,Effekt", ,,Wirkung", ,,führt zu" u.ä. vor. Oft ist die Hypothese über einen kausalen Effekt auch mit Hilfe der ***Ceteris-paribus-Klausel*** formuliert, derzufolge bestimmte Effekte nur bei Konstanthaltung aller potentiellen Störvariablen postuliert werden. Hinter dieser Klausel verbirgt sich der auf John Stuart Mill (1862) zurückgehende Kausalitätsbegriff, daß, wenn alle anderen Variablen konstant sind und nur X variiert, die Kovariation zwischen X und der später auftretenden Variablen Y auf X als Ursache zurückzuführen ist.

Wie wir sehen werden, ist damit ein sehr strenger Kausalitätsbegriff involviert. Mit der Ceteris-paribus-Klausel bezieht man sich außerdem auf eine in der Regel nicht herstellbare Realität. In den meisten Fällen ist die Konstanthaltung aller potentiellen Störvariablen schlechterdings nicht möglich. Muß man außerdem mit Interaktionen rechnen, stellt sich die Frage, auf welchen Werten man die jeweiligen Störvariablen konstanthalten soll, da Interaktion ja bedeutet, daß die Effekte nicht auf jeder Stufe der konstantgehaltenen Variablen gleich sind. Diesen Problemen kann man entgehen, wenn man die durch die Ceteris-paribus-Klausel formulierte Aussage über eine kausale Abhängigkeit ***indirekt*** überprüft. Erdfelder und Bredenkamp (in diesem Band, Kap. 14, Abschnitt 2.4) zeigen nämlich, daß aus einer derartigen Aussage unter bestimmten Voraussetzungen noch Konsequenzen ableitbar sind, die sich in einem randomisierten Experiment überprüfen lassen.

Eine Alternative dazu ist, die Ceteris-paribus-Klausel auch schon in der Theorie oder Hypothese durch die ***Ceteris-distributionibus-paribus-Klausel zu*** ersetzen, derzufolge nur noch die ***Verteilungen*** aller potentiellen Störvariablen konstantzuhalten sind. Dies erreicht man durch die Kontrolltechnik der Randomisierung. Durch die dabei vorgenommene zufällige Aufteilung der Vpn auf die Versuchsbedingungen stellt man die stochastische Unabhängigkeit der Treatmentvariablen mit allen potentiellen Störvariablen her, deren Werte bereits vor der experimentellen Behandlung festliegen. (Dabei beachte man, daß hier ***nicht*** von Abhängigkeiten in einer Stichprobe, sondern von den wahren Abhängigkeiten die Rede ist.) Die durch die Randomisierung hergestellte stochastische Unabhängigkeit impliziert aber, daß die Verteilungen der potentiellen Störvariablen in allen experimentellen Bedingungen gleich sind. Dadurch werden die Effekte der anderen potentiellen Störvariablen gleichmäßig auf die Versuchsbedingungen verteilt. Durch die Randomisierung werden also die Störvariablen nicht konstantgehalten, auch werden deren Effekte nicht eliminiert. Stattdessen werden die ***Verteilungen*** der potentiellen Störvariablen innerhalb der Treatment-Bedingungen gleichgehalten. Der Nachteil ist allerdings, daß man im randomisierten Experiment nur ***durchschnittliche*** Effekte - gemittelt über die nichtbeachteten potentiellen Störvariablen - analysiert. Diese Gedanken sollen in den folgenden Abschnitten näher erläutert werden.

2.3.1 Vorbereitende Definitionen

Um die obigen Vorstellungen zu präzisieren, ist es zunächst angebracht, den ***allgemeinen*** Begriff der ***Konfundierung*** zu definieren, wie er im Kontext von Regressionsmodellen verwendet werden kann. (Im Abschnitt 2.2 hatten wir uns auf den Fall einer ***linearen*** Regression beschränkt.) Dabei spielt zunächst neben X nur ein einziger weiterer Regressor W eine Rolle.

Unkonfundiertheit: Seien Y eine numerische sowie X und W beliebige Zufallsvariablen auf einem gemeinsamen W-Raum. Dann heißt X mit W bzgl. der Regression ***E(YIX) unkonfundiert*** genau dann, wenn gilt:

(u_1) X und W sind stochastisch unabhängig,
oder
(u_2) es existiert eine numerische Funktion ***f*** derart, daß für ***f(W)*** gilt:
$E(Y|X, W) = E(Y|X) + f(W)$.

X heißt mit W bzgl. $E(Y|X)$ *konfundiert*, wenn weder (u_1) noch (u_2) gilt.

Im Fall (u_1) sind die bedingten Wahrscheinlichkeiten $P(W{=}w|X = x)$ mit den unbedingten Wahrscheinlichkeiten $P(W = w)$ identisch und im Fall (u_2) stimmt die einfache Regression $E(Y|X)$ bis auf eine additive Konstante mit den bedingten Regressionen $E_{W=w}(Y|X) = E(Y|X) + f(w)$ überein.

Bei gewöhnlichen Regressionsmodellen ist die Unkonfundiertheit im allgemeinen ***nicht*** gewährleistet, es sei denn, die Bedingung (u_1) oder die Bedingung (u_2) gilt aus ***empirischen*** Gründen. Im Rahmen eines randomisierten Experiments dagegen ist die Unkonfundiertheit ***immer*** erfüllt, falls es sich bei W um eine Variable handelt, deren Werte bereits ***vor*** der Applikation des Treatments festliegen, da durch die Randomisierung die Unabhängigkeit von X und W ***bergestellt*** wird. Dabei beachte man wieder, daß hier nicht von der Unabhängigkeit in einer Stichprobe die Rede ist.

Konsequenzen für die Interpretation: Die Unkonfundiertheit hat wichtige Konsequenzen für die Interpretation der durch die Regression $E(Y|X)$ beschriebenen Abhängigkeit. Es läßt sich nämlich zeigen, daß die bedingten Erwartungswerte $E(Y|X=x)$ dann als ***Mittelung*** der bedingten Erwartungswerte $E(Y|X=x, W=w)$ - gemittelt über die Verteilung von W - interpretiert werden können, d.h. im Falle der Unkonfundiertheit von X und W gilt für jeden Wert $E(Y|X=x)$ der Regression $E(Y|X)$:

$$E(Y|X=x) = \sum_w E(Y|X=x, W=w) \cdot P(W=w), \tag{2.17}$$

wobei die Summierung über alle Werte w von W vorzunehmen ist. Daß die Unkonfundiertheit die Gültigkeit der Gleichung 2.17 impliziert, erkennt man folgendermaßen: Da die Gleichung

$$E(Y|X=x) = \sum_w E(Y|X=x, W=w) \cdot P(W=w|X=x) \tag{2.18}$$

allgemeingültig ist, sieht man sofort, daß im Fall der stochastischen Unabhängigkeit von X und W die Gleichung 2.17 erfüllt ist, denn dann gilt $P(W=w|X=x) = P(W=w)$ [s. Bedingung (u,)]. Gleichung 2.17 ist aber auch dann erfüllt, wenn für alle Werte von X und W gilt: $E(Y|X=x, W=w) = E(Y|X=x) + f(w)$ [s. Bedingung (u_2)]. Dies aber sind die beiden Bedingungen, von denen zumindest eine erfüllt ist, wenn die Unkonfundiertheit gilt.

Ist die Unkonfundiertheit erfüllt, resultieren die bedingten Erwartungswerte $E(Y|X=x)$ demzufolge aus der Mittelung der bedingten Erwartungswerte $E(Y|X=x, W=w)$ über die Verteilung von W. Dies ist weitaus mehr, als bei einem gewöhnlichen Regressionsmodell gilt. Dort kann es nämlich z.B. durchaus sein, daß für ***jeden*** festen Wert einer potentiellen Störvariablen W eine ***negative*** lineare regressive Abhängigkeit besteht, aber dennoch eine positive lineare regressive Abhängigkeit beobachtet wird, wenn man nur die Regression $E(Y|X)$ betrachtet (s. dazu ein ausführliches Beispiel bei Steyer, 1992, Kap.3). Dies ist genau der bereits früher behandelte Fall $\alpha_1 > 0$ und $\beta_1 < 0$ (s. die Gleichungen 2.7 und 2.10).

Warum ist es so wichtig, daß man diese Mittelungsinterpretation vornehmen kann? Angenommen, die psychologische Hypothese postuliert bei konstantgehaltenen Störvariablen - repräsentiert durch die mehrdimensionale Zufallsvariable $\boldsymbol{W} = (W_1, \ldots, W_m)$ - einen ***positiven regressiv linearen Effekt*** von X auf Y, dann läßt sich diese Hypothese wie folgt formulieren (s. Erdfelder & Bredenkamp, Kap. 14, Abschnitt 2.4, in diesem Band):

$$E(Y|X, W) = g_0(W) + g_1(W) \cdot X,$$
wobei $g_1(w) > 0$ für alle Werte w von W. (2.19)

Kann man nun $\boldsymbol{W}$ nicht beobachten - und man wird nur selten alle potentiellen Störvariablen beobachten können -, dann läßt sich diese Hypothese dennoch ***indirekt*** überprüfen, indem man durch Randomisierung dafür sorgt, daß X und $\boldsymbol{W}$ stochastisch unabhängig sind. Dann gilt nämlich auch, daß die Funktion $g_1(W)$ von X regressiv unabhängig ist, d.h.

$$E[g_1(W)|X] = E[g_1(W)] = \alpha_1, \qquad (2.20)$$

wobei α_1 der Steigungskoeffizient der linearen Regression $E(Y|X) = \alpha_0 + \alpha_1 \cdot X$ ist. Der Erwartungswert einer Variablen, die nur positive Werte annehmen kann [hier: die Werte der Funktion g,(W)], ist aber notwendigerweise ebenfalls positiv. Aus der obigen psychologischen Hypothese kann man unter der Voraussetzung eines randomisierten Experiments (Steyer, 1992, zeigt, daß dafür auch die schwächere Annahme der Gleichung 2.17 reicht) also ableiten, daß auch der Koeffizient a_1 der einfachen linearen Regression $E(Y|X) = \alpha_0 + a_1 \cdot X$ positiv ist. Entscheidet man aufgrund eines randomisierten Experiments dann, daß α_1 ***nicht*** positiv ist, kann damit die psychologische Hypothese als falsifiziert gelten (s. Erdfelder & Bredenkamp, Kap. 14, Abschnitt 2.4 in diesem Band). Die Gültigkeit der Gleichung 2.17 ist demnach nicht nur aus ästhetischen Gründen wünschenswert, sondern auch ***notwendige*** Voraussetzung, wenn man bestimmte psychologische Hypothesen im Sinne der Falsifikation empirisch überprüfen will.

