Sonderdruck aus: Psychologische Rundschau, 57 (4), 213–223

Mehrebenenanalysen in der psychologischen Forschung

Vorteile und Möglichkeiten der Mehrebenenmodellierung mit Zufallskoeffizienten

John B. Nezlek, Michela Schröder-Abé und Astrid Schütz

Zusammenfassung. Daten mit Mehrebenenstruktur liegen vor, wenn für jede Versuchsperson eine Vielzahl von Messungen erhoben wird (z. B. in Tagebuchstudien) oder wenn Individuen in Gruppen analysiert werden. Derartige Daten können mit Hilfe von Mehrebenenanalysen ausgewertet werden. Der vorliegende Artikel erläutert das Prinzip der Mehrebenenmodellierung mit Zufallskoeffizienten, nennt Vorteile gegenüber herkömmlichen Analysestrategien und liefert Beispiele aus empirischen Studien. Außerdem wird diskutiert, welche Aspekte bei Mehrebenenanalysen zu beachten sind. Hierzu gehören die Zentrierung von Prädiktoren, die Fixierung von Koeffizienten und die Zahl der Analyseebenen.
Schlüsselwörter: Hierarchische Datenstrukturen, Messwiederholung, Mehrebenenanalysen, Mehrebenenmodellierung, Zufallskoeffizienten, Zufallseffekte, ereigniskontingente Daten, intervallkontingente Daten, Tagebuchstudien

Multilevel analyses in psychological research. Advantages and potential of multilevel random coefficient modeling

Abstract. Multilevel data structures are commonplace in psychological research, e. g., in diary studies with numerous observations per participant, or when studying several groups of individuals. These data can be analyzed using multilevel random coefficient modeling (MRCM). This article explains the logic of MRCM and the advantages of MRCM over more traditional techniques. In addition, some examples of MRCM analyses are described, and various factors that need to be considered when conducting MRCM analyses are discussed (e. g., centering of predictors, fixing coefficients, and number of levels of analysis).
Key words: nested data, repeated measures, multilevel analyses, random coefficients, random effects, event contingent data, interval contingent data

„Once you know that hierarchies exist, you see them everywhere." Diese Aussage von Kreft und de Leeuw (1998, S. 1) verweist auf die Bedeutung von Datenstrukturen mit mehreren Ebenen in der Psychologie und verwandten Disziplinen. Eine Mehrebenenstruktur liegt vor, wenn Daten einer Analyseebene hierarchisch in einer zweiten geschachtelt sind. Dies betrifft Daten mit Messwiederholungen, bei denen pro Person mehrere Messzeitpunkte vorliegen, sowie Daten von Individuen in Gruppen. In beiden Fällen sind die einzelnen Beobachtungen nicht unabhängig voneinander, was bei der Datenanalyse zu berücksichtigen ist. Andernfalls können Schätzungen von Effekten (Zusammenhängen) und Varianzen verfälscht werden sowie inkorrekte Signifikanzbefunde auftreten.

Im vorliegenden Beitrag werden verschiedene Methoden zur Analyse von Mehrebenendaten dargestellt und verglichen. Dabei wird argumentiert, dass die *Mehrebenenmodellierung mit Zufallskoeffizienten* (multilevel random coefficient modeling, MRCM) am besten für Mehrebenendaten geeignet ist. Diese Analysestrategie wird dann ausführlich an Hand mehrerer Beispiele erläutert. Ein wichtiges Ziel dieses Artikels ist es, eine Alternative zu herkömmlichen Analysestrategien aufzuzeigen und so die Formulierung von Hypothesen und Analyse von Daten aus einer neuen Perspektive zu ermöglichen.

Die Analyse von Individuen in Gruppen ist die wahrscheinlich häufigste Form von Mehrebenendesigns – sie war der Anlass für die Entwicklung der hier vorgestellten Techniken. Bei der Analyse muss beachtet werden, dass Individuen in Gruppen geschachtelt sind, also z. B. dass Schüler und Schülerinnen Mitglieder von Klassenverbänden sind. Aus diesem Grund muss die Leistung eines Schülers oder einer Schülerin als Funktion von Einflüssen auf individueller Ebene (individuelle Unterschiede zwischen Schülern, z. B. kognitive Fähigkeiten) und auf Gruppenebene (Unterschiede zwischen Schulklassen, z. B. Erfahrung der Lehrkraft) analysiert werden. Merkmale auf Klassenebene sind für alle Schüler in einer bestimmten Schulklasse gleich, können aber über Schulklassen hin-

Wesentliche Teile dieser Arbeit entstanden während einer Gastprofessur von John Nezlek an der TU Chemnitz. Wir danken der DFG für die Förderung dieser Zusammenarbeit (FO 327/2-2). Für hilfreiche Hinweise zu einer früheren Version des Artikels danken wir Bernd Marcus, Frank Papenmeier, Katrin Rentzsch, Almut Rudolph und Aline Vater.

DOI: 10.1026/0033-3042.57.4.213

weg variieren. Häufig ist auch von Interesse, ob die Beziehungen zwischen Variablen der individuellen Ebene auf Gruppenebene variieren, z. B. ob der Zusammenhang zwischen kognitiven Fähigkeiten und Leistungen in verschiedenen Schulklassen gleich groß ist. Falls dies nicht der Fall ist, interessiert, ob diese Variabilität durch Merkmale auf Gruppenebene (z. B. Klassengröße) erklärt werden kann. Da letztgenannte Fragen Beziehungen betreffen, werden in derartigen Analysen nicht Mittelwerte, sondern Zusammenhänge (auf Basis von Kovarianzen) als abhängige Variablen untersucht.

Marsh, Köller und Baumert (2001) untersuchten beispielsweise, wie sich die durchschnittlichen Mathematikleistungen in den jeweiligen Schulklassen auf das Fähigkeitsselbstkonzept von Schülern und Schülerinnen im Bereich Mathematik auswirken. Dabei wurde ein so genannter „Big-Fish-Little-Pond-Effekt" gefunden. Das heißt, eine Schülerin nimmt ihre Fähigkeiten positiver wahr, wenn sie sich in einer relativ schwachen Klasse befindet (großer Fisch im kleinen Teich); eine Schülerin mit den gleichen Fähigkeiten würde diese jedoch negativer einschätzen, wenn sie Mitglied einer relativ leistungsstarken Klasse ist (kleiner Fisch im großen Teich). Diese Untersuchung ist ein anschauliches Beispiel für die Notwendigkeit von Mehrebenenanalysen, da auf unterschiedlichen Analyseebenen unterschiedliche Zusammenhänge zwischen den Variablen bestehen: Auf Klassenebene zeigt sich ein negativer Zusammenhang zwischen Leistung und Fähigkeitsselbstkonzept, d. h. bei Konstanthaltung der individuellen Leistung findet man einen negativen Effekt der mittleren Klassenleistung auf das Fähigkeitsselbstkonzept. Auf Ebene der einzelnen Schüler korreliert Leistung hingegen positiv mit dem Fähigkeitsselbstkonzept (Schüler mit besseren Leistungen weisen ein höheres Fähigkeitsselbstkonzept auf).

Mehrebenenstrukturen entstehen auch bei der Erhebung von Mehrfachmessungen. In diesem Fall sind Beobachtungen in Personen geschachtelt. Beispiel hierfür sind die in den letzten 20 Jahren in der Persönlichkeits- und Sozialpsychologie zunehmend verbreiteten Studien, die auf intensiven Messwiederholungen basieren. Dabei führen die Teilnehmer und Teilnehmerinnen z. B. ein Tagebuch, in dem sie einmal oder mehrmals täglich Eintragungen vornehmen. Untersucht wurden beispielsweise soziale Ressourcen (Bachmann, 1998), der Zusammenhang von Arbeits- und Freizeitaktivitäten mit Wohlbefinden (Sonnentag, 2001) oder der Einfluss sozialer Interaktionen auf die Entwicklung der Persönlichkeit (Asendorpf & Wilpers, 1998). Auch in der Klinischen Psychologie wird dieses Forschungsdesign zunehmend in der Prozessdiagnostik eingesetzt oder um in Therapieevaluationsstudien den Behandlungsverlauf zu erfassen (vgl. Wilz & Brähler, 1997). Mehrebenenmodelle sind für derartige Veränderungsmessungen besonders gut geeignet (vgl. Eid, 2003). So kann der Behandlungsverlauf durch individuelle Merkmale der Patientinnen und Patienten vorhergesagt und dadurch die Therapie besser individuell abgestimmt werden (Lutz, Lowry, Kopta, Einstein & Howard, 2001).

Sind Beobachtungen innerhalb von Personen geschachtelt, so richten sich Analysen meist auf intraindividuelle Zusammenhänge und auf die Frage, wie diese als Funktionen von interindividuellen Unterschieden variieren. Beispielsweise beschreiben Personen in einer Studie täglich erlebte negative und positive Ereignisse sowie ihren Zustandsselbstwert. Eine mögliche Fragestellung betrifft das Ausmaß, in dem die tägliche Selbstwertschätzung in Abhängigkeit von stressreichen Ereignissen variiert und ob diese Kovariation stärker für negative als für positive Ereignisse ist. Zusätzlich können interindividuelle Differenzen einbezogen werden, um festzustellen, ob z. B. der Zustandsselbstwert bei einigen Personen stärker fluktuiert als bei anderen (beispielsweise in Abhängigkeit von impliziter Selbstwertschätzung; vgl. Schröder, Schumny & Schütz, in Vorbereitung).

Zusammenhänge auf unterschiedlichen Analyseebenen

Das wichtigste Grundprinzip von Mehrebenenanalysen ist, dass Phänomene auf unterschiedlichen Analyseebenen *gleichzeitig* untersucht werden. Dieser Aspekt ist deshalb wichtig, weil Zusammenhänge auf unterschiedlichen Analyseebenen mathematisch voneinander unabhängig sind. Auf dieses Problem wurde bereits von mehreren Autoren hingewiesen (vgl. Affleck, Zautra, Tennen & Armeli, 1999; Asendorpf, 1995, 2000; Nezlek, 2001; Schmitz, 2000). Asendorpf (2000) stellt dabei zusätzlich den Bezug von Datenanalysen auf Aggregatebene und nomothetischer Perspektive bzw. von Analysen auf der Individualebene und idiographischer Perspektive her (vgl. dazu auch Laux, 1995).

Das folgende Beispiel illustriert, wie wichtig es ist, die Analyseebenen zu berücksichtigen, da die Zusammenhänge auf den einzelnen Analyseebenen eines Datensatzes unterschiedlich sein können. Nehmen wir an, dass in einer Studie zum Zusammenhang von Produktivität und Kohäsion diese beiden Variablen in drei Arbeitsteams gemessen werden. In den Abbildungen 1 bis 4 sind hypothe-

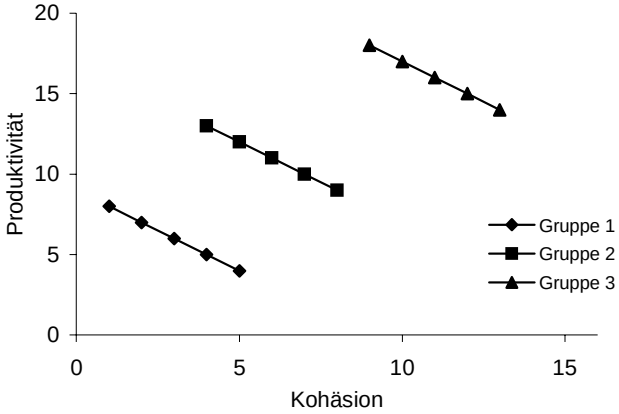

Abbildung 1. Negativer Zusammenhang auf individueller Ebene, positiver Zusammenhang auf Gruppenebene.

