Zuerst erschienen in: Psychologische Beiträge 14 (1972), 1, pp. 16-37

Probabilistische Modelle experimentalpsychologischer Versuchssituationen

Eine vollständige multifaktorielle Verallgemeinerung des mehrdimensionalen Testmodells von *Rasch* und ihre Anwendung in der experimentellen Psychologie

von Wilhelm F. Kempf, Erlangen

1. Einleitung

Die Anwendung quantitativer Methoden in der Psychologie wirft eine Reihe von wissenschaftstheoretischen Problemen auf, deren Diskussion noch lange nicht als abeschlossen gelten kann. Eines der Hauptprobleme betrifft die Frage, wie psychologische Variable quantifiziert werden können, ein anderes die Stichprobenabhängigkeit statistischer Aussagen. Auf Grund meßtheoretischer Überlegungen zeigt sich, daß beide Probleme eng miteinander verknüpft sind.

Ausgehend von Wittgensteins erkenntnistheoretischer Auffassung, wonach Erkenntnis als Repräsentation empirischer Strukturen durch logisch-sprachliche Strukturen zu verstehen ist, kann man Messung als Repräsentation von Objekten mittels (reeller) Zahlen definieren. Für die Satzsysteme der kognitiven Wissenschaften gelten genaue Bedingungen ihrer Anwendbarkeit (operative Bedingungen) sowie genaue Bedingungen ihrer sprachlich-formalen Richtigkeit (operationale Bedingungen), deren Einhaltung oder Nicht-Einhaltung jederzeit festgestellt werden kann. Dasselbe gilt auch für die numerische Repräsentation von Objekten. Ihre operativen und operationalen Bedingungen wurden von Suppes u. Zinnes (1963) einer eingehenden Diskussion unterzogen. Für die Quantifizierbarkeit psychologischer Variablen ergeben sich daraus strenge Konsequenzen.

2. Abgeleitete Messung

Während die Festlegung einer fundamentalen Skala das Vorliegen unmittelbar beobachteter empirischer Relationen zwischen den Meßobjekten voraussetzt, die vor jeder Messung und unabhängig von dieser bestehen, handelt es sich bei psychologisch interessierenden Variablen zumeist um hypothetische Konstrukte, die sich einer direkten Beobachtung entziehen. Das allgemeinste Prinzip bei der Messung einer solchen Größe besteht in der Beobachtung anderer Variablen, die mit der zu messenden in einem gesetzmäßigen Zusammenhang stehen, und ihrer Quantifizierung im Hinblick auf

Konstanzer Online-Publikations-System (KOPS)
URL: http://www.ub.uni-konstanz.de/kops/volltexte/2008/-6029/
URN: http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-60297

eine Theorie. Das klassische Beispiel für eine solche ‚abgeleitete' Messung stammt aus der Physik, wo Kraft als das Produkt von Masse und Beschleunigung definiert wird. Eine solche Beziehung kann allerdings nicht durch geeignete Definition von Skalen allein erzielt werden, sondern setzt das Vorliegen eines Naturgesetzes, im speziellen Fall der Proportionalität zwischen Masse und Beschleunigung bei Gleichheit der Kraft, voraus. Erst durch die Beobachtung dieser Gesetzmäßigkeit wird die Annahme einer latenten Dimension ‚Kraft' und ihre Quantifizierung gerechtfertigt.

Die allgemeine Theorie der abgeleiteten Messung nach SUPPES u. ZINNES (1963) kann erweitert werden, so daß sie auch multidimensionale abgeleitete Messungen und simultane Mehrfachmessungen mit einschließt:

Den Ausgangspunkt bei der Konstruktion einer solchen Skala bildet ein erweitertes abgeleitetes Meßsystem. Darunter verstehen wir eine geordnete Folge

$$\mathscr{B} = \{B_1, \ldots, B_N, f_1, \ldots, f_m\}, \tag{2.1}$$

worin B_1 bis B_N nichtleere Mengen von Meßobjekten und f_1 bis f_m auf einer dieser Mengen oder auf cartesischen Produkten dieser Mengen definierte reellwertige Funktionen.

Wir definieren: seien $\mathscr{B} = \{B_1, \ldots, B_N, f_1, \ldots, f_m\}$ ein erweitertes abgeleitetes Meßsystem, $g_i = g_1, \ldots, g_M$ auf einer der Mengen B_1 bis B_N oder auf cartesischen Produkten dieser Mengen definierte Abbildungen auf den n_i-dimensionalen euklidischen Raum, und R eine Relation zwischen f_1 bis f_m und g_1 bis g_M, dann ist die geordnete Folge

$$\{\mathscr{B}, g_1, \ldots, g_M, R\} \tag{2.2}$$

eine (erweiterte) abgeleitete Skala. Die Relation R heißt die repräsentierende Relation der Skala, g_1 bis g_M sind die abgeleiteten numerischen Zuordnungsvorschriften.

Die Skaleneigenschaft einer (abgeleiteten) Messung wird durch die relative Eindeutigkeit der abgeleiteten numerischen Zuordnungsvorschriften g_1 bis g_M definiert: Abbildungen ψ_i $(i = 1, \ldots, M)$ des n_i-dimensionalen euklidischen Raumes in sich selbst heißen zulässige Transformationen, wenn die Abbildungen

$$g_i' =: \psi_i \cdot g_i \qquad (i = 1, \ldots, M) \tag{2.3}$$

die Eigenschaft haben, daß auch die geordnete Folge

$$\{\mathscr{B}, g_1', \ldots, g_M', R\} \tag{2.4}$$

eine abgeleitete Skala ist, also $R(f_1, \ldots, f_m, g_1', \ldots, g_M')$ erfüllt ist. Durch die Klasse der zulässigen Transformationen ist der Skalentyp einer Messung festgelegt. Im Falle eindimensionaler Messung unterscheidet man hauptsächlich zwischen Nominalskalen, Ordinalskalen, Intervallskalen, Rationalskalen und absoluten Skalen.

Für Nominalskalen, die lediglich eine Benennung der Meßobjekte leisten, sind beliebige umkehrbar eindeutige Transformationen zulässig. Ordinalskalen, die die Meßobjekte in eine Rangordnung bringen, sind bis auf streng monoton wachsende Transformation invariant. Für Intervallskalen, bei denen sowohl der Nullpunkt als auch die Maßeinheit willkürlich gewählt sind, sind beliebige positive lineare Transformationen zulässig, für Rationalskalen, die über einen absoluten Nullpunkt verfügen, beliebige Ähnlichkeitstransformationen. Die absolute Skala mit festem Nullpunkt und fester Maßeinheit erlaubt keinerlei Transformationen.

Diese Einteilung der Skalentypen kann auch auf den Fall mehrdimensionaler Messung übertragen werden. Man denke sich die zulässigen Transformationen ψ_i aus n_i Abbildungen $\psi_{i1}, \ldots, \psi_{in_i}$ der Menge der reellen Zahlen in sich selbst zusammengesetzt, wobei ψ_{ip} $(p = 1, \ldots, n_i)$ die Transformation der p-ten Meßdimension bezeichnet. Der Funktionstyp von ψ_{ip} gibt die Skaleneigenschaft der p-ten Meßdimension an.

3. Spezifische Objektivität

Bei der Analyse psychologischer Testsituationen kann man davon ausgehen, daß das Anliegen der Psychologie nicht so sehr darin besteht, eine Messung an sich vorzunehmen, als vielmehr Vergleiche von Objekten (z.B. Vpn) durchzuführen, wobei man Objetivität der Vergleiche anstrebt. Dasselbe gilt auch für experimentalpsychologische Versuchssituationen. Hier ist man allerdings weniger am Vergleich der Vpn interessiert, als am Vergleich der Effekte der systematisch variierten Versuchsbedingungen auf das Verhalten der Vpn. Wir werden darauf später noch zurückkommen.

Dem Begriff des Vergleiches wird in der Testpsychologie durch die Definition einer Menge (M_1) von Meßobjekten (Vpn), die verglichen werden, und einer Menge (M_2) von ‚Instrumenten' (Items), mit deren Hilfe die Vergleiche durchgeführt werden, ein spezifischer Sinn gegeben (vgl. RASCH 1966a).

Der ‚Kontakt' eines bestimmten Objektes mit einem bestimmten Instrument erzeugt ein Resultat, das qualitativ oder quantitativ sein kann. Aus den beobachteten Resultaten sollen Aussagen über die Relation zwischen Objekten hergeleitet werden. Damit sie als wissenschaftlich gelten können, dürfen die aus den Resultaten des Kontaktes zweier Objekte $v_1, w_1 \in M_1$ mit einer Menge von Instrumenten $v_2, w_2, \ldots, z_2 \in M_2$ hergeleiteten Aussagen über die Relation zwischen den Objekten nicht davon abhängen, welche Instrumente zufällig aus M ausgewählt wurden. Außerdem müssen die Vergleiche transitiv sein, d.h. zu dem gleichen Ergebnis führen, wenn zwei Objekte direkt verglichen werden, wie wenn der Vergleich über andere Objekte als Zwischenstufe erfolgt. Vergleiche, die diese beiden Forderungen erfüllen, nennt RASCH (1966a) ‚spezifisch objektiv'.

4. Die meßtheoretische Analyse psychologischer Testsituationen

Die Struktur eines psychologischen Meßmodells ist durch die spezifische Art von Daten, die in der Psychologie verfügbar sind, bereits in einem gewissen Sinne determiniert. Die Vorgangsweise bei seiner theoretischen Konzeption kann nach KEMPF (1970) in drei Stufen charakterisiert werden:

1. Festlegung der Grundannahmen des Modells und Prüfung seiner formalen Widerspruchsfreiheit (operationale Bedingungen).
2. Untersuchung der speziellen Eigenschaften des Modells
   a. Herleitung der operativen Bedingungen seiner empirischen Gültigkeit und Angabe von Methoden, mit deren Hilfe sie geprüft werden können (Repräsentationsproblem), und
   b. Untersuchung der Skaleneigenschaft der Messung (Eindeutigkeitsproblem).
3. Untersuchung des Modells auf seine empirische Gültigkeit.

Wir haben bereits oben erwähnt, daß das Resultat des Kontaktes einer Vp mit einem Item qualitativ oder quantitativ sein kann. Im folgenden wollen wir von qualitativen Resultaten ausgehen und sagen, daß jeder Kontakt einer Vp mit einem Item zu einer endlichen oder abzählbar unendlichen Anzahl verschiedener Resultate führen kann. Dies bedeutet insofern keine Einschränkung der Theorienbildung, als man auch bei quantitativen Resultaten – z.B. bei der Messung der Reaktionszeit – zwangsläufig zu einer qualitativen Betrachtung kommt: ungeachtet der Feinheit der Skala, auf der man die quantitativen Resultate mißt, muß man sich festlegen, wieviele Dezimalen man berücksichtigen will.

Psychologische Tests sind im einfachsten Fall aus Items aufgebaut, die nur Alternativantworten zulassen: sie können ‚gelöst' werden oder nicht. Setzt man

$$a_{v_1 v_2} = \begin{cases} 1 \text{ wenn Vp } v_1 \text{ Item } v_2 \text{ löst} \\ 0 \text{ wenn Vp } v_1 \text{ Item } v_2 \text{ nicht löst} \end{cases} \qquad (4.1)$$

$$(\forall\, (v_1, v_2) \in M_1 \times M_2),$$

dann ist das erweiterte abgeleitete Meßsystem, das einem solchen Test entspricht, durch die geordnete Folge

$$\mathfrak{S} = \{M_1, M_2, f\} \qquad (4.2)$$

gegeben, worin M_1 eine nichtleere Menge von Vpn und M_2 eine nichtleere Menge von Items ist. f ist eine, über der Produktmenge $M_1 x M_2$ definierte Abbildung, deren Wertbereich die Menge $\{0, 1\}$ ist[1]. Jedem geordneten Paar $(v_1, v_2) \epsilon M_1 x M_2$ wird gemäß (4.1) ein Bildlelement $a_{v_1 v_2} \epsilon \{0, 1\}$ zugeordnet. Formal[2]:

$$f: M_1 x M_2 \longrightarrow \{0, 1\}, \qquad (4.3)$$
$$(v_1, v_2) \longrightarrow a_{v_1 v_2}.$$

Ausgehend von der Feststellung, daß sich nicht nur Vpn in ihren Einstellungen unterscheiden, sondern auch Items verschieden ‚schwierig' sein können, entwickelt GUTTMAN (1950) ein Meßmodell für die simultane Messung der Einstellung von Vpn und der Schwierigkeit von Items. Ein leichtes Item aus einem Fragebogen zur Erfassung der Einstellung gegenüber farbigen Minderheiten könnte z.B. lauten: „Es macht mir nichts aus, wenn in unserer Stadt auch Neger wohnen." Ein schwierigeres Item wäre: „Ich habe nichts dagegen, daß meine Tochter einen Neger heiratet."

Die syntaktische Formulierung von GUTTMANS Modell lautet: Gegeben ist ein erweitertes abgeleitetes Meßsystem der Form (4.2). Gesucht werden Abbildungen

$$g_R: M_R \longrightarrow \mathbf{R} \qquad (4.4)$$
$$v_R \longrightarrow x_{v_R} \qquad (R = 1, 2)$$

derart, daß

$$a_{v_1 v_2} = 1 \quad \text{genau dann, wenn} \quad x_{v_1} \geqslant x_{v_2} \qquad (4.5)$$
$$(\forall (v_1, v_2) \epsilon M_1 x M_2)$$

erfüllt ist. x_{v_1} kann als die Einstellung der Vp v_1 interpretiert werden, x_{v_2} als die Schwierigkeit des Items v_2. Es gelten die beiden Sätze

Repräsentationssatz: Sei r_{v_2} die Anzahl der Personen, die Item v_2 lösen, dann gibt es Abbildungen g_R, die die gewünschte Eigenschaft (4.5) haben, genau dann, wenn die positive Beantwortung eines Items v_2 durch Vp v_1 immer auch die positive Beantwortung aller (leichteren oder gleich leichten) Items w_2 mit $r_{w_2} \geqslant r_{v_2}$ durch Vp v_1 voraussetzt.

1 $\{0, 1\}$ bezeichnet die Menge, bestehend aus den Zahlen 0 und 1.

2 Allgemein bezeichnet die Schreibweise

$$f: A \longrightarrow B,$$
$$a \longrightarrow f(a)$$

eine Abbildung f des Definitionsbereiches A in den Wertebereich B, welche jedem Element a aus A ein Bildelement $f(a) \epsilon B$ zuordnet.

Eindeutigkeitssatz: Die Abbildungen g_R sind bis auf gemeinsame streng monoton wachsende Transformationen bestimmt.

Können Items gefunden werden, die die geforderte empirische Gesetzmäßigkeit (4.5) erfüllen, so wird das Einstellungskontinuum durch jedes dieser Items in zwei Teile geteilt. Zwei Personen v_1 und w_1 mit $x_{v_1} < x_{v_2} \leqslant x_{w_1}$ werden durch das Item v_2 eindeutig in eine Rangreihe gebracht. Jedes andere Item w_2 mit $x_{v_1} < x_{w_2} \leqslant x_{w_1}$ leistet dasselbe. Das Einstellungskontinuum wird durch einen Test aus k Items in k + 1 Rangklassen geteilt. Jede andere Teilmenge von Items aus M_2 führt zum selben Ergebnis, d.h. es gibt kein Item, das eine Umkehr von Rangpositionen bewirkt. Die Vergleiche sind auch transitiv. Also ist die Forderung nach spezifischer Objektivität erfüllt. Diese Feststellung ist allerdings nur *cum grano salis* zu verstehen: die Genauigkeit der Aussagen, die über die Relation zwischen Vpn getroffen werden können, hängt sehr wohl von der Auswahl der Items aus M_2 ab. Enthält ein Test kein Item v_2 mit $x_{v_1} < x_{v_2} \leqslant x_{w_1}$, so kann mit Hilfe dieses Tests keine Aussage über die Relation zwischen den beiden Personen v_1 und w_1 getroffen werden. Unabhängigkeit der *Meßgenauigkeit* von der Auswahl der Instrumente wäre jedoch eine Forderung, die nicht einmal in der Physik erfüllt werden kann.