Additive Dekomponierbarkeit: Wem die oben erläuterte „Mittelungsinterpretation“ zu wenig erscheint, der kann zusätzlich noch die Bedingung der „Additiven Dekomponierbarkeit“ fordern, die eine Interaktion zwischen X und einer potentiellen Störvariablen $\boldsymbol{W}$ verbietet. Diese Bedingung ist z. B. erfüllt, wenn Gleichung 2.10 gilt. In diesem Fall setzt sich die Regression also ***additiv*** aus den Effekten der beiden Regressoren zusammen. Der interpretative Vorteil wurde ausführlich im Abschnitt 2.2.5 behandelt.

Eine allgemeine Definition der Additiven Dekomponierbarkeit, die nicht auf den Fall der multiplen ***linearen*** Regression beschränkt ist, kann man wie folgt formulieren: Seien Y eine numerische und X und W beliebige Zufallsvariablen auf einem gemeinsamen W-Raum. Dann heißt die Regression $E(Y|X, W)$ ***additiv dekomponierbar*** genau dann, wenn gilt:

(a) es existieren zwei numerische Funktionen g(X) und h(W) von X bzw. von ***W***, für die gilt:

$$E(Y|X, W) = g(X) + h(W). \tag{2.21}$$

Konsequenzen für die Interpretation: Man kann nun zeigen, daß, wenn X und ***W*** bzgl. $E(Y|X)$ unkonfundiert sind und die Additive Dekomponierbarkeit erfüllt ist, sich die Regression $E(Y|X, W)$ durch die Gleichung

$$E(Y|X, W) = E(Y|X) + f(W) \tag{2.22}$$

darstellen läßt, wobei $f(W)$ wiederum eine Funktion von ***W*** ist. Gilt aber die Gleichung 2.22, so hat dies wichtige Konsequenzen für die Interpretation der durch die Regression $E(Y|X)$ beschriebenen Abhängigkeit. Für jeden festen Wert w von ***W*** gilt dann nämlich:

$$E_{W=w}(Y|X) = E(Y|X) + f(w), \tag{2.23}$$

d.h. die durch $E(Y|X)$ beschriebene Abhängigkeit gilt für jeden Wert w von W, wenn man von der additiven Konstanten $f(w)$ absieht. Das Ignorieren von W ist in diesem Fall unbedenklich.

Repräsentiert W bspw. die Zugehörigkeit zu einer Subpopulation (z. B. Männer oder Frauen), so ist die regressive Abhängigkeit des Regressanden Y vom Regressor X in der Regel in jeder dieser Subpopulationen gleich und auch gleich wie in der Gesamtpopulation, abgesehen von der additiven Konstanten $f(w)$. Anschaulich bedeutet das, daß die (W=w)-bedingten Regressionen $E_{W=w}(Y|X)$ parallel verlaufen müssen, und auch parallel zur (unbedingten) Regression $E(Y|X)$.

Die Additive Dekomponierbarkeit läßt nicht zu, daß zwischen X und der Störvariablen ***W*** eine Interaktion im varianzanalytischen Sinn besteht. Dies ist eine Bedingung, die nicht einmal im randomisierten Experiment hergestellt werden kann, denn trotz zufälliger Zuweisung der Vpn auf die Versuchsbedingungen ist es möglich, daß die experimentelle Behandlung (repräsentiert durch die Zufallsvariable X) auf Y bei Männern anders wirkt als bei Frauen (***W*** repräsentiere die Geschlechtsvariable). Lediglich die Unkonfundiertheit mit den oben beschriebenen interpretativen Konsequenzen (die ,,Mittelungsinterpretation“) läßt sich durch die Randomisierung herstellen.

2.3.2 *Schwache kausale regressive Abhängigkeit*

Im randomisierten Experiment gilt die Unkonfundiertheit nicht nur für eine einzige, sondern für alle potentiellen Störvariablen, jedenfalls dann, wenn man die Menge der potentiellen Störvariablen mit der Menge derjenigen Variablen gleichsetzt, deren Werte bereits vor der Applikation des Treatments festliegen. Da im randomisierten Experiment die Unkonfundiertheit erfüllt ist, gilt in diesem Fall auch die ***Schwache Kausalität,*** die in diesem Kontext nichts anderes meint, als daß X mit ***jeder*** potentiellen Störvariablen bzgl. Y unkonfundiert ist, und der Regressor X dem Regressanden ***Y vorgeordnet*** ist.

Anstelle der Unkonfundiertheit kann man zur Definition der Schwachen ***Kausalität*** auch die etwas schwächere Bedingung verwenden, daß Gleichung **2.17** für ***alle*** potentiellen Störvariablen gilt. Beschränkt man sich der Einfachheit halber auf den Fall diskreter potentieller Störvariablen (der allgemeine Fall wird bei Steyer, 1992, behandelt), so kann man die ***schwache kausale regressive Abhängigkeit*** also wie folgt definieren: Seien Y eine numerische und X eine beliebige Zufallsvariable auf einem gemeinsamen W-Raum. Dann heißt Y von ***X schwach kausal regressiv abhängig*** genau dann, wenn gelten:

(w_1) Für alle potentiellen Störvariablen gilt Gleichung 2.17.
(w_2) X ist Y ***vorgeordnet.***

Die Bedingung (w_2) ist im randomisierten Experiment natürlich ebenfalls automatisch erfüllt, da man den Regressanden Y erst nach der Setzung der durch X repräsentierten experimentellen Bedingungen beobachtet (zur Formalisierung des Begriffs der Vorgeordnetheit s. Steyer, 1992, Kap. 6). Man beachte, daß Unabhängigkeit hier als Spezialfall der Abhängigkeit betrachtet wird.

2.3.3 Starke kausale regressive Abhängigkeit

Die interpretativen Probleme, die durch mögliche Interaktionen entstehen können, wurden oben ausführlich diskutiert. Will man nun Abhängigkeiten betrachten, bei denen angenommen wird, daß keine Interaktionen auftreten, muß man neben der Unkonfundiertheit auch die Additive Dekomponierbarkeit postulieren und zwar für ***alle*** potentiellen Störvariablen. Dies führt aber zur Bedingung der ***Starken Kausalität,*** die besagt, daß außer der Vorgeordnetheit auch für ***jede*** potentielle Störvariable die Additive Dekomponierbarkeit gilt sowie, daß X und ***W*** bzgl. $E(Y|X)$ unkonfundiert sind.

Eine etwas schwächere Definition der Starken Kausalität läßt sich wie folgt formulieren: Seien Y eine numerische und X eine beliebige Zufallsvariable auf einem gemeinsamen W-Raum. Dann heißt ***Y*** von ***X stark*** kausal ***regressiv abhängig*** genau dann, wenn gelten:

(S_1) Für alle potentiellen Störvariablen existiert eine numerische Funktion f derart, daß für f(W) die Gleichung 2.22 gilt.
(S_2) X ist ***Y vorgeordnet.***

Wie oben bereits bemerkt, läßt sich die Starke Kausalität nicht mehr durch Techniken der Versuchsplanung herstellen. Man bedenke aber, daß diese Bedingung immer noch schwächer ist, als die weithin verbreitete Vorstellung, alle wichtigen Variablen in die Regression einzubeziehen.

2.3.4 Zusammenfassende Bemerkungen

Wegen dem Problem der multiplen Determination und dem Meßfehlerproblem werden in den Sozialwissenschaften häufig Regressionsmodelle verwendet. Damit ist man jedoch mit dem Grundproblem von Regressionsmodellen konfrontiert: Mit jedem zusätzlichen Regressor können eine Vielzahl neuer Abhängigkeiten eines gegebenen Regressanden Y von einem gegebenen Regressor X betrachtet werden. Regressionsmodelle und stochastische Modelle im allgemeinen sind daher für die Entwicklung, Formulierung und Prüfung von Theorien nur dann von Interesse, wenn sie ***zu kausalen*** Regressionsmodellen vervollständigt werden.

Ein kausales unterscheidet sich von einem nichtkausalen Regressionsmodell im wesentlichen durch zweierlei:

(a) Durch eine prozeßhafte Betrachtung des zugrundeliegenden Zufallsexperiments, wodurch eine Vorgeordnetheit zwischen den Variablen und Ereignissen angenommen werden kann.
(b) Eine Kausalitätsbedingung wie z.B. die Schwache Kausalität, in der eine bestimmte Beziehung zwischen der betrachteten Regression und den potentiellen Störvariablen postuliert wird.

Im randomisierten Experiment ist die Schwache Kausalität ***immer*** erfüllt. In nichtrandomisierten Studien dagegen sind sowohl die Schwache als auch die Starke Kausalität falsifizierbar, wenn man statistische Entscheidungen zwischenschaltet. In randomisierten Experimenten dagegen kann nur die Starke Kausalität falsch sein, da diese Interaktionen zwischen der betrachteten Treatmentvariablen X und potentiellen Störvariablen ausschließt. Bei ***schwachen*** kausalen regressiven Abhängigkeiten dagegen, die in jedem randomisierten Experiment vorliegen, sind solche Interaktionen zugelassen.