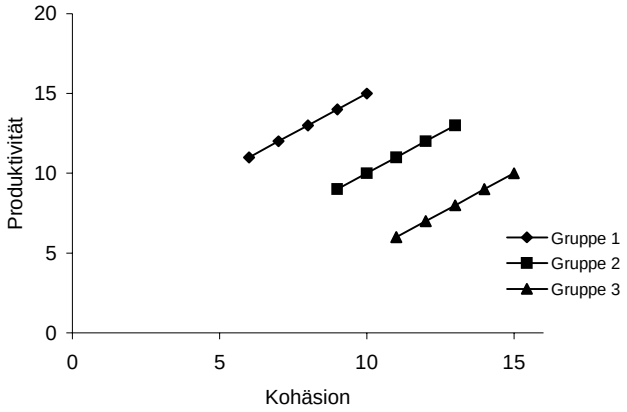

Abbildung 2. Positiver Zusammenhang auf individueller Ebene, negativer Zusammenhang auf Gruppenebene.

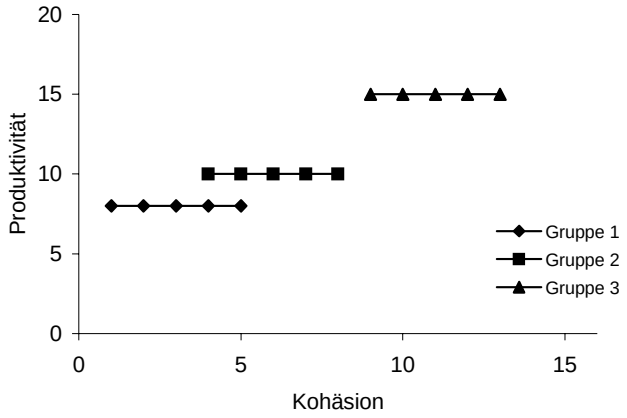

Abbildung 3. Kein Zusammenhang auf individueller Ebene, positiver Zusammenhang auf Gruppenebene.

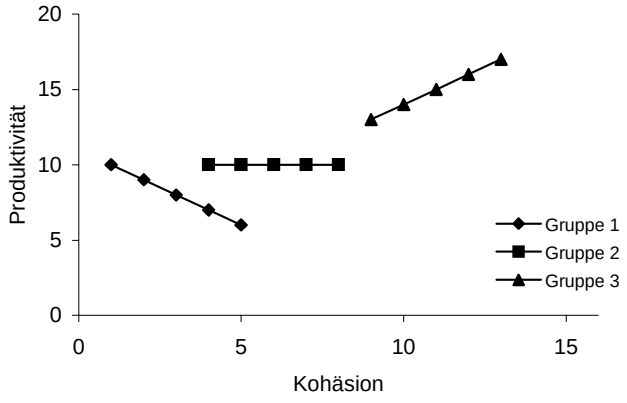

Abbildung 4. Variierende Zusammenhänge auf individueller Ebene, positiver Zusammenhang auf Gruppenebene.

tische Daten dargestellt. Im ersten Datensatz (Abbildung 1) zeigt sich ein negativer Zusammenhang auf der individuellen Ebene (innerhalb der Gruppen) und ein positiver Zusammenhang auf Gruppenebene (zwischen den Gruppen). In jeder der drei Gruppen gilt also, dass Personen, die über höhere Kohäsion berichten, weniger produktiv sind. Betrachtet man den Zusammenhang jedoch über die Gruppen hinweg, so zeigt sich, dass höhere Kohäsion mit höherer Produktivität einhergeht. Im zweiten Datensatz ist der Zusammenhang auf individueller Ebene (innerhalb der Gruppen) positiv und auf Gruppenebene negativ (Abbildung 2). Im dritten Datensatz findet sich in keiner der drei Gruppen ein Zusammenhang auf individueller Ebene, jedoch ein positiver Zusammenhang auf Gruppenebene (Abbildung 3). Im vierten Datensatz schließlich besteht ebenfalls ein positiver Zusammenhang auf Gruppenebene, die Beziehungen auf individueller Ebene variieren jedoch über die drei Gruppen hinweg (Abbildung 4).

Wie diese Beispiele verdeutlichen, kann jeder Zusammenhang auf individueller Ebene in Kombination mit jedem Zusammenhang auf Gruppenebene auftreten. Je nach Betrachtungsebene würde die Frage nach dem Zusammenhang der Variablen unterschiedliche Antworten ergeben, von denen jede zwar richtig ist, aber nur einen Teil der Informationen wiedergibt. Es wäre ein Fehlschluss, von einem Zusammenhang auf einer Analyseebene auf einen Zusammenhang auf einer anderen Ebene zu schließen, beispielsweise von interindividuellen auf intraindividuelle Zusammenhänge. Derartige Fehlschlüsse sind nicht selten, wenn es darum geht, gruppenstatistisch gewonnene Gesetzmäßigkeiten auf einzelne Individuen anzuwenden (vgl. Schmitz, 2000). Mehrebenenanalysen bieten den Vorteil, eine vollständige Betrachtung auf allen Ebenen zu ermöglichen.

Probleme und Defizite herkömmlicher Analysestrategien

Daten mit Mehrebenenstrukturen wurden und werden mit einer Vielzahl unterschiedlicher Techniken analysiert, die jedoch nicht alle gleichermaßen akkurate Ergebnisse (d. h. Parameterschätzungen für eine Population auf der Basis von Stichproben, die aus dieser Population gezogen wurden) liefern. Herkömmliche Analysestrategien, v. a. Standard-Regressionsanalysen, können (im Gegensatz zur weiter unten ausführlich erläuterten Mehrebenenmodellierung mit Zufallskoeffizienten) zu Ungenauigkeiten bzw. Fehlern führen. In der Vergangenheit wurden Mehrebenendaten mit traditionellen Analysen ausgewertet und dabei nach einer Aggregations- oder Disaggregationsstrategie vorgegangen. In Disaggregationsanalysen werden Eigenschaften der Analyseeinheiten der übergeordneten Ebene (z. B. Arbeitsgruppen) den Analyseeinheiten der unteren Ebene (z. B. Personen) zugeordnet und die Analysen auf der untergeordneten Ebene durchgeführt. In Aggregationsanalysen werden die Variablen für jede übergeordnete Analyseeinheit (z. B. Gruppe) zusammengefasst und Analysen auf der Gruppenebene durchgeführt.

Hauptnachteil von Disaggregationsstrategien ist die Annahme, dass die Zusammenhänge zwischen den Variablen der individuellen Ebene für alle Gruppen gleich sind. Das Auspartialisieren von Dummy-Variablen bei so genannten „Least Squares Dummy Variables"-Analysen (LSDV) kontrolliert zwar Unterschiede zwischen den Gruppen in den Mittelwerten der Prädiktoren, berücksichtigt jedoch nicht die Möglichkeit, dass Zusammenhänge von Variablen der individuellen Ebene, die innerhalb von Gruppen bestehen, über Gruppen hinweg variieren können. Beispielsweise findet sich in Abbildung 4 ein perfekt negativer Zusammenhang zwischen den beiden Variablen in Gruppe 1, kein Zusammenhang in Gruppe 2 und ein perfekt positiver Zusammenhang in Gruppe 3. Eine LSDV-Analyse würde zwischen den beiden Variablen hingegen einen Zusammenhang von Null schätzen, was offensichtlich den Zusammenhang zwischen Produktivität und Kohäsion in diesem Datensatz nicht korrekt beschreibt. Damit zeigt sich, dass selbst ausgereiftere Disaggregationsstrategien wie die LSDV-Analyse nicht geeignet sind, um Daten mit einer Mehrebenenstruktur zu analysieren.

In Aggregationsanalysen werden die Analysen auf der Gruppenebene durchgeführt, nachdem die Variablen der unteren Analyseebene für jede übergeordnete Analyseeinheit (z. B. Gruppe) zusammengefasst wurden. Ein Vorteil der Aggregations- gegenüber der Disaggregationsstrategie ist, dass untersucht werden kann, wie die auf einer Analyseebene bestehenden Zusammenhänge auf einer anderen Ebene variieren. In einer Studie zu Schulleistungen könnten beispielsweise Zusammenhänge zwischen zwei Variablen auf Schülerebene (z. B. IQ und Mathematikleistungen) für jede Klasse berechnet werden und Unterschiede in diesen Korrelationen analysiert werden (z. B. in Abhängigkeit von Eigenschaften der Lehrkraft).

Bei vielen Aggregationsanalysen ist man jedoch mit einem weiteren Problem konfrontiert: Unterschiedliche Klassen bestehen beispielsweise aus einer unterschiedlichen Anzahl von Schülerinnen und Schülern und statistische Kennwerte aus größeren Klassen könnten reliabler sein als aus kleineren Klassen, u. a. da sie auf mehr Beobachtungen beruhen. Zusätzlich lässt sich dies folgendermaßen veranschaulichen: Nimmt man z. B. zwei Gruppen an, deren Einzelmesswerte (beispielsweise für schulische Leistungen) für Gruppe A 3, 3, 4, 5, 5 und für Gruppe B 1, 1, 4, 7, 7 betragen, so erhält man für beide Gruppen einen Mittelwert von 4. Allerdings ist dieser Mittelwert für Gruppe A ein genauerer Indikator der Schulleistungen als für Gruppe B. Sofern herkömmliche Analysen aggregierter Mittelwerte solche Besonderheiten (im Kontext der Mehrebenenanalysen bezeichnet man dies auch als die Reliabilität der aggregierten Beobachtungen) nicht explizit durch Gewichtung berücksichtigen, reagieren sie nicht auf derartige Unterschiede. Die genannten Punkte sind von besonderer Bedeutung, wenn Zusammenhänge zwischen Variablen geschätzt werden sollen, denn die Reliabilität eines Kovarianzmaßes wird durch die Reliabilität der beiden Messungen (oder deren Fehlen) beeinflusst. Zur Lösung dieses Problems wurden so genannte „weighted-least-squares"-Analysen vorgeschlagen, die aggregierte Beobachtungen nach ihrer Größe und der Reliabilität ihrer statistischen Kennwerte gewichten (z. B. Kenny, Kashy & Bolger, 1998).

Doch auch für „weighted-least-squares"-Analysen, ebenso wie für alle anderen bisher beschriebenen herkömmlichen Analysestrategien (Aggregation oder Disaggregation), besteht das Problem, dass Fehler nicht korrekt konzeptualisiert werden: Koeffizienten der Regressionsgleichungen werden als feste Effekte (fixed effects) und nicht als Zufallseffekte (random effects) behandelt, d. h. es wird kein Parameter für den Zufallsfehler eines Koeffizienten geschätzt, was zu inakkuraten Parameterschätzungen und Signifikanztests führen kann. Je umfassender der Inferenzraum, d. h. je größer die Population ist, die ein Koeffizient beschreiben soll, ist, desto mehr Gründe gibt es, einen Koeffizienten als zufällig zu modellieren (vgl. dazu Littell, Milliken, Stroup & Wolfinger, 1996, S. 230 ff.). Beispielsweise ist es in den meisten Studien, in denen für jeweils eine Person Daten über eine Reihe von Tagen hinweg gesammelt werden, unwichtig, an welchen Tagen genau die Studie stattfindet. Man nimmt an, dass die Tage zufällig aus einer Population von Tagen gezogen werden und die üblichen Lebensumstände der Teilnehmer repräsentieren. Koeffizienten, die auf Stichproben anderer Tage basieren, wären ebenso valide (wenn auch nicht von genau gleichem numerischen Wert) wie die, die auf der gezogenen Stichprobe basieren. Demzufolge sind die Koeffizienten zufällig in dem Sinne, dass sie als Stichprobe aus der Population möglicher Koeffizienten für jeden Teilnehmer gezogen werden. Daher sollten Koeffizienten zur Beschreibung von Zusammenhängen auf intraindividueller Ebene als zufällig und nicht als fest behandelt werden. Jede Einzelbeobachtung wird also durch zwei Arten von Fehlern beeinflusst, zum einen durch die Stichprobenziehung von Tagen und zum anderen durch die Stichprobenziehung von Teilnehmern. Entsprechend sind auch die jeweiligen Koeffizienten, die die Tage und die Teilnehmer beschreiben, fehlerbehaftet. Herkömmliche Analysestrategien können jedoch nicht zwei Unbekannte gleichzeitig schätzen, was letztlich zu irreführenden Parameterschätzungen führen kann.

