Zur Begründbarkeit von Metriken in der empirischen Sozialforschung

Götz Rohwer

[Vortragsmanuskript]

Bochum, November 1997

Einleitung

Statistische Methoden werden häufig mit dem Attribut „quantitativ" versehen und von „qualitativen Methoden" der Sozialforschung unterschieden. Es ist jedoch oft nicht ganz klar, was mit dieser Unterscheidung — und insbesondere mit dem Wort „quantitativ" — gemeint ist und in welcher Weise es überhaupt zutreffend ist, statistische Methoden als „quantitativ" zu charakterisieren.

In der Umgangssprache gibt es einen direkten Zusammenhang zwischen den Worten „quantitativ" und „messen"; diejenigen Eigenschaften sind quantitativ, bei denen sinnvoll eine Größe gemessen werden kann. Prototypische Beispiele sind etwa Längen, Gewichte und Zeitdauern. Wir haben die Vorstellung, daß es Dinge gibt, die durch unterschiedliche Arten von Eigenschaften charakterisiert werden können. Einige dieser Eigenschaften lassen nur qualitative Unterscheidungen zu, z.B. Farbunterschiede; einige andere Eigenschaften sind quantifizierbar, z.B. Länge und Gewicht. Orientiert man sich an dieser durch die Umgangssprache nahegelegten Unterscheidung, hat es die empirische Sozialforschung hauptsächlich mit qualitativen Sachverhalten zu tun, nämlich mit qualitativ unterschiedlichen Merkmalen, Zuständen, Fähigkeiten und Verhaltensweisen sozialer Akteure.

Demgegenüber hat sich in der Methodenliteratur ein sehr unspezifischer Gebrauch des Wortes „messen" eingebürgert. Die meisten Methodenlehrbücher orientieren sich an der Idee einer regelgeleiteten Repräsentation von Objekten durch Zahlen. Dementsprechend heißt es (ich zitiere aus einem neueren Lehrbuch): „Unter 'Messen' wird allgemein die Zuordnung von Zahlen ('Meßwerten') zu Objekten gemäß festgelegten Regeln verstanden."[1] Definitionen dieser Art – in Verbindung mit der durch die Umgangssprache nahegelegten Assoziation zwischen „messen" und „quantifizieren" – erzeugen dann die Vorstellung, daß das wesentliche Merkmal eines quantifizierenden Umgangs mit der sozialen Realität darin liege, sie durch Zahlen zu repräsentieren. Und da statistische Methoden Zahlen verwenden, erscheinen sie dementsprechend als quantitative Methoden.

Man kann sich jedoch leicht klarmachen, daß dieser Gedankengang nur das Resultat eines unglücklichen Sprachgebrauchs ist und zur Aufklärung der Frage, in welcher Weise statistische Methoden eine Quantifizierung voraussetzen oder anstreben, nichts beiträgt. Denn daß Zahlen

[1] R. Schnell, P. B. Hill, E. Esser, Methoden der empirischen Sozialforschung (3. Aufl.). München 1992, S. 143.

zur Repräsentation von Objekten und ihrer Merkmale verwendet werden, hat mit Quantifizierung zunächst noch gar nichts zu tun. Als Minimalbedingung sollte gefordert werden, daß man davon sprechen kann, daß Unterschiede zwischen Objekten mehr oder weniger stark ausgeprägt sein können. Diese Minimalbedingung wird durch den Begriff der Metrik zu fassen versucht. Seine genaue (mathematische) Definition ist für die folgenden Überlegungen nicht erforderlich, es genügt eine kurze Erläuterung. Voraussetzung ist eine gewisse Menge von Objekten. Eine Metrik ist dann definiert als eine Funktion, die jeweils zwei dieser Objekte eine Zahl zuordnet, die sich als ein Indikator für die Art und das Ausmaß ihres Unterschieds interpretieren läßt. In diesem Sinne ermöglicht eine Metrik einen quantitativen Vergleich von Objekten.

Meine Frage ist nun, ob und ggf. wo quantifizierende Metriken in der empirischen Sozialforschung – und insbesondere bei der Anwendung statistischer Methoden – überhaupt benötigt werden. Ich werde in zwei Schritten vorgehen. In einem ersten Schritt möchte ich zeigen, daß statistische Methoden zwar eine quantitative Beschreibung von Gesamtheiten von Objekten anstreben – und infolgedessen Metrikkonstruktionen für Vergleiche von Gesamtheiten eine zentrale Rolle spielen –, daß dafür jedoch keineswegs auch eine Metrik für den Vergleich der individuellen Objekte selbst vorausgesetzt zu werden braucht. — In einem zweiten Schritt möchte ich dann etwas genauer auf einige Bestrebungen in der empirischen Sozialforschung eingehen, Metriken nicht nur für Verteilungsvergleiche, sondern auch für einen quantitativen Vergleich sozialer Untersuchungsobjekte zu konstruieren. Ich möchte zeigen, daß die Konstruktion *solcher* Metriken tatsächlich problematisch ist und daß im übrigen auch ziemlich unklar ist, wie solche Metriken für informative Beschreibungen gesellschaftlicher Verhältnisse genutzt werden könnten.

1 Metriken für Verteilungsvergleiche

Ich beginne mit meiner ersten Überlegung, die sich auf statistische Methoden bezieht. Die Gesamtheit dieser Methoden ist zwar äußerst breit und vielfältig; es gibt jedoch stets einen einfachen Ausgangspunkt in Gestalt einer real oder virtuell existierenden Menge individueller Objekte, die sich durch gewisse Eigenschaften charakterisieren lassen. Dann erfolgt allerdings sogleich ein Perspektivenwechsel: nicht mehr die individuellen Objekte, sondern ihre Gesamtheit bildet den Erkenntnisgegenstand. Dies kommt am deutlichsten in dem Grundbegriff der Statistik,

dem Begriff der statistischen Verteilung, zum Ausdruck. Die Statistik interessiert sich für Merkmalsverteilungen in Gesamtheiten, nicht für deren individuelle Mitglieder. Zum Beispiel für die Einkommensverteilung bei Haushalten, nicht jedoch für die individuellen Haushalte; oder für die Verteilung der Noten bei Abschlußprüfungen, nicht jedoch für die individuellen Prüfungsergebnisse. Um das zu betonen, ist vielleicht der Hinweis nützlich, daß unsere Sprache zweideutig ist, wenn im Plural über Individuen gesprochen wird. Mit einem Satz über eine Gesamtheit von Individuen kann eine Aussage über jedes ihrer individuellen Mitglieder gemeint sein; zum Beispiel: alle Menschen sind sterblich. Andererseits kann jedoch auch eine Aussage über die Gesamtheit *als Gesamtheit* gemeint sein; zum Beispiel: die Einkommensungleichheit hat in den vergangenen Jahren in der Bundesrepublik zugenommen. Statistische Aussagen sind immer Aussagen vom zweiten Typ: nicht Aussagen über individuelle Objekte, sondern Aussagen über Gesamtheiten von Objekten.

Nun, was wird für solche Aussagen vorausgesetzt? Nur daß es eine Menge von Objekten gibt, die sich durch gewisse Eigenschaften charakterisieren lassen. Dabei kann es sich gleichermaßen um qualitative oder quantitative Eigenschaften handeln. Der statistische Verteilungsbegriff ist tatsächlich neutral gegenüber dieser Unterscheidung. Es ist insbesondere nicht erforderlich, die individuellen Objekte irgendwie anders zu charakterisieren als wir es mit unseren üblichen sprachlichen Mitteln immer schon tun. Ein quantifizierender Umgang mit der sozialen Realität wird also insoweit nicht vorausgesetzt.

Nun ist allerdings richtig, daß sich die Statistik nicht damit begnügt, Merkmalsverteilungen zu ermitteln; sie bilden tatsächlich nur ihren Ausgangspunkt. Die meisten Fragestellungen der Statistik richten sich darauf, zwei oder mehr Verteilungen zu vergleichen und Einsichten in ihre Zusammenhänge zu gewinnen. Und hierfür spielen dann Metriken eine wichtige Rolle. — Bevor ich darauf etwas näher eingehe, möchte ich jedoch betonen, daß selbst an dieser Stelle (also für einen Vergleich von Merkmalsverteilungen) Metriken nicht unbedingt erforderlich sind. Metriken dienen dem Zweck, quantitative Vergleiche vornehmen zu können. Viele sinnvolle Vergleiche können jedoch vorgenommen werden, ohne explizit zu quantifizieren, indem nach Ordnungsrelationen gesucht wird. Zum Beispiel könnten wir uns dafür interessieren, ob die Einkommensungleichheit in den vergangenen Jahren tatsächlich zugenommen hat. Diese Frage setzt offensichtlich nicht voraus, daß wir die Unterschiede zwischen Einkommensverteilungen quantifizieren können, sondern nur einen ordinalen Begriff von Einkommensungleichheit; wir müssen in der

Lage sein, von zwei Einkommensverteilungen zu sagen (sofern sie nicht gleich sind), daß die eine oder die andere eine größere Einkommensungleichheit aufweist. Und es könnte natürlich auch sein, daß wir nur in der Lage sind, eine partielle Ordnung festzustellen.

Es ist jedoch richtig, daß die meisten Standardverfahren der Statistik einen quantifizierenden Vergleich von Verteilungen anstreben und demzufolge eine Metrik für Verteilungsvergleiche voraussetzen. Wichtig ist nun, daß solche Metriken nicht empirisch gewonnen werden können. Die empirisch zugängliche Verfassung der Realität liefert keinerlei Hinweise darauf, wie Gesamtheiten von individuellen Objekten quantitativ verglichen werden sollten. Metriken für Verteilungsvergleiche müssen vielmehr erfunden und konstruiert werden. Tatsächlich verdankt sich die Vielfalt der unterschiedlichen statistischen Methoden in erster Linie der Tatsache, daß im Prinzip beliebig viele Metriken für Verteilungsvergleiche erfunden werden können. Ich will hier zur Illustration nur kurz auf zwei unterschiedliche Ansätze hinweisen.

a) Den ersten Ansatz lernt jeder bereits in der elementaren deskriptiven Statistik kennen, dort wo es darum geht, Verteilungen durch Mittelwerte, Quantile, Streuungsmaße usw. zu charakterisieren. Jede dieser Kennzahlen kann offenbar (und wird tatsächlich auch) dazu verwendet, Verteilungen zu vergleichen. Dabei sollte man sich jedoch klarmachen, daß man dann stets eine bestimmte Metrik für den Verteilungsvergleich verwendet und daß sich demzufolge die Frage stellt, warum man die eine und nicht stattdessen eine andere Metrik verwenden sollte. Das ist natürlich eine Zweckmäßigkeitsfrage, die nicht mit statistischen Argumenten entschieden werden kann. Sie muß vom Anwender statistischer Methoden entschieden werden; aber es ist meistens ziemlich unklar, wie an dieser Stelle vernünftige, begründbare Entscheidungen getroffen werden können.

b) Vielleicht kann aber die statistische Methodenlehre dennoch eine gewisse Hilfestellung anbieten. Die Problematik bei der Verwendung der üblichen Maßzahlen für Verteilungsvergleiche besteht ja (zumindest teilweise) darin, daß diese Maßzahlen gar nicht im Hinblick auf die Zielsetzung geschaffen worden sind, Metriken für Verteilungsvergleiche zu begründen. Es liegt deshalb nahe, sich auch explizit der Frage zu stellen, wie solche Metriken konstruiert werden können. Einen überlegenswerten Zugang zu dieser Frage liefert das Konzept der Substitutionsmetriken, das ich kurz erläutern möchte. Wie ich zeigen möchte, läßt sich diese Klasse von Metriken auf einfache Weise als eine Verallgemeinerung einer sehr elementaren Idee des Messens, nämlich des Zählens, begründen.

Jeder versteht, was es heißt, eine Menge von Dingen zu zählen, um schließlich sagen zu können: diese Menge besteht aus soundsoviel Dingen (Einheiten, Elementen). Dadurch wird auf elementare Weise ein Größenbegriff begründet; genauer gesagt: eine Familie von Größenbegriffen, denn je nachdem, welche Art von Einheiten gezählt wird, entsteht ein spezifischer Größenbegriff; z.B. „Anzahl Beschäftigte“ oder „Anzahl Familienmitglieder“. Ich werde sie im folgenden *Zählgrößen* nennen.

Folgt man der traditionellen Meßtheorie, wie sie insbesondere von Hermann von Helmholtz formuliert worden ist, lassen sich alle elementaren Größenbegriffe auf Zählgrößen zurückführen; z.B. Längen, Gewichte und Zeitdauern. Der Unterschied besteht nur darin, daß nicht die Anzahl von Dingen in einer Menge von Dingen gezählt wird, sondern geeignet definierte Einheiten. Vorausgesetzt wird, daß sich Einheiten zu Maßstäben zusammensetzen lassen und daß dann diese Maßstäbe mit den zu messenden Objekten verglichen werden können. Zum Messen von Zeitdauern bedient man sich zum Beispiel Uhren, also Apparate, die regelmäßige Folgen von Zeiteinheiten erzeugen können. Um dann die zeitliche Dauer eines Vorgangs zu messen, muß man nur zählen, wieviel Zeiteinheiten während des Vorgangs verstreichen.

Nun kann jedoch die elementare Idee des Zählens noch auf andere Weise zur Begründung von Metriken verwendet werden. Gegeben sei eine Menge von Objekten: A, B, C usw. Wir möchten jeweils zwei dieser Objekte miteinander vergleichen, zum Beispiel A und B. Wir können dann fragen: was müssen wir tun, um A in B zu verwandeln? Die Idee ist: je einfacher es ist, A in B zu verwandeln, desto kleiner ist der Unterschied; und umgekehrt: je schwieriger es ist, A in B zu verwandeln, desto größer ist der Unterschied. Um diesen Gedanken zu präzisieren, ist es natürlich erforderlich, eine genaue Definition der Operationen anzugeben, durch die Objekte in andere Objekte überführt werden können. Um ohne mathematische Notationen auszukommen, will ich mich jedoch darauf beschränken, die grundsätzliche Idee anhand der bereits erwähnten Aufgabe zu erläutern, zwei Einkommensverteilungen zu vergleichen.

Die heuristisch leitende Frage ist in diesem Fall: Was müßten wir tun, um eine Einkommensverteilung in eine andere Einkommensverteilung zu transformieren? Die Antwort liegt auf der Hand. Sind zwei Einkommensverteilungen, A und B, gegeben, müßten wir einem Teil der Einkommensbezieher in der Verteilung A andere Einkommen geben, bis schließlich ihre Einkommensverteilung mit B identisch wird.

In diesem Beispiel kann man offenbar sehr leicht eine Elementaroperation angeben, nämlich einen Einkommensbezieher aus einer in eine andere Einkommensklasse zu bringen. Man kann dann zählen, wie viele solcher Elementaroperationen erforderlich sind, um eine Einkommensverteilung A in eine Einkommensverteilung B umzuwandeln. Da das im allgemeinen auf viele unterschiedliche Weisen geschehen kann, erscheint es zweckmäßig, nach der geringsten Anzahl von Elementaroperationen zu suchen, die für den Transformationsprozeß erforderlich sind. Diese minimale Anzahl von Elementaroperationen kann dann als Größe des Unterschieds zwischen den Einkommensverteilungen interpretiert werden. Die Interpretation entspricht dann unserem heuristischen Leitfaden: Sind nur wenige Operationen erforderlich, sind sich die beiden Verteilungen sehr ähnlich; und je mehr Einkommensbeziehern ein neues Einkommen zugewiesen werden muß, desto größer ist der Unterschied der beiden Verteilungen.

Dies ist ein einfaches Beispiel für eine Substitutionsmetrik; genauer gesagt: für eine einfache (ungewichtete) Substitutionsmetrik. Denn wir haben einfach nur die Anzahl der erforderlichen Elementaroperationen gezählt. Das Beispiel zeigt aber bereits, wie eine noch etwas flexiblere Klasse von Metriken definiert werden kann. Denn eine Elementaroperation besteht hier darin, einen Einkommensbezieher aus einer in eine andere Einkommensklasse zu befördern. Also liegt es nahe, den Aufwand (das Gewicht) einer Elementaroperation davon abhängig zu machen, wie groß der Unterschied zwischen den beiden Einkommensklassen ist. Dies wäre dann ein Beispiel für eine gewichtete Substitutionsmetrik.

Dieses Beispiel läßt sich nun sehr weitgehend verallgemeinern. Es ist klar, daß sich die eben beschriebene Methode nicht nur zum Vergleich von Einkommensverteilungen eignet, sondern daß mit ihr für beliebige Arten von Verteilungen Metriken gebildet werden können. Es handelt sich insoweit um ein allgemeines Konstruktionsprinzip für Metriken. Ihr Vorteil liegt darin, daß in diesem Fall eine Metrik nicht als Beiprodukt ganz anderer Überlegungen entsteht, sondern daß man sich bei ihrer Konstruktion explizit der Frage stellen muß, wie zwei Verteilungen sinnvoll verglichen werden können. Natürlich liefert die Idee der Substitutionsmetriken keine eindeutige Antwort, sondern nur einen formalen Rahmen, um über Metrikkonstruktionen nachdenken zu können. Aber mehr kann von der statistischen Methodenlehre gar nicht erwartet werden. Die Entscheidung für eine bestimmte Metrik bleibt immer mehr oder weniger willkürlich.

Dann stellt sich die Frage, wie mit der verbleibenden Willkür umgegangen werden sollte. Sie einfach zu ignorieren, ist sicherlich nicht befriedigend; und ebenso unbefriedigend ist es, blindlings einer der Konventionen zu folgen, die sich in der empirischen Sozialforschung zur scheinbaren Vermeidung von Willkür eingebürgert haben. Eine perfekte Lösung für dieses Problem gibt es vermutlich nicht. Immerhin gibt es einen Gedankengang, mit dem (wie ich glaube) etwas angemessener auf das Willkürproblem reagiert werden kann. Das Problem besteht ja darin, daß unterschiedliche Metriken zu unterschiedlichen substantiellen Schlußfolgerungen führen können, etwa bei der Frage, ob die Einkommensungleichheit tatsächlich in den vergangenen Jahren zugenommen hat oder nicht. Anstatt Schlußfolgerungen nur auf eine mehr oder weniger willkürlich ausgewählte Metrik zu gründen, erscheint es sinnvoll, die Abhängigkeit der substantiellen Forschungsergebnisse von der jeweils vorausgesetzten Metrik explizit zu thematisieren; und insbesondere der Frage nachzugehen, wie sich die Schlußfolgerungen bei unterschiedlichen Metriken verändern würden. Dann wird es vielleicht möglich sein, zumindest in einigen Fällen Ergebnisse zu finden, die gegenüber der Wahl unterschiedlicher Metriken vergleichsweise robust sind.

2 Metriken für Individualvergleiche

Soviel zur ersten Hälfte meiner Überlegungen. Es sollte deutlich geworden sein, daß statistische Methoden im Hinblick auf die Quantifizierung individueller sozialer Objekte neutral sind. Die individuellen Objekte und ihre Eigenschaften werden als gegeben vorausgesetzt. Die typischen Fragestellungen richten sich darauf, wie Unterschiede und Beziehungen zwischen Gesamtheiten solcher Objekte, die durch Merkmalsverteilungen charakterisierbar sind, beschrieben werden können. Für *diesen* Zweck werden dann ggf. Metriken konstruiert.

Nun gibt es allerdings in der empirischen Sozialforschung auch Bestrebungen, Metriken auch unmittelbar auf der Ebene der individuellen Untersuchungsobjekte einzuführen, um deren quantitativen Vergleich zu ermöglichen. Soweit solche Metriken auf der Grundlage bereits verfügbarer quantitativer Merkmale gebildet werden, ist das meistens unproblematisch. Zum Beispiel können Haushalte anhand der Höhe ihres verfügbaren Einkommens unterschieden und verglichen werden. Im allgemeinen verlangt jedoch die Zielsetzung, individuelle Objekte zu vergleichen, daß dafür mehr als nur eines ihrer Merkmale berücksichtigt wird. Spätestens dann ist jedoch nicht mehr offensichtlich, wie eine Metrik konstruiert werden kann, die einen sinnvollen Vergleich erlaubt.

Ich möchte betonen, daß sich diese Zielsetzung – individuelle Objekte zu vergleichen – auf signifikante Weise von den gewöhnlichen Fragestellungen der Statistik unterscheidet. Statistische Fragestellungen, die an der Idee von Merkmalsverteilungen ansetzen, abstrahieren vollständig von den individuellen Objekten und setzen insbesondere keine Metrik für individuelle Objektvergleiche voraus. Die Frage, wie Unterschiede zwischen individuellen Objekten erfaßt und ggf. quantifiziert werden können, ist nicht nur für die Statistik ganz irrelevant; es ist auch schwer, ein Erkenntnisinteresse zu finden, daß diese Frage für die empirische Sozialforschung relevant machen könnte. Die Frage, wie Unterschiede zwischen Individuen erfaßt werden können, spielt zwar *innerhalb* der gesellschaftlichen Interaktionsprozesse eine wichtige Rolle — zum Beispiel bei der Einstellung von Arbeitskräften oder beim Verkauf von Vorschlägen zur Partnerwahl. Und daraus kann die empirische Sozialforschung sicherlich interessante Forschungs*themen* gewinnen. Aber zu untersuchen, wie in der gesellschaftlichen Praxis Metriken konstruiert und verwendet werden, sollte nicht damit verwechselt werden, daß empirische Sozialforscher selbst neue Metriken zum Vergleich von Individuen erfinden. Die Unterscheidung ist m.E. wichtig, gerade auch in Anbetracht der Tatsache, daß sich die gesellschaftlichen Akteure die von Sozialwissenschaftlern und Psychologen erfundenen Metriken zueigen machen können (eine Art von „Praxisbezug“, auf den die Sozialwissenschaften nicht immer stolz sein können).

Ich will jedoch an dieser Stelle auf die Frage nach Erkenntnisinteressen, die Metriken für den Vergleich von Individuen sinnvoll machen könnten, nicht näher eingehen; sondern mich stattdessen mit zwei Vorfragen beschäftigen. Zunächst mit der Frage, ob und ggf. wie begründbare Metriken für Individualvergleiche konstruiert werden können; und im Anschluß dann – allerdings nur sehr kurz – mit der Frage, wie auf der Grundlage von Metriken Beschreibungen von Gesamtheiten individueller Objekte gewonnen werden können.

Zunächst zur ersten Frage. Der Ausgangspunkt besteht in einer Menge von Individuen (individueller Objekte), die jeweils durch eine Reihe von Eigenschaften charakterisiert sind. Es soll eine Metrik konstruiert werden, die diese Individuen quantitativ vergleichbar macht. Dafür braucht man zunächst einen in seiner Bedeutung fixierbaren Gesichtspunkt, im Hinblick auf den die Objekte verglichen werden sollen. Meistens liefert der Kontext eine Vorgabe; zum Beispiel wenn man Haushalte im Hinblick auf ihr Einkommen oder Arbeitslose im Hinblick auf die Dauer ihrer Arbeitslosigkeit vergleichen möchte. Solche Beispiele sind einfach, weil man dann an bereits verfügbare quantitative Eigenschaften der Individuen anknüpfen kann. In den genannten Beispielen etwa an Einkommensgrößen und Zeitdauern. Metrikkonstruktionen sind jedoch bereits in diesen einfachen Fällen problematisch, und zwar aus zwei Gründen. Erstens können durch Rückgriff auf eine quantitative Eigenschaft beliebig viele korrespondierende Metriken konstruiert werden. Es ist zwar meistens üblich, die einfachen Differenzen zu verwenden; also zum Beispiel bei Einkommensvergleichen die Differenz der Einkommen. Aber in diesem Fall könnte man auch vorschlagen, daß unterschiedliche Einkommenspositionen proportional zum jeweils erzielten Einkommensniveau erfaßt werden sollten, was dann darauf hinauslaufen würde, die Differenzen zwischen logarithmierten Einkommen zu verwenden; und das wäre dann eine ganz andere Metrik für Einkommensvergleiche. — Zweitens ist die Konstruktion von Metriken nur dann vergleichsweise einfach, wenn *nur eine* quantitative Eigenschaft in Betracht gezogen wird. In vielen Fällen ist das jedoch unzureichend. Als Beispiel kann wieder die Aufgabe dienen, Haushalte im Hinblick auf ihre Einkommenspositionen zu vergleichen. Es erscheint plausibel, daß bei einem solchen Vergleich nicht nur das Einkommen, sondern weitere Merkmale der Haushalte berücksichtigt werden sollten; insbesondere die Haushaltsgröße. Dann wird jedoch zunächst ganz unklar, wie ein sinnvoller Vergleich vorgenommen werden sollte. Man könnte zum Beispiel das Pro-Kopf-Einkommen verwenden. Aber das wäre nur eine von vielen verschiedenen Möglichkeiten, gegen die außerdem eingewendet werden kann, daß sich mit zunehmender Haushaltsgröße gewisse Einsparungsmöglichkeiten ergeben. Es ist also zumindest nicht offensichtlich, wie sog. Äquivalenzeinkommen zum Vergleich von Haushalten definiert werden sollten; und über diese Frage gibt es tatsächlich eine ausgedehnte und kontroverse Diskussion (übrigens auch in der gesellschaftlichen Praxis selbst, nämlich bei der Festlegung von Sozialhilfesätzen). Ich will hier jedoch nicht auf Einzelheiten eingehen, sondern mit diesem Beispiel nur verdeutlichen, daß Metriken für Individualvergleiche nicht schon durch die Auswahl der zu berücksichtigenden Eigenschaften (in diesem Beispiel: Haushaltseinkommen und Haushaltsgröße) gegeben sind, sondern daß eine explizite (und meistens kontrovers diskutierbare) Konstruktion erforderlich ist.

Es ist vielleicht hilfreich, an dieser Stelle zwei grundsätzlich unterschiedliche Ansätze zur Konstruktion von Metriken für Individualvergleiche zu unterscheiden.

a) Der erste Ansatz wird durch das eben genannte Beispiel – Äquivalenzeinkommen für Haushaltsvergleiche – illustriert. Die Metrikkonstruktion erfolgt hier in zwei Schritten. In einem ersten Schritt wird *eine* *neue quantitative Eigenschaft* konstruiert, die zur Charakterisierung der zu vergleichenden Objekte dienen soll (in unserem Beispiel sind es die Äquivalenzeinkommen). Dann wird in einem zweiten Schritt ausgehend von diesen neuen Eigenschaften eine Metrik gebildet. Charakteristisch für diesen Ansatz ist also ein zweistufiges Vorgehen, wobei die erste Stufe darin besteht, daß eine vorab noch nicht verfügbare Skala für die quantifizierende Charakterisierung der individuellen Objekte gebildet wird. Man kann insofern von Metrikkonstruktionen sprechen, die auf einer vorgängigen Skalierung der Objekte beruhen.

Dieser Ansatz ist tatsächlich nicht nur in der empirischen Sozialforschung ziemlich verbreitet. Ein einfaches Beispiel aus der gesellschaftlichen Praxis ist etwa die Bildung von Gesamtnoten zum Vergleich von Ergebnisse aus schriftlichen und mündlichen Prüfungen. Es ist üblich, Gesamtnoten als einen irgendwie gewichteten Durchschnitt aus den Einzelnoten zu berechnen und dann die Gesamtnoten zu vergleichen. Das ist natürlich ziemlich willkürlich, hat aber immerhin den Vorteil, leicht durchschaubar zu sein. Bei den in der empirischen Sozialforschung verwendeten Skalierungsverfahren ist das jedoch oft nicht der Fall; sondern es werden quantifizierbare Eigenschaften von Individuen konstruiert (wie z.B. quantitative Indikatoren für Umweltbewußtsein oder Intelligenzquotienten), deren Zusammenhang mit beobachtbaren Merkmalen und Fähigkeiten der Individuen bestenfalls partiell durchschaubar ist. Der Aufgabe, Metriken begründbar zu machen, wird dadurch sicherlich nicht gedient.

Viele der in der empirischen Sozialforschung üblichen Skalierungsmethoden haben indessen noch einen weiteren problematischen Aspekt. Akzeptieren wir vorläufig (mit allen Vorbehalten) die Absicht, neue quantifizierbare Eigenschaften begründen zu wollen. Es erscheint evident, daß für solche Eigenschaftskonstruktionen nur auf bereits bekannte, den Objekten sinnvoll zurechenbare Eigenschaften zurückgegriffen werden sollte. Zahlreiche Skalierungsverfahren widersprechen jedoch dieser Anforderung, indem sie vermeintlich individuell zurechenbare Eigenschaften aus Merkmalen ableiten, die zunächst nur für eine Gesamtheit von Objekten definiert sind, d.h. aus Verteilungseigenschaften. Um nicht in eine technische Diskussion einzelner Verfahren abzugleiten, will ich das Problem nur anhand eines Beispiels verdeutlichen. Dazu beziehe ich mich wieder auf die Frage, wie Äquivalenzeinkommen für Haushaltsvergleiche definiert werden können. Einige Autoren haben versucht, eine Definition auf empirischem Wege zu finden, indem zunächst ein nichtlineares Regressionsmodell für den Zusammenhang zwischen Haushaltsgröße und Einkommen geschätzt wird und dann die für jede Haushaltsgröße geschätzten durchschnittlichen Einkommen als eine Skala für Äquivalenzeinkommen verwendet werden; jeder individuelle Haushalt kann dann durch seine Abweichung von diesen „mittleren Äquivalenzeinkommen“ charakterisiert werden.[2] Aber was dabei herauskommt, kann nicht mehr sinnvoll als eine den individuellen Haushalten zurechenbare Eigenschaft interpretiert werden. Haushalte werden dann vielmehr durch ihre relative Position in einer Gesamtheit von Haushalten charakterisiert; und die in der Skalierungsliteratur übliche Rhetorik, Eigenschaften individueller Objekte meßbar machen zu können, wird hinfällig. — Vor allem ist aber der Anspruch problematisch, auf diese Weise eine empirische Begründung für Metriken finden zu können. In diesem Beispiel ist das offensichtlich, weil das Verfahren eigentlich nur darin besteht, eine Definition von Äquivalenzeinkommen zu finden, die die Ungleichheit der resultierenden Einkommensverteilung minimiert. Der Anspruch, quantifizierbare Eigenschaften – oder allgemeiner: Metriken – *empirisch* begründen zu können, ist jedoch (wie ich bereits im ersten Teil meiner Überlegungen angedeutet habe) grundsätzlich irreführend. Metriken sollen dazu dienen, Objekte quantitativ vergleichbar zu machen. Die Frage ist, warum und ggf. wie das getan werden *sollte*. Das ist letztlich eine normative Frage, die durch empirische Argumente nicht entschieden werden kann. Die von manchen Wissenschaftstheoretikern immer noch verbreitete Meinung, daß solche Fragen nur willkürlich entschieden werden könnten, führt dann allerdings in eine Sackgasse.

b) Eine Alternative zu Metrikkonstruktionen, die auf vorgängigen Skalierungsverfahren beruhen, besteht darin, sich der Aufgabe, individuelle Objekte zu vergleichen, unmittelbar zu stellen. Kontrastierend kann man von einstufigen Verfahren sprechen, mit deren Hilfe für jeweils zwei Objekte eine Aussage über die Größe ihres Unterschieds gemacht werden kann. Wiederum bedarf es natürlich einer Leitidee, inbezug auf die der Vergleich vorgenommen werden soll.

Bei diesen Verfahren kann man subjektive und objektive Varianten unterscheiden. Als „subjektiv“ bezeichne ich ein Verfahren, wenn es seine Daten aus Meinungsäußerungen gewinnt. Im Kern bestehen diese subjektiven Verfahren darin, eine Reihe von Personen danach zu befragen, wie sie die Unterschiede zwischen Objekten wahrnehmen bzw. einschätzen,

[2] Vgl. B. Buhmann, L. Rainwater, G. Schmaus, T. M. Smeeding, Equivalence Scales, Well-Being, Inequality, and Poverty: Sensitivity Estimates Across Ten Countries Using the Luxembourg Income Study (LIS) Database (Review of Income and Wealth 34, 1988, 115 – 142).

und aus den Befragungsergebnissen dann eine Metrik zu konstruieren. Solche Verfahren sind hauptsächlich von Psychologen (oder besser gesagt: Psychometrikern) entwickelt worden, sie werden jedoch gelegentlich auch in der empirischen Sozialforschung verwendet. Exemplarisch kann auf die Versuche verwiesen werden, eine Metrik für berufliche Prestigeunterschiede aus der subjektiven Einschätzung solcher Unterschiede durch einige Befragungspersonen zu gewinnen. Es erscheint mir jedoch offensichtlich, daß dieser Ansatz bestenfalls Informationen über die befragten Subjekte, nicht jedoch über die vorgeblich zu untersuchenden Sachverhalte (in diesem Fall Berufsgliederungen) erbringen kann.[3] Und ich werde deshalb auf diese subjektiven Metrikkonstruktionen nicht näher eingehen.

Man kann sich nun überlegen, ob es auch objektive Verfahren zur Metrikkonstruktion gibt, die gleichwohl ohne eine vorgängige Skalierung auskommen können. Dabei ist mit dem Wort „objektiv" hier gemeint, daß die Metrik durch ein Verfahren begründet werden kann, dessen Ergebnisse ohne Rekurs auf subjektive Meinungsäußerungen begründet werden können. — Einen interessanten Zugang zu dieser Frage liefert wiederum die Idee der Substitutionsmetriken. Als Beispiel beziehe ich mich auf Lebensverläufe. Die Frage, wie Lebensverläufe verglichen werden können, spielt in der neueren Lebensverlaufsforschung eine wichtige Rolle. Tatsächlich gibt es noch keine befriedigende Antwort; und das ist auch kaum überraschend, da Lebensverläufe hochkomplexe Gebilde sind. Dies ist nicht nur dann der Fall, wenn wir mit umgangssprachlichen Formulierungen an Lebensverläufe denken, sondern auch noch, wenn wir uns die schematischen Vereinfachungen der statistischen Lebensverlaufsforschung zueigen machen.

Lassen Sie mich kurz sagen, wie dieses schematische Lebensverlaufskonzept beschaffen ist (das übrigens wegen seiner Allgemeinheit die Entwicklung beliebiger Objekte erfassen kann). Ausgangspunkt ist die Fixierung einer Zeitachse und eines Zustandsraums. Ein Lebensverlauf entsteht dann daraus, daß der Zustandsraum auf der zugrundeliegenden Zeitachse auf bestimmte Weise durchwandert wird. Jeder individuelle Lebensverlauf kann dann durch eine Folge von Zuständen repräsentiert werden. Welche Aspekte realer Lebensverläufe durch dieses schematische Konzept erfaßt werden können, hängt natürlich von der Konzeption des Zustandsraums ab. In der statistischen Lebensverlaufsforschung werden

[3] Eine überzeugende Kritik an diesen Metrikkonstruktionen für berufliches Prestige findet sich in Ray Pawsons Aufsatz „Desperate Measures" (British Journal of Sociology 33, 1982, 35 – 63).

bisher meistens nur sehr einfache Zustandsräume betrachtet. Zum Beispiel beschäftigen sich viele Untersuchungen mit Erwerbsgeschichten, in denen oft nur vier verschiedene Zustände unterschieden werden: „Ausbildung und Fortbildung", „Erwerbstätigkeit", „Arbeitslosigkeit" und „aus dem Erwerbsleben ausgeschieden". Aber bereits bei einer so kleinen Anzahl von Zuständen gibt es eine äußerst große Anzahl unterschiedlicher Verlaufsmöglichkeiten. Nehmen wir dieses Beispiel mit vier Zuständen und eine Zeitachse, die in 200 Monate eingeteilt ist, also etwa 17 Jahre dauert. Dann gibt es rein rechnerisch bereits 1.6 Milliarden unterschiedliche Verlaufsmöglichkeiten. Es ist klar, daß die Idee statistischer Verteilungen, wenn man sie unmittelbar auf solche Verläufe anwendet, in eine Sackgasse führt. Hat man z.B. eine Stichprobe von 1000 oder auch 5000 Lebensverläufen (durchaus typische Größenordnungen in der gegenwärtigen Lebensverlaufsforschung), findet man regelmäßig, daß sich die Lebensverläufe aller oder fast aller Stichprobenmitglieder voneinander unterscheiden. Und zwar auch dann noch, wenn man nur sehr wenige Zustandsunterscheidungen vornimmt.

In der neueren Lebensverlaufsforschung ist deshalb die Vorstellung, mit statistischen Methoden auch ganze Lebensverläufe beschreiben zu können, in den Hintergrund getreten. Stattdessen konzentriert sich das Forschungsinteresse darauf, Zustandsänderungen zu erfassen und zu erklären. Typische Fragestellungen sind zum Beispiel: wovon hängt es ab, ob eine Person arbeitslos wird; oder: wovon hängt es ab, ob ein Arbeitsloser wieder eine neue Arbeit findet; usw. Aber so wichtig und interessant solche Fragestellungen auch sind, es ist klar, daß durch ihre Bearbeitung nur begrenzt Einsichten in ganze Lebensverläufe gewonnen werden können. Es besteht sogar eine gewisse Gefahr, daß durch die Konzentration auf die Erklärung einzelner Zustandsveränderungen die Lebensverläufe als Gesamtprozesse aus dem Blickfeld geraten.

Die Frage, wie auch Lebensverläufe als Gesamtprozesse einer statistischen Beschreibung zugänglich gemacht werden können, behält deshalb ihre Bedeutung. Um einen Zugang zu finden, ist dann vor allem die Frage wichtig, ob eine Metrik begründet werden kann, mit deren Hilfe sich Unterschiede zwischen Lebensverläufen sinnvoll erfassen lassen. Obwohl hier Skepsis sicherlich angebracht ist, liefert die Idee der Substitutionsmetriken immerhin einen überlegenswerten Vorschlag. Die heuristische Idee wird durch die Frage geliefert: Wie aufwendig ist es, einen Lebensverlauf in einen anderen Lebensverlauf zu transformieren? Wiederum ist es erforderlich, die Elementaroperationen zu fixieren, die für diesen Transformationsprozeß verwendet werden können. Da die schematischen

Lebensverläufe Folgen von Zuständen sind, liegt es nahe, als Elementaroperation die Substitution eines Zustands durch einen anderen Zustand während einer Zeiteinheit anzunehmen. Dann kann man einfach zählen, wieviele solcher Elementaroperationen mindestens erforderlich sind, um einen Lebensverlauf in einen anderen zu transformieren.

Man kann sich jedoch leicht klarmachen, daß diese simple Substitutionsmetrik für einen Vergleich von Lebensverläufen unzureichend ist. Denn läßt man als Elementaroperationen nur eine Substitution von Zuständen während vergleichbarer Zeiteinheiten zu, können zwei Lebensverläufe als sehr unterschiedlich erscheinen, obwohl sie sich nur minimal in den Aufenthaltsdauern in den verschiedenen Zuständen unterscheiden. In den vergangenen Jahren sind deshalb wesentlich elaboriertere Substitutionsmetriken für einen Vergleich von Zustandssequenzen entwickelt worden, die z.B. auch ein Einfügen und Löschen von Zuständen erlauben. Diese Entwicklung wird zur Zeit hauptsächlich von mathematischen Biologen vorangetrieben, die eine Metrik zum Vergleich von Genen benötigen, um deren Evolutionsmuster aufzuspüren. Für Anwendungen in der empirischen Sozialforschung hat sich besonders der amerikanische Soziologe Andrew Abbott eingesetzt. Die Verwendung solcher Metriken für einen Vergleich von Lebensverläufen steht allerdings noch ganz am Anfang.

3 Das Repräsentationsproblem

Es sollte (zumindest skizzenhaft) deutlich geworden sein, daß bereits die Konstruktion von Metriken für Individualvergleiche mit zahlreichen Schwierigkeiten verbunden ist. Zusätzliche Schwierigkeiten treten auf, wenn man danach fragt, wie solche Metriken in der empirischen Sozialforschung sinnvoll verwendet werden können.

Zunächst sollte noch einmal auf einen wichtigen Unterschied hingewiesen werden. Im ersten Teil meiner Überlegungen habe ich zu zeigen versucht, daß Metriken eine wichtige und zumindest potentiell sinnvolle Bedeutung im Rahmen der Aufgabe haben, Gesamtheiten von Objekten anhand von Merkmalsverteilungen zu vergleichen. In diesem Zusammenhang geht es tatsächlich darum, die individuell unterschiedlichen Verteilungen zu vergleichen und Zusammenhänge zwischen ihnen festzustellen. Anders verhält es sich jedoch bei der Konstruktion von Metriken für den Vergleich individueller sozialer Objekte. Denn die empirische Sozialforschung ist im allgemeinen gar nicht an den individuellen Objekten und ihren Unterschieden interessiert — und braucht insoweit auch gar keine Metriken für Individualvergleiche. Sie ist vielmehr daran interessiert, gesellschaftliche Verhältnisse, Gesamtheiten von Individuen, zu beschreiben (und natürlich darüberhinaus Modelle zu finden, anhand derer sich besser verstehen läßt, wie diese Verhältnisse funktionieren).

Damit entsteht folgende allgemeine Problemstellung. Gegeben ist eine Gesamtheit individueller Objekte und eine für sie definierte Metrik. Wie kann dann, ausgehend von dieser Metrik, eine sinnvolle Beschreibung der Gesamtheit gegeben werden?

Es gibt eine Reihe von Verfahren, die eine Antwort versprechen; zum Beispiel Verfahren der Clusteranalyse, der Korrespondenzanalyse und der multidimensionalen Skalierung. Sobald man sich jedoch genauer mit diesen Verfahren beschäftigt, wird ziemlich unklar, welche Art von Einsichten mit ihrer Hilfe gewonnen werden können. Eine zentrale Schwierigkeit besteht offenbar darin, daß die durch eine Metrik gegebene Information nicht unmittelbar an die Idee einer statistischen Merkmalsverteilung anschlußfähig ist. Das Verteilungskonzept setzt individuell zurechenbare Merkmale voraus; eine Metrik liefert jedoch zunächst etwas ganz anderes, nämlich einen paarweisen Vergleich von Objekten. — Man kann im wesentlichen zwei Umgangsformen mit diesem Problem unterscheiden.

a) Die eine Gruppe von Ansätzen strebt an, Metriken durch den Objekten individuell zurechenbare Eigenschaften zu repräsentieren. Man kann pauschal von „Skalierungsansätzen“ sprechen. Der Grundgedanke läßt sich bereits ohne Rückgriff auf spezielle Verfahren verdeutlichen. Ist nämlich eine Metrik für eine gewisse Menge von Objekten gegeben, kann man offenbar jedes individuelle Objekt auch durch seine Beziehungen zu allen übrigen Objekten charakterisieren. Auf diese Weise läßt sich eine Metrik für n Objekte durch $n-1$ Variablen repräsentieren. So betrachtet besteht das Ziel von Skalierungsverfahren eigentlich nur darin, mit weniger Variablen dieser Art auszukommen. Allgemein gesprochen geht es darum, eine gewisse, möglichst kleine Anzahl von Variablen zu finden, so daß eine ihrer korrespondierenden Metriken die vorgegebene Metrik möglichst gut repräsentiert. Am einfachsten, aber zugleich am gröbsten, wird dieser Gedankengang bei den clusteranalytischen Verfahren verfolgt. Sie liefern am Schluß nur eine einzige Variable, die für jedes individuelle Objekt angibt, zu welcher der durch das Verfahren gebildeten Objektgruppen es gehört. Andere Verfahren, wie z.B. die verschiedenen Varianten der multidimensionalen Skalierung, zielen darauf ab, nicht nur diskrete Indikatorvariablen, sondern metrische Variablen zu konstruieren, um die vorgegebene Metrik zu approximieren.

Diese Unterschiede in den Verfahren sind jedoch für den momentanen Gedankengang nicht wesentlich. Wesentlich ist vielmehr die ihnen gemeinsame Intention, aus Metriken individuell zurechenbare Eigenschaften zu gewinnen, denen eine inhaltliche Interpretation gegeben werden kann. Sie werden dann häufig als „latente Variablen" bezeichnet, also Variablen, deren Werte nicht beobachtet werden können, sondern durch Rechenverfahren ermittelt werden müssen. Hat man solche latenten Variablen konstruiert, können dann die üblichen, am Verteilungsbegriff ansetzenden statistischen Verfahren verwendet werden. Man kann dann nicht nur die Verteilung dieser latenten Variablen ermitteln, sondern sie schließlich auch als erklärende Variablen in statistischen Modellen verwenden. In einem neueren Lehrbuch heißt es dazu lapidar: „Wir sprechen [...] von einer *manifesten Variablen*, wenn eine Variable direkt beobachtbar ist (z.B. Anzahl gelöster Testaufgaben), und von einer *latenten Variablen*, wenn wir annehmen, daß sie einer manifesten Variablen als hypothetisches Konstrukt zugrunde liegt."[4]

Der Gedankengang ist ziemlich merkwürdig. Zunächst geht man von einer Metrik aus, also von einem rein relational definierten Sachverhalt. Dann werden daraus individuell zurechenbare Eigenschaften konstruiert und als latente Variablen interpretiert. Aber bereits dieser erste Konstruktionsschritt ist fragwürdig. Denn die Werte dieser latenten Variablen charakterisieren nicht Eigenschaften der individuellen Objekte, sondern Eigenschaften der durch die Metrik charakterisierten Gesamtheit von Objekten. Es ist reine Willkür, daraus individuell zurechenbare Eigenschaften zu bilden, und widerspricht im übrigen einem Grundsatz der neuzeitlichen Wissenschaft, nämlich dem Ockham'schen Leitsatz, daß man der Realität nicht ohne wirklich gute Gründe bloß ausgedachte Eigenschaften andichten sollte. — Noch merkwürdiger wird es, wenn man sich klarmacht, daß diese latenten Variablen ja nicht nur die vorgegebene Metrik repräsentieren sollen, sondern daß sie – wie das gerade angeführte Zitat angedeutet hat – schließlich dazu dienen sollen, beobachtbare Sachverhalte zu erklären. Ich will zwar nicht behaupten, daß bei dieser Vorgehensweise grundsätzlich nur Pseudo-Erklärungen entstehen können. Aber jeder Versuch, solche Erklärungen zu rechtfertigen, steht offenbar vor dem Problem, wie den latenten Variablen auch unabhängig von der Metrik, durch die sie konstruiert worden sind, eine reproduzierbare Bedeutung gegeben werden kann.

b) Obwohl die neuere Lehrbuchliteratur überwiegend diese Skalierungsansätze propagiert, ist die Praxis der empirischen Sozialforschung nicht so konsequent. Sie begnügt sich fast immer mit der Idee, daß es ausreichend sei, ausgehend von einer Metrik für Individualvergleiche die „Struktur" der Gesamtheit von Objekten zu charakterisieren. Werden zum Beispiel Clusteranalysen durchgeführt, läuft das Verfahren darauf hinaus, die schließlich erzeugten Objektgruppen „inhaltlich zu interpretieren", d.h. für die Gruppenzugehörigkeit von Objekten inhaltlich interpretierbare Eigenschaften zu erfinden. Ganz ähnlich verhält es sich bei den Anwendungen der Korrespondenzanalyse (zum Beispiel bei Bourdieu, der sie als Instrument zur Ermittlung seiner „feinen Unterschiede" einsetzt). Strenggenommen besteht der Unterschied zu den Skalierungsansätzen jedoch nur darin, daß die Konstruktion der latenten Variablen informell bleibt und daß man sich damit begnügt, ihnen eine gewisse Bedeutung zu geben.

[4] J. Bortz, N. Döring, Forschungsmethoden und Evaluation, Berlin 1995, S. 6.

Nimmt man demgegenüber die Formulierung ernst, daß die vorausgesetzte Metrik dazu dienen soll, Aussagen über die „Struktur" der Gesamtheit von Objekten zu gewinnen, bedarf es zunächst einer Klärung, was mit diesem Strukturbegriff gemeint sein soll. Die einzige aus meiner Sicht plausible Antwort liegt dann jedoch darin, daß die Struktur der Gesamtheit *durch ihre Metrik gegeben* ist. Die Formulierung, eine Struktur „in der Metrik" zu finden, wird dann allerdings unsinnig. Eine sinnvolle Aufgabe liegt dann (wie mir scheint) nur noch darin, zu versuchen, die gegebene Metrik durch eine andere, möglichst einfacher überschaubare Metrik zu repräsentieren. Und natürlich kann man sich schließlich der viel wichtigeren Frage zuwenden, wie denn die durch die Metrik gegebene Struktur der Objekte zustande gekommen ist und sich verändert. Das wäre dann die Antithese zum Skalierungsansatz: Metriken nicht als Pseudo-Meßinstrumente zu vergegenständlichen, sondern sie als zu erklärende soziale Sachverhalte wahrzunehmen.