Zu einer vollständigeren und mathematisch exakteren Darstellung der Theorie kausaler Regressionsmodelle sei auf Steyer (1992) verwiesen. Dort findet man auch kommentierte Hinweise auf die umfangreiche Literatur zu anderen (auch anderen ***stochastischen)*** Theorien kausaler Abhängigkeit. Am interessantesten sind dabei m. E. die Weiterentwicklungen der Theorie von Suppes (1970, 1981) durch Stegmüller (1983) und vor allem durch Spohn (1980, 1983, 1990, 1991, im Druck). Weiter sei auf die Darstellung ***bedingter*** kausaler Regressionsmodelle bei Steyer (1992) hingewiesen, die wegen der multivariaten Natur empirischer Phänomene in vielen Fällen nützlicher sein können, als die hier dargestellten einfachen kausalen Regressionsmodelle.

3. Stochastische Meßmodelle

Stochastische Meßmodelle sind spätestens seit Thurstone (1931) Kern der psychometrischen Literatur (Fischer, 1974; Formann, 1984; Gulliksen, 1950; Hambleton & Swaminathan, 1985; Kubinger, 1988; Langeheine & Rost, 1988; Lewis, 1986; Lord & Novick, 1968; Rost, 1988; Rost & Strauß, 1992; Steyer, Wender & Widaman, 1993). Allerdings stehen in diesen Arbeiten statistische Probleme des Schätzens von Parametern und Testens von Hypothesen im Vordergrund. Meßtheoretische Fragen, wie die Untersuchung des Skalenniveaus u.ä., wurden bisher weitgehend vernachlässigt (s. jedoch Fischer, 1974, 1988; Steyer, 1989; Steyer & Eid, 1993). Die meßtheoretischen Fragen sind dagegen Kern der Arbeiten von Falmagne (1985), Krantz, Luce, Suppes und Tversky (1971), Luce, Krantz, Suppes und Tversky (1990), Roberts (1979) und Suppes, Krantz, Luce und Tversky (1989).

In diesem Abschnitt soll nun exemplarisch gezeigt werden, wie man in stochastischen Meßmodellen ***Konstrukte konstruieren*** kann. In einen Meßmodell werden die Verknüpfungen zwischen theoretischen und empirischen Begriffen expliziert. Ohne ein Meßmodell wären aus Aussagen über theoretische Begriffe keine logischen Implikationen für die Empirie möglich, d.h. erst durch ein Meßmodell kann eine Theorie zu einer ***empirischen*** Theorie werden, in dem Sinn, daß aus ihr Aussagen über die Empirie logisch ableitbar sind. AUS der Sicht einer deduktivistischen Methodologie (s. Erdfelder & Bredenkamp, Kap. 14 in diesem Band) sind also Meßmodelle von zentraler wissenschaftstheoretischer Bedeutung.

Zur Illustrierung eines solchen Meßmodells wähle ich das mathematisch einfachste und bekannteste stochastische Meßmodell, nämlich das ***Modell essentiell T-äquivalenter Variablen.*** Die analoge Argumentation kann man aber z.B. mit dem Rasch-Modell (s. dazu Steyer, 1989, oder ausführlicher, Steyer & Eid, 1993) und mit jedem anderen stochastischen Meßmodell führen. Die Auswahl des Modells essentiell **τ-äquivalenter** Variablen zur exemplarischen Darstellung eines stochastischen Meßmodell ist nicht nur wegen seiner Einfachheit und Verbreitung, sondern auch dadurch begründet, daß seine Struktur als Meßmodell lange Zeit verkannt wurde. So wurden z.B. die Modelle der klassischen Testtheorie von Suppes und Zinnes (1963) als atheoretisches ,,pointer measurement“ (p.21) abgetan. Dieser Vorwurf trifft m.E. jedoch nur auf viele unbedachte Anwendungen, vielleicht auch auf atheoretische Darstellungen (s. z.B. Lienert, **1961),** nicht jedoch auf die Modelle selbst zu, wie in diesem Abschnitt gezeigt werden soll.

3.1 Grundbegriffe der Klassischen Theorie psychometrischer Tests

Bevor das eigentliche Meßmodell eingeführt werden kann, müssen zuvor einige Grundbegriffe bereitgestellt werden, die als Antwort auf das ***Meßfehlerproblem*** entwickelt wurden.

Angenommen wir wollen aufgrund eines Testergebnisses Aussagen über eine bestimmte Persönlichkeitseigenschaft der getesteten person machen. Würden wir nun den Test mehrmals vorlegen, dann würden wir feststellen, daß die Person bei der ersten Vorgabe andere Antworten auf die Fragen (Items) eines Persönlichkeitstests gibt, als bei der zweiten oder dritten Meßgelegenheit. Auch bei der Verrechnung der Antworten zu einem Gesamttestwert würden wir in der Regel feststellen, daß der resultierende Testwert bei jeder Meßgelegenheit ein anderer wäre. Demnach müßten wir dann auch verschiedene Aussagen über die betreffende Persönlichkeitseigenschaft der Person machen, je nachdem, ob wir die erste, zweite oder dritte Meßgelegenheit betrachten.

Zur Erklärung des Sachverhalts, daß die Testwerte der Person bei jeder der drei Meß-Gelegenheiten anders ausfallen, sind prinzipiell drei Möglichkeiten denkbar:

(a) Die zu messende Eigenschaft verändert sich zwischen den Messungen.
(b) Die Unterschiede kommen durch Meßfehler zustande.
(c) Sowohl Meßfehler als auch Veränderungen der Eigenschaft sind für die Unterschiede verantwortlich.

Erklärung (a) scheint nur sinnvoll, wenn die beobachteten Veränderungen der Testwerte systematisch sind, wenn die betreffende Person z.B. auf ***allen*** Items des Tests höhere Werte als bei einer früheren Vorgabe erreicht. Veränderungen der Antworten zwischen zwei Meßgelegenheiten fallen jedoch oft unsystematisch und widersprüchlich aus. Bei einigen Items wird die Person gleiche Antworten geben, bei anderen erzielt sie höhere und bei wieder anderen niedrigere Werte. In derartigen Fällen, die sehr oft auftreten, scheint die Erklärung (a) unbefriedigend.

Bei den beiden Erklärungsmöglichkeiten (b) und (c) wird vorausgesetzt, daß sich der beobachtete Testwert aus einem Meßfehler und einem Wert zusammensetzt, der die tatsächliche Eigenschaft der Person repräsentiert. Für jede Person wird also ein Wert gedacht, der ***nicht*** mit einem Meßfehler behaftet ist, wofür sich die Bezeichnung wahrer ***Wert*** eingebürgert hat. Prinzipiell läßt sich dies auf verschiedene Weise präzisieren.

Wahrer Wert und Fehler: Die in der Psychologie am häufigsten verwendete Präzisierung der Vorstellung eines ***wahren Werts*** einer Person ist seine Definition als Erwartungswert einer Testwertvariablen Y_i bzgl. der intraindividuellen (d.h. der Person-bedingten) Verteilung von Y_i (Gulliksen, 1950; Lord & Novick, 1968). Betrachtet man ein Zufallsexperiment, bei dem eine Person zufällig aus einer Menge U von Personen (der Population) gezogen wird und ihre Werte in m Tests oder Testteilen festgestellt werden, so kann man neben den m Testwertvariablen Y_i, die durch die entsprechenden Auswertungsvorschriften der Tests definiert sind, die zugehörigen True-Score-Variablen τ_i betrachten, deren Werte die eben definierten wahren Werte der Person sind. Entsprechend werden die ***Meßfehlervariablen*** ε_i als Differenz $Y_i - \tau_i$ definiert. Bei dieser Präzisierung der Idee eines wahren Wertes wird also jede Testwertvariable Y_i additiv in ihre True-Score- und ihre Fehlervariable dekomponiert, d.h. für $i = 1, \ldots, m$ gilt:

$$Y_i = \tau_i + \varepsilon_i. \tag{3.1}$$

Aus den oben beschriebenen Definitionen der True-Score- und der Fehlervariablen lassen sich u. a. die folgenden Eigenschaften ableiten (s. z.B. Knoche, 1990; Novick, 1966; Steyer, 1989; Tack, 1980; Zimmerman, 1975):

$$E(\varepsilon_i) = 0, \tag{3.2}$$

$$Cov(\varepsilon_i, \tau_j) = 0, \tag{3.3}$$

$$Var(Y_i) = Var(\tau_i) + Var(\varepsilon_i), \tag{3.4}$$

$$E(\varepsilon_i | \tau_j) = 0. \tag{3.5}$$

Diese Gleichungen sind Spezialfälle der Eigenschaften des Residuums bzgl. einer bedingten Erwartung oder Regression (s. z.B. Steyer, 1992 oder Steyer & Eid, 1993). Keine dieser Gleichungen kann in einer empirischen Anwendung falsch sein, es sei denn, man ginge von einer anderen Definition der True-Score- und der Fehlervariablen aus. Dies ist ein Beispiel für den in der Einleitung angesprochenen Sachverhalt, daß ein Teil einer Theorie schon aus logischen Gründen wahr ist. An dieser Stelle kann also eine Theorienrevision nicht ansetzen, wohl dagegen bei den zugrundegelegten Definitionen und Annahmen. Letztere sind allerdings rein formaler Natur und stellen für Anwendungen keine wesentliche Restriktionen dar, da sie lediglich das betrachtete empirische Phänomen strukturieren und so einer mathematischen Behandlung zugänglich machen. Die einzige Voraussetzung, die zu den Gleichungen 3.1 bis 3.5 führt, ist die Endlichkeit der Varianz der Y-Variablen.