Mehrebenenmodellierung mit Zufallskoeffizienten

Zunehmend wird gefordert, Daten mit einer Mehrebenenstruktur mittels Mehrebenenmodellierung mit Zufallskoeffizienten (multilevel random coefficient modeling, MRCM) auszuwerten (Bryk & Raudenbush, 1992; Kenny, Kashy & Bolger, 1998; Kreft & de Leeuw, 1998). Ausführliche Einführungen zu MRCM bieten Bryk und Raudenbush (1992), Kreft und de Leeuw (1998) und Snijders und Bosker (1999); in deutscher Sprache u. a. Engel (1998), Ditton (1998), Langer (2004), Richter und Naumann (2002) und von Saldern (1986). Mit MRCM sind genauere Analysen möglich als mit vergleichbaren herkömmlichen Techniken, wie z. B. Standard-Regressionsanalysen, denn bei letzteren können abhängige Werte zu nichterwartungstreuen

und inkonsistenten Schätzungen führen. Da bei MRCM Maximum-Likelihood-Prozeduren zur Schätzung der Koeffizienten eingesetzt werden (vgl. Nezlek, 2001), bieten sie den großen Vorteil, Zufallsfehler auf allen Analyseebenen gleichzeitig modellieren zu können. Damit ist es, wie weiter hinten erläutert, möglich, Effekte als zufällig zu modellieren.

Bei der Mehrebenenmodellierung mit Zufallskoeffizienten werden Zusammenhänge (d. h. Varianzen und Kovarianzen) auf mehreren Ebenen gleichzeitig untersucht. Der vorliegende Artikel konzentriert sich auf Analysen mit zwei Ebenen, da solche Datenstrukturen relativ häufig vorkommen und eine Vielfalt an Forschungsfragen betreffen. Theoretisch ist die Anzahl der Ebenen in einer Datenstruktur nicht beschränkt, dennoch ist Zurückhaltung bei der Hinzunahme weiterer Analyseebenen zu empfehlen, worauf wir später noch eingehen.

Konzeptuell kann man sich MRCM als eine Reihe geschachtelter Regressionsanalysen vorstellen, in denen die Koeffizienten einer Analyseebene zu abhängigen Variablen auf der nächsten Analyseebene werden.[1] Mehrebenenanalysen werden üblicherweise durch separate Gleichungen für jede Analyseebene beschrieben. Programme, die MRCM durchführen, lösen jedoch eine einzige Gleichung, die alle Analyseebenen gleichzeitig enthält.

Analyse von Individuen in Gruppen

Mehrebenendesigns werden häufig dann eingesetzt, wenn Teilnehmerinnen und Teilnehmer einer Studie unterschiedlichen Gruppen angehören. Im Folgenden wird erläutert, wie die Mehrebenenmodellierung mit Zufallskoeffizienten in diesem Fall angewendet wird (vgl. dazu auch Nezlek & Zyzniewski, 1998). Angenommen sei eine Datenstruktur mit zwei Ebenen für Mitarbeiter in Arbeitsgruppen. Die Analysen beinhalten Gleichungen auf der Ebene der einzelnen Mitarbeiter (Ebene 1) und Gleichungen auf der Ebene der Arbeitsgruppen (Ebene 2). Die einfachsten Analysen, die als Nullmodell (totally unconditional model) bezeichnet werden, modellieren nur den Mittelwert auf jeder Ebene. Die Gleichungen 1 und 2 stellen eine solche Analyse dar:

Ebene 1: $y_{ij} = \beta_{0j} + r_{ij}$ (Gl. 1)

Ebene 2: $\beta_{0j} = \gamma_{00} + u_{0j}$ (Gl. 2)

In diesem Modell wird die kontinuierliche Variable y für i Individuen in j Gruppen gemessen. Die Regressionskonstante (intercept) β_{0j} gibt den Mittelwert von y für jede Gruppe j an. Die Variable y wird auf Ebene 1 als Funktion der Regressionskonstante für jede Gruppe (β_{0j}) und des Fehlers (r_{ij}) modelliert (Gleichung 1). Die Varianz von r_{ij} entspricht der Varianz der abhängigen Variable auf Ebene 1.

Die Regressionskonstante β_{0j} wird in diesem Modell nur als Funktion des Gesamtmittelwerts (γ_{00}) und des Fehlers (u_{0j}) modelliert (Gleichung 2). Die Varianz von u_{0j} ist die Varianz der abhängigen Variable auf Ebene 2. Die Gesamtvarianz von y entspricht der Summe der Varianzen auf Ebene 1 und Ebene 2. Derartige Nullmodelle (auch als unkonditionierte Modelle bezeichnet) geben also Aufschluss darüber, wie die Varianz auf die einzelnen Analyseebenen verteilt ist. Somit zeigen sie, auf welchen Ebenen konditionierte Modelle (d. h. Modelle mit zusätzlichen Prädiktoren) informativ sein können. Zusätzlich benötigt man die Varianzen für die Berechnung von Effektstärken. Aus diesen Gründen wird die Berechnung unkonditionierter Modelle, obwohl sie keine Hypothesen testen, als erster Schritt (vor der Berechnung konditionierter Modelle) empfohlen.

Das Hauptinteresse der meisten Forscherinnen und Forscher wird jedoch auf den Ergebnissen konditionierter Modelle liegen. Ein einfaches Beispiel ist das folgende Modell, welches auf Ebene 1 unkonditioniert und auf Ebene 2 konditioniert ist, d. h. zusätzliche Prädiktoren enthält. Angenommen, für die einzelnen Mitarbeiter in den verschiedenen Arbeitsgruppen würde die Produktivität gemessen und zusätzlich sei die Zeitdauer des Bestehens der Gruppen bekannt, so kann man dies durch folgendes Modell (Gleichungen 3 und 4) abbilden:

Ebene 1: $y_{ij} = \beta_{0j} + r_{ij}$ (Gl. 3)

Ebene 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}\,(\text{Zeit}) + u_{0j}$ (Gl. 4)

In diesem Modell ist Produktivität die abhängige Variable (y_{ij}). Die Gleichung für Ebene 1 entspricht der im unkonditionierten Modell, während in die Gleichung für Ebene 2 die Zeitdauer des Bestehens der Gruppe (Zeit) als Prädiktor einbezogen wurde (Gleichung 4). Geprüft wird, ob die Zeitdauer des Bestehens mit der mittleren Produktivität einer Gruppe zusammenhängt, indem die Signifikanz des γ_{01} (Zeit)-Koeffizienten getestet wird.

Der Vorteil dieser Analyse gegenüber einer Analyse aggregierter Mittelwerte (d. h. Berechnung der mittleren Produktivität für jede Gruppe und Korrelation dieser Mittelwerte mit der Zeitdauer) ist die Präzisionsgewichtung bei MRCM (Bryk & Raudenbush, 1992). Dabei werden Unterschiede in den Gruppengrößen sowie die oben beschriebenen Reliabilitätsunterschiede der Gruppenmittelwerte zur Gewichtung herangezogen.

Auch in Modellen der Ebene 1 können Prädiktoren ergänzt werden. Beispielsweise könnte man zusätzlich zur Produktivität die wahrgenommene Kohäsion messen. Die zugehörige Fragestellung bezieht sich auf den Zusammenhang zwischen Produktivität und Kohäsion auf individueller Ebene (d. h. innerhalb von Gruppen) und wird in folgendem Modell (Gleichung 5) dargestellt:

Ebene 1: $y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{Kohäsion}) + r_{ij}$ (Gl. 5)

[1] Diese hierarchische Struktur erklärt, warum MRCM manchmal „Hierarchische Lineare Modelle“ (HLM) genannt werden (neben anderen Bezeichnungen wie z. B. „Varianzkomponenten-Modelle“). Es wird jedoch in Anlehnung an de Leeuw und Kreft (1995) empfohlen, die Bezeichnung HLM nicht zu verwenden. Damit soll vermieden werden, dass Verwechslungen auftreten zwischen der allgemeinen Art der Analyse (MRCM) und einer speziellen, vielfach eingesetzten Technik bzw. einem häufig verwendeten Programm (HLM; Raudenbush, Bryk, Cheong & Congdon, 2000). Aus diesem Grund wird im Artikel einheitlich der Begriff MRCM verwendet.

In diesem Modell wird für jede der j Gruppen ein Koeffizient (β_{1j}) geschätzt, der den Zusammenhang von Produktivität und Kohäsion innerhalb jeder Gruppe repräsentiert. Wie in der Literatur zu linearen Modellen werden auch im Kontext der Mehrebenenanalysen solche Koeffizienten als Regressionssteigungen (slopes) bezeichnet (vgl. z. B. Bryk & Raudenbush, 1992). Der Zusammenhang von Produktivität und Kohäsion wird auf statistische Signifikanz hin überprüft (Weicht die mittlere Regressionssteigung für Kohäsion bedeutsam von 0 ab?), indem auf Ebene 2 eine Erweiterung des Basismodells erfolgt (Gleichungen 6 und 7):

Ebene 2: Regressionskonstante: $\beta_{0j} = \gamma_{00} + u_{0j}$ (Gl. 6)

Ebene 2: Kohäsion: $\beta_{1j} = \gamma_{10} + u_{1j}$ (Gl. 7)

Der Mittelwert der Regressionssteigungen wird durch γ_{10} repräsentiert. Wenn γ_{10} sich von Null unterscheidet, gilt dies auch für den Mittelwert der Regressionssteigungen von Kohäsion. Zu beachten ist, dass sowohl β_0 als auch β_1 als Zufallseffekte modelliert werden – beiden ist ein Fehlerterm (u_{0j} und u_{1j}) zugeordnet. Um einen Effekt zu „fixieren", würde man den Zufallsfehlerterm aus dem Modell entfernen (vgl. dazu aber den Abschnitt zu Modellen mit festen vs. Zufallseffekten weiter unten).

Ähnlich zur Analyse von Regressionskonstanten können Prädiktoren auch zu Gleichungen der Ebene 2 hinzugefügt werden, die Regressionssteigungen betreffen. Dieser Ansatz wird als „slopes as outcomes"-Analyse bezeichnet. Dabei werden Interaktionseffekte modelliert, nämlich Interaktionen von Prädiktoren auf der Individual- und der Kontextebene (so genannte cross-level interactions). Beispielsweise könnte von Interesse sein, wie der Zusammenhang von Produktivität und Kohäsion (die Regressionssteigung β_{1j} des Modells der Ebene 1, vgl. Gleichungen 5 und 7) als Funktion der Zeitdauer, seit der die Gruppe besteht, variiert. Das Modell in Gleichung 8 stellt eine solche Analyse dar:

Ebene 2: Kohäsion: $\beta_{1j} = \gamma_{10} + \gamma_{11}$ (Zeit) + u_{1j} (Gl. 8)

In diesem Modell wird die Hypothese geprüft, indem der γ_{11} (Zeit)-Koeffizient auf Signifikanz getestet wird. Auch hier liegt der Vorteil dieser Analyse gegenüber herkömmlichen Analysen (d. h. Berechnung der Zusammenhänge zwischen Produktivität und Kohäsion für jede Gruppe und deren anschließende Korrelation mit der Variable Zeitdauer) in der bereits beschriebenen Präzisionsgewichtung. Diese ist für die Analyse von Regressionssteigungen sogar noch wichtiger als für Regressionskonstanten, da die Reliabilität einer Regressionssteigung eine gemeinsame Funktion der Reliabilitäten der beiden Variablen ist, die zur Regressionssteigung beitragen.

Messwiederholungsdesigns

Die Mehrebenenmodellierung mit Zufallskoeffizienten eignet sich für eine Vielfalt von Datenstrukturen. Neben der oben erläuterten Anwendung für Daten von Individuen in Gruppen ist MRCM auch für Messwiederholungsdesigns angebracht, die zunehmend in der Persönlichkeits-, Sozial- und Gesundheitspsychologie Anwendung finden. In derartigen Studien wird eine größere Anzahl von Beobachtungen pro Person erhoben, weswegen man auch von „intensive repeated measures"-Designs spricht. In diesem Fall sind Personen die Analyseeinheiten der übergeordneten Ebene (Ebene 2) und die wiederholten Beobachtungen, die für jede Person vorliegen, sind Analyseeinheiten der (untergeordneten) Ebene 1 (im Gegensatz zu Analysen für Individuen in Gruppen, in denen die Individuen auf Ebene 1 modelliert werden). Wheeler und Reis (1990) unterscheiden dabei Designs, in denen Daten nach dem Auftreten bestimmter Ereignisse (ereigniskontingent) oder dem Verstreichen einer bestimmten Zeit (intervallkontingent) aufgezeichnet werden (zur Analyse derartiger Daten mittels MRCM vgl. Nezlek, 2001).

Tagebücher zur Protokollierung sozialer Interaktionen (z. B. Asendorpf & Wilpers, 1999), die oft nach dem Vorbild des von Wheeler und Nezlek (1977) entwickelten Rochester Interaction Record erstellt werden, stellen ein typisches Beispiel für die ereigniskontingente Datenerhebung dar. Kritisches Ereignis sind soziale Interaktionen, d. h. die Teilnehmerinnen und Teilnehmer beschreiben alle sozialen Interaktionen, die sie innerhalb eines Zeitraumes (meist zwei Wochen) erleben. Bei der Datenanalyse (vgl. dazu Nezlek, 2003) betrachtet man Unterschiede zwischen Interaktionstypen (z. B. mit Freunden vs. Partnern, bzw. am Arbeitsplatz oder zu Hause) und Zusammenhänge zwischen Interaktionen und Persönlichkeitsmerkmalen, z. B. Neurotizismus (Schröder & Schütz, 2004) oder emotionale Intelligenz (Lopes et al., 2004).

In Tagebuchstudien beschreiben Modelle der Ebene 1 intraindividuelle Zusammenhänge und Modelle der Ebene 2 untersuchen interindividuelle Unterschiede in diesen intraindividuellen Zusammenhängen. Mit anderen Worten, geprüft wird, wie interindividuell variierende Prädiktoren intraindividuelle Zusammenhänge moderieren. Beispielsweise ist es möglich, auf Ebene 1 soziale Interaktionen als Funktion der Anzahl der anwesenden Personen zu klassifizieren (z. B. Dyaden oder Gruppen). Unterschiede in Reaktionen auf Dyaden – vs. Gruppeninteraktionen können mit Hilfe des folgenden Modells (Gleichung 9) auf Ebene 1 untersucht werden:

Ebene 1: $y_{ij} = \beta_{0j} + \beta_{1j}$ (Dyaden-Kontrast) + r_{ij} (Gl. 9)

In dieser Analyse ist y eine Bewertung der sozialen Interaktion (z.B. wahrgenommene Intimität) und Dyaden-Kontrast eine kontrastkodierte Variable, die für Dyaden als 1 und für Gruppen als –1 kodiert wird. Dabei repräsentiert die Regressionssteigung den Unterschied in der wahrgenommenen Intimität zwischen Dyaden und Gruppen.

Die Regressionssteigung der Ebene 1 kann dann wiederum auf Ebene 2 analysiert werden. Mit Hilfe des Modells in Gleichung 10 wird z. B. geprüft, ob im Mittel ein Unterschied in der wahrgenommenen Intimität zwischen Dyaden und Gruppen besteht:

Ebene 2: Dyaden-Kontrast: $\beta_{1j} = \gamma_{10} + u_{1j}$ (Gl. 10)

Es ist aber auch möglich, die Regressionssteigung als Funktion einer Variable der Personenebene (Ebene 2), wie z. B. Ängstlichkeit, zu analysieren (Gleichung 11):

Ebene 2: Dyaden-Kontrast:
$$\beta_{1j} = \gamma_{10} + \gamma_{11} \text{ (Ängstlichkeit)} + u_{1j} \qquad \text{(Gl. 11)}$$

Mit dem Modell in Gleichung 11 kann beispielsweise geprüft werden, ob der Unterschied in wahrgenommener Intimität zwischen dyadischen Interaktionen und Gruppeninteraktionen für ängstlichere Personen größer ist als für weniger ängstliche.

Werden pro Individuum die wiederholten Messungen jeweils nach dem Verstreichen einer bestimmten Zeit erhoben, so spricht man von intervallkontingenter Datenerhebung. Ein Beispiel hierfür sind Untersuchungen zu Veränderungen psychischer Zustände (states), wie Stimmung oder Selbstwertschätzung, die mit Zustandsmaßen ein- oder mehrmals täglich über einen längeren Zeitraum (z. B. 14 Tage) erhoben werden. Die Analysen konzentrieren sich auf intraindividuelle Zusammenhänge zwischen Daten auf Tagesebene (z. B. tägliches Zustandsselbstwertgefühl und tägliche Ereignisse) und darauf, wie diese intraindividuellen Zusammenhänge als Funktion interindividueller Unterschiede (wie Depressivität; vgl. Schröder, Schumny & Schütz, in Vorbereitung) variieren.

Die Modelle zur Analyse intervallkontingenter Daten ähneln strukturell denen zur Analyse ereigniskontingenter Daten. Die Kovariation zwischen Selbstwertschätzung und täglichen Ereignissen innerhalb von Personen kann auf Ebene 1 mit folgendem Modell dargestellt werden (Gleichung 12):

Ebene 1:
$$y_{ij} = \beta_{0j} + \beta_{1j} \text{ (Pos. Ereignisse)} + \beta_{2j} \text{ (Neg. Ereignisse)} + r_{ij} \qquad \text{(Gl. 12)}$$

Mit einer derartigen Analyse kann man testen, ob der mittlere Koeffizient für die täglichen Ereignisse sich bedeutsam von Null unterscheidet, ob also ein Zusammenhang zwischen positiven oder negativen Ereignissen und Selbstwertschätzung besteht.

Zusätzlich können interindividuelle Unterschiede in diesen Zusammenhängen untersucht werden:

Ebene 2: Positives Ereignis:
$$\beta_{1j} = \gamma_{10} + \gamma_{11} \text{ (Depressivität)} + u_{1j} \qquad \text{(Gl. 13)}$$

In diesem Modell (Gleichung 13) wird die Regressionssteigung für die positiven Ereignisse beispielsweise als Funktion von Depressivität dargestellt.

Allgemeine Überlegungen und praktische Hinweise

Unabhängig vom Design ist bei der Durchführung von Mehrebenenanalysen die Berücksichtigung verschiedener Punkte notwendig, von denen einige im Folgenden diskutiert werden sollen.

Modelle mit festen vs. Zufallseffekten

Für die korrekte Modellierung des Fehlers in einem Mehrebenenmodell ist es zunächst wichtig, die Terminologie zu verstehen. Koeffizienten können auf drei verschiedene Arten modelliert werden: als zufällig, fest oder konstant. Wird ein Koeffizient als zufällig modelliert, so bedeutet dies, dass (zusätzlich zu einem festen Effekt, d. h. der Schätzung des Koeffizienten für den Mittelwert) ein Zufallsfehlerterm für diesen Koeffizienten geschätzt wird. Wenn ein Koeffizient als fest modelliert wird, bedeutet dies, dass für diesen Effekt kein Zufallsfehlerterm geschätzt wird. Es bedeutet jedoch *nicht*, dass keine Variabilität in diesem Koeffizienten besteht (in diesem Fall würde es sich um einen konstanten Effekt handeln, der von den meisten Programmen für Mehrebenenanalysen nicht geschätzt wird). Wenn eine Regressionssteigung als fester Effekt modelliert wird und in das Modell ein Prädiktor auf Ebene 2 einbezogen wird, beschreibt man die Regressionssteigung als nicht-zufällig varriierend, d. h. sie variiert (in Abhängigkeit vom Prädiktor), jedoch ohne einen Zufallsfehlerterm.[2] Wenn die Regressionssteigung hingegen mit Zufallsfehlerterm modelliert wird, nennt man sie zufällig variierend.

In der weiter oben beschriebenen Beispielstudie zur Produktivität und Kohäsion in Arbeitsgruppen wurde auf Ebene 2 die mittlere Regressionssteigung zwischen Produktivität und Kohäsion geschätzt (γ_{10}). Im Beispiel (Gl. 7) wurde die Regressionssteigung als Zufallseffekt modelliert, was daran erkennbar ist, dass zusätzlich der Fehlerterm u_{1j} in das Modell aufgenommen wurde. Entfernt man den Fehlerterm u_{1j} aus dem Modell, so liegt eine Modellierung der Regressionssteigung als fester Effekt vor – ein Vorgehen, das als „Fixierung/Festsetzung eines Effekts" bezeichnet wird. Aber auch wenn die Regressionssteigung als fester Effekt modelliert wird, können Unterschiede zwischen Gruppen in den Regressionssteigungen analysiert werden, z. B. in Abhängigkeit von Zeit als Prädiktor auf Ebene 2 (Gl. 8).

Obwohl man meist an festen Effekten interessiert ist (z. B. zur Beantwortung der Frage, ob ein Mittelwertskoeffizient sich bedeutsam von Null unterscheidet), ist bei der Modellierung der Fehlerterme Vorsicht geboten, da sich diese stark auf die Ergebnisse von Analysen (inklusive Signifikanztests) auswirken kann. In den meisten Fällen empfiehlt es sich daher, Effekte als zufällig zu modellieren. In Studien, in denen, wie weiter oben beschrieben, für die Teilnehmer eine Zufallsstichprobe von Tagen gezogen wird oder in denen Teilnehmer Gruppen zugeordnet werden, gibt es zwei Quellen für Stichprobenfehler. Folglich sind beide Arten von Fehlern auch in die Modelle aufzunehmen. Gelegentlich ist es jedoch schwierig, Zufallsfehlerterme reliabel zu schätzen. Bei der Analyse wird dann

[2] Um Missverständnisse bezüglich der Begrifflichkeiten zu vermeiden, sei auf folgenden Unterschied hingewiesen: Im Gegensatz zu MRCM bedeutet der Begriff „feste Effekte" in Programmen zu Strukturgleichungsmodellen, dass es sich um konstante Effekte handelt, d. h. der Wert eines Maßes ist für alle Analyseeinheiten gleich.

ein p-Wert, der deutlich größer ist als .05, für den Zufallsfehlerterm ausgegeben. In diesem Sonderfall sollte der Zufallsfehlerterm entgegen der oben gegebenen Empfehlungen aus dem Modell entfernt, also nicht geschätzt werden. Detailliertere Erläuterungen zu dieser Problematik gibt Nezlek (2001).

Optionen der Zentrierung

Zentrierung bezieht sich auf die „Lokation", d. h. den Referenzwert eines Prädiktors bei der Parameterschätzung. Beispielsweise werden in herkömmlichen Regressionsanalysen Variablen häufig mittelwertszentriert, und die Koeffizienten basieren auf Abweichungen vom Stichprobenmittelwert. Im Kontext der MRCM bestehen unterschiedliche Zentrierungsoptionen auf Ebene 1 und Ebene 2, die im Folgenden erläutert werden.

Für die Ebene 1 existieren drei Optionen: Beibehaltung der ursprünglichen Metrik (d. h. keine Zentrierung), Zentrierung um den Gesamtmittelwert und Zentrierung um die Gruppenmittelwerte. In der folgenden Darstellung bezieht sich der Begriff Gruppe auf die Analyseeinheit der Ebene 2. Häufig ist dies eine tatsächliche Gruppe, wie z. B. eine Schulklasse oder eine Gruppe von Mitarbeitern. Wie weiter oben beschrieben, können aber auch pro Person Daten von mehreren Messzeitpunkten vorliegen. In diesem Fall stellen Individuen die Analyseeinheiten der Ebene 2 dar und werden zum Zwecke der Zentrierung ebenfalls als „Gruppen" bezeichnet.

Wenn Prädiktoren nicht zentriert werden (also „unzentriert" in die Analyse eingehen), basieren die Regressionssteigungen auf Zusammenhängen zwischen einer abhängigen Variablen und Prädiktoren, die als Abweichungen von Null repräsentiert sind. In derartigen Analysen repräsentiert die Regressionskonstante für jede Gruppe den erwarteten Wert der abhängigen Variablen für den Fall, dass der Prädiktor gleich Null ist. Wenn Prädiktoren um den Gesamtmittelwert zentriert werden, beruhen die Regressionssteigungen auf Zusammenhängen zwischen einem Kriterium und Prädiktoren, die in Form von Abweichungen vom Gesamtmittelwert repräsentiert sind. Hier entspricht die Regressionskonstante für jede Gruppe dem erwarteten Wert der abhängigen Variablen für den Fall, dass der Prädiktor gleich dem Gesamtmittelwert für diesen Prädiktor ist. Wenn Prädiktoren um die Gruppenmittelwerte zentriert werden, beruhen die Regressionssteigungen auf Zusammenhängen zwischen einer abhängigen Variable und Prädiktoren, die in Form von Abweichungen vom Mittelwert jeder Gruppe dargestellt sind. Hierbei repräsentiert die Regressionskonstante für jede Gruppe den erwarteten Wert der abhängigen Variablen für den Fall, dass ein Prädiktor gleich dem Mittelwert des Prädiktors für diese Gruppe ist.

Ein Beispiel soll die Unterschiede zwischen den drei Optionen verdeutlichen: In einer Studie könnten z. B. die schulischen Leistungen und die Intelligenz jedes einzelnen Schulkindes in zehn unterschiedlichen Klassen gemessen und der Zusammenhang dieser beiden Variablen untersucht werden. Auf Ebene 1 ist Intelligenz der Prädiktor (x) und schulische Leistung das Kriterium (y). Wird x (Intelligenz) nicht zentriert, so repräsentiert die Regressionskonstante für jede Analyseeinheit (hier: jede Klasse) den erwarteten Wert für y (schulische Leistung), wenn x gleich Null ist. Dies wäre kein angemessenes Modell für das beschriebene Beispiel, da Intelligenz als Prädiktor diesen Wert nicht annehmen kann. Wird x um den Gesamtmittelwert zentriert, so repräsentiert die Regressionskonstante für jede Gruppe (Klasse) den erwarteten Wert für y (Schulleistung), wenn x (Intelligenz) gleich dem Gesamtmittelwert ist, d. h. die erwartete Leistung für eine Person mit einer im Vergleich zur Gesamtstichprobe durchschnittlichen Intelligenz. Wird x um die Gruppenmittelwerte zentriert, so stellt die Regressionskonstante für jede Gruppe (Klasse) den erwarteten Wert für y (Schulleistung) dar, wenn x (Intelligenz) dem Mittelwert von x für die jeweilige Gruppe entspricht.

Im Folgenden sollen die wichtigsten Unterschiede zwischen der Zentrierung um den Gesamtmittelwert und der Zentrierung um die Gruppenmittelwerte erläutert werden: Im Gegensatz zur Zentrierung um die Gruppenmittelwerte werden die Regressionskonstanten bei Zentrierung um den Gesamtmittelwert um Unterschiede zwischen den Gruppen in den Mittelwerten der Prädiktoren bereinigt. Wenn Prädiktoren um den Gesamtmittelwert zentriert werden, tragen daher Unterschiede zwischen den Gruppen in den Mittelwerten der Prädiktoren zu Parameterschätzungen (einschließlich der Schätzungen von Fehlern) bei. Dies ist nicht der Fall, wenn Prädiktoren um die Gruppenmittelwerte zentriert werden. Die Zentrierung um die Gruppenmittelwerte ähnelt funktionell am stärksten einem Vorgehen, bei dem für jede Analyseeinheit ein eigenes Standard-Regressionsmodell geschätzt und anschließend die mittleren Koeffizienten aus diesen Modellen untersucht werden. Einige Forscher empfehlen bei der Zentrierung der Prädiktoren um die Gruppenmittelwerte, dass die Mittelwerte der Prädiktoren der Ebene 1 als zusätzliche Prädiktoren auf Ebene 2 aufgenommen werden. Wenn diese Mittelwerte nicht einbezogen werden, ist diese Varianz nicht im Modell enthalten und das Modell könnte falsch spezifiziert sein.

Die Zentrierung auf der Ebene 2 ist einfacher, denn es existieren lediglich zwei Optionen: keine Zentrierung und Zentrierung um den Gesamtmittelwert. Ähnlich wie auf Ebene 1 stellt bei Analysen ohne Zentrierung die Regressionskonstante den erwarteten Wert der abhängigen Variablen dar, wenn die Prädiktoren gleich Null sind. Die Koeffizienten für die Prädiktoren beziehen sich dann auch auf diese Regressionskonstante. Werden Prädiktoren um den Gesamtmittelwert zentriert, entspricht die Regressionskonstante dem erwarteten Wert, wenn die Prädiktoren gleich dem Gesamtmittelwert sind. Dies ähnelt der Prozedur, die in den meisten Standard-Regressionsanalysen genutzt wird.

Optionen der Zentrierung werden in Bryk und Raudenbush (1992, S. 25–29), Kreft und de Leeuw (1998, S. 106–114) sowie in Kreft, de Leeuw und Aiken (1995) disku-

tiert. Obwohl Prinzipien der Zentrierung für unterschiedliche Arten von Analysen variieren, ist es dennoch möglich, einige grobe Richtlinien zu geben. Erstens ist es bei Modellen mit unzentrierten Prädiktoren wahrscheinlicher als bei Modellen mit zentrierten Prädiktoren, dass auf Grund hoher Korrelationen zwischen Regressionskonstanten und -steigungen Probleme bei der Parameterschätzung auftreten. Häufig ist es daher notwendig, Prädiktoren zu zentrieren, um diese Korrelationen zu reduzieren. Zweitens sollten Variablen zentriert werden, wenn Null keinen sinnvollen Wert für eine Variable der Ebene 1 darstellt, z.B. wenn die Variable auf einer Skala von 1 bis 7 gemessen wird, auf der Null kein möglicher Wert ist. Ungeachtet all dieser Empfehlungen ist es wichtig, dass die Entscheidungen bezüglich der Zentrierung auf der Basis der Datenstruktur und der interessierenden Hypothesen getroffen werden, denn wie Bryk und Raudenbush (1992, S. 27) bemerken: „No single rule covers all cases."

Hinzufügen von Analyseebenen

Obwohl es konzeptuell sinnvoll sein kann, weitere Analyseebenen hinzuzufügen, besitzen Datenstrukturen in vielen Fällen nicht genügend Information um die notwendigen Parameter zu schätzen. In einer Studie von Nezlek (2004) wurden beispielsweise über einen Zeitraum von zwei Wochen täglich Daten von 75 Teilnehmern aus vier Ländern gesammelt. Ursprünglich legten diese Daten ein Modell mit drei Ebenen (Tage, Personen und Länder) nahe, jedoch waren anfängliche Analysen mit drei Ebenen nicht erfolgreich. Vier Länder reichten nicht aus, um die Variable „Land" als Zufallseffekt zu modellieren. Aus diesem Grund wurden Modelle mit nur zwei Analyseebenen (Tage geschachtelt in Personen) verwendet und „Land" als Variable auf Personenebene modelliert. Es ist schwierig, Empfehlungen bezüglich der Frage zu geben, ab welcher Zahl von Analyseeinheiten weitere Ebenen hinzugefügt werden können. Nach der persönlichen Erfahrung der Autoren werden mindestens 10 Einheiten pro Ebene benötigt, jedoch ist dieser Wert auch stark von Faktoren wie den Kovarianzen zwischen den Messwerten abhängig.

Strukturgleichungsmodelle und MRCM

Der vorliegende Artikel hat Mehrebenenanalysen im Kontext der Mehrebenenmodellierung mit Zufallskoeffizienten diskutiert, es ist jedoch auch möglich, Mehrebenendaten mit Hilfe von Strukturgleichungsmodellen zu analysieren. Eine bekannte Anwendung von Strukturgleichungsmodellen für Mehrebenendaten sind Wachstumskurvenanalysen. Die Vor- und Nachteile von Strukturgleichungsmodellen vs. MRCM werden bei Schnabel, Little und Baumert (2000) diskutiert. Zwar existieren bislang keine klaren Richtlinien, wann welcher Ansatz zu bevorzugen sei, einige Hinweise können aber gegeben werden.

Wie weiter unten im Absatz zu Software dargestellt wird, ist es für Anwender, die mit Strukturgleichungsmodellen und Mehrebenenanalysen (noch) nicht vertraut sind, wahrscheinlich leichter, MRCM-Analysen und -Programme zu verstehen und durchzuführen als Strukturgleichungsmodelle. Ist man hingegen sehr an Fehlerstrukturen und latenten Variablen interessiert, so bieten Strukturgleichungsmodelle gewisse Vorteile. Obwohl mit MRCM auch latente Variablen analysiert werden können (vgl. z. B. Vansteelandt, Van Mechelen & Nezlek, 2005), bieten Strukturgleichungsmodelle mehr Flexibilität. Andererseits sind MRCM-Ansätze besser für den Umgang mit fehlenden Daten und unregelmäßigeren Datenstrukturen geeignet.

Softwareoptionen

Mit der zunehmenden Verbreitung von Mehrebenenanalysen stieg auch die Anzahl an Software-Paketen, die zur Durchführung von MRCM genutzt werden können. Zwei Programme wurden speziell für die Analyse von Mehrebenendaten entwickelt: HLM (Raudenbush, Bryk, Cheong & Congdon, 2000) und MLwiN (Rabash et al., 2000). Es ist aber auch möglich, Mplus, SAS, LISREL und SPSS für diese Analysen zu verwenden. Für die meisten Benutzer sind wahrscheinlich HLM oder MlwiN die besten Optionen, da in beiden Programmen die Mehrebenenmodelle mit Hilfe der Standardnomenklatur spezifiziert werden, wodurch eine korrekte Spezifikation der Modelle (einschließlich der Zentrierung und der Fehlerterme) deutlich erleichtert wird. HLM ist dabei das benutzerfreundlichere Programm, MlwiN bietet hingegen einige zusätzliche Optionen, die in HLM nicht möglich sind, wie z. B. Datentransformationen innerhalb des Programms, Bootstrapping und Simulationen.

Für erfahrene SAS-Anwender mag das Programm PROC MIXED geeignet sein, um Mehrebenenanalysen durchzuführen. Eine gute Einführung dazu bietet Singer (1998) und eine sehr umfassende Diskussion und Beschreibung zahlreicher Anwendungen von PROC MIXED wird von Littell, Milliken, Stroup und Wolfinger (1996) geboten. In ähnlicher Weise mag es für erfahrene Anwender von LISREL bzw. Mplus sinnvoll sein, die ihnen vertrauten Programme auch zur Durchführung von Mehrebenenanalysen einzusetzen. Im Falle des Einsatzes der letztgenannten Programme, die für die unterschiedlichsten Zwecke verwendet werden, wird jedoch empfohlen, Aspekten, die in Mehrebenenanalysen wichtiger sind als in anderen Analysen (wie z. B. der Zentrierung), besondere Aufmerksamkeit zu schenken.

Probleme und Grenzen

Nach der Nennung von Vorteilen der Mehrebenenmodellierung mit Zufallskoeffizienten soll auch auf Grenzen des Verfahrens hingewiesen werden. Zum einen ist dabei auf die Anforderungen hinzuweisen, die an die Daten bzw. die Stichprobe zu stellen sind. Es ist wichtig, dass eine ausreichende Anzahl von Untersuchungseinheiten pro Analyseebene zur Verfügung steht, um für Modellparameter und vor allem Standardfehler brauchbare Schätzungen abgeben zu können. Diese Forderung stößt in mancher Studie

an praktische bzw. ökonomische Grenzen bei der Datenerhebung. Mehrebenenanalysen wurden traditionell für Anwendungen mit sehr großen Stichproben entwickelt, Erkenntnisse zu Minimalanforderungen an die Stichproben liegen jedoch erst vereinzelt vor. Kreft (1996; zitiert nach Hox, 1998) befürwortet beispielsweise eine „30/30-Faustregel“ für einfache 2-Ebenen-Modelle. Bei Interesse an Interaktionen zwischen den Ebenen bzw. an Varianzanteilen wird die mindestens erforderliche Anzahl der Untersuchungseinheiten auf Ebene 2 mit N = 50 (Hox, 1998) bzw. N = 100 (Snijders & Bosker, 1993) angegeben. Neuere Ergebnisse aus Simulationsstudien legen auf Ebene 2 eine Stichprobengröße von mindestens N = 50 nahe (Maas & Hox, 2004). Neben derartigen praktischen Problemen ist zum anderen eine eher konzeptuelle und grundsätzliche Grenze der Mehrebenenmodellierung zu erwähnen: Komplexere Beziehungen zwischen Variablen, wie z. B. Kausalketten, sind mit Hilfe der Mehrebenenmodellierung nicht ohne weiteres abbildbar, hier sind eher Strukturgleichungsmodelle angebracht.

Fazit

Im vorliegenden Artikel wurde erläutert, wie die Mehrebenenmodellierung mit Zufallskoeffizienten zur Analyse von Messwiederholungsdesigns (z. B. Tagebuchstudien) sowie von Gruppendaten angewendet werden kann. Trotz der genannten Probleme ist die MRCM in vielen Fällen die Methode der Wahl zur Analyse von Daten mit Mehrebenenstruktur. Sie bieten gegenüber herkömmlichen Methoden wie Standard-Regressionsanalysen den Vorteil, die Struktur der Daten (u.a. die Abhängigkeit der Messungen und die Quellen des Messfehlers) angemessen zu berücksichtigen.

Literatur

Affleck, G., Zautra, A., Tennen, H. & Armeli, S. (1999). Multilevel daily process designs for consulting and clinical psychology: A preface for the perplexed. *Journal of Consulting and Clinical Psychology, 67*, 746–754.

Asendorpf, J. B. (1995). Persönlichkeitspsychologie: Das empirische Studium der individuellen Besonderheit aus spezieller und differentieller Perspektive. *Psychologische Rundschau, 46*, 235–247.

Asendorpf, J. B. (2000). Idiographische und nomothetische Ansätze in der Psychologie. *Zeitschrift für Psychologie, 208*, 72–90.

Asendorpf, J. B. & Wilpers, S. (1998). Personality effects on social relationships. *Journal of Personality and Social Psychology, 74*, 1531–1544.

Asendorpf, J. B. & Wilpers, S. (1999). KIT: Kontrolliertes Interaktionstagebuch zur Erfassung sozialer Interaktionen, Beziehungen und Persönlichkeitseigenschaften. *Diagnostica, 45*, 82–94.

Bachmann, N. (1998). *Die Entstehung von sozialen Ressourcen abhängig von Individuum und Kontext: Ergebnisse einer Multilevel-Analyse.* Münster: Waxmann.

Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Ditton, H. (1998). *Mehrebenenanalyse. Grundlagen und Anwendungen des Hierarchisch Linearen Modells.* Weinheim: Juventa.

Eid, M. (2003). Veränderungsmessung und Kausalanalysen. In M. Jerusalem & H. Weber (Hrsg.), *Psychologische Gesundheitsforschung. Diagnostik und Prävention* (S. 105–120). Göttingen: Hogrefe.

Engel, U. (1998). *Einführung in die Mehrebenenanalyse. Grundlagen, Auswertungsverfahren und praktische Beispiele.* Opladen: Westdeutscher Verlag.

Hox, J. (1998). Multilevel modeling: When and why. In I. Balderjahn, R. Mathar & M. Schader (Eds.), *Classification, data analysis and data highways* (pp. 147–154). New York: Springer.

Kenny, D. A., Kashy, D. A. & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 233–265). New York: McGraw Hill.

Kreft, I. G. G. & de Leeuw, J. (1998). *Introducing multilevel modeling*. Newbury Park, CA: Sage Publications.

Kreft, I. G. G., de Leeuw, J. & Aiken, L. (1995). The effects of different forms of centering on hierarchical linear models. *Multivariate Behavioral Research, 30*, 1–22.

Laux, L. (1995). Die Integration idiographischer und nomothetischer Persönlichkeitspsychologie. In A. Kruse & R. Schmitz-Scherzer (Hrsg.), *Psychologie der Lebensalter* (S. 15–23). Darmstadt: Steinkopff.

Langer, W. (2004). *Mehrebenenanalyse: Eine Einführung für Forschung und Praxis.* Wiesbaden: Verlag für Sozialwissenschaften.

de Leeuw, J. & Kreft, I. G. G. (1995). Questioning multilevel models. *Journal of Educational and Behavioral Statistics, 20*, 171–189.

Littell, R. C., Milliken, G. A., Stroup, W. W. & Wolfinger, R. D. (1996). *SAS System for mixed models.* Cary, NC: SAS Institute.

Lopes, P. N., Brackett, M. A., Nezlek, J. B., Schütz, A., Sellin, I. & Salovey, P. (2004). Emotional intelligence and daily social interaction. *Personality and Social Psychology Bulletin, 30*, 1018–1034.

Lutz, W., Lowry, J., Kopta, S. M., Einstein, D. A. & Howard, K. I. (2001). Prediction of dose-response relations based on patient characteristics. *Journal of Clinical Psychology, 57*, 889–900.

Maas, C. J. M. & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58*, 127–137.

Marsh, H., Köller, O. & Baumert, J. (2001). Reunification of East and West German school systems: Longitudinal multilevel modeling study of the Big-Fish-Little-Pond-Effect on academic self-concept. *American Educational Research Journal, 38*, 321–350.

Nezlek, J. B. & Zyzniewski, L. E. (1998). Using hierarchical linear modeling to analyze grouped data. *Group Dynamics: Theory, Research, and Practice, 2*, 313–320.

Nezlek, J. B. (2001). Multilevel random coefficient analyses of event and interval contingent data in social and personality psychology research. *Personality and Social Psychology Bulletin, 27*, 771–785.

Nezlek, J. B. (2003). Using multilevel random coefficient modeling to analyze social interaction diary data. *Journal of Social and Personal Relationships, 20*, 437–469.

Nezlek, J. B. (2004, August). A Cross-cultural study of relationships between daily events and daily well-being. In J. B. Nezlek & R. Sorrentino (Eds.), *Cross-Cultural Studies of Daily Experience.* Symposium conducted at the XVII International Congress of the International Association for Cross-Cultural Psychology, Xi’an, China.

Rabash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I., Lewis,

T. (2000). *A user's guide to MLwiN*. London: Centre for Multilevel Modelling, University of London.
Raudenbush, S., Bryk, A., Cheong, Y. F. & Congdon, R. (2000). *HLM5*. Chicago, IL: Scientific Software.
Richter, T. & Naumann, J. (2002). Mehrebenenanalysen mit hierarchisch-linearen Modellen. *Zeitschrift für Medienpsychologie, 14* (N.F.2) 4, 155–159.
Schmitz, B. (2000). Auf der Suche nach dem verlorenen Individuum: Vier Theoreme zur Aggregation von Prozessen. *Psychologische Rundschau, 51* (2), 83–92.
Schnabel, K. U., Little, T. D. & Baumert, J. (2000). Modeling longitudinal and multilevel data. In T. D. Little, U. K. Schnabel & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 9–14). Mahwah, NJ: Erlbaum.
Schröder, M. & Schütz, A. (2004, September). *Positive selbstbezogene Kognitionen in sozialen Interaktionen. Zusammenhänge mit Persönlichkeitseigenschaften und Situationsmerkmalen.* Vortrag auf dem 44. Kongress der Deutschen Gesellschaft für Psychologie in Göttingen.
Schröder, Schumny & Schütz. *Implizite und explizite Selbstwertschätzung und Umgang mit emotional relevanten Alltagsereignissen.* Manuskript in Vorbereitung.
Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23,* 323–355.
Snijders, T. & Bosker, R. (1993). Modeled variance in two-level models. *Journal of Educational Statistics, 18,* 237–259.
Snijders, T. & Bosker, R. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modelling.* London: Sage Publications.
Sonnentag, S. (2001). Work, recovery activities, and individual well-being: A diary study. *Journal of Occupational Health Psychology, 6* (3), 196–210.
von Saldern, M. (1986). *Mehrebenenanalyse. Beiträge zur Erfassung strukturierter Realität.* Weinheim: PVU.
Vansteelandt, K., Van Mechelen, I. & Nezlek, J. B. (2005). The co-occurrence of emotions in daily life: A multilevel approach. *Journal of Research in Personality, 39,* 325–335.
Wheeler, L. & Nezlek, J. (1977). Sex differences in social participation. *Journal of Personality and Social Psychology, 35,* 742–754.
Wheeler, L. & Reis, H. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality, 59,* 339–354.
Wilz, G. & Brähler, E. (Hrsg.) (1997). *Tagebücher in Therapie und Forschung. Ein anwendungsorientierter Leitfaden.* Göttingen: Hogrefe.

Prof. Dr. John B. Nezlek

Department of Psychology
College of William and Mary
P.O. Box 8795
Williamsburg, Va 23187-8795
USA
E-Mail: john.nezlek@wm.edu

Dipl.-Psych. Michela Schröder-Abé

Institut für Psychologie
Technische Universität Chemnitz
09107 Chemnitz
E-Mail: michela.schroeder@phil.tu-chemnitz.de

Prof. Dr. Astrid Schütz (Korrespondenzadresse)

Institut für Psychologie
Technische Universität Chemnitz
09107 Chemnitz
E-Mail: astrid.schuetz@phil.tu-chemnitz.de

Hartmut Ditton

Mehrebenenanalyse

Grundlagen und Anwendungen des Hierarchisch Linearen Modells

Juventa Verlag Weinheim und München 1998

Der Autor

Hartmut Ditton, Jg. 1956, Dr. phil. habil., ist Privatdozent im Fach Erziehungswissenschaft an der Katholischen Universität Eichstätt mit den Arbeitsschwerpunkten Forschungsmethoden, empirische Bildungs-, Schul- und Unterrichtsforschung.

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Ditton, Hartmut:
Mehrebenenanalyse : Grundlagen und Anwendungen des hierarchisch linearen Modells / Hartmut Ditton. - Weinheim ; München : Juventa Verlag, 1998
(Juventa-Paperback)
ISBN 3-7799-1049-7

Umschlaggestaltung: Atelier Warminski, 63654 Büdingen
Umschlagabbildung: Logo des Multilevel Model Project
Printed in Germany

ISBN 3-7799-1049-7

Vorwort

Der Erkenntnisfortschritt in der sozialwissenschaftlichen Forschung ist nicht zuletzt von der Entwicklung der Forschungsmethoden und statistischen Analyseverfahren abhängig. Im Bereich der Methoden wurden in den zurückliegenden Jahren bedeutende Fortschritte erreicht, die auch durch die rasante Entwicklung der Computertechnik begünstigt wurden. Auf heutigen Standard-PCs sind Analysen möglich, die noch in den 70er Jahren jenseits der Leistungsfähigkeit von Großrechnern lagen. Zugleich wurde die Benutzerfreundlichkeit entscheidend verbessert und die Fehleranfälligkeit der Anwendungen reduziert. Faktorenanalysen, ehemals eher ein Geheimverfahren für Experten, gehören heute zum Standard selbst einfacher Statistikpakete und können von jedem Anfänger eines Statistikkurses mit wenigen "Mausklicks" durchgeführt werden.

Ein anspruchsvolles statistisches Analyseverfahren, das bereits seit langem diskutiert wird, ist die Mehrebenenanalyse. Hier geht es darum, die oftmals vorliegende hierarchische Struktur von Untersuchungsdaten in der Analyse entsprechend einem theoretischen Modell und statistisch adäquat zu analysieren. Insbesondere im Bereich der Erziehungswissenschaft ergeben sich eine Vielzahl von Anwendungsfällen. Z.B. sollte berücksichtigt werden, daß die Leistungsentwicklung von Schülern nicht nur von deren individuellen Merkmalen abhängt, sondern ebenso von Faktoren des Unterrichts und der besuchten Schule. Eine Analyse derartiger Beziehungsstrukturen geht über die Möglichkeiten von Standard-Statistikverfahren hinaus.

Für Mehrebenenanalysen sind komplexe Berechnungen der Beziehungen zwischen Variablen mehrerer Analyseebenen nötig. Erst seit den 80er Jahren wurden anwendbare Algorithmen und Programme zur Analyse entwickelt. Die Fortschritte in der Weiterentwicklung der Verfahren sind seither beträchtlich und es gibt inzwischen mehrere Software-Pakete, die spezifisch für die Analyse von Mehrebenendaten geeignet sind.

Ein Blick in die internationale Forschung führt zu dem Schluß, daß Mehrebenenanalysen inzwischen bereits zum Standardrepertoire angewandter Verfahren gehören. Besonders zu Fragen schulischer Wirkungen und schulischer Effektivität liegt bereits eine sehr große Zahl an Arbeiten vor, wie eine Durchsicht einschlägiger Zeitschriften unschwer zeigt (International Journal of Educational Research, School Effectiveness and School Improvement, Educational Research and Evaluation, Sociology of Education).

Im deutschsprachigem Raum werden die Möglichkeiten der neuen Verfahren zur Mehrebenenanalyse dagegen bisher kaum genutzt und scheinen immer noch als eine Art Geheimwissenschaft betrachtet zu werden. Einer der vielfältigen Gründe dafür könnte auch das Fehlen deutschsprachiger Literatur zu den Verfahren sein. Der vorliegende Band soll dazu beitragen, diese Lücke zu schließen.

Der Text hat primär einen einführenden und anwendungsbezogenen Charakter. Entsprechend dem Untertitel stehen elementare Grundlagen und Anwendungen der Mehrebenenanalyse im Vordergrund. Auf Beweisführungen und eine differenzierte Herleitung der Verfahren wird daher weitgehend verzichtet und statt dessen auf die Primärliteratur verwiesen. Die Darstellung hält sich an die Notation des Hierarchisch Linearen Modells von ANTHONY BRYK und STEPHEN RAUDENBUSH. Von daher sollte für den interessierten Leser, der weitergehende Informationen sucht, der Zugang zu den Primärquellen kein Problem darstellen. Überdies dürfe damit auch die Grundlage geschaffen sein, um das HLM-Programm bedienen und die Ergebnisse in den Ausdrucken nachvollziehen zu können. Zum HLM Programm selbst finden sich Ausführungen in Kapitel 5. Sonderanwendungen und spezifische Erweiterungen des Basismodells werden in Grundzügen in Kapitel 6 behandelt.

Es ist zu erwarten, daß die weitere Entwicklung der Modelle und Programme zur Mehrebenenanalyse rasant voranschreiten wird. Für jeweils aktuellste Informationen sei auf das *Multilevel Models Project* unter Leitung von HARVEY GOLDSTEIN an der Universität London verwiesen. Auf der Homepage des Projekts sind eine Fülle an Informationen und Verweise auf aktuelle Arbeiten zu finden. Über den aktuellen Stand der Software informieren am besten die manuals und Begleitunterlagen zu den jeweiligen Programmen - bes. HLM und ML/3, die über Scientific Software und ProGAMMA zu erhalten sind.

Der vorliegende Band soll auch dazu anregen, selbst Mehrebenenanalysen mit Beispieldaten zu rechnen, um sich mit den Verfahren und Vorgehensweisen vertraut zu machen und den möglichen Gewinn für eigene Forschungsarbeiten abschätzen zu können. Diesbezüglich sei darauf hingewiesen,

1 Fragestellung und Ansatz der Mehrebenenanalyse

Dieses Kapitel gibt einen einführenden Überblick zu den Fragestellungen, Besonderheiten und Anwendungsmöglichkeiten von Mehrebenenanalysen. Um die Spezifik von Mehrebenenverfahren zu verdeutlichen, ist es sinnvoll, die Abgrenzung gegenüber anderen Analyseverfahren herauszustellen. Außerdem wird in kurzen Zügen die Entwicklung der Mehrebenenanalyse bis hin zu den heute verfügbaren Verfahren aufgezeigt.

Charakterisieren lassen sich Mehrebenenanalysen in Bezug auf die *Ordnung der Gegenstände oder Objekte*, über die in der Analyse Aussagen gemacht werden sollen. Die Gegenstände können in solche erster, zweiter und höherer Ordnung so eingeteilt werden, daß die Gegenstände einer jeweils niedrigeren Ordnung als Elemente der Gegenstände der nächst höheren Ordnung definierbar sind. Die Gegenstände zweiter und höherer Ordnung sind somit als Mengen von Gegenständen der niedrigeren Ordnung darstellbar (vgl. HUMMELL 1972, 12ff.). Die Gegenstände erster Ordnung sind die jeweils individuellen Einheiten, über die in einer Untersuchung Aussagen gemacht werden sollen. In den Sozialwissenschaften handelt es sich sehr häufig um einzelne Personen. Gegenstände höherer

Ordnung können dann z.B. Cliquen, Gruppen, Organisationen, Gemeinden usw. sein, denen die Individuen angehören.

Das Charakteristikum von Mehrebenenanalysen besteht nun darin, "*daß Objekte verschiedener Ordnung gleichzeitig zum Gegenstand der Untersuchung werden*" (HUMMELL 1972, S. 13). Diese Definition ist allerdings noch recht weit gefaßt. Im spezifischen Sinn eines konkreten statistischen Verfahrens kann von einer Mehrebenenanalyse erst dann gesprochen werden, *wenn Gegenstände verschiedener Ordnung in einer Analyse simultan verrechnet werden* und somit hinsichtlich der Wirkungen auf eine abhängige Variable neben Merkmalen der individuellen Einheiten auch Merkmale kollektiver Einheiten Berücksichtigung finden.

Anwendungsfälle für Mehrebenenanalysen sind folglich *Datensätze mit einer hierarchischen Struktur*, bei denen Daten für individuelle und kollektive Einheiten vorliegen, die in Bezug aufeinander zu analysieren sind. Die sich aus der Datenstruktur und der Spezifik der Fragestellung ergebenden Besonderheiten werden nachfolgend behandelt.

1.1 Hierarchisch strukturierte Daten in der sozialwissenschaftlichen Forschung

Zahlreiche Datensätze, die in der sozialwissenschaftlichen Forschung analysiert werden, haben eine hierarchische Struktur. Sehr häufig liegen Variablen für Untersuchungseinheiten vor, die in

natürlicher Weise *übergeordneten* Einheiten zugeordnet werden können. Ein anschauliches Beispiel dafür sind Daten für einzelne Schüler, die nach den Schulklassen oder Schulen, die sie besuchen, zusammengefaßt werden können. Ein weiteres Beispiel sind Daten für Beschäftigte, die nach den Abteilungen oder Betrieben, in denen sie arbeiten, gruppiert werden können. Oftmals werden schon bei der Stichprobenziehung nicht Individuen, sondern Cluster (z.B. Schulen, Schulklassen oder Betriebe) ausgewählt, mit dem Ziel, die Zusammengehörigkeit der Probanden und die zwischen ihnen bestehenden Beziehungen im Hinblick auf die Untersuchungsfrage in Rechnung zu stellen.

Obwohl derartig geschachtelte Datenstrukturen sehr häufig sind und geradezu als Grundphänomen sozialwissenschaftlicher Untersuchungen angesehen werden können, sind elaborierte und den Datenstrukturen angemessene Analysemethoden erst seit Mitte der 80er Jahre entscheidend weiterentwickelt und in breiterem Rahmen verfügbar gemacht worden. Überwiegend werden jedoch bis heute hierarchisch organisierte Daten unangemessen analysiert. Das hat erhebliche Auswirkungen auf die Qualität der gewonnenen Ergebnisse. Auf die Frage danach, was die Probleme bei einer Analyse solcher Datensätze sind und worin die Spezifik einer Mehrebenenanalyse besteht, beziehen sich die folgenden Ausführungen.

Um den Kern vorwegzunehmen: *Analysen hierarchischer Daten unter Ignorierung ihrer Mehrebenenstruktur können zu gravierenden Fehlern führen und unter Umständen sogar völlig unbrauchbar sein*! Sie können die tatsächlichen Beziehungen in

den Untersuchungseinheiten falsch abbilden und die Forschung grundlegend in die Irre führen. Die dahinterstehende Problematik hat mehrere Aspekte.

Die üblicherweise auch für Mehrebenendaten verwendeten Statistikverfahren (wie Regressions- und Varianzanalyse) sind für Mehrebenendaten ungeeignet. Das liegt daran, daß die Voraussetzungen für die Anwendung dieser Verfahren bei hierarchisch strukturierten Daten nicht erfüllt sind. Wie schon erwähnt, werden in vielen Untersuchungen bei der Stichprobenziehung keine unabhängigen individuellen Einheiten ausgewählt, sondern zusammengehörige Cluster. Dies ist sehr häufig deshalb angebracht, weil auch Beziehungen und Beziehungsstrukturen zwischen den Individuen innerhalb der Einheiten, denen sie angehören, zu berücksichtigen sind und weil Wirkungen des gemeinsamen Kontextes überprüft werden sollen. Insofern ist diese Form der Stichprobenziehung bei einer Vielzahl von Fragestellungen angemessen, sie wirft aber unter dem Aspekt der Anwendung statistischer Analyseverfahren, die von der Annahme unabhängiger Stichprobenelemente ausgehen, Probleme auf. Bei der Ziehung von Clustern liegen nämlich naturgemäß keine unabhängigen Beobachtungen vor und mit einiger Wahrscheinlichkeit sind als Konsequenz daraus die Schätzwerte für die Standardfehler, die mit statistischen Routineverfahren ermittelt werden, fehlerhaft.

Daß keine unabhängigen Beobachtungen vorliegen, hat eine besondere Bedeutung. Die Beobachtungen (Fälle, Personen) innerhalb der Aggregateinheiten (z.B. Schulklassen, Schulen, Betriebe) sind ein-

ander ähnlicher als bei einer Zufallsstichprobe von Individuen zu erwarten ist. Sie sind gemeinsamen Einflüssen oder Erfahrungen ausgesetzt, die für die Einheiten eines Aggregats gerade typisch sind. Die Schüler einer Schulklasse haben z.B. gemeinsam mehr oder weniger kompetente Lehrer, eine bestimmte Regelung ihrer Beziehungen miteinander und vieles mehr, was sie von den Individuen anderer Aggregateinheiten unterscheidet. Die sozialwissenschaftliche Forschung interessiert sich sehr oft explizit für diese Besonderheiten sozialer Einheiten (Schulen, Betriebe usw.) und die Differenzen zwischen den einzelnen Einheiten.

Aufmerksamkeit verdient in dieser Hinsicht besonders das Zusammenwirken von individuell-persönlichen und Aggregatfaktoren. Erziehungswissenschaftliche Fragestellungen sind z.B. häufig darauf ausgerichtet, den Einfluß von Aggregatbedingungen auf den individuellen Lernerfolg und -fortschritt zu ermitteln, so etwa in der Frage nach der Wirksamkeit unterschiedlicher Lehr- und Lernmethoden, der Qualität des Unterrichts oder der Qualität von Schulen, aber z.B. auch in der Frage nach den Wirkungen unterschiedlicher Schulsysteme (z.B. integriertes vs. gegliedertes Schulsystem). Hierbei stellt sich auch die Frage, ob Bedingungen des Unterrichts, der Schule oder des Schulsystems für alle Schüler gleichförmig oder verschiedenartig wirken - etwa in Abhängigkeit von Individualmerkmalen wie Geschlecht, Leistungsstand oder soziale Herkunft der Schüler. Ein anderes Beispiel: In einer organisationssoziologischen Studie soll geprüft werden, von welchen Faktoren die Häufigkeit von krankheitsbedingten Arbeitssausfällen abhängt. Relevant könnten hierbei

sowohl individuelle Faktoren sein (z.B. Alter, gesundheitlicher Zustand) als auch Merkmale der betrieblichen Situation (Betriebsklima, Unternehmenskultur). Möglicherweise ist außerdem die Wirkung von Faktoren der betrieblichen Situation für verschiedene Beschäftigtengruppen von unterschiedlicher Bedeutung.

In Untersuchungszusammenhängen wie den beispielhaft genannten und in zahlreichen Fragestellungen der Sozialwissenschaften sind somit drei Arten von Effekten von Interesse:

- Effekte von Individualvariablen (z.B. Geschlecht, Alter, sozialer Status)
- Effekte von Aggregatvariablen (Merkmale von Schulklassen, Schulen oder Betriebseinheiten, Betrieben)
- Das Zusammenwirken von Aggregat- und Individualvariablen.

Um dies auf ein häufig verwendetes Beispiel in diesem Band zu beziehen: In einer Schuluntersuchung soll der Leistungsstand von Schülern untersucht werden. Zur Erklärung der Schülerleistungen könnten eine Vielzahl von Variablen eine bedeutsame Rolle spielen. Aus der Gruppe der Individualmerkmale vor allem die kognitiven Fähigkeiten der Schüler, ihre soziale Herkunft, möglicherweise das Geschlecht und die Motivation. Aus der Gruppe der Aggregatmerkmale können zudem relevant sein: Merkmale der Lehrkraft oder des Unterrichts, z.B. des Unterrichtsstils, der Kompetenz der Lehrkraft oder auch die Sozialbeziehungen zwischen den Schülern. Daran sollte bereits deutlich werden, daß beide Aspekte, also die Merkmale der einzelnen Schüler und der Schulklassen (vor allem des Unterrichts) aufeinander bezogen zu analysieren sind und das Untersuchungsdesign daher auf eine Mehrebenenanalyse hin angelegt werden sollte. Durch eine Mehrebenenanalyse wird es außerdem möglich, auch Effekte der dritten Gruppe, nämlich Interaktionseffekte von Individual- und Aggregatmerkmalen, in einem dafür geeigneten Design zu testen: Es könnte z.B. sein, daß männliche und weibliche oder gute und schlechte Schüler oder Schüler unterschiedlicher sozialer Herkunft ganz unterschiedlich auf Unterrichtsfaktoren reagieren. Bestimmte Unterrichtsmethoden oder -bedingungen könnten für leistungsschwächere Schüler förderlich sein und ganz andere Unterrichtsmethoden für leistungsstärkere Schüler. Solche Fragestellungen sind mit hierarchischen Modellen prüfbar, deren Anwendung von daher in erster Linie *inhaltlich* angezeigt ist.

Mehrebenenanalysen spielen in den unterschiedlichsten sozialwissenschaftlichen Disziplinen eine Rolle. Was hierbei als *Individual-* und was als *Aggregateinheit* zu verstehen ist, kann sehr unterschiedlich sein. Der in Mehrebenenanalysen verwendete Begriff *Individualeinheit meint die individuelle Einheit der untersten Ordnung in einer Untersuchung*, die keineswegs eine einzelne Person sein muß. Individualeinheiten können ebenso Schulen und Aggregateinheiten z.B. die verschiedenen Schulbezirke sein, in denen diese Schulen liegen. Weitere Beispiele für Individual- und Aggregateinheiten sind: Schulbezirke in Bundesländern, Betriebe in Regionen oder auch einzelne Regionen

innerhalb der Länder. Kurz: Ein Mehrebenendesign kennzeichnet jedwede mögliche Form von hierarchisch geschachtelten Daten, in denen die Einheiten auf der ersten Ebene einer übergeordneten Ebene angehören. Der Begriff Individualeinheit bezeichnet die Einheit auf der niedrigsten Analyseebene, das niedrigste Aggregierungsniveau der Daten. Höhere Ebenen können alle sein, in denen sich die Einheiten einer untergeordneten Ebene zusammenfassen lassen, denen sie ganz natürlich - oder doch in sinnvoller Weise aufgrund von Gemeinsamkeiten - zugeordnet werden können.

Mehrebenenanalysen sind nicht auf zwei Ebenen, die im folgenden Text im Vordergrund stehen, begrenzt. Schon die heute verfügbaren Modelle und die Software dazu erlauben es, Analysen mit *mehr als zwei Ebenen* durchzuführen. Daß im folgenden das Zwei-Ebenen-Modell den größten Raum einnimmt, liegt daran, daß bereits damit die allgemeinen Grundlagen, Bedingungen und Möglichkeiten der Mehrebenenanalyse aufgezeigt werden können. Die Erweiterung für Modelle mit drei Analyseebenen wird im sechsten Kapitel vorgestellt.

Eine spezifische Anwendungsmöglichkeit der Mehrebenenanalyse ist die Analyse von *Längsschnittdaten*, die ebenfalls eine hierarchische Struktur aufweisen. Hier sind die Beobachtungen zu mehreren Zeitpunkten (als erste Ebene) den Individuen (als zweite Ebene) zuordenbar und möglicherweise werden zudem personenspezifische Umweltvariablen (dritte Ebene) als relevante Größen in die Untersuchung einbezogen. Prüfbar wäre z.B. wieweit die kognitive Entwicklung mit der sprachlichen kovariiert und inwieweit Merkmale

der Person oder des Kontextes auf den Entwicklungsverlauf wirken. Auf die Anwendung von Mehrebenenanalysen für Längsschnittdaten wird ebenfalls in Kapitel 6 eingegangen.

Ein weiterer, allerdings sehr spezifischer Anwendungsfall für ein Mehrebenendesign sind *Metaanalysen*, in denen Ergebnisse empirischer Untersuchungen als Fälle auf der ersten Ebene und Spezifika der Untersuchungsdesigns als Aggregatmerkmale betrachtet werden. Es könnte etwa geprüft werden, ob die Stichprobengröße von Bedeutung für die gefundenen Ergebnisse ist oder ob signifikante Differenzen in den Untersuchungsergebnissen im Vergleich zwischen ex-post-facto und experimentellen Untersuchungen bestehen. Aufgrund des eigenständigen Status von Metaanalysen wird darauf im folgenden nicht eingegangen (vgl. dazu: RAUDENBUSH und BRYK 1985).

Obwohl es eine breite Fülle hierarchischer Datenstrukturen und Anwendungsfälle in der sozialwissenschaftlichen Forschung gibt, werden Mehrebenendaten nicht immer mit adäquaten Verfahren analysiert. Dafür gibt es mehrere Gründe. Für die Analyse hierarchisch strukturierter Daten sind die traditionellen statistischen Verfahren wenig geeignet und somit entfällt die Möglichkeit der Datenanalyse mit den verbreiteten Software-Paketen weitgehend bzw. lassen sich korrekte Analysen nur durch eine trickreiche Überlistung der üblichen Verfahren realisieren. Erst seit ca. Mitte der 80er Jahre wurden ausreichend elaborierte Analysetechniken und handhabbare Software entwickelt, die nun erst eine größere Verbreitung erlangen. Mehrebenenanalysen stellen außerdem hohe An-

forderungen an das Untersuchungsdesign. Die Stichprobe muß eine genügend große Anzahl von Elementen auf jeder Analyseebene beinhalten; für eine Drei-Ebenen-Analyse im schulischen Bereich müßten z.B. Daten für genügend viele Schüler in genügend vielen Schulklassen in genügend vielen Schulen vorliegen. Das verlangt eine komplexe Anlage der Untersuchung und erfordert einen erheblichen Aufwand, nicht zuletzt auch in der Datenhaltung.

Daten mit einer hierarchischen Struktur zu analysieren bedeutet, daß die einbezogenen Variablen unterschiedlichen Aggregatniveaus angehören. In der Regel wird als abhängige Variable ein Individualmerkmal untersucht, zu dessen Erklärung bzw. Vorhersage weitere Individualmerkmale und zudem noch Aggregatmerkmale (z.B. der Unterricht in der Schulklasse, Merkmale der Schule, oder auch des Wohngebietes) herangezogen werden. Die Forderung dabei ist, der hierarchischen Datenstruktur auch in den angewandten Analyseverfahren gerecht zu werden.

Ebenso kann umgekehrt danach gefragt werden, welche Folgen es hat, wenn Mehrebenendaten mit ungeeigneten Verfahren analysiert werden? Diesbezüglich hat CRONBACH (1976) in einem bekannten und vielzitierten Aufsatz harsche Kritik an der Praxis der sozialwissenschaftlichen Forschung artikuliert: Sie habe durch die vorherrschende Ignorierung des Mehrebenencharakters ihrer Daten mehr an Zusammenhängen und bedeutsamen Beziehungen verdeckt als sie aufgedeckt habe. CRONBACH verweist auf das Problem, daß die vorherrschende Forschungspraxis durch die Verwendung von ungeeigneten Analyseverfahren für Mehrebenendaten nicht in der Lage ist, das Netzwerk der individuell-sozialen Bedingungen abzubilden. Die Verwendung unangemessener Methoden führe in zahlreichen Studien zu falschen Schlußfolgerungen, und zwar deshalb, weil die Wirkungen von Individual- und Aggregateinflüssen nicht oder statistisch unzureichend auseinandergehalten werden. Um diese Kritik nachvollziehen zu können, ist es sinnvoll, die üblichen Strategien der Analyse von Mehrebenendaten zu betrachten.

1.2 Probleme und Defizite traditioneller Analysestrategien

Wegen des Fehlens geeigneter statistischer Analyseverfahren für hierarchisch strukturierte Daten wurden in der Vergangenheit überwiegend traditionelle Verfahren (besonders Standard-Regressionsanalysen) eingesetzt, um Daten mit einer Mehrebenenstruktur zu analysieren. Daß dies zu gravierenden Fehlern führen kann, wurde schon erwähnt. Im folgenden sollen die einzelnen Aspekte genauer behandelt werden. Bezugspunkt ist ein Beispiel mit zwei Schulklassen, für die Daten zur sozialen Herkunft und zu den schulischen Leistungen der Schüler vorliegen. Ziel jeder Analyse sollte es natürlich sein, die Wirklichkeit bestmöglich abzubilden. Auf traditionellem Weg kann dies für die hier gewählten Daten nicht geleistet werden, vielmehr ist eine Mehrebenenanalyse erforderlich. Anhand des Beispiels sollen zugleich zentrale Aspekte der Mehrebenenanalyse verdeutlicht werden.

Ignorierung der Mehrebenenstruktur

Ein sehr verbreiteter Weg ist es, die Mehrebenenstruktur der Daten bei der Analyse schlicht zu ignorieren, d.h. so zu tun, