

Skalierung

Ingwer Borg und Peter H. Schönemann

Der Begriff *Skalierung* wird in der Wissenschaft in vielfältiger Weise verwendet. Im allgemeinen wird mit Skalierung eine Maßstabsänderung bezeichnet, etwa beim Ausschneiden und Vergrößern eines Teils einer geographischen Karte, beim Einpassen einer Abbildung in einen vorgegebenen Rahmen, bei der Überführung einer Temperaturmessung von Grad Celsius auf Grad Fahrenheit oder bei der Adjustierung der Koeffizienten in einem Gleichungssystem. Solche Skalierungen sollen interessante Informationen zugänglicher machen oder die Qualität dieser Informationen (z.B. die Genauigkeit der Lösung eines Gleichungssystems) erhöhen.

Den meisten Methoden der Skalierung, in jedem Fall aber dem Begriff der *Skala* (*scala* = Treppe), liegt die Vorstellung einer Meßlatte oder eines Thermometers zugrunde, mit dem man Gegenstände der Empirie (z.B. Personen) so vermessen kann, daß aus den entstehenden Skalenwerten Informationen entnommen werden können (z.B. durch Vergleich der Skalenwerte selbst, durch Vergleich von Differenzen oder Verhältnissen der Skalenwerte, oder durch weitere Verrechnung der Skalenwerte), die sich in der Empirie als gültig erweisen. Welche Relationen und Operationen auf der Skala empirisch gültige Einsichten vermitteln, muß methodisch und/oder empirisch geklärt werden.

Im folgenden geben wir eine kurze Übersicht über die Hauptthemen der Skalierung.

1 Skalierung als Datenerhebung

Vor allem in der mehr sozialwissenschaftlich orientierten Psychologie wird unter Skalierung vielfach die direkte Erhebung von Meßwerten verstanden. Ein Beispiel ist die *Likert-Skalierung*, bei der Befragten eine Aussage wie „Ich sehe der Zukunft mit Optimismus entgegen“ vorgelegt wird mit der Aufforderung, hierauf durch Auswahl einer Zahl aus einer Kategorienskala wie etwa „1 = trifft voll zu, 2 = trifft zu, 3 = teils-teils, 4 = trifft nicht zu, 5 = trifft gar nicht zu“ (*Rating-Skala*; siehe dazu McReynolds & Ludwig, 1987) zu antworten. Hierbei finden verschiedene Skalen Verwendung, z.B. fünf- oder siebenstufige Skalen, Skalen mit bzw. ohne „neutrale“ Mittelkategorie, Skalen mit und ohne verbale Kennzeichnung der Kategorien, bipolare Skalen (wie oben) und unipolare Skalen (wie etwa „0%, 10%, ..., 100%“) usw. Welche Auswirkungen derartige Variationen auf die sich ergebenden Daten haben, wie zuverlässig und valide *ratings* sind und wie *Rating-Urteile* zu erklären sind, wurde extensiv untersucht (Guilford, 1954, Kapitel 11; Schuman & Presser, 1981; Haubensak, 1985).

Zweck einer solchen Datenerhebung ist es, statt freier und möglicherweise verbosere Antworten des Befragten einen für alle Befragten gleichen Bereich zulässiger Antworten festzulegen. Die Antworten sind zudem als Zahlen direkt der statistischen Analyse zugänglich.

Einen etwas engeren Fokus hat die Größenschätzung (*magnitude estimation*), bei der die Vpn gebeten werden, das Verhältnis eines Objektes i relativ zu einem Anker k hinsichtlich einer bestimmten Eigenschaft anzugeben. Zum Beispiel können i und k Töne sein, die Eigenschaft „wahrgenommene Lautstärke“, und das Urteil kann lauten „ i ist 3.5 Mal so laut wie k “. Im Kontext der Wahrnehmungspsychologie findet man, daß der Zusammenhang der hierbei resultierenden Größenschätzskala und der entsprechenden physikalischen Verhältnisse im allgemeinen durch eine Potenzfunktion beschrieben werden kann (Stevens, 1975; vgl. auch Irtel, in diesem Band).

2 Skalierung als anschauliche Darstellung von Daten

Der weitaus größte Teil der Skalierung beschäftigt sich mit Fragen zur Weiterverarbeitung gegebener numerischer Daten. Die naheliegendste Frage ist dabei die, wie man die in der Praxis oft sehr umfangreichen Daten so darstellen kann, daß sie besonders anschaulich werden.

Nehmen wir an, wir hätten bei 50 Personen jeweils 10 Eigenschaften erfaßt. Anstatt die Zahlenwerte in der 50×10 Datenmatrix zu analysieren, könnten wir jede ihrer Zeilen als Punkt eines zehndimensionalen Raums, aber z.B. auch als *Chernoff-Gesicht* (Chernoff, 1973) darstellen, d.h. vermittelt eines Cartoon-Gesichtes, in dem die verschiedenen Gesichtsmerkmale die Variablen repräsentieren. So kann man etwa durch die Krümmung des Mundes (vom „U“ bis zum „umgekehrten U“) die Werte der Variablen X_1 , durch den Winkel der Augenbrauen eine Variable X_2 usw. darstellen. Mit Chernoff-Gesichtern dieser Art, die leicht auf dem Computer zu produzieren sind, lassen sich die Meßwerte der 50 Personen in anschaulicher Form darstellen.

Die Darstellung von Daten durch derartige *Ikonen* ist immer möglich. Man spricht daher auch von *trivialer Skalierung* (Borg & Staufenbiel, 1993), weil die Skalierung selbst keine weiteren Einsichten bringt. Der Nutzen der Skalierung liegt ausschließlich in ihrer Anschaulichkeit. Ermöglicht sie z.B. eine Typologisierung, die sich in den Daten selbst bestätigen läßt, dann hat sie ihren Zweck erfüllt. Dieses Kriterium läßt sich empirisch prüfen (Borg & Staufenbiel, 1992).

3 Skalierung als Indexbildung

Eine weitere Variante des Begriffs Skalierung ist, im typischen Beispiel, die Messung bestimmter Eigenschaften einzelner Personen mit Hilfe von Item-Batterien. Die einzelnen Antworten werden zu einem Gesamtwert verrechnet, der den Skalenwert darstellt (*summated scale*). Derartige Skalenwerte sollten differenzierter, zuverlässiger und umfassender sein als Antwortwerte für Einzelitems, vorausgesetzt, alle Items messen das gleiche Konstrukt. Der einfachste Fall ist ein Fähigkeitstest, bei dem die Summe der richtigen Lösungen bei den einzelnen Testaufgaben den Grad der Fähigkeit indiziert.

Der Konstruktion (d.h. Zusammenstellung) der Item-Batterie wird besondere Aufmerksamkeit geschenkt. Items werden formuliert und empirisch auf ihre Homogenität geprüft, d.h. es wird getestet, ob sie eindimensional sind (i.S. der Faktorenanalyse oder eines bestimmten Testmodells). Nicht passende Items werden so lange eliminiert, bis die verbleibende Itembatterie sich empirisch als homogen erweist (vgl. den Beitrag über *Latent-Trait-Modelle* von Roskam und den Beitrag von Stumpf, in diesem Band).

Homogenität ist aber nicht zwingend, wenn, wie z.B. bei der Intelligenz, die einzelnen Dimensionen des Konstruktes positiv korreliert sind. Dann ist es eine Frage des praktischen Nutzens, ob man Subskalen bildet oder alle Werte zu einem einzigen Index verrechnet.

Empirisch wird weiter untersucht, wie zuverlässig (z.B. stabil bei Meßwiederholungen) und valide (z.B. nützlich für die Vorhersage anderer Meßwerte) die Skalenwerte sind.

Skalierung im Sinne einer solchen Indexbildung wird aus angewandter Motivation heraus betrieben. Wichtig ist vor allem, daß die Skalenwerte zur Vorhersage anderer Meßwerte taugen. Die Werte eines Fähigkeitstests sollen sich beispielsweise zur Vorhersage einer späteren Prüfungs- oder Berufsleistung eignen. Was genau dabei gemessen wird, interessiert bei dieser Ingenieur-Perspektive nur begrenzt.

4 Skalierung als bedingtes Messen

Ein wesentlicher Teil der älteren Skalierungsliteratur (Coombs, 1964; Torgerson, 1958) beschäftigt sich mit der Frage, ob und wie man Beobachtungen über den Zusammenhang von Objekten aus einer oder zwei Mengen als Abstände zwischen Punkten einer Skala darstellen kann. Der Zusammenhang zwischen Daten und Skala wird als Modell formuliert.

So gehen etwa die *Fechner-Modelle* (Baird & Noma, 1978) davon aus, daß die interessierenden Objekte einer Menge (z.B. bestimmte politische Parteien) paarweise untereinander verglichen werden und daß für jedes Paar (i, j) ein Wert $p(i, j)$ vorliegt, der angibt, wie stark Objekt i über Objekt j in einem bestimmten Sinn (z.B. Wahlbevorzugung) dominiert. Die Werte $p(i, j)$ müssen zwischen 0.00 und 1.00 liegen und werden als Wahrscheinlichkeiten gedeutet. Empirisch werden die $p(i, j)$ -Parameter meist über Anteilswerte geschätzt, also z.B. als der (durch 100 dividierte) Prozentsatz der Personen aus einer Stichprobe von Befragten, die i gegenüber j vorziehen.

Die Fechner-Modelle formulieren eine Beziehung zwischen den $p(i, j)$ -Daten und den Skalenwerten $s(i)$ bzw. $s(j)$ für die Objekte i und j . Die Beziehung lautet: (a) Wenn $p(i, j) = 0.5$, dann auch $s(i) - s(j) = 0$; (b) wenn $p(i, j) > p(k, l)$, dann auch $s(i) - s(j) > s(k) - s(l)$, für alle i, j, k, l . Je extremer also die Wahrscheinlichkeit $p(i, j)$ ist (das heißt z.B.: je häufiger Partei i gegenüber j präferiert wird), desto „weiter oben“ auf der Skala s sollte i relativ zu j liegen. Gelten die Beziehungen (a) und (b), dann kann man die $(n)(n - 1)/2$ unabhängig voneinander beobachteten Daten aus nur n Skalenwerten rekonstruieren („erklären“). Dabei wird angenommen, daß $p(i, j) = 1 - p(j, i)$. Selten wird auch $p(j, i)$ empirisch geschätzt, so daß die Gültigkeit dieser Annahme meist nicht geprüft werden kann.

Man kann nun fragen, unter welchen Umständen die Darstellbarkeit von p -Werten

im Sinne der Fechner-Modelle überhaupt möglich ist. Dabei ergibt sich, daß dies genau dann der Fall ist, wenn für alle i, j, k eine Art Dreiecksgleichung (*Darstellungsrelation*) gilt: $f[p(i, k)] = f[p(i, j)] + f[p(j, k)]$, mit f als irgendeiner streng monotonen Funktion, für alle i, j, k . Da $f[p(i, j)]$ der Differenz $s(i) - s(j)$ im Modell der Skala entspricht, heißt das also, daß sich die beobachteten p -Werte „in etwa“ (d.h. bis auf Verzerrungen durch f , die die Ordnung der p -Werte aber nicht verändern darf) wie Differenzwerte einer Skala addieren lassen sollten. Daraus folgt, daß die p -Werte untereinander eine bestimmte Ordnungsstruktur aufweisen müssen, damit sie skalierbar i.S. des Fechner-Modells sind (Mausfeld, 1985).

Spezifiziert man f – wie etwa in Thurstones *Case-V-Modell* (Thurstone, 1927) als kumulative Gauß-Funktion – dann ist die Dreiecksgleichung sogar numerisch exakt prüfbar. Nehmen wir, um dies zu zeigen, der Einfachheit halber an, f sei die lineare Funktion $f(p) = p - 0.5$. Weiter sei $p(\text{CDU}, \text{SPD}) = 0.60$ und $p(\text{CDU}, \text{FDP}) = 0.90$. Dann muß $p(\text{FDP}, \text{SPD}) = 0.20$ gelten, wie man durch Einsetzen in die Darstellungsrelation findet. Auf der Skala liegt die CDU somit 0.10 Einheiten über der SPD und diese wiederum 0.30 Einheiten über der FDP.

In der Praxis findet man, daß die Darstellungsrelation selten exakt gilt. Dann stellt sich die Frage, wie man trotzdem eine Skala findet, die die empirischen p -Schätzungen so gut wie möglich darstellt. Das „so gut wie möglich“ wird formuliert in einer *Verlustfunktion*, meist einem Kleinste-Quadrate-Kriterium, das durch eine entsprechende Vorgehensweise (Algorithmus) bei der Berechnung der Skala zu minimieren ist.

Skalierung dieser Art wird als *bedingtes Messen* (vgl. Schönemann & Borg, 1983) definiert, weil sich ähnliche Fragen stellen wie beim fundamentalen Messen (vgl. Niederée & Narens, in diesem Band). Ausgegangen wird allerdings nicht von einem empirischen Relativ, sondern von numerischen Daten, denen bestimmte Eigenschaften zugeschrieben werden. (Insofern ist das Messen „bedingt“: Die Aussagen hängen immer davon ab, ob bzw. wie weit diese Eigenschaften gelten.) Die Fragen sind:

- Welcher Art müssen die Daten sein (formale Eigenschaften, Struktur), damit sie in einem bestimmten Modell dargestellt werden können? (*Existenzfrage*)
- Bis auf welche Transformationen liegt die Repräsentation dann fest? (*Eindeutigkeitsfrage*)
- Welche Informationen kann man dann der Repräsentation entnehmen? (*Deutbarkeitsfrage*)
- Wie konstruiert man die Skala dann, wenn sie existiert? (*Konstruktionsfrage*)

Im Grunde tritt bei der Behandlung dieser Fragen der Zweck der Skalierung als Meßmethode in den Hintergrund gegenüber der Untersuchung der Bedingungen, unter denen die Darstellung der Daten in einem jeweils besonderen Sinn möglich ist. Die Skalierungsmodelle selbst werden interpretiert als Theorien darüber, wie die empirischen Beobachtungen zustande kommen. Das Thurstone *Case-V-Modell* beispielsweise ist eine Theorie der Unterschiedswahrnehmung bei eindimensional variiierenden Objekten.

Die klassischen Skalierverfahren von Thurstone (1927), Guttman (1944), Coombs (1950) u.a. (zusammenfassend dargestellt z.B. in Torgerson, 1958) basieren allesamt auf *falsifizierbaren Modellen*. Daten, die in diesen Modellen darstellbar sind, haben also ganz besondere Eigenschaften, die nicht für alle Daten gelten. Die Darstellbarkeit selbst weist diese Eigenschaften nach.

In der Forschungspraxis wurden allerdings selten mehr als die Algorithmen bzw. die Computerprogramme zur Herstellung einer optimalen Skala beachtet. Das hat vielfach dazu geführt, daß die Anwender optimale Skalen wiederholt berechnet und nicht passende Informationen so lange eliminiert haben, bis eine im formalen Sinn akzeptable Skala entstand. Was diese Skala dann eigentlich noch messen soll und wie, geht dabei mehr oder weniger verloren.

5 Multidimensionale Skalierung

Wenn man empirische Relationen nicht durch die Eigenschaften eindimensionaler Skalen erklären kann, dann liegt es nahe, das Modell der Skala zu verallgemeinern zu einer multidimensionalen Struktur. Die multidimensionale Skalierung (MDS; Borg & Lingoes, 1987; Borg & Groenen, im Druck; Kruskal & Wish, 1978) stellt Ähnlichkeits- oder Unähnlichkeitsdaten für Paare von Objekten, $s(i, j)$, durch Distanzen $d(i, j)$ zwischen Punkten eines m -dimensionalen Raums, der *MDS-Konfiguration*, dar.

Die $s(i, j)$'s können direkt erhoben (z.B. als numerische Einschätzungen der wahrgenommenen Ähnlichkeit von jedem Objektpaar) oder errechnet werden (z.B. als Korrelationen über Eigenschaftsprofile von je zwei Objekten). Die beabsichtigte Darstellung ist dann $s(i, j) \mapsto f(s(i, j)) = d(i, j)$. Die Funktion f bestimmt das MDS-Modell. Es besagt, welche Transformationen auf den Daten $s(i, j)$ zulässig sind (siehe hierzu Velleman & Wilkinson, 1993). Eine mögliche Spezifikation von f ist die *Intervall-MDS*, bei der $f(s(i, j)) = a + b \cdot s(i, j)$ gilt, wobei a und b reellwertige Konstanten sind. Das meistverwendete Modell ist die *Ordinal-MDS*, bei der f eine monotone Funktion ist, so daß für Unähnlichkeitswerte gilt: wenn $s(i, j) < s(k, l)$, dann $d(i, j) < d(k, l)$. Dieses Modell wird manchmal auch als *nichtmetrische MDS* bezeichnet, ein zu unscharfer Begriff, weil darunter auch die *Nominal-MDS* fällt, bei der in der Darstellung nur gelten soll, daß wenn $s(i, j) \neq s(k, l)$, dann $d(i, j) \neq d(k, l)$ (Takane, Young & DeLeeuw, 1977).

Wählt man nun einen Raum bestimmter Dimensionalität (m), dann läßt sich die Distanz $d(i, j)$ für jedes Paar (i, j) von Punkten berechnen mittels $d(i, j) = [\sum_{a=1}^m (x_{ia} - x_{ja})^2]^{1/2}$, wobei x_{ia} der Koordinatenwert von Punkt i auf der Dimension a der MDS-Konfiguration ist. Die Funktion f ist so zu bestimmen, daß die Modellgleichung „möglichst genau“ gilt. Was damit gemeint ist, spezifiziert die Verlustfunktion. Will man größere Fehler, $e(i, j) = f(s(i, j)) - d(i, j)$, stärker gewichten als kleine, dann ist ein Quadratsummen-Kriterium sinnvoll, das als *Roh-Stress* bezeichnet wird: $\sum_{i < j} [f(s(i, j)) - d(i, j)]^2$. *Roh-Stress* läßt sich allerdings dadurch minimieren, daß man die Koordinatenwerte der Objekte verkleinert. Um dies zu verhindern, kann man den *Roh-Stress* durch $\sum_{i < j} d(i, j)^2$ dividieren. Zieht man noch die Quadratwurzel, ergibt sich der *Stress*-Wert (Kruskal, 1964).

Für die meisten MDS-Modelle läßt sich eine Konfiguration, die *Stress* minimiert, nur *iterativ* finden. Dabei wird zunächst irgendeine Startkonfiguration festgelegt.

Dann werden die Punkte in kleinen Schritten so lange hin- und hergeschoben, bis der *Stress* nicht mehr zu verkleinern ist (siehe hierzu Borg & Groenen, im Druck).

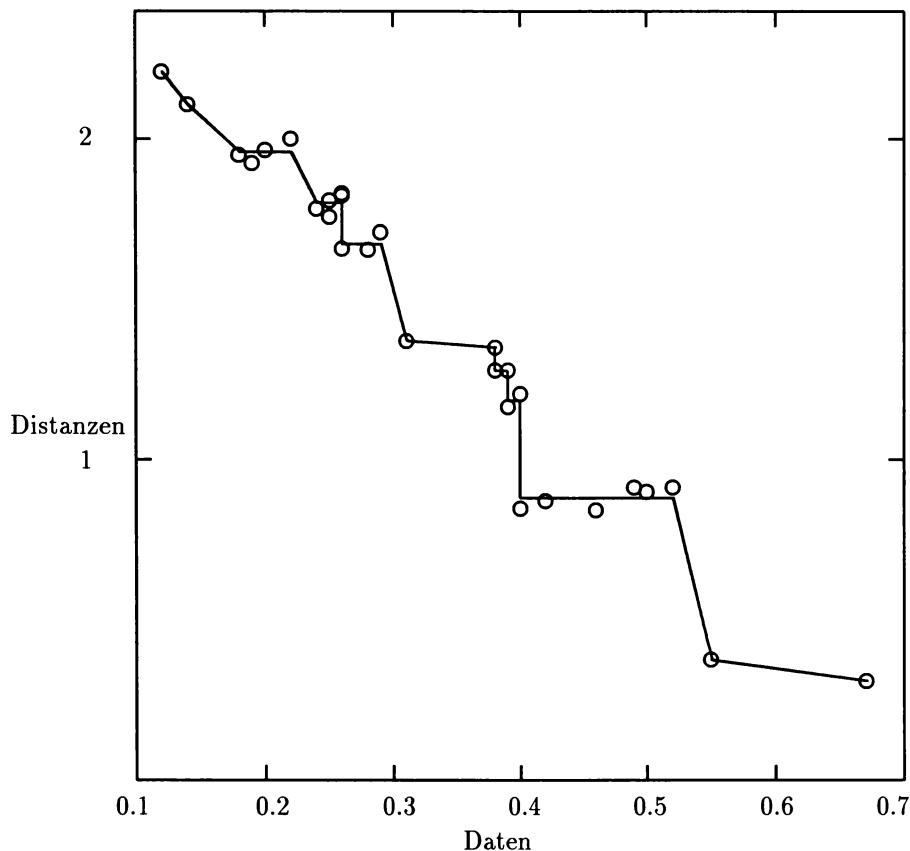


ABBILDUNG 1. Shepard-Diagramm, gehörig zur MDS-Darstellung in Abbildung 1 des Kapitels zur Facettentheorie (Borg, in diesem Band).

Im Kapitel zur Facettentheorie (Borg, in diesem Band) ist ein Beispiel für die ordinale MDS gezeigt, bei der Korrelationen in Distanzen einer zweidimensionalen MDS-Konfiguration abgebildet werden. Wie gut dies möglich ist, zeigt der *Stress*-Wert, der hier 0.01505 ist. Für ihn fordert man meist als Minimalkriterium, daß er deutlich kleiner sein soll als der Wert, den man für Zufallsdaten erwarten kann. Entsprechende Erwartungswerte liegen aus Computersimulationen vor (Borg & Lingoes, 1987). Informativer als der *Stress*-Wert ist das *Shepard-Diagramm* (vgl. Abbildung 1). In diesem Diagramm sieht man Daten-Distanz-Paare als Punkte und die Funktion f als Regressionskurve durch diesen Punkteschwarm. Im vorliegenden Fall einer ordinalen MDS von Ähnlichkeitsdaten ist die Regressionskurve monoton abfallend („treppenförmig nach unten“). Der Roh-*Stress* entspricht der Summe der vertikalen, quadrierten Abstände der Punkte von dieser Regressionskurve. In Abbildung

1 sind diese Abstände offenbar sehr klein. Der weitgehend lineare Verlauf der Regressionskurve macht zudem deutlich, daß eine Intervall-MDS zu einer sehr ähnlichen MDS-Konfiguration führen würde.

Das hier verwendete Beispiel der MDS zeigt eine Anwendung, die facettheoretisch geleitet ist und danach fragt, ob der MDS-Raum *Regionen* aufweist, die den Unterscheidungen der Facetten entsprechen. Im klassischen Fall der MDS hat man dagegen eine Theorie dafür, welche *Dimensionen* den beobachteten Relationen zugrunde liegen. Die subjektive Unähnlichkeitsstärke zwischen zwei Rechtecken könnte z.B. der Differenz ihrer Breiten plus der Differenz ihrer Höhen entsprechen. Zur Prüfung dieser Modellvorstellung würde man also die beobachteten numerischen Unähnlichkeitsurteile zu erklären versuchen durch die *City-Block-Distanz* von Punkten (d.h. die Summe der Abstände dieser Punkte entlang der Dimensionen) einer Ebene (Borg & Leutner, 1983; Schönemann, 1994; zur *City-Block-Distanz*, siehe auch Meiser & Humburg, in diesem Band).

Eine so klare dimensionale Theorie liegt allerdings selten vor. Oft überträgt man diese Modellvorstellung trotzdem ohne weitere Begründung auf die Analyse irgendwelcher (Un-)Ähnlichkeitsdaten. So werden z.B. nicht selten die Korrelationen von Fragebogenitems durch Abstände von Punkten eines mehrdimensionalen Raums (möglichst niedriger Dimensionalität) dargestellt. Anschließend wird nach der Bedeutung „der Dimensionen“ gefragt. Die Bedeutung soll sich durch Interpretation der üblichen Koordinatenachsen ergeben, ähnlich wie in der explorativen Faktorenanalyse (vgl. Schönemann & Borg, in diesem Band).

In der Praxis sind MDS-Darstellungen selten höher als zweidimensional. Das zeigt, daß die MDS oft nur zur Veranschaulichung der Datenstruktur verwendet wird. Dann empfiehlt es sich aber, *jede* Art regelhafter Muster der MDS-Punktekonfiguration zu beachten, insbesondere auch regionale Muster, bei denen die Punkte, die verschiedenen apriorischen Kategorisierungen zugeordnet sind, in verschiedenen Gebieten des MDS-Raums fallen (Borg, 1992).

Abschließend sei noch ein Spezialfall der MDS erwähnt, das *Unfolding-Modell*. Hierbei liegen Relationen zwischen zwei Mengen vor, z.B. die Präferenzstärken verschiedener Personen für die Parteien im Bundestag. Im *Unfolding-Raum* bilden die Parteien eine Punktmenge, die Personen eine andere. Letztere werden *Idealpunkte* genannt, weil laut Modell eine Partei für eine Person umso attraktiver wird, je näher sie am Punkt dieser Person liegt. Das *Unfolding-Modell* postuliert dabei, daß die Personen die Parteien in gleicher Weise wahrnehmen, aber andere Vorstellungen davon haben können, was „ideal“ ist. Technisch läßt sich das *Unfolding* mit den meisten MDS-Programmen durchführen. Die dabei entstehenden Darstellungen sind aber im multidimensionalen Fall nicht ohne weiteres interpretierbar, weil sie zahlreiche Unbestimmtheiten aufweisen (Borg & Lingoes, 1987).

6 Multiple Skalierung

In der Skalierung ist es oft üblich, die erforderlichen Daten durch Aggregation über verschiedene Personen zu gewinnen. Man unterstellt dabei, daß diese Personen im wesentlichen die gleichen Wahrnehmungen haben und damit als Replikationen ein und derselben Person betrachtet werden können. Trifft diese Annahme nicht zu,

dann kann sich Multidimensionalität als Artefakt ergeben.

In der multiplen Skalierung fragt man daher, ob die Personenstichprobe zerlegt werden kann in Teilstichproben, die jeweils durch eine eindimensionale Skala erklärt werden können (vgl. auch Rost & Erdfelder, in diesem Band). Norpoth (1979) berichtet z.B. die Ergebnisse von Umfragen, bei denen jede Person die verschiedenen politischen Parteien in eine Rangfolge der Bevorzugung bringen sollte (wie z.B. „CDU vor FDP vor SPD vor Grüne vor PDS vor NPD“). Zur Erklärung dieser Daten benötigt man einen mindestens zweidimensionalen *Unfolding*-Raum, obwohl die Vermutung nahe liegt, daß Parteienpräferenz im wesentlichen von der Lage des Idealpunktes auf einer links-rechts Dimension abhängt. Erzeugt man nun zwei Dimensionen, die sich nur dadurch unterscheiden, daß die FDP einmal links, einmal rechts von der CDU liegt, dann kann man mit diesen Dimensionen fast alle Daten erklären: Die Personen lassen sich zwei Teilgruppen zuordnen, die jeweils eindimensional skalierbar sind und sich nur hinsichtlich ihrer Wahrnehmung der FDP unterscheiden (Borg & Staufenbiel, 1993).

Ein ähnliches Beispiel berichten Coombs, Coombs und Lingoes (1978) für die typische Abfolge der Konsumgüter, die sich Befragte eines Entwicklungslandes kaufen würden, wenn sie könnten. Insgesamt sind diese Daten nicht skalierbar i.S. eines Fechner-Modells. Man kann aber die Stichprobe der Befragten zerlegen in wenige Teilstichproben, deren Wünsche jeweils skalierbar sind.

Technisch führt man die Suche nach derartigen Teilstichproben entweder inhaltlich geleitet durch oder mit Hilfe entsprechender Computerprogramme (Coombs et al., 1978). Die hierbei auftretenden Fragen erfordern allerdings eine Reihe von komplexen Abwägungen. So muß man spezifizieren, was man damit meint, möglichst wenige Teilstichproben zu finden, die jeweils noch akzeptabel gut i.S. eines bestimmten Modells darstellbar sind.

7 Nicht-dimensionale Skalierung

Skalierungsmodelle suchen im allgemeinen eine optimale Repräsentation der Daten auf einer Dimension bzw. in einem multidimensionalen Raum, fast immer dem der vertrauten euklidischen Geometrie. Gelegentlich werden aber auch Netzwerke (Graphen) skaliert oder sogar rein algebraische oder statistische Modelle. Bei letzteren ist der Übergang zur Datenanalyse fließend.

Ein Beispiel für *dimensionsloses* Skalieren (*Dual-Skalierung*) ist das *conjoint measurement* (*verbundene Messung*; vgl. auch Niederée & Narens, in diesem Band). Im einfachsten Fall liegen hierbei Beobachtungen unter Variation von zwei unabhängigen Variablen vor. Man bittet z.B. eine Person, neun Autos, beschrieben durch gute/mittlere/bescheidene Fahrleistungen und hohe/mittlere/geringe ständige Kosten, in eine Präferenz-Rangreihe zu bringen. In der Skalierung wird dann versucht, die Beobachtungen zu erklären durch eine additive Verknüpfung von (zu findenden) Skalenwerten für Fahrleistung und Kosten: Die Funktionswerte Nutzen = $f(\text{Fahrleistung}) + g(\text{Kosten})$ sollen so geordnet sein wie die Rangdaten. Das ist nur dann möglich, wenn die Daten eine bestimmte Struktur erfüllen, in der die Faktoren Kosten und Fahrleistung nicht (disordinal) interagieren. Diese strukturellen Voraussetzungen lassen sich global dadurch prüfen, daß man die Daten mit einem

entsprechenden *Conjoint-Measurement*-Programm (siehe etwa LINGOES, 1973) skaliert und prüft, ob die Verlustfunktion gleich Null ist. Differenziertere Einsichten dahingehend, welche der Voraussetzungen verletzt sind, erlaubt im Prinzip die axiomatische Vorgehensweise, vorausgesetzt allerdings, die unabhängigen Variablen sind hinreichend fein abgestuft (ADAMS & FAGOT, 1959; SCHÖNEMANN, 1994).

Sind die Voraussetzungen erfüllt oder annähernd erfüllt, interessieren vor allem die *Teilnutzen*-Funktionen f und g . Im Auto-Beispiel kann man z.B. fragen, ob eine Verringerung der Kosten eine Verschlechterung der Fahrleistungen überkompensiert oder ob eine solche Überkompensierung über den ganzen Bereich der Skalen gilt usw.

8 Weiterführende Literatur

Der Klassiker der Skalierungsliteratur ist TORGERTSON (1958). Er ist heute für Anwender etwas überholt, weil Skalierungen „per Hand“ infolge der breiten Verfügbarkeit entsprechender Computerprogramme nicht mehr nötig sind. Hier empfiehlt sich eher das Buch von BORG und STAUFENBIEL (1993), das die meisten der obigen Themen anwendungsorientiert behandelt. Hierzu gibt es auch ein Begleitbuch (Staufenbiel, 1993), das die vollständigen Daten und Steuerdateien für fast alle Beispiele enthält, so daß Leser die Verfahren durch Probieren am Computer explorieren können. Die Sichtweise des Skalierens als bedingten Messens wird in mathematisch etwas anspruchsvollere Weise dargestellt in Schönemann und Borg (1983). Zur Dual-Skalierung siehe NISHISATO (1980). Weiterhin gibt es in der preiswerten Sage-Reihe „Quantitative Applications in the Social Sciences“ eine ganze Reihe von Bänden, die einzelne Themen der Skalierung behandeln, so z.B. SPECTOR (1992; *summed scales*), ANDRICH (1988; Rasch-Modelle), LODGE (1981; Größenschätzmethode), KRUSKAL und WISH (1978; MDS), McIVER und CARMINES (1981; eindimensionale Skalierung) oder DeVellis (1991; Skalenkonstruktion).

Literaturverzeichnis

- ADAMS, E. W. & FAGOT, R. F. (1959). A model of riskless choice. *Behavioral Science*, 4, 1–10.
- ANDRICH, D. (1988). *Rasch models for measurement*. Newbury Park: Sage.
- BAIRD, J. C. & NOMA, E. (1978). *Fundamentals of scaling and psychophysics*. New York: Wiley.
- BORG, I. (1992). *Grundlagen und Ergebnisse der Facettentheorie*. Bern: Huber.
- BORG, I. & GROENEN, P. (im Druck). *Modern multidimensional scaling*. Santa Clara: Springer.
- BORG, I. & LEUTNER, D. (1983). Dimensional models for the perception of rectangles. *Perception and Psychophysics*, 34, 257–269.
- BORG, I. & LINGOES, J. C. (1987). *Multidimensional similarity structure analysis*. New York: Springer.
- BORG, I. & STAUFENBIEL, T. (1992). The performance of snow flakes, suns, and factorial suns in the graphical representation of multivariate data. *Multivariate Behavioral Research*, 27, 43–55.
- BORG, I. & STAUFENBIEL, T. (1993). *Theorien und Methoden der Skalierung*. (2., rev. Aufl.) Bern: Huber.

- Chernoff, H. (1973). The use of faces to represent points in k -dimensional space graphically. *Journal of the American Statistical Association*, 68, 361–368.
- Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, 57, 145–158.
- Coombs, C. H. (1964). *A theory of data*. New York: Wiley.
- Coombs, C. H., Coombs, L. C. & Lingo, J. C. (1978). Stochastic cumulative scales. In S. Shye (Ed.), *Theory construction and data analysis in the behavioral sciences* (pp. 280–298). San Francisco: Jossey-Bass.
- DeVellis, R. F. (1991). *Scale development*. Newbury Park: Sage.
- Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.
- Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9, 139–150.
- Haubensak, G. (1985). *Absolutes und vergleichendes Urteil: Eine Einführung in die Theorie psychischer Bezugssysteme*. Berlin: Springer.
- Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, 115–129.
- Kruskal, J. B. & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills: Sage.
- Lingo, J. C. (1973). *The Guttman-Lingo nonmetric program series*. Ann Arbor: Mathesis Press.
- Lodge, M. (1981). *Magnitude scaling*. Newbury Park: Sage.
- Mausfeld, R. (1985). *Grundzüge der Fechner-Skalierung*. Frankfurt: Lang.
- McIver, J. P. & Carmines, E. G. (1981). *Unidimensional scaling*. Newbury Park: Sage.
- McReynolds, P. & Ludwig, K. (1987). On the history of rating scales. *Personality and Individual Differences*, 8, 281–283.
- Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.
- Norpoth, H. (1979). Dimensionen des Parteienkonflikts und Präferenzordnungen der deutschen Wählerschaft. *Zeitschrift für Sozialpsychologie*, 10, 350–362.
- Schönemann, P. H. (1994). The reasonable ineffectiveness of mathematics in the social sciences. In Borg, I. & Mohler, P. P. (Eds.) (1994). *Trends and perspectives in empirical social research* (pp. 149–160). New York: DeGruyter.
- Schönemann, P. H. & Borg, I. (1983). Grundlagen der metrischen mehrdimensionalen Skaliermethoden. In H. Feger & J. Bredenkamp (Hrsg.), *Messen und Testen* (= Enzyklopädie der Psychologie, Themenbereich B, Serie I, Band 3, S. 247–345). Göttingen: Hogrefe.
- Schuman, H. & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. New York: Academic Press.
- Spector, P. E. (1992). *Summated rating scale construction*. Newbury Park: Sage.
- Staufenbiel, T. (1993). *Materialband zu 'Theorien und Methoden der Skalierung': Analysen und Graphiken mit SYSTAT/SYGRAPH*. Mannheim: ZUMA.
- Stevens, S. S. (1975). *Psychophysics*. New York: Wiley.
- Takane, Y., Young, F. W. & DeLeeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least-squares method with optimal scaling features. *Psychometrika*, 42, 7–67.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 78, 392–402.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- Velleman, P. F. & Wilkinson, L. (1993). Nominal, ordinal, and ratio typologies are misleading. *The American Statistician*, 47, 65–72. (Leicht überarbeitet auch in I. Borg & P. P. Mohler (Eds.) (1994), *Trends and perspectives in empirical social research* (pp. 161–177). New York: De Gruyter).