Man beachte, daß die in den älteren Darstellungen der Klassischen Testtheorie (Gulliksen, 1950; Lehmann, 1983; Lord & Novick, 1968; Wottawa, 1980) als Axiom betrachtete Annahme der Unkorreliertheit der Fehlervariablen ε_i untereinander keine Folgerung aus den oben vorgenommenen Definitionen der τ_i bzw. ε_i ist, d.h. es gilt nicht unbedingt $Cov(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$. Die Unkorreliertheit der Fehlervariablen ***untereinander*** kann in empirischen Anwendungen also durchaus falsch sein (s. dazu Tack, 1980; Zimmerman & Williams, 1977). Tatsächlich ist die Unkorreliertheit der Fehlervariablen zwar für viele mathematische Ableitungen bequem, aber nicht unbedingt für alle Zwecke notwendig.

Reliabilität: Eine notwendige Voraussetzung für die Brauchbarkeit einer Meßwertvariablen Y_i und des ihr zugrundeliegenden Meßinstruments ist, daß ihre Varianz nicht ausschließlich aus Fehlervarianz besteht. Der andere Extremfall, in dem die Fehlervarianz $Var(\varepsilon_i)$ gleich 0 ist, ist zwar erstrebenswert, aber in den Sozialwissenschaften eher die Ausnahme als die Regel. Eine einfache Kenngröße, die das Ausmaß der Fehlerbehaftetheit bzw. ***Unreliabilität*** einer Variablen Y_i angibt, ist der Varianzanteil $Var(r_i)$ / $Var(Y_i)$.

Dividiert man beide Seiten der Gleichung 3.4 durch die Varianz von Y_i, so erhält man: 1 = Var(Yi) / Var(Yi) = [Var(ti) / Var(Yi)] + [Var(ei) / Var(Yi)]. Den Kennwert

$$Rel(Y_i) := 1 - Var(\varepsilon_i) \;/\; Var(Y_i) = Var(\tau_i) \;/\; Var(Y_i) \tag{3.6}$$

bezeichnet man als ***Reliabilität.*** Man kann zeigen, daß es sich dabei um einen speziellen Determinationskoeffizienten handelt (s. Gl.2.5). Es gilt nämlich außer Gleichung 3.6 auch: $\boldsymbol{Rel(Y_i)}$ = $Var[E(Y_i|r_i]$ / $\boldsymbol{Var(Y_i)}$. Damit sind nun die Grundbegriffe bereitgestellt, mit denen sich das Modell essentiell τ-äquivalenter Variablen formulieren läßt.

3.2 Das Modell essentiell τ-äquivalenter Variablen

Ausgangspunkt sind m Testwertvariablen Y_i sowie deren True-Score- und Fehlervariablen τ_i bzw. ε_i. Die ***Annahme der essentiellen τ-Äquivalenz*** läßt sich nun wie folgt formulieren: Zu jedem Paar i, j = 1, . . ., m existiert eine reelle Konstante h_{ij}, für die gilt:

$$\tau_i = \tau_j + \lambda_{ij}. \tag{3.7}$$

Diese Annahme besagt, daß die True-Score-Variablen aller m betrachteten Y-Variablen identisch sind bis auf eine Verschiebung um eine additive Konstante (d.h. eine Translation), nämlich h_{ij}.

3.2.1 Existenz

Betrachten wir nun den Fall j = 1 und bezeichnen τ_1 mit η und $-A_{i1}$ mit λ_i, dann resultiert aus der obigen Gleichung für alle i = 1, . . ., m:

$$\tau_i = \eta - \lambda_i. \tag{3.8}$$

Diese Gleichung bezeichne ich als das ***Fundamentalgesetz der subtraktiven Parametrisierung des Modells essentiell giquivalenter Variablen.*** Ich ziehe hier die subtraktive Parametrisierung gegenüber der (logisch äquivalenten) additiven Parametrisierung $\tau_i = \eta + \kappa_i$ vor, weil sich h_i in Gleichung 3.8 in vielen Anwendungen als ***Schwierigkeit*** des Tests interpretieren läßt. Aus der Annahme der essentiellen τ-Äquivalenz folgt also die ***Existenz*** einer latenten Variablen η, die für alle Y-Variablen gemeinsam ist, und eines für die Variable Y_i spezifischen Kennwerts L_i. Setzt man nämlich die Gleichungen 3.8 und 3.1 zusammen, dann erhält man für alle $i = 1, \ldots, m$:

$$Y_i = \eta - \lambda_i + \varepsilon_i. \tag{3.9}$$

Dabei ist die latente Variable η eine theoretische Größe, deren Werte in den üblichen Anwendungen die Fähigkeit, Einstellung oder Eigenschaft der gezogenen Person charakterisiert, wohingegen λ_i ein Kennwert für den durch die Variable Y_i repräsentierten Test(tei1) ist.

3.2.2 Eindeutigkeit und Bedeutsamkeit

Offenbar ist aber η durch die Annahme der essentiellen τ-Äquivalenz nicht völlig ***eindeutig*** definiert, denn wir haben η oben willkürlich mit z_1 gleichgesetzt. Wir hätten η ebensogut mit z_2 oder mit $z_1 + \alpha$ gleichsetzen können, wobei α eine beliebige reelle Zahl sein kann. Auch dann kommt man zu den Gleichungen 3.8 und 3.9. Demnach ist η nur eindeutig bis auf Translationen, d.h. Verschiebungen um eine additive Konstante definiert. Mit anderen Worten, η ist ***differenzskaliert.*** Dieses Skalenniveau reicht aus, damit bspw. Aussagen über die Varianz von η und über die Reliabilität der Y_i im meßtheoretischen Sinn ***bedeutsam*** sind. Der Wahrheitswert dieser Aussagen ist also ***invariant unter den zulässigen Transformationen.*** Auf dem gleichen Weg kann man sich überlegen, daß auch die Größen λ_i differenzskaliert sind.

Alle obigen Überlegungen basieren auf einer einzigen Annahme, der essen***tiellen τ-Äquivalenz.*** Mit der Gültigkeit dieser Annahme stehen und fallen z.B. alle Aussagen über die Existenz und Eindeutigkeit der theoretischen Variablen η. Daher ist es angebracht zu überlegen, ob und wann diese Annahme in empirischen Anwendungen erfüllt sein kann und wie man sie gegebenenfalls überprüfen kann.

In der Regel werden zwei Parallelformen eines Tests gerade so konstruiert, daß die durch die betreffenden Auswertungsvorschriften definierten beiden Testwertvariablen Y_1 und Y_2 essentiell τ-äquivalent sind. Dabei ist allerdings zu bedenken, ob nicht jede lineare Transformation $Y'_i = a + b\,Y_i$, $i = 1, 2$, die gleiche Information wie die ursprünglichen beiden Variablen Y_i enthält. In diesem Fall wäre zu überlegen, ob man nicht ein Meßmodell konstruiert, in dem nicht die beiden ***Variablen Y_i,*** sondern die beiden ***Familien*** aller linearen Transformation der beiden Y_i den Ausgangspunkt darstellen. Dies hätte den Effekt, daß die einzuführende theoretische Größe η nur intervallskaliert wäre.

3.2.3 *Testbarkeit*

Die erste empirisch prüfbare Konsequenz aus dem Modell essentiell τ-äquivalenter Variablen ist die Gleichheit der Erwartungswerte der Differenzvariablen $Y_i - Y_j$ in Subpopulationen $U^{(1)}, U^{(2)} \subset U$. Dabei wird keine weitere Annahme als die der essentiellen t-Aquivalenz vorausgesetzt (zum Beweis s. Steyer, 1988, 1989, oder auch Steyer & Eid, 1993):

$$E^{(1)}(Y_i - Y_j) = E^{(2)}(Y_i - Y_j), \quad i, j = 1, ..., m. \tag{3.10}$$

In Anwendungen kann man also die Annahme der essentiellen τ-Äquivalenz prüfen, indem man untersucht, ob die Erwartungswerte der Differenzvariablen $Y_i - Y_j$, i, j = **1,** ***m,*** in zwei ***verschiedenen*** Subpopulationen identisch sind. Die obige Gleichung ist eine Hypothese über die Gleichheit der Erwartungswerte von m $(m - 1)/2$ (Differenz-)Variablen. Fügt man die Annahmen eines Stichprobenmodells, insbesondere die Annahmen der Normalverteilung und der Varianzhomogenität, hinzu, dann ist im Fall m = 2 das statistische Standardverfahren zur Prüfung der o.g. Hypothese der t-Test und im Fall $m > 2$ der multivariate t-Test, der sich z.B. mit der SPSS-Prozedur MANOVA leicht durchführen läßt.

Der zweite Typ empirisch überprüfbarer Folgerungen aus der Annahme der essentiellen τ-Äquivalenz betrifft die Kovarianzen der betrachteten Y-Variablen. Dabei wird allerdings die Bedingung der essentiellen τ-Äquivalenz in Konjunktion mit einer zusätzlichen Annahme geprüft, nämlich der Unkorreliertheit der Fehlervariablen ε_i untereinander:

$$Cov(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j. \tag{3.11}$$

Da η eine Translation (Verschiebung um eine additive Konstante) von einer True-Score-Variablen τ_i ist, von der wir schon nach Gleichung 3.3 wissen, daß sie unkorreliert mit den Fehlervariablen ist, gilt außerdem für $i = 1, \ldots, m$:

$$Cov(\varepsilon_i, \eta) = 0. \tag{3.12}$$

Diese beiden Gleichungen führen nun zu der in der folgenden Gleichung angegebenen Struktur der Kovarianzen der Y-Variablen, wobei $i, j = 1, \ldots, m$:

$$Cov(Y_i, Y_j) = \begin{cases} Var(\eta), & i \neq j, \\ Var(\eta) + Var(\varepsilon_i), & i = j. \end{cases} \tag{3.13}$$

Gemäß dieser Gleichung haben verschiedene essentiell τ-äquivalente Variablen Y_i und Y_j, $i \neq j$, jeweils die gleiche ***Kovarianz,*** falls die Fehler unkorreliert sind. Diese Kovarianz ist zugleich die Varianz von η. Dabei beachte man, daß hier nicht von den Stichprobenkovarianzen die Rede ist, sondern von den wahren (Populations-)Kovarianzen.