Auf Grund seiner deterministischen Konzeption ist GUTTMANS Modell nur von theoretischer Bedeutung. Die experimentalpsychologische Erfahrung lehrt, daß psychologische Gesetze nicht mit derselben Präzision an manifesten Variablen feststellbar sind wie etwa physikalische. Deterministische Datenmodelle, welche Relationen zwischen beobachteten Daten beschreiben und somit feste Relationen zwischen manifesten Variablen postulieren, sind daher in der Psychologie unrealistisch. Sie treffen Voraussetzungen, die zwar in speziell konstruierten Beispielen, in der Praxis jedoch nie erfüllt sind. Es ist daher zweckmäßig, keine deterministischen, sondern stochastische Modelle zu konstruieren, deren Aussagen sich nicht auf beobachtete Daten, sondern auf die Parameter von Zufallsverteilungen beziehen.

Das stochastische Modell eines Tests aus dichotomen Items sieht in seiner allgemeinsten Form folgendermaßen aus:

Gegeben ist eine nichtleere Menge M_1 von Vpn und eine nichtleere Menge M_2 von Items, für welche jeweils zwei Antwortmöglichkeiten bestehen. Über der Produktmenge $M_1 x M_2$ ist eine reellwertige Funktion

$$p: M_1 x M_2 \longrightarrow \rangle 0, 1 \langle, \qquad (4.6)$$

$$(v_1, v_2) \longrightarrow p(v_1, v_2)$$

definiert[3]. $p(v_1, v_2)$ bezeichnet die Wahrscheinlichkeit, mit der Vp v_1 Item v_2 lösen kann. Gesucht werden Abbildungen

3 $\rangle 0, 1 \langle$ bezeichnet das offene Intervall $I = \{a \in \mathbf{R} : 0 < a < 1\}$

$$g_R: \; M_R \longrightarrow \mathcal{R}_{n_R} \qquad (4.7)$$
$$v_R \longrightarrow x_{v_R} = (x_{v_R}^{(1)}, \ldots, x_{v_R}^{(h)}, \ldots, x_{v_R}^{(n_R)})$$
$(R = 1,2)$

die mit den Lösungswharscheinlichkeiten $p(v_1, v_2)$ durch eine eindeutige (jedoch nicht notwendigerweise eineindeutige) Funktion $f(x_{v_1}, x_{v_2})$ verknüpft sind:

$$p(v_1, v_2) = f(x_{v_1}, x_{v_2}) \qquad (4.8)$$
$(\forall\, (v_1, v_2) \in M_1 x M_2)$.

Wie bei GUTTMANS Skalogrammanalyse werden die für das Itemlösungsverhalten relevanten Eigenschaften der Vpn und der Items simultan gemessen. Jeder Vp v_1 wird ein Parameter x_{v_1} zugeordnet. Dieser kann ein Skalar sein ($n_1 = 1$, eindimensionale Messung), z.B. wenn die Beantwortung der Items eines Einstellungsfragebogens seitens der Vpn ausschließlich von der in Frage stehenden Einstellung abhängt. Im allgemeinen Fall ist x_{v_1} ein Vektor ($n_1 > 1$, mehrdimensionale Messung). Dies ist z.B. dann der Fall, wenn die Beantwortung der Items nicht nur von der betreffenden Einstellung selbst, sondern auch von deren Intensität etc. abhängt. Desgleichen wird jedem Item v_2 ein Parameter x_{v_2} zugeordnet, der ebenfalls ein Skalar ($n_2 = 1$, eindimensionale Messung) oder ein Vektor ($n_2 > 1$, mehrdimensionale Messung) sein kann. Letzteres ist z.B. dann der Fall, wenn sich die Items nicht nur hinsichtlich ihrer Schwierigkeit, sondern auch in ihrer Trennschärfe, etc. unterscheiden.

Wie RASCH (1966) zeigt, ist die Dimensionalität der Parameter im Falle spezifischer Objektivität immer kleiner als die Anzahl der Reaktionsmöglichkeiten. Tests aus dichotomen Items können daher nur eindimensionale Eigenschaftskontinua spezifisch objektiv erfassen. Notwendig und hinreichend für spezifische Objektivität ist das einsimensionale Testmodell von RASCH: ausgehend von einem erweiterten abgeleiteten Meßsystem der Form: $\{M_1, M_2, p\}$ werden Abbildungen

$$g_R: \; M_R \longrightarrow \mathcal{R}_+ \qquad (4.9)$$
$$v_R \longrightarrow x_{v_R}$$
$(R = 1, 2)$

gesucht derart, daß

$$p\,(v_1, v_2) = \frac{x_{v_1} x_{v_2}}{1 + x_{v_1} x_{v_2}} \qquad (4.10)$$
$(\forall\, (v_1, v_2) \in M_1 x M_2)$

erfüllt ist. Nur unter Voraussetzung der empirischen Gültigkeit des Modells

(4.10) ist es möglich, Personen unabhängig von der Auswahl der Items aus M_2 hinsichtlich ihrer latenten Eigenschaft x_{v_1} zu vergleichen. Auf Grund der Symmetrie zwischen Item- und Personenparametern in (4.10) können auch die Items unabhängig von der Auswahl der Personen aus M_1 hinsichtlich ihrer ‚Leichtigkeit' x_{v_2} verglichen werde.

FALMAGNE (unveröffentlichtes Manuskript, zit. n. MICKO, 1970) bewies für das Modell (4.10) einen Repräsentationssatz und einen Eindeutigkeitssatz[4]:

Repräsentationssatz: Abbildungen g_R, die mit der Funktion p in der Form (4.10) zusammenhängen, gibt es genau dann, wenn die Funktionswerte der Funktion

$$\begin{aligned} c: M_1 x M_2 &\longrightarrow \mathbb{R}_+ \\ (v_1, v_2) &\longrightarrow c(v_1, v_2) \end{aligned} \tag{4.11}$$

mit

$$c(v_1, v_2) = \frac{p(v_1, v_2)}{1 - p(v_1, v_2)} \tag{4.12}$$

der kritischen Bedingung

$$\frac{c(v_1, v_2)}{c(w_1, v_2)} = \frac{c(v_1, w_2)}{c(w_1, w_2)} \tag{4.13}$$

$(\forall\, v_1, w_1 \in M_1, v_1, w_2 \in M_2)$

genügen.

Eindeutigkeitssatz: Die Abbildungen g_R sind bis auf Ähnlichkeitstransformationen

$$\begin{aligned} \psi_R: \mathbb{R}_+ &\longrightarrow \mathbb{R}_+ \\ x_{v_R} &\longrightarrow x'_{v_R} = \beta_R x_{v_R} \end{aligned} \tag{4.14}$$

$(R = 1,2)$

bestimmt. Zwischen den Maßkonstanten β_R besteht die Beziehung $\beta_1 \beta_2 = 1$.

Eine vollständige Einführung in RASCHS eindimensionales Testmodell kann im Rahmen des vorliegenden Artikels nicht gegeben werden. Ausführliche Darstellungen findet der interessierte Leser bei RASCH (1966b), FISCHER (1968) und STENE (1968).

Unsere bisherigen Überlegungen bezogen sich auf Tests, deren Items genau zwei Antwortmöglichkeiten zulassen. Im folgenden verallgemeinern wir auf Items mit $n \geqslant 2$ Reaktionsalternativen. Wir haben bereits erwähnt,

4 FALMAGNES Beweisführung braucht an dieser Stelle nicht wiedergegeben werden. Sie ergibt sich durch Spezialisierung der Beweisführung auf S. 32–36 auf N = m = 2.

daß der Kontakt einer Vp mit einem Item im allgemeinen Fall zu einer endlichen oder abzählbar unendlichen Anzahl von Resultaten führen kann. O.B.d.A. können wir voraussetzen, daß die Resultate so klassifiziert werden, daß jeder Kontakt einer Vp mit einem Item zu genau einem Resultat führt und daß es m verschiedene mögliche Resultate gibt. Sei $p^{(l)}(v_1, v_2)$ die Wahrscheinlichkeit, mit welcher der Kontakt zwischen Vp v_1 und Item v_2 zu dem Resultat l führt, dann gilt

$$\sum_{l=1}^{m} p^{(l)}(v_1, v_2) = 1 \tag{4.15}$$

$(\forall (v_1, v_2) \in M_1 x M_2)$.

Folglich genügt es bei der Konstruktion des stochastischen Modells eines Tests mit m Antwortmöglichkeiten pro Item, die Wahrscheinlichkeiten für $m-1$ der möglichen Resultate zu betrachten. O.B.d.A. seien dies die Resultate $1, 2, \ldots, m-1$. Die Wahrscheinlichkeit $p^{(m)}(v_1, v_2)$, mit welcher der Kontakt zwischen Vp v_1 und Item v_2 zu dem Resultat m führt, ist dann durch die Wahrscheinlichkeiten $p^{(1)}(v_1, v_2), \ldots, p^{(m-1)}(v_1, v_2)$ vollständig determiniert:

Gegeben sind eine nichtleere Menge M_1 von Vpn und eine nichtleere Menge M_2 von Items. Über der Produktmenge $M_1 x M_2$ sind $m-1$ Funktionen

$$\begin{aligned} p^{(l)} : M_1 x M_2 &\longrightarrow \,\rangle 0,1 \langle \\ (v_1, v_2) &\longrightarrow p^{(l)}(v_1, v_2) \end{aligned} \tag{4.16}$$

$(l = 1, \ldots, m-1)$

definiert.

Gesucht werden Abbildungen

$$\begin{aligned} g_R : M_R &\longrightarrow \mathfrak{R}^{n_R} \\ \check{v}_R &\longrightarrow x_{v_R} = (x_{v_R}^{(1)}, \ldots, x_{v_R}^{(h)}, \ldots, x_{v_R}^{(n_R)}) \end{aligned} \tag{4.17}$$

$(R = 1, 2)$

derart, daß die Reaktionswahrscheinlichkeiten $p^{(l)}(v_1, v_2)$ durch eindeutige Funktionen $f^{(l)}(x_{v_1}, x_{v_2})$ mit den Parametern x_{v_1} und x_{v_2} verknüpft sind:

$$p^{(l)}(v_1, v_2) = f^{(l)}(x_{v_1}, x_{v_2}) \tag{4.18}$$

$(\forall (v_1, v_2) \in M_1 x M_2, l = 1, \ldots, m-1)$.

Notwendig und hinreichend für spezifische Objektivität ist im Falle von $m \geqslant 2$ Antwortmöglichkeiten pro Item das mehrdimensionale Testmodell von Rasch, das die Reaktionen der Vpn durch die Annahme von $m-1$ latenten Eigenschaften der Personen erklärt. Je stärker die l-te Eigenschaft bei einer Vp ausgeprägt ist, desto größer ist die Wahrscheinlichkeit, daß sie die l-te Antwortmöglichkeit wählt. Ein Beispiel: ein Item aus einem Fragebogen zur Erfassung von ‚Introversion-Extraversion' könnte lauten: „Es fällt mit schwer, Gesprächsstoff zu finden, wenn ich jemanden kennenlerne." Die Antwort „ja" weist auf Introversion hin, „nein" auf Extraversion. Antwortet eine Vp mit „ich weiß nicht", so kann das entweder darauf zurückzuführen sein, daß die Vp weder stark introvertiert noch extravertiert ist, oder aber darauf, daß die Vp Entscheidungen zu vermeiden sucht. Gleichzeitig unterscheiden sich die Items in der ‚Leichtigkeit' ihrer Antwortmöglichkeiten. Die syntaktische Formulierung des Modells lautet:

Gegeben ist ein erweitertes abgeleitetes Meßsystem der Form $\{M_1, M_2, p^{(1)}, \ldots, p^{(m-1)}\}$. Gesucht werden abgeleitete numerische Zuordnungsvorschriften

$$\begin{aligned} g_R: \; M_R &\longrightarrow \mathcal{R}^+_{m-1} \\ v_R &\longrightarrow x_{v_R} = (x^{(1)}_{v_R}, \ldots, x^{(l)}_{v_R}, \ldots, x^{(m-1)}_{v_R}) \end{aligned} \tag{4.19}$$

$(R = 1, 2)$

die der repräsentierenden Relation

$$p^{(l)}(v_1, v_2) = \frac{x^{(l)}_{v_1} x^{(l)}_{v_2}}{1 + \sum\limits_{h=1}^{m-1} x^{(h)}_{v_1} x^{(h)}_{v_2}} \tag{4.20}$$

$(\forall\, (v_1, v_2) \in M_1 x M_2, \; l = 1, \ldots, m-1)$

genügen.

Der Beweis der folgenden beiden Sätze ergibt sich durch Spezialisierung der Beweisführung auf S. 32–36.

Repräsentationssatz: Abbildungen g_R, die mit den Funktionen $p^{(l)}$ in der Form (4.20) zusammenhängen, gibt es genau dann, wenn die Funktionswerte der Funktionen

$$\begin{aligned} c^{(l)}: M_1 x M_2 &\longrightarrow \mathcal{R}_+ \\ (v_1, v_2) &\longrightarrow c^{(l)}(v_1, v_2) \end{aligned} \tag{4.21}$$

$(l = 1, \ldots, m-1)$

mit

$$c^{(l)}(v_1, v_2) = \frac{p^{(l)}(v_1, v_2)}{1 - \sum\limits_{h=1}^{m-1} p^{(h)}(v_1, v_2)} \tag{4.22}$$

der kritischen Bedingung

$$\frac{c^{(l)}(v_1, v_2)}{c^{(l)}(w_1, v_2)} = \frac{c^{(l)}(v_1, w_2)}{c^{(l)}(w_1, w_2)} \tag{4.23}$$

und

$$\frac{c^{(l)}(v_1,v_2)}{c^{(l)}(v_1,w_2)} = \frac{c^{(l)}(w_1,v_2)}{c^{(l)}(w_1,w_2)}$$

$(\forall\, v_1, w_1 \in M_1, v_2, w_2 \in M_2; l = 1, \ldots, m-1)$

genügen.

Eindeutigkeitssatz: Die Abbildungen g_R sind bis auf Transformationen der Form

$$\begin{aligned} \psi_R : \mathcal{R}^+_{m-1} &\longrightarrow \mathcal{R}^+_{m-1} \\ x_{v_R} &\longrightarrow x'_{v_R} \\ x'_{v_R} &= (\beta_R^{(1)} x_{v_R}^{(1)}, \ldots, \beta_R^{(l)} x_{v_R}^{(l)}, \ldots, \beta_R^{(m-1)} x_{v_R}^{(m-1)}) \end{aligned} \tag{4.24}$$

$(R = 1, 2)$

bestimmt. Zwischen den Maßstabskonstanten $\beta_R^{(l)}$ besteht die Beziehung $\beta_1^{(l)}\beta_2^{(l)} = 1$; $(l = 1, \ldots, m-1)$.

5. Die meßtheoretische Analyse experimentalpsychologischer Versuchssituationen

Der logische Aufbau psychologischer Experimente ist dem Aufbau einer Testsituation weitgehend äquivalent. Vpn werden unter Variation der Versuchsbedingungen mit Instrumenten (z.B. Testitems, sinnlose Silben etc.) in Kontakt gebracht. Aus den beobachteten Resultaten der Kontakte zwischen Personen und Instrumenten sollen Aussagen über den Einfluß der Versuchsbedingungen auf die Reaktionen der Vpn hergeleitet werden. Anders formuliert: Der Einfluß der Versuchsbedingungen auf die Reaktionen der Vpn soll gemessen werden.

Als Beispiel erwähnen wir ein einfaches Experiment zur Untersuchung des Einflusses der Variablen Beleuchtung und Lärm auf die Testleistung in einem Zahlenreihentest, in dem die Vpn verschiedene Zahlenreihen unter verschiedenen Beleuchtungs- und Lärmbedingungen bearbeiten. Wir können dann unser Experiment in einzelne experimentelle Situationen (‚Kontakte' einer Vp mit einer Zahlenreihe, einer Beleuchtungsbedingung und einer Lärmbedingung) zerlegen und jede einzelne Situation durch Angabe der an ihr beteiligten Vp, Zahlenreihe, Beleuchtungs- und Lärmbedingung beschreiben. Allgemein kann man sagen, daß jedes psychologische Experiment aus einer Anzahl solcher experimenteller Situationen (Kontakte von Personen, Instrumenten und Versuchsbedingungen) besteht. Diese werden nach N Gesichtspunkten, die sich auf Merkmale der Vpn, der Instrumente oder der Versuchsbedingungen beziehen können, klassifiziert. Jede Situation wird durch eine Kombination von N Merkmalen beschrieben, indem man für jede Klassifikation angibt, in welche Klasse die Situation fällt. Seien M_1 eine nichtleere Menge von Vpn, M_2 eine nichtleere Menge von Instrumenten und $M_3, M_4, \ldots, M_N$ nichtleere Mengen von Versuchsbedingungen, dann beschreibt die geordnete Folge $(v_1, v_2, v_3, \ldots, v_N)$ eine Versuchssituation, in der eine Vp $v_1 \in M_1$ unter den Bedingungen $v_3 \in M_3, v_4 \in M_4, \ldots, v_N \in M_N$ mit einem Instrument $v_2 \in M_2$ in Kontakt gebracht wird.

Wie alle empirischen Theorien, sollen die aus psychologischen Experimenten gewonnenen Aussagen ein gewisses Maß an Allgemeingültigkeit aufweisen, d.h. sie sollen nicht nur für einzelne Personen oder Situationen gelten, sondern für wohldefinierte Populationen (M_1) von Personen und Mengen $(\bar{M})$ von Situationen. Die Menge der Situationen, für welche die Gültigkeit der Theorie gefordert wird, stimmt allerdings nicht immer mit der Produktmenge $M_1 x M_2 x \ldots x M_N$ überein, da der Fall eintreten kann, daß manche Merkmalskombinationen $(v_1, \ldots, v_N)$ – auch theoretisch – nicht möglich sind. M ist daher im allgemeinen Teilmenge von $M_1 x M_2 x \ldots x M_N$. Eine empirische Theorie soll aber auch die Herleitung singulärer Aussagen gestatten. Sie muß daher nicht nur für die jeweilige Population von Individuen bzw. Menge von Situationen ‚durchschnittlich' richtig sein, sondern für jede Subpopulation von M_1 und jede Teilmenge aus $\bar{M}$, also auch für jede einzelne Person aus M_1 und jede einzelne Situation aus M gelten. Aus diesem Grund, und da experimentalpsychologische Aussagen nicht anders als aus Beobachtungen an Individuen in Situationen gewonnen werden können, reicht es bei der Konstruktion mathematischer Modelle psychologischer Experimente nicht aus, den durchschnittlichen Effekt der Versuchsbedingungen in einer Stichprobe von Personen und Instrumenten mathematisch zu beschreiben. Vielmehr muß gefordert werden, daß das Modell auch das Zustandekommen der Resultate der einzelnen Situationen zu beschreiben vermag.

Bei geeigneter Klassifikation der Ergebnisse führt jede experimentelle Situation zu genau einem von m möglichen Resultaten. Im oben erwähnten Beispiel könnten dies die beiden Ereignisse „Zahlenreihe gelöst" und „Zahlenreihe nicht gelöst" sein. Aus den beobachteten Ergebnissen sollen Aussagen über den Effekt der experimentell variierten Bedingungen auf die Resultate der experimentellen Situationen hergeleitet werden. An ihrem Zustandekommen können aber neben den (möglichen) Einflüssen der zu untersuchenden Versuchsbedingungen auch individuelle Merkmale der verwendeten Instrumente, wie z.B. die Schwierigkeit der Zahlenreihen, und individuelle Reaktionstendenzen der Vpn – manche Vpn können Zahlenreihen besser lösen als andere – beteiligt sein. Zur Beschreibung der Resultate sind daher gleichzeitig allgemeine ‚Treatment-Parameter', welche den Effekt der Versuchsbedingungen beschreiben, ‚Instrument-Parameter' und ‚Personen-Parameter' erforderlich. Das allgemeine stochastische Modell eines psychologischen Experimentes muß daher lauten:

Gegeben sind eine nichtleere Menge M_1 von Vpn, eine nichtleere Menge M_2 von Instrumenten und nichtleere Mengen M_3 bis M_N von Versuchsbedingungen. Jede Versuchssituation $(v_1, \ldots, v_N)$ führt zu genau einem von m möglichen Resultaten. Über der Menge der Situationen M sind $m-1$ reellwertige Funktionen

$$p^{(l)}: \overline{M} \longrightarrow \rangle 0, 1 \langle, \qquad (v_1, \ldots, v_N) \longrightarrow p^{(l)}(v_1, \ldots, v_n) \tag{5.1}$$
$$(l = 1, \ldots, m-1)$$

definiert. $p^{(l)}(v_1, \ldots, v_N)$ bezeichnet die Wahrscheinlichkeit, mit der in der Situation $(v_1, \ldots, v_N)$ das l-te Resultat eintritt. Gesucht werden Abbildungen

$$g_R: M_R \longrightarrow \mathcal{R}_{n_R}, \qquad v_R \longrightarrow x_{v_R} = (x_{v_R}^{(1)}, \ldots, x_{v_R}^{(h)}, \ldots, x_{v_R}^{(n_R)}) \tag{5.2}$$
$$(R = 1, \ldots, N)$$

die mit den Wahrscheinlichkeiten $p^{(l)}(v_1, \ldots, v_N)$ durch eindeutige Funktionen $f^{(l)}(x_{v_1}, \ldots, x_{v_N})$ verknüpft sind

$$p^{(l)}(v_1, \ldots, v_N) = f^{(l)}(x_{v_1}, \ldots, x_{v_N}) \tag{5.3}$$
$$(\forall\, (v_1, \ldots, v_N) \in \overline{M},\ l = 1, \ldots, m-1).$$

Indem zur Beschreibung der Daten gleichzeitig allgemeine Treatment-Parameter, Instrument-Parameter und Personenparameter erforderlich sind, bewirkt jede Vergrößerung der Vpn-Stichprobe die Einführung eines neuen, im Hinblick auf die Frage nach dem Effekt der zu untersuchenden Versuchsbedingungen „inzidentellen" Parameters. Auf Grund eines Satzes von NEYMANN u. SCOTT (1948) können jedoch die Treatment-Parameter nicht in statistisch befriedigender Weise geschätzt werden, wenn ihre Schätzung zugleich mit der Schätzung inzidenteller Parameter erfolgen muß, da dann die Anzahl der zu schätzenden Parameter nicht gegen einen endlichen Wert konvergiert, während die Anzahl der Beobachtungen gegen unendlich strebt. Nur Modelle, die eine Trennung der Parameter erlauben, gestatten statistisch einwandfreie Parameterschätzungen und infolge der Parametertrennung einwandfreie Stichprobenunabhängigkeit der Ergebnisse. Wie aus Arbeiten von RASCH (1966a) und ANDERSEN (1970) hervorgeht, ist die Separierbarkeit der Parameter untrennbar mit der spezifischen Objektivität verbunden. An Hand eines Repräsentationssatzes zeigen wir, daß die spezifische Objektivität erhalten bleibt, wenn man die Testmodelle von RASCH (1966a) wie folgt verallgemeinert:

Gegeben sind N nichtleere Mengen M_1 bis M_N. Über der Menge der experimentellen Situationen $M \subseteq M_1 x M_2 x \ldots x M_N$ sind m - 1 reellwertige Funktionen

$$p^{(l)} : \bar{M} \longrightarrow \rangle 0,1 \langle \qquad (5.4)$$
$$(v_1, \ldots, v_N) \longrightarrow p^{(l)}(v_1, \ldots, v_N)$$
$$(l = 1, \ldots, m-1)$$

definiert. $p^{(l)}(v_1, \ldots, v_N)$ bezeichnet die Wahrscheinlichkeit, mit der in der Versuchssituation $(v_1, \ldots, v_N)$ das Reultat l eintritt. Gesucht werden N Abbildungen auf den positiven (m - 1)-dimensionalen euklidischen Raum

$$g_R : M_R \longrightarrow \mathcal{R}^+_{m-1} \qquad (5.5)$$
$$v_R \longrightarrow x_{v_R} = (x^{(1)}_{v_R}, \ldots, x^{(l)}_{v_R}, \ldots, x^{(m-1)}_{v_R})$$
$$(R = 1, \ldots, N)$$

derart, daß

$$p^{(l)}(v_1, \ldots, v_N) = \frac{\prod_{R=1}^{N} x^{(l)}_{v_R}}{1 + \sum_{h=1}^{m-1} \prod_{R=1}^{N} x^{(h)}_{v_R}} \qquad (5.6)$$
$$(\forall\, (v_1, \ldots, v_N) \in M, l = 1, \ldots, m-1)$$

erfüllt ist. Eine äquivalente Schreibweise (Beweis s.S. 32) führt eine umkehrbar eindeutige Transformation der Funktionen $p^{(l)}$ durch:

$$\Phi : \rangle 0,1 \langle \longrightarrow \mathcal{R}_+ \qquad (5.7)$$
$$p^{(l)}(v_1, \ldots, v_N) \longrightarrow c^{(l)}(v_1, \ldots, v_N)$$
$$(l = 1, \ldots, m-1)$$

mit

$$c^{(l)}(v_1, \ldots, v_N) = \frac{p^{(l)}(v_1, \ldots, v_N)}{1 - \sum_{h=1}^{m-1} p^{(h)}(v_1, \ldots, v_N)} \qquad (5.8)$$
$$(l = 1, \ldots, m-1)$$

und fordert, daß die Abbildungen g_R mit den Funktionen $c^{(l)} = \Phi \cdot p^{(l)}$ in dem Zusammenhang

$$c^{(l)}(v_1, \ldots, v_N) = \sum_{R=1}^{N} x^{(l)}_{v_R} \qquad (5.9)$$
$$(\forall\, (v_1, \ldots, v_N) \in \bar{M}, l = 1, \ldots, m-1)$$

stehen. Es gelten die beiden Sätze:

Repräsentationssatz: Abbildungen g_R, die mit den Funktionen $c^{(l)}$ in der Form (5.9) und folglich mit den Funktionen $p^{(l)}$ in der Form (5.6) zusammenhängen, gibt es genau dann, wenn die Funktionswerte $c^{(l)}(v_1,\ldots,v_N)$ der kritischen Bedingung

$$\frac{c^{(l)}(a_1,\ldots,b_{R-1},w_R,c_{R+1},\ldots,d_N)}{c^{(l)}(a_1,\ldots,b_{R-1},v_R,c_{R+1},\ldots,d_N)} = \tag{5.10}$$

$$= \frac{c^{(l)}(i_1,\ldots,j_{R-1},w_R,o_{R+1},\ldots,q_N)}{c^{(l)}(i_1,\ldots,j_{R-1},v_R,o_{R+1},\ldots,q_N)}$$

$(\forall\, a_1,i_1 \in M_1,\ldots,b_{R-1},j_{R-1} \in M_{R-1}, w_R,v_R \in M_R, c_{R+1},o_{R+1} \in M_{R+1},\ldots$
$\ldots, d_N,q_N \in M_N; l = 1,\ldots,m-1)$

genügen. (Beweis s. S. 33.)

Eindeutigkeitssatz: Die Abbildungen g_R sind bis auf Transformationen der Form

$$\begin{aligned} \psi_R : \mathcal{R}^+_{m-1} &\longrightarrow \mathcal{R}^+_{m-1} \\ x_{v_R} &\longrightarrow x'_{v_R} \\ x'_{v_R} &= (\beta_R^{(1)} x_{v_R}^{(1)},\ldots,\beta_R^{(l)} x_{v_R}^{(l)},\ldots,\beta_R^{(m-1)} x_{v_R}^{(m-1)}) \end{aligned} \tag{5.11}$$

bestimmt. Zwischen den Maßstabskomponenten $\beta_R^{(l)}$ besteht die Beziehung

$$\prod_{R=1}^{N} \beta_R^{(l)} = 1, \tag{5.12}$$

$(l = 1,\ldots,m-1)$.

(Beweis s. S. 35.)

Die allgemeinste Interpretation des erweiterten Modells (5.6)[5] besagt, daß jede Versuchsituation $(v_1,\ldots,v_N)$ nach N Gesichtspunkten klassifiziert und jede Klasse v_R jeder Klassifikation M_R durch einen $(m-1)$-dimensionalen Parameter x_{v_R} charakterisiert wird. Z.B. könnten M_1 eine Menge von Vpn, M_2 eine Menge von Testitems und M_3 bis M_N Mengen von Versuchsbedingungen sein; $(v_1,v_2,v_3,\ldots,v_N)$ beschreibt dann eine Situation, in der Vp v_1 das Item v_2 unter den Bedingungen v_3 bis v_N bearbeitet. Die l-te Dimension $x_{v_R}^{(l)}$ des Parameters x_{v_R} beschreibt den Einfluß der Klasse v_R der Klassifikation M_R auf die Wahscheinlichkeit, mit der die Versuchssituation zum l-ten Resultat führt. Je größer $x_{v_R}^{(l)}$, desto größer die Wahrscheinlichkeit des l-ten Resultates.

5 Ein Spezialfall des Modells (5.6) für m = 2 mögliche Resultate pro experimenteller Situation wurde bereits von Micko (1970) diskutiert und mit Erfolg auf empirische Daten angewendet.

Dabei ist zu beachten, daß das Modell ausschließlich generelle Parameter enthält. Jede Klasse v_R jeder Klassifikation M_R wird in allen experimentellen Situationen, welche bezüglich M_R in die Klasse v_R fallen, durch denselben Parameter x_{v_R} charakterisiert. Wechselwirkungen zwischen Klassifikationen sind nicht vorgesehen. Auf unser Experiment zur Untersuchung des Einflusses der Variablen Lärm und Beleuchtung auf die Testleistung in einem Zahlenreihentest übertragen, bedeutet dies, daß z.B. das Ausmaß der Leistungsverringerung, welche durch bestimmte schlechte Lichtverhältnisse gegenüber bestimmten anderen, besseren Lichtverhältnissen hervorgerufen wird, für alle Personen aus M_1 und alle Zahlenreihen aus M_2 gleich groß angenommen wird und auch nicht davon abhängen darf, wie stark die gleichzeitig experimentell eingeführte Lärmbelästigung der Vpn ist. Kann diese Gesetzmäßigkeit bei der experimentellen Überprüfung der empirischen Gültigkeit des Modells nicht nachgewiesen werden, so wird es erforderlich, die gewählte Klassifikation der Versuchssituationen zu ändern. Zwei Fälle sind denkbar

1. Zwischen zwei Klassifikationen M_3 und M_4[6] besteht vollständige Interaktion. Der Effekt der Zugehörigkeit zu jeder Klasse $v_3 \epsilon M_3$ hat in Verbindung mit jeder Klasse $v_4 \epsilon M_4$ ein anderes Ausmaß. Auf unser Beispiel übertragen: Das Ausmaß der Leistungsverringerung, welche durch die Lichtverhältnisse $v_3 \epsilon M_3$ gegenüber anderen Lichtverhältnissen $w_3 \epsilon M_3$ hervorgerufen wird, ist in Verbindung mit jeder der experimentell eingeführten Lärmbedingungen $v_4 \epsilon M_4$ ein anderes. In diesem Fall hat es keinen Sinn, von getrennten Einflüssen von Lärm und Beleuchtung zu sprechen. Wir ersetzen daher die Mehrfachklassifikation der experimentellen Situationen nach Beleuchtungs- und Lärmbedingung durch eine kombinierte Einfachklassifikation M_3'. Für jede experimentelle Situation wird angegeben, welche *Kombination* $v_3' \epsilon M_3'$ einer Lärmbedingung mit einer Beleuchtungsbedingung für sie zutrifft. Dann wird durch das Modell jeder Kombination einer Lärmbedingung mit einer Beleuchtungsbedingung ein Parameter zugeordnet, also der generelle Einfluß der Kombinationen von Lärm- und Beleuchtungsbedingungen gemessen.
2. Zwischen zwei Klassifikationen M_3 und M_4[7] besteht teilweise Interaktion. Der Effekt der Zugehörigkeit experimenteller Situationen zu jeder Klasse $v_3 \epsilon M_3$ hat zwar nicht in Verbindung mit jeder Klasse $v_4 \epsilon M_4$ dasselbe Ausmaß, es gibt aber Teilmengen von Klassen aus M_4, für die der Effekt ein genereller ist. Auf unser Beispiel übertragen, ist dies z.B. dann der Fall, wenn das Ausmaß der Leistungsverringerung, welche durch die Lichtverhältnisse $v_3 \epsilon M_3$ gegenüber anderen Lichtverhältnissen $w_3 \epsilon M_3$ hervorgerufen wird, in Verbindung mit machen Lärmbedingungen dasselbe ist, sich

6 O.B.d.A.
7 O.B.d.A.

aber vom Ausmaß der Leistungsverminderung unterscheidet, welches hervorgerufen wird, wenn die Beleuchtungsbedingungen in Verbindung mit bestimmten anderen Lärmbedingungen auftreten. Es ist dann zwar noch sinnvoll, von getrennten Einflüssen von Lärm und Beleuchtung zu sprechen, der Einfluß der Beleuchtungsbedingungen kann aber nicht mehr als ein genereller, sondern nur noch als ein bedingt genereller Effekt angesehen werden. Wir behalten daher die Klassifikation von experimentellen Situationen nach Lärmbedingungen (M_4) bei, ersetzen aber die Klassifikation nach Beleuchtungsbedingungen (M_3) durch eine kombinierte Klassifikation (M_3'), indem wir für jede experimentelle Situation angeben, welche Beleuchtungsbedingung für sie zutrifft und welcher Teilmenge von Lärmbedingungen die gleichzeitig zutreffende Lärmbedingung angehört. Dann wird durch das Modell jeder Lärmbedingung $v_4 \in M_4$ ein Parameter zugeordnet und jeder Kombination $v_3' \in M_3'$ einer Beleuchtungsbedingung mit einer Teilmenge von Lärmbedingungen ein weiterer Parameter. Es werden also der generelle Effekt der Lärmbedingungen und die bedingt generellen Effekte der Beleuchtungsbedingungen, welche in Verbindung mit bestimmten Teilmengen von Lärmbedingungen auftreten, gemessen.

Interaktionen zwischen drei oder mehr Klassifikationen können in analoger Weise behandelt werden.

6. Beweisanhang

Beweis zu Seite 29:

Einsetzen von (5.6) in (5.8) ergibt

$$c^{(l)}(v_1, \ldots, v_N) = \frac{\dfrac{\prod_{R=1}^{N} x_{v_R}^{(l)}}{1 + \sum_{h=1}^{m-1} \prod_{R=1}^{N} x_{v_R}^{(h)}}}{1 - \sum_{b=1}^{m-1} \dfrac{\prod_{R=1}^{N} x_{v_R}^{(b)}}{1 + \sum_{h=1}^{m-1} \prod_{R=1}^{N} x_{v_R}^{(h)}}}$$

$$= \prod_{R=1}^{N} x_{v_R}^{(l)}$$

Umgekehrt folgt aus (5.8)

$$p^{(l)}(v_1,\dots,v_N) = c^{(l)}(v_1,\dots,v_N)\,(1-\sum_{h=1}^{m-1} p^{(h)}(v_1,\dots,v_N))$$

$$= \frac{c^{(l)}(v_1,\dots,v_N)}{\dfrac{1-\sum\limits_{h=1}^{m-1} p^{(h)}(v_1,\dots,v_N)+\sum\limits_{h=1}^{m-1} p^{(h)}(v_1,\dots,v_N)}{1-\sum\limits_{h=1}^{m-1} p^{(h)}(v_1,\dots,v_N)}}$$

$$= \frac{c^{(l)}(v_1,\dots,v_N)}{1+\dfrac{\sum\limits_{h=1}^{m-1} p^{(h)}(v_1,\dots,v_N)}{1-\sum\limits_{h=1}^{m-1} p^{(h)}(v_1,\dots,v_N)}}$$

$$= \frac{c^{(l)}(v_1,\dots,v_N)}{1+\sum\limits_{h=1}^{m-1} c^{(h)}(v_1,\dots,v_N)}$$

Also ist Φ umkehrbar eindeutig. Einsetzen von (5.9) in obige Gleichung ergibt schließlich (5.6). Die Schreibweisen (5.6) und (5.9) sind daher äquivalent.

Beweis des Repräsentationssatzes (zu Seite 30):

Wie man durch Einsetzen von (5.9) in (5.10) leicht nachprüfen kann, ist die kritische Bedingung (5.10) erfüllt, wenn es Abbildungen g_R gibt, die der repräsentierenden Relation (5.9) genügen. Daß auch die Umkehrung gilt, zeigen wir mittels vollständiger Induktion.

Zunächst definieren wir Abbildungen g_R, indem wir aus jeder der Mengen M_2 bis M_N ein beliebiges Element herausgreifen, (o.B.d.A. seien das die Elemente v_R, $(R = 2,\dots,N)$), und ihm ein beliebiges $(m-1)$-Tupel positiver reeller Zahlen $(x_{v_R}^{(1)},\dots,x_{v_R}^{(m-1)})$ zuordnen. Einem beliebigen Element der Menge M_1, (o.B.d.A. sei dieses Element v_1), ordnen wir den Vektor $(x_{v_1}^{(1)},\dots,x_{v_1}^{(m-1)})$ zu, worin

$$x_{v_1}^{(l)} =: \frac{c^{(l)}(v_1,\dots,v_N)}{\prod\limits_{R=2}^{N} x_{v_R}^{(l)}}\;;$$

$(l = 1,\dots,m-1)$.

Jedem anderen Element w_R, $(R = 1,\dots,N)$ ordnen wir den Vektor $(x_{w_R}^{(1)},\dots,x_{w_R}^{(m-1)})$ zu, worin

$$x_{w_R}^{(l)} =: \frac{c^{(l)}(v_1, \ldots, v_{R-1}, w_R, v_{R+1}, \ldots, v_N)}{c^{(l)}(v_1, \ldots, v_{R-1}, v_R, v_{R+1}, \ldots, v_N)};$$

$(l = 1, \ldots, m-1)$.

Damit sind unsere Abbildungen g_R eindeutig definiert. Zu zeigen ist, daß sie mit den Funktionen $c^{(l)}$ in der Form (5.9) zusammenhängen, wenn die kritische Bedingung (5.10) erfüllt ist.

Setzen wir $R = 1$, dann gilt für beliebige Elemente w_1 aus M_1:

$$c^{(l)}(w_1, v_2, \ldots, v_N) = \frac{x_{w_1}^{(l)}}{x_{v_1}^{(l)}} c^{(l)}(v_1, \ldots, v_N)$$

$$= \frac{\dfrac{x_{w_1}^{(l)} c^{(l)}(v_1, \ldots, v_N)}{c^{(l)}(v_1, \ldots, v_N)}}{\prod_{s=2}^{N} x_{v_s}^{(l)}} = x_{w_1}^{(l)} \cdot \prod_{s=2}^{N} x_{v_s}^{(l)};$$

$(l = 1, \ldots, m-1)$.

Die Abbildungen g_R stehen also in dem Zusammenhang (5.9) mit den Funktionen $c^{(l)}$, solange wir beliebige Elemente w_1 der Menge M_1 und die Elemente v_1 bis v_N der Mengen M_2 bis M_N betrachten.

Induktionsannahme: Die Abbildungen g_R stehen in dem Zusammenhang (5.9) mit den Funktionen $c^{(l)}$, solange wir beliebige Elemente w_1 bis w_R der Mengen M_1 bis M_R und die Elemente v_{R+1} bis v_N aus M_{R+1} bis M_N betrachten, d.h.:

$$c^{(l)}(w_1, \ldots, w_R, v_{R+1}, \ldots, v_N) = \prod_{t=1}^{R} x_{w_t}^{(l)} \prod_{s=R+1}^{N} x_{v_s}^{(l)};$$

$(l = 1, \ldots, m-1)$.

Ist die kritische Bedingung (5.10) erfüllt, dann gilt:

$$\frac{c^{(l)}(w_1, \ldots, w_R, w_{R+1}, v_{R+2}, \ldots, v_N)}{c^{(l)}(w_1, \ldots, w_R, v_{R+1}, v_{R+2}, \ldots, v_N)} =$$

$$= \frac{c^{(l)}(v_1, \ldots, v_R, w_{R+1}, v_{R+2}, \ldots, v_N)}{c^{(l)}(v_1, \ldots, v_R, v_{R+1}, v_{R+2}, \ldots, v_N)} =$$

$$= \frac{x_{w_{R+1}}^{(l)}}{x_{v_{R+1}}^{(l)}};$$

$(l = 1, \ldots, m-1)$

und daher laut Induktionsannahme

$$c^{(l)}(w_1,\dots,w_R,w_{R+1},v_{R+2},\dots,v_N)\,x^{(l)}_{v_{R+1}} =$$
$$= x^{(l)}_{w_{R+1}} \prod_{t=1}^{R} x^{(l)}_{w_t} \prod_{x=R+1}^{N} x^{(l)}_{v_s}$$

$(l = 1,\dots,m-1)$

und daher

$$c^{(l)}(w_1,\dots,w_R,w_{R+1},v_{R+2},\dots,v_N) = \prod_{t=1}^{R+1} x^{(l)}_{w_t} \prod_{s=R+2}^{N} x^{(l)}_{v_s};$$

$(l = 1,\dots,m-1)$.

Setzen wir $R = N-1$, so folgt die Gültigkeit von (5.9) für beliebige Auswahl von Elementen aus M_1 bis M_N; also gibt es Abbildungen g_R, die (5.9) erfüllen, genau dann, wenn die Funktionen $c^{(l)}$ der kritischen Bedingung (5.10) genügen.

Beweis des Eindeutigkeitssatzes (zu Seite 30):

Seien g_R und g'_R, $(R = 1,\dots,N)$, Abbildungen, die mit den Funktionen $c^{(l)}$, $(l = 1,\dots,m-1)$, in der Form (5.9) zusammenhängen, dann definieren wir:

$$\beta^{(l)}_{v_R} =: x^{(l)}_{v_R}/x'^{(l)}_{v_R};$$

$(l = 1,\dots,m-1)$.

Aus irgendeiner der Mengen M_R, $(R = 1,\dots,N)$ wählen wir zwei beliebige Elemente aus; o.B.d.A. seien das die Elemente v_R und w_R. Wir erhalten dann:

$$\beta^{(l)}_{v_R} = x^{(l)}_{v_R}/x'^{(l)}_{v_R} \qquad \text{und} \qquad \beta^{(l)}_{w_R} = x^{(l)}_{w_R}/x'^{(l)}_{w_R};$$

$(l = 1,\dots,m-1)$.

Aus (5.9) folgt:

$$\frac{c^{(l)}(v_1,\dots,v_{R-1},w_R,v_{R+1},\dots,v_N)}{c^{(l)}(v_1,\dots,v_{R-1},v_R,v_{R+1},\dots,v_N)} = \frac{x^{(l)}_{w_R}}{x^{(l)}_{v_R}} = \frac{x'^{(l)}_{w_R}}{x'^{(l)}_{v_R}};$$

$(l = 1,\dots,m-1)$,

und schließlich:

$$\beta^{(l)}_{w_R} = \frac{x^{(l)}_{w_R}}{x'^{(l)}_{w_R}} = \frac{x^{(l)}_{v_R}}{x'^{(l)}_{v_R}} = \beta^{(l)}_{v_R} = \beta^{(l)}_R,$$

$(l = 1,\dots,m-1)$.

Ferner gilt:

$$c^{(l)}(v_1, \ldots, v_N) = \prod_{R=1}^{N} x_{v_R}^{(l)} = \prod_{R=1}^{N} x_{v_R}^{\prime(l)} = \prod_{R=1}^{N} x_{v_R}^{(l)} \prod_{R=1}^{N} \beta_R^{(l)}$$

$(l = 1, \ldots, m-1)$

und daher:

$$\prod_{R=1}^{N} \beta_R^{(l)} = 1$$

$(l = 1, \ldots, m-1)$.

Zusammenfassung

Das mehrdimensionale Testmodell von RASCH wird von 2 auf N Parameter verallgemeinert. Das verallgemeinerte Modell gestattet, experimentalpsychologische Versuchssituationen nach N Gesichtspunkten, die sich auf Merkmale der Personen, der verwendeten Instrumente oder der Versuchsbedingungen beziehen, zu klassifizieren und durch eine Kombination von N Parametern, die spezifisch objektiv bestimmbar sind und damit Stichprobenunabhängigkeit der Ergebnisse gewährleisten, zu beschreiben.

Literatur

ANDERSEN, E. B.: Conditional Inference for Multiple Choice Questionnaires. Copenhagen School of Economics and Business Administration. Report No. 8 (1970).

FISCHER, G. H.: Neue Entwicklungen in der psychologischen Testtheorie. In: Fischer, G. H. (Hrsg.): Psychologische Testtheorie. Bern: Huber (1968).

GUTTMAN, L.: The Basis for Scalogramm Analysis. In: Stauffer, S. A. et al.: Measurement and Prediction. Princeton: Princeton University Press (1950).

KEMPF, W. F.: Grundlagenuntersuchungen zum Thematischen Apperzeptionstest. Phil. Diss. Wien (1970).

MICKO, H. C.: Eine Verallgemeinerung des Meßmodells von Rasch mit einer Anwendung auf die Psychophysik der Reaktionen. Psychologische Beiträge XII, 4–22 (1970).

NEYMANN, J. u. SCOTT, E. L.: Consistent estimates based on partially consistent observations. Econometrica Vol. 16, 1–32 (1948).

RASCH, G.: An informal report on a theory of objectivity in comparisons. Proceedings of the NUFFIC international summer session in science at "Het Oude Hof", The Hague, 14.–28. July (1966a).

RASCH, G.: An Individualistic Approach to Item Analysis. In: Lazarsfeld, P. F. u. Henry, N. W. (Hrsg.): Readings in Mathematical Social Science. Chicago: Science Research Associates (1966b).

STENE, J.: Einführung in Raschs Theorie psychologischer Messung. In: Fischer, G. H (Hrsg.): Psychologische Testtheorie. Bern: Huber (1968).

SUPPES, P. u. ZINNES, J. L.: Basic Measurement Theory. In: Luce, R. D., Bush, R. R. u. Galanter, E. (Hrsg.): Handbook of Mathematical Psychology, Vol. I. New York: Wiley (1963).

Anschrift des Verfassers:

Dr. Wilhelm Kempf
Psychologisches Institut der Universität Erlangen-Nürnberg
Abteilung Experimentelle Psychologie
8520 Erlangen
Bismarckstraße 1