Zur Illustration sei der Spezialfall $m = 3$ betrachtet. Die Kovarianzmatrix Σ der Y-Variablen kann dann gemäß Gleichung 3.13 wie folgt notiert werden:

$$\Sigma = \begin{bmatrix} Var(\eta) + Var(\varepsilon_1) & Var(\eta) & Var(\eta) \\ Var(\eta) & Var(\eta) + Var(\varepsilon_2) & Var(\eta) \\ Var(\eta) & Var(\eta) & Var(\eta) + Var(\varepsilon_3) \end{bmatrix} .. \tag{3.14}$$

Als theoretische Parameter kommen hier also die Varianzen von η und die Varianzen der Fehlervariablen vor. Gemäß Gleichung 3.13 impliziert das Modell essentiell τ-äquivalenter Variablen mit unkorrelierten Fehlern die Gleichheit der Kovarianzen $Cov(Y_1, Y_2) = Cov(Y_1, Y_3) = Cov(Y_1, Y_3)$. Im Fall m = 3 hat das Modell also bereits empirisch testbare Konsequenzen für die Kovarianten der Y-Variablen, die ja, falls kein Modell essentiell τ-äquivalenter Variablen vorläge, beliebige Werte annehmen könnten.

Zur praktischen Durchführung eines solchen Tests kann man ein Programm zur Analyse von Kovarianzstrukturen (zur Einführung in Kovarianzstrukturmodelle s. z.B. Bollen, 1989, Hayduk, 1987, oder Saris & Stronkhorst, 1984) wie z.B. LISREL 7 (Jöreskog & Sörbom, 1989) verwenden. Dies wird ausführlich von Steyer und Eid (1993) demonstriert, die auch eine weitere testbare Konsequenz des Modells essentiell τ-äquivalenter Variablen behandeln.

3.2.4 Identifizierbarkeit

Die Gleichung 3.13 zeigt nicht nur, daß die Kovarianzmatrix der Y-Variablen eine bestimmte Struktur haben muß, wenn die beiden Modellannahmen gelten, sondern auch, wie man die theoretischen Parameter, die Varianzen von η und %$_i$ aus den Varianzen und Kovarianzen der Y-Variablen bestimmen (identifizieren) kann.

Die nächsten beiden Gleichungen geben an, wie die Varianzen ***Var(q)*** und ***Var(E$_i$)*** der latenten Variablen aus den Varianzen und Kovarianzen der beobachtbaren Variablen Y_i bestimmt werden können, wenn die Annahmen der essentiellen τ-Äquivalenz und unkorrelierter Fehler gültig sind. Unter diesen beiden Annahmen gelten für i, j = 1, . . ., *m:*

$$Var(\eta) = Cov(Y_i, Y_j), \quad i \neq j, \tag{3.15}$$

$$Var(\varepsilon_i) = Var(Y_i) - Cov(Y_i, Y_j), \quad i \neq j. \tag{3.16}$$

Gemäß diesen Gleichungen sind die Varianzen $Var(\eta)$ und $Var(\varepsilon_i)$ im Modell essentiell τ-äquivalenter Variablen bereits bei m = 2 Y-Variablen identifiziert, da nur von zwei verschiedenen Indizes *i* und j die Rede ist.

Die ***Varianzanteile*** der Y_i, die durch die latente Variable η determiniert sind, die Reliabilitäten also, können nach der folgenden Formel bestimmt werden, wenn die Annahmen der essentiellen τ-Äquivalenz und unkorrelierter Fehler gelten. Dann folgt nämlich für ***i, j*** = 1, . . ., ***m:***

$$Rel(Y_i) = \frac{Cov(Y_i, Y_j)}{Var(Y_i)}, \quad i \neq j. \tag{3.17}$$

Sind außerdem auch noch die Fehlervarianzen gleich, d.h. gilt $Var(E_i) = Var(\varepsilon_j)$, i ≠ j, dann folgt auch

$$Kor(Y_i, Y_j) = Rel(Y_i). \tag{3.18}$$

Dies ist die Gleichung, die den Verfahren zur Bestimmung der Reliabilität über die ***Paralleltest-*** und über die ***Retestkorrelation*** zugrundeliegt.

3.3 Zusammenfassende Bemerkungen

Stochastische Meßmodelle dienen zwei wichtigen Zwecken: Zum einen wird in einem solchen Modell die theoretische, vom Meßfehler bereinigte Größe konstruiert und zum anderen erlauben solche Modelle abzuschätzen, wie groß der Meßfehler ist, mit dem man bei einer Messung rechnen muß. Die einzelnen Schritte bei der Konstruktion eines Konstrukts im Rahmen eines stochastisehen Meßmodells kann man wie folgt zusammenfassen: Ausgangspunkt sind Zufallsvariablen Y_i, $i = 1, \ldots, m$, deren Werte beobachtbares Verhalten der betrachteten Beobachtungseinheit (hier: einer Person) repräsentieren. Die Modellannahmen bestehen aus Aussagen über bestimmte Eigenschaften der Verteilungen dieser Zufallsvariablen (hier: ihrer person-bedingten Erwartungswerte - der True scores - und deren Residuen). Daraus läßt sich die ***Existenz*** einer oder mehrerer theoretischer Größen (hier η und die Parameter λ_i) deduzieren, und man kann feststellen, wie ***eindeutig*** diese theoretischen Größen durch die Modellannahmen definiert sind. Dies ist gleichbedeutend mit der Feststellung ihres ***Skalenniveaus.*** Daraus läßt sich ableiten, welche Aussagen, die die theoretischen Größen involvieren, ***bedeutsam*** sind, in dem Sinn, daß der Wahrheitswert dieser Aussagen invariant ***unter den zulässigen Transformationen*** ist. In einem weiteren Schritt kann man untersuchen, wie sich die theoretischen Größen aus den Verteilungskennwerten der Ausgangsvariablen Y_i bestimmen oder ***identifizieren*** lassen. Um die Überprüfung des Modells zu ermöglichen, läßt sich schließlich feststellen, welche ***empirisch testbaren Konsequenzen*** sich aus den Modellannahmen ableiten lassen.

Die statistischen Tests und Schätzungen der Parameter basieren dann auf weiteren Annahmen wie z.B. der unabhängigen identischen Verteilung in der Stichprobe, die durch das ***N-malige*** Durchführen des oben behandelten Zufallsexperiments entsteht. Oft (aber nicht immer, s. z.B. Browne, 1984) sind auch Annahmen über bestimmte Verteilungsklassen (wie z.B. Normalverteilung) nötig. Diese statistischen Annahmen und Modelle sind zwar für Anwendungen unerläßlich, aber für die Logik des eigentlichen stochastischen Modells, das oben skizziert wurde, das sich auf das zufällige Ziehen ***einer einzigen*** Beobachtungseinheit bezieht, irrelevant.

4. Zusammenfassung und Ausblick

In diesem Beitrag wurde dargelegt, wie in stochastischen Modellen ihre beiden Aufgaben, die Explikation von ***Meßmodellen*** und von ***Abhängigkeitsbegriffen*** gelöst werden können. Beide Aufgaben dienen dem für eine deduktivistische Methodologie zentralen Ziel, die ***logische Ableitung von Aussagen über die Empirie aus der Theorie zu ermöglichen.*** Nur dann, wenn eine solche logische Ableitbarkeit möglich ist, sind theoretische Aussagen auch falsifizierbar. Um die Deduzierbarkeit empirischer Aussagen zu gewährleisten, sind mindestens zwei Dinge vonnöten: ***Erstens*** müssen die in der Theorie vorkommenden theoretischen Begriffe (wie z. B. ,,Frustration“ und ,,Aggression“) mit Beobachtbarem verknüpft werden, was durch Meßmodelle geschieht (s. dazu Abschnitt 3). ***Zweitens*** müssen aber auch die postulierten Abhängigkeiten zwischen den theoretischen Begriffen in einer Sprache formuliert werden, die logische Ableitungen für beobachtbare Sachverhalte erlaubt. Die umgangssprachliche Formulierung ,,Frustration führt zu Aggression“ gestattet vielleicht Folgerungen nach Plausibilitätsüberlegungen, aber keinerlei logische Ableitungen, die m.E. in einer entwickelten wissenschaftlichen Disziplin zu fordern sind. Genau an dieser Stelle können wir die in Abschnitt 2 behandelten stochastischen Abhängigkeitsbegriffe verwenden. Sie stellen die Sprache zur Formulierung der Aussagen über die ***Abhängigkeiten*** zwischen den theoretischen Begriffen bereit. Dies ist nicht nur für die Wissenschaft, sondern auch für die Gesellschaft von großer Bedeutung, wenn man die möglichen Anwendungsfelder in der Epidemiologie (,,die radioaktiven Emissionen ***führen*** zu Leukämie“), Ökologie („die $C0_2$-Abgase ***führen*** zu Waldsterben“), Sozialpolitik (,,Arbeitslosigkeit ***führt*** zu Alkoholismus“), Verkehrspolitik („Verbreiterung der Straße ***reduziert*** die Unfallgefahr“) etc. bedenkt.

In der Psychologie sind meist nur ***stochastische*** Modelle realistisch, und zwar sowohl als Meßmodelle als auch als Abhängigkeitsmodelle. Die stochastische Natur dieser Modelle kompliziert zwar den Prozeß der Falsifikation durch die Notwendigkeit, eine ***Entscheidung*** über die Gültigkeit einer Hypothese aufgrund statistischer Untersuchungen und Überlegungen zu treffen (s. Ostmann & Wutke, Kap. 16 in diesem Band), aber dennoch bleibt die Falsifizierbarkeit das entscheidende Kriterium für eine Abgrenzung wissenschaftlicher empirischer Theorien von vorwissenschaftlichen Theorien. Letztere sind zwar notwendige Schritte im Wissenschaftsprozeß, sind aber m.E. nur als Zwischenlösungen anzusehen.

4.1 Stochastische Meßmodelle

Stochastische Meßmodelle explizieren die Verknüpfung zwischen theoretischen und empirischen Begriffen. Ohne eine klare Verknüpfung ihrer theoretischen Begriffe mit Beobachtbarem wären Disziplinen wie z.B. die Psychologie, Soziologie oder Wirtschaftswissenschaften keine empirischen Wissenschaften, sondern Philosophie. Neben der Verknüpfung zwischen Theorie und Empirie können Meßmodelle einige allgemeine Probleme der Sozialwissenschaften lösen, die im folgenden kurz diskutiert werden.

Das Meßfehlerproblem: Sozialwissenschaftliche Beobachtungen und Messungen sind in der Regel meßfehlerbehaftet. Stochastische Meßmodelle können diese Meßfehler berücksichtigen und darüber Auskunft geben, wie ***stark*** diese Meßfehlerbehaftetheit ist. (Siehe hierzu die Meßfehlervarianz und die Reliabilität in Abschnitt 3.)

Das Problem situationaler Spezifität: Neben dem Meßfehlerproblem erschwert auch die Veränderung und Variabilität der zu messenden Eigenschaften und Zustände von Personen das Messen in der Psychologie. Das entsprechende gilt auch für die Messung von Eigenschaften anderer Objekte wie z. B. Gruppen oder sozialer Institutionen. Wie können wir z.B. abschätzen, welcher Anteil eines Testwerts auf Meßfehler, welcher Anteil auf die Besonderheit der Situation, in der der Test vorgelegt wird, und welcher Anteil auf die zu messende überdauernde Eigenschaft des betrachteten Objekts zurückgeht? Während für das Meßfehlerproblem schon seit einigen Jahrzehnten Lösungen existieren, gibt es erst in jüngster Zeit befriedigende Lösungsmöglichkeiten für das Problem der situationalen Spezifität von Messungen, die auch für nichtexperimentelle Beobachtungsstudien brauchbar sind (s. z. B. Schmitt & Steyer, 1990; Steyer, 1987, 1988; Steyer & Schmitt, 1990a, b; Steyer, Ferring & Schmitt, 1992).

Das Problem der situationalen Spezifität ist grundsätzlich nur in solchen Modellen zu lösen, in denen wiederholte Messungen berücksichtigt werden. Im Rahmen von Strukturgleichungsmodellen sind dazu die Bücher von Möbus und Schneider (1986) sowie die Artikel von Möbus und Nag1 (1983) und von Jöreskog und Sörbom (1977) zu nennen. Auch im Rahmen der probabilistischen Testtheorie gibt es dazu Ansätze (s. z.B. Andersen, 1985, 1988; Fischer, 1989; Langeheine & van de Pol, 1990a, b; Rost & Spada, 1983).

Das Problem der Methodenspezifität: In vielen Anwendungen der obigen Modelle hat sich herausgestellt, daß die Berücksichtigung von Meßfehlern und situativen Effekten noch immer nicht differenziert genug ist, um der tatsächlichen Komplexität psychologischer Messungen gerecht zu werden. Es stellte sich nämlich heraus, daß es auch noch ***Meßmethoden-spezifische Effekte*** gibt.

Selbst wenn Tests konstruiert werden, um die gleiche Eigenschaft zu messen, gelingt dies oft nicht perfekt. Neben der zu messenden Eigenschaft hängen die Testwerte auch noch von systematischen Effekten ab, die für den Test (das Meßinstrument) spezifisch sind. Dies kommt dadurch zum Ausdruck, daß die Testwertvariablen Y_{ik} und Y_{il}, die durch den gleichen Test i zu zwei verschiedenen Meßgelegenheiten k und l erhoben werden, höher miteinander korrelieren, als es durch Modelle erklärbar wäre, die nur Meßfehler, und situative Effekte berücksichtigen. Jöreskog (1979), Saris und van Meurs (1990) sowie Steyer, Ferring und Schmitt (1992) behandeln daher Modelle, in denen neben den o.g. Problemen auch das Problem der ***Methoden-Spezifität*** berücksichtigt wird.

Das Problem heterogener Populationen: Ein weiteres Problem, das man inzwischen im Rahmen stochastischer Meßmodelle lösen kann, ist das ***Problem heterogener Populationen.*** Oft zeigt sich, daß sich Personen in ihren ***Antwortstilen*** bei der Beantwortung von Items eines Fragebogens voneinander unterscheiden. Dies kann man dadurch in einem stochastischen Meßmodell berücksichtigen, daß man davon ausgeht, daß eine heterogene Population vorliegt, bei der in jeder Subpopulation eine andere Beziehung zwischen den Antworten auf die Items und den latenten Variablen besteht. So bedeutet „volle Zustimmung“ bei einer Person mit einer Tendenz zu extremen Antworten in bezug auf die zu messende Einstellung nicht das gleiche, wie die gleiche Antwort einer Person, die keine solche Tendenz zu Extremantworten aufweist (s. dazu Rost, 1991; Rost & Langeheine, 1991).

4.2 Stochastische Abhängigkeitsbegriffe

Neben den oben diskutierten Problemen, die im Rahmen stochastischer Meßmodelle gelöst werden können, sind stochastische Modelle aber auch zur Formulierung von Hypothesen über ***Abhängigkeiten*** zwischen theoretischen Variablen, aber auch zwischen empirischen Variablen unerläßlich. Die Relevanz stochastischer Abhängigkeitsbegriffe für eine deduktivistische Methodologie ergibt sich aus folgender Überlegung: Kommen an nur einer Stelle stochastisehe Begriffe wie Erwartungswert, Wahrscheinlichkeit, Korrelation o.ä. vor, und dies trifft auf den größten Teil empirischer Untersuchungen zu, so sind nur dann Deduktionen möglich, wenn die theoretischen Aussagen selbst bereits in Termini der Wahrscheinlichkeitstheorie formuliert sind oder wenigstens direkt in diese Sprache übersetzt werden können.

Beispiele für verschiedene in psychologischen Hypothesen vorkommende Abhängigkeitsbegriffe wurden ausführlich im Abschnitt 2 behandelt. Es waren dies die Begriffe der einfachen regressiven Abhängigkeit, der korrelativen Abhängigkeit, der partiellen regressiven Abhängigkeit, der bedingten regressiven

Abhängigkeit, sowie der starken und der schwachen kausalen regressiven Abhängigkeit. Auch die mit den verschiedenen Arten stochastischer Abhängigkeit behandelten Präzisierungen von Abhängigkeitsbegriffen dienen letztlich dazu, dem Hauptanliegen der deduktivistischen Methodologie, der logischen Ableitbarkeit empirisch falsifizierbarer Aussagen, Rechnung zu tragen.

4.3 Ausblick

Stochastische Modelle dienen nicht nur der Lösung spezieller Probleme innerhalb der Mathematischen Psychologie, obwohl sie auch dort zweifelsohne eine wichtige Rolle spielen (s. z.B. Colonius, 1984; DeGreef & van Buggenhaut, 1984; Fischer, ***1993;*** Holling, ***1989;*** Roskam & Suck, 1987). In diesem Beitrag sollte stattdessen gezeigt werden, daß sie eine zentrale und unverzichtbare Funktion für alle Bereiche der „gewöhnlichen" empirischen Sozialwissenschaften einnehmen, jedenfalls dann, wenn man sich zu einer deduktivistischen Methodologie bekennt, deren Kernpostulat die Falsifizierbarkeit empirischer wissenschaftlicher Theorien ist.

Dazu sei angemerkt, daß man auch den Strukturalismus (Balzer, Moulines & Sneed, ***1987;*** Gähde, 1983; Gähde & Stegmüller, 1986; Gähde, Jagodzinski & Steyer, 1993; Stegmüller, 1979, 1986; Stephan, ***1990;*** Westermann, 1987; Westmeyer, 1989, 1992) als einen wichtigen Ansatz im Rahmen einer deduktivistischen Methodologie empirischer Wissenschaften ansehen kann, ist doch sein Hauptanliegen, die ***logische Struktur wissenschaftlicher Theorien (s.*** Sneed, 1971) zu explizieren. Auch wenn sich in den Untersuchungen der strukturalistischen Wissenschaftstheorie herausgestellt hat, daß bestimmte Theorieteile in der Regel für sich genommen nicht falsifizierbar sind, ist damit das Falsifizierbarkeitspostulat keineswegs ad acta gelegt. In Termini des Strukturalismus bedeutet es nichts anderes, als daß der ***empirische Gehalt*** (s. Diederich, 1981) einer empirischen wissenschaftlichen Theorie nicht leer sein darf.

Literatur

Andersen, E. B. **(1985).** Estimating latent correlations between repeated testings. ***Psychometrika,*** 50, **3-16.**

Andersen, E.B. (1989). Comparison of latent structure models. In R. Langeheine & J. Kost (Eds.), ***Latent trait and latent class models*** (pp. 207-229). New York: Plenum.

Andersen, E. B. (1990). ***The statistical analysis of categorical data.*** Berlin: Springer.

Ash, R. A. (1972). ***Real analysis and probability.*** New York: Academic Press.

Austin J.T. & Wolfle, L.M. (1991). Annotated bibliography of structural equation modelling: Technical work. ***British Journal of Mathematical and Statistical Psychology,* 44, 93-152.**

Balzer, W., Moulines, C.U. & Sneed, J.D. **(1987).** ***An architectonic for*** science. Dordrecht: Reidel.

Bartussek, D. **(1970).** Eine Methode zur Bestimmung von Moderatoreffekten. ***Diagnostica, 16, 57-76.***

Bauer, H. **(1978).** ***Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*** (3. Auflage). Berlin: de Gruyter.

Bentler, P. M. (1986). Structural modeling and Psychometrika: An historical perspective on growth and achievements. ***Psychometrika, 51, 35-51.***

Bishop, Y.M.M., Fienberg, S.E. & Holland, ***P.*** W. (1977). ***Discrete multivariate analysis.*** Boston: MIT Press.

Bollen, K.A. (1989). ***Structural equations with latent variables. New*** York: Wiley.

Bosch, K. ***(1986). Elementare Einführung in die Wahrscheinlichkeitsrechnung.*** Braunschweig: Vieweg.

Bredenkamp, J. (1984a). Theoretische und experimentelle Analysen dreier Wahrnehmungstäuschungen. ***Zeitschrift für Psychologie,* 192, 47-61.**

Bredenkamp, J. (1984b). Theoretische und experimentelle Analysen einiger Wahrnehmungstäuschungen. ***Archiv für Psychologie, 136,* 281-291.**

Breiman, L. **(1968).** ***Probability.*** Reading, MA: Addison-Wesley.

Browne, M.W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. ***British Journal of Mathematical*** and ***Statistical Psychology,*** **37, 62-83.**

Cheek, J.M. (1982). Aggregation, moderator variables, and the validity of personality tests: A peer rating study. ***Journal of Personality and Social Psychology, 43,* 1254-1269.**

Colonius, H. (1984). ***Stochastische Theorien individuellen Wahlverhaltens.*** Berlin: Springer.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. Psychological ***Bulletin,*** 69, 161-182.

DeGreef, E. & van Buggenhaut, J. (1984) (Eds.). ***Trends in Mathematical Psychology.*** Amsterdam: North-Holland.

De Gruijter, D. N. M. & van der Kamp, L.J. T. (1984). ***Statistical models in psychological and educational testing.*** Lisse: Swets & Zeitlinger.

Diederich, W. (1981). ***Strukturalistische Rekonstruktionen: Untersuchungen zur Be***deutung, ***Weiterentwicklung*** und ***interdisziplinären Anwendung des strukturalistischen Konzepts wissenschaftlicher Theorien.*** Braunschweig: Vieweg.

Erdfelder, E. & Steyer, R. (1984). Zur Psychophysik einiger Größentäuschungen. Psy***chologische Beiträge, 26, 639-646.***

Falmagne, ***C.*** **(1985).** ***Elements of psychophysical theory.*** New York: Oxford University Press.

Fischer, ***G.*** H. ***(1974). Einführung in die Theorie psychologischer Tests.*** Bern: Huber.

Fischer, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. ***Psychometrika, 46, 59-77.***

Fischer, G. H. (1983). Neuere Testtheorie. In H. Feger & J. Bredenkamp (Hrsg.), Mes-***sen und Testen. Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie 1: Forschungsmethoden der Psychologie, Band 3 (S. 604-692).*** Göttingen: Hogrefe.

Fischer, G. H. (1988). Spezifische Objektivität: Eine wissenschaftstheoretische Grundlage des Rasch-Modells. In K. Kubinger (Hrsg.), ***Moderne Testtheorie (S.*** 87-11). Weinheim: Psychologie Verlags Union.

Fischer, G. H. (1989). An IRT-based model for dichotomous longitudinal data. Psy-***chometrika,*** 54, 599-624.

Fischer, G. H. (1993). Proceedings of the 22nd Meeting of the European Mathematical Psychology Group in Vienna. New York: Springer.

Formann, A. (1984). ***Die Latent-Class-Analyse.*** Weinheim: Beltz.

Gähde, U. (1983). ***T-Theoretizität und Holismus.*** Frankfurt am Main: Lang.

Gähde, U. & Stegmüller, W. (1986). An argument in favor of the Duhem-Quine-Thesis: From the structuralist point of view. In: L.E. Hahn & P.H. Schilpp (Eds.), ***The Philosophy*** of ***W.*** v. ***Quine.*** The library of living philosophers, Vol. 15, La Salle, Il.: Open Court Publishing Company.

Gähde, U., Jagodzinski, W. & Steyer, R. (1992). On a structuralist reconstruction of latent state-trait theory. In H. Westmeyer (Ed.), ***The structuralist program in psychology: Foundations and applications*** (pp. 105-119). Toronto: Hogrefe-Huber.

Gänssler, P. & Stute, W. ***(1977). Wahrscheinlichkeitstheorie.*** Berlin: Springer.

Ghiselli, E. E. (1963). Moderating effects and differential reliability and validity. ***Journal of Applied Psychology, 47,*** 81-86.

Goodman, L. A. & Kruskal, W. H. (1979). ***Measures of association for cross-classifications.*** New York: Springer.

Gulliksen, H. (1950). ***Theory of mental tests.*** New York: Wiley.

Härtter, E. (1987). ***Wahrscheinlichkeitsrechnung, Statistik und mathematische*** Grund-*lagen. Begriffe, Definitionen.* Göttingen: Vandenhoeck & Ruprecht.

Hager, W. (1987). Grundlagen einer Versuchsplanung zur Prüfung empirischer Hypothesen in der Psychologie. In ***G.*** Lüer (Hrsg.), ***Allgemeine experimentelle Psychologie (S.*** 43-264). Stuttgart: Fischer.

Hager, W. (1992). Eine Strategie zur Entscheidung über psychologische Hypothesen. ***Psychologische Rundschau,*** 43, 18-29.

Hambleton, R. H. & Swaminathan, H. (1985). ***Item response theory.*** Boston: Kluwer-Nijhoff.

Hartung, J., Elpelt, B. & Klösener, K.-H. (1989). ***Statistik. Lehr- und Handbuch der angewandten Statistik*** (7. Auflage). München: Oldenbourg.

Hayduk, L. A. (1987). ***Structural equation modeling with LISREL: Essentials and advances.*** Baltimore: Johns Hopkins University Press.

Hinderer, K. (1980). ***Grundbegriffe der Wahrscheinlichkeitstheorie.*** Berlin: Springer.

Holling, H. (1989). Psychische Beanspruchung durch Wartezeiten in der Mensch-Computer-Interaktion. Berlin: Springer.

Jöreskog, K. G. (1979). Statistical estimation of structural equation models in longitudinal-developmental investigations. In J. R. Nesselroade & P. B. Baltes (Eds.), ***Longitudinal research in the study of behavior and development*** (pp.303-352). New York: Academic Press.

Jöreskog, K. G. & Sörbom, D. (1977). Statistical models and methods for analysis of longitudinal data. In D.J. Aigner & A. S. Goldbergen (Eds.), ***Latent variables in socio-economic models*** (pp.285-325). Amsterdam: North-Holland.

Jöreskog, K. G. & Sörbom, D. (1989). ***Lisrel 7. A guide to the program und its applications*** (2nd edition). Chicago, IL: SPSS Inc.

Kenny, D. A. & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. ***Psychological Bulletin, 96,*** 201-210.

Knoche, ***N.*** (1990). ***Modelle der empirischen Pädagogik.*** Mannheim: B.I.-Wissenschaftsverlag.

Knoke, D. & Burke, P. J. (1980). ***Log-linear models.*** Beverly Hills, CA: Sage.

Kolmogoroff, A. ***(1970). Grundbegriffe der Wahrscheinlichkeitsrechung.*** Berlin: Sprin***ger.*** (Nachdruck der Originalarbeit von 1933).

Krantz, D.H., Luce, R.D., Suppes, P. & Tversky, A. (1971). ***Foundations of measurement. Vol I. Additive and polynomial representations.*** New York: Academic Press.

Krauth, J. & Lienert, G.A. (1973). ***KFA - Die Konfigurationsfrequenzanalyse.*** Freiburg: Alber.

Krickeberg, K. & Ziezold, H. ***(1979). Stochastische Methoden.*** Berlin: Springer.

Kubinger, K. (1988). ***Moderne Testtheorie.*** Weinheim: Psychologie Verlags Union.

Langeheine, R. (1982). Log-lineare Modelle. In J. van Koolwijk & M. Wieken-Mayser (Hrsg.), ***Techniken der empirischen*** Sozialforschung. ***Band*** 8. Kausalanalyse (S.122-195). München: Oldenbourg.

Langeheine, R. & Rost, J. (Eds.). (1988). ***Latent trait und latent class models. New*** York: Plenum.

Langeheine, R. & van de Pol, E (1990a). Veränderungsmessung bei kategorialen Daten. ***Zeitschrift für Sozialpsychologie, 21,*** 88-100.

Langeheine, R. & van de Pol, F. (1990b). A unifying framework for Markov modeling in discrete space and discrete time. ***Sociological Methods & Research, 18, 416-441.***

Lehmann, G. (1983). Testtheorie. In H. Feger & J. Bredenkamp (Hrsg.), ***Messen*** und ***Testen. Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie Z: Forschungsmethoden der Psychologie, Band 3 (S. 427-543).*** Göttingen: Hogrefe.

Lewis, C. (1986). Test theory and Psychometrika: The past twenty-five years. Psycho***metrika, 51,*** 11-22.

Lienert, G.A. (1961). ***Testaufbau*** und ***Testanalyse.*** Weinheim: Beltz.

Loeve, M. ***(1977). Probability theory*** I (4th edition). New York: Springer.

Loeve, M. (1978). ***Probability theory*** II (4th edition). New York: Springer.

Lord, F. M. & Novick, M. R. (1968). ***Statistical theories of mental test*** scores. Reading, MA: Addison-Wesley.

Luce, R.D., Krantz, D., Suppes, P. & Tversky, A. (1990). ***Foundations of measurement. Vol. III. Representation, axiomatization, and invariance.*** San Diego: Academic Press.

Mill, J. ***S.*** (1862). ***System der deductiven*** und ***inductiven Logik. Erster Theil*** (Übersetzung von J. Schiel. 2. deutsche, nach der fünften des Originals erweiterte Auflage. Original erschienen 1843). Braunschweig: Vieweg.

Möbus, C. & Nagl, W. (1983). Messung, Analyse und Prognose von Veränderungen. In J. Bredenkamp & H. Feger (Hrsg.), ***Hypothesenprüfung - Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie I: Forschungsmethoden der Psychologie, Band*** 5 (239-470). Göttingen: Hogrefe.

Moosbrugger, H. (1981). Zur differentiellen Validität bei nichtlinearen Test-Kriteriumszusammenhängen. ***Zeitschrift für Differentielle und Diagnostische Psychologie, 2,*** 219-234.

Müller, P. H. (Hrsg.) (1975). ***Lexikon der Stochastik*** (2. Auflage). Berlin: Akademie-Verlag.

Mulaik, S.A. (1986). Factor analysis and Psychometrika: Major developments. Psy***chometrika, 51, 22-33.***

Novick, M. R. (1966). The axioms and principal results of classical test theory. ***Journal of Mathematical Psychology, 3,*** 1-18.

Penner, L.A. & Wymer, W.E. (1983). The moderator variable approach to behavioral predictability: Some of the variables some of the time. ***Journal of Research in Personality, 17, 339-353.***

Plachky, D. (1981). ***Stochastik*** II. Wiesbaden: Akademische Verlagsgesellschaft.

Poppen K. R. (1984). ***Logik*** der Forschung (8. Auflage). Tübingen: Mohr. (1. Auflage 1934).

Rasch, ***G.*** (1980). ***Probabilistic models for some intelligence and attainment tests*** (2nd edition). Chicago: University of Chicago Press. (1. Auflage erschienen 1960 in Kopenhagen: Nielsen & Lydiche).

Renyi, A. (1977). ***Wahrscheinlichkeitsrechnung.*** Berlin: VEB Deutscher Verlag der Wissenschaften.

Roberts, F. (1979). ***Measurement theory.*** Reading, MA: Addison-Wesley.

Roskam, E.E. & Suck, R. (1987) (Eds.). ***Pro gress in Mathematical Psychology - I.*** Amsterdam: North-Holland.

Rost, J. (1988). ***Quantitative und qualitative probabilistische Testtheorie.*** Bern: Huber.

Rost, J. (1991). A logistic mixture distribution model for polytomous item responses. ***British*** Journal ***of Mathematical and Statistical Psychology, 44, 75-92.***

Rost, J. & Langeheine, R. (1991). Mischverteilungsmodelle: Die Methodologie der kommenden Jahre. In D. Frey (Hrsg.), ***Bericht über den 37. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1990*** (S. 590–596). Göttingen: Hogrefe.

Rost, J. & Spada, H. (1983). Die Quantifizierung von Lerneffekten anhand von Testdaten. *Zeitschrift für Differentielle und Diagnostische Psychologie, 4,* 29–49.

Rost, J. & Strauß, B. (1992). Psychometrics and test theory-recent developments. *German Journal of Psychology, 16,* **91-119.**

Saris, W. & van Meurs, A. ***(1990). Evaluation of measurement instruments by meta-analysis and multitrait-multimethod*** studies. Amsterdam: North-Holland.

Saris, W. E. & Stronkhorst, L. E. (1984). Causa1 ***modelling in nonexperimentaf research.*** Amsterdam: Sociometric Research Foundation.

Saunders, D. R. (1956). Moderator variables in prediction. ***Educational und Psychological Measurement, 16, 209-222.***

Schmitt, M. J. (1990). ***Konsistenz als Persönlichkeitseigenschaft!*** Berlin: Springer.

Schmitt, M. J. & Steyer, R. (1990) Beyond intuition and classical test theory: A reply to Epstein. ***Methodika,*** **4, 101-107.**

Schmitt, M. & Götz-Baltes, B. (1992). Common and uncommon moderator concepts: Comment on Wermuth's ,,Moderating Effects in Multivariate Normal Distributions". ***Methodika, 6,*** 1-4.

Sneed, J. D. ***(1971). The logical structure of mathematical physics.*** Dordrecht: Reidel.

Spohn, W. (1980). Stochastic independence, causa1 independence, and shieldability. Journal ***of Philosophical Logic, 9, 73-99.***

Spohn, W. (1983). Deterministic and probabilistic reasons and causes. ***Erkenntnis, 19,*** **371-396.**

Spohn, W. (1990). Direct and indirect causes. ***Topoi, 9,*** **125-145.**

Spohn, W. (1991). A reason for explanation: Explanations provide stable reasons. In W. Spohn, B.C. van Fraassen & B. Skyrmes (Eds.), ***Existence and explanation*** (pp. 165-196). Dordrecht: Kluwer.

Spohn, W. (im Druck). On Reichenbach's principle of the common cause. In W.C. Salmon & G. Wolters (Eds.), ***Logic, language, and the structure of scientific theories.*** Pittsburgh, PA: University of Pittburgh Press.

Stegmüller, W. (1979). ***The structuralist view of theories.*** Berlin: Springer.

Stegmüller, W. **(1983).** ***Erklärung, Begründung, Kausalität*** (Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band 1). Berlin: Springer.

Stegmüller, W. **(1986).** ***Theorie und Erfahrung: 3. Teilband. Die Entwicklung des neuen Strukturalismus seit 1973*** (Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band II). Berlin: Springer.

Stephan, E. ***(1990).*** Zur ***logischen*** Struktur ***psychologischer Hypothesen.*** Berlin: Springer.

Steyer, R. (1987). Konsistenz und Spezifität: Definition zweier zentraler Begriffe der Differentiellen Psychologie und ein einfaches Modell zu ihrer Identifikation. ***Zeitschrift für Differentielle und Diagnostische Psychologie, 8, 245-258.***

Steyer, R. ***(1988). Experiment, Regression und Kausalität. Die logische Struktur kausaler Regressionsmodelle.*** Unveröff. Habilitationsschrift, Universität Trier, FB ***I*** - Psychologie.

Steyer, R. (1989). Models of classical psychometric test theory as stochastic measurement models: Representation, uniqueness, meaningfulness, identifiability, and testability. ***Methodika, 3, 25-60.***

Steyer, R. (1992). ***Theorie kausale; Regressionsmodelle.*** Stuttgart: Gustav Fischer Verlag.

Steyer, R. & Eid, M. (1993). ***Messen und Testen.*** Berlin: Springer.

Steyer, R. & Schmitt, M.J. (1990a). The effects of aggregation across and within occasions on consistency, specificity, and reliability. ***Methodika, 4, 58-94.***

Steyer, R. & Schmitt, M. J. (1990b). Latent state-trait models in attitude research. Qua-***lity*** and ***Quantity, 24, 427-445.***

Steyer, R. & Schmitt, M.J. (1993a). An introduction to latent state-trait theory. In R. Steyer, H. Gräser, & K.F. Widaman (Eds.), ***Consistency und specificity: Latent state-trait models in Differential Psychology*** (pp. 1-19). New York: Springer.

Steyer, R. & Schmitt, M. J. (1993b). Models of latent state-trait theory. In R. Steyer, H. Gräser, & K. E Widaman (Eds.), ***Consistency and specificity: Latent state-truit models in Differential Psychology*** (pp. 21-50). New York: Springer.

Steyer, R., Wender, K. E & Widaman, K.F. (Eds.). (1992). Proceedings of the 7th European Psychometric Meeting. Stuttgart: Gustav Fischer Verlag.

Suppes, P. ***(1970). A probabilistic theory of causality.*** Amsterdam: North-Holland.

Suppes, P. (1981). Scientific causa1 talk: A reply to Martin. ***Theory and Decision, 13, 363-380.***

Suppes, P. & Zinnes, J.L. (1963). Basic measurement theory. In: R. D. Luce, R. R. Bush & E. Galanter (Eds.), ***Handbook of Mathematical Psychology. Vol. Z*** (pp. 3-76). New York: Wiley.

Tack, W. H. (1980). Zur Theorie psychometrischer Verfahren. Formalisierung der Erfassung von Situationsabhängigkeit und Veränderung. ***Zeitschrift für Differentielle und Diagnostische Psychologie, 1,*** 87-108.

Thurstone, L. L. ***(1931). The reliability and validity of tests.*** Ann Arbor, MI: Edwards Brothers.

Torgerson, W.S. (1986). Scaling and Psychometrika: Spatial and alternative representations of similarity data. ***Psychometrika, 51, 57-63.***

Tutz, G. (1989). ***Latent Trait-Modelle für ordinale Beobachtungen.*** Berlin: Springer.

Viertl, R. ***(1990). Einführung in die Stochastik.*** Berlin: Springer.

Westermann, R. (1987). ***Strukturalistische Theorienkonzeption und empirische Forschung in der Psychologie.*** Berlin: Springer.

Westmeyer, H. (Hrsg.) (1989). Psychological theories from a structuralist point of view. New York: Springer.

Westmeyer, H. (Hrsg.) (1992). ***The structuralist program in psychology: Foundations und applications.*** Toronto: Hogrefe-Huber.

Wermuth, N. (1989). Moderating effects in multivariate normal distributions. ***Methodika, 3, 74-93.***

Wottawa, H. (1980). Grundriß ***der Testtheorie.*** München: Juventa Verlag.

Zimmerman, D. W. (1975). Probability spaces, Hilbert spaces, and the axioms of test theory. ***Psychometrika, 40,*** 395-412.

Zimmerman, D. W. & Williams, R. H. (1977). The theory of test validity and correlated errors of measurement. ***Journal of Mathematical Psychology, 16,*** 135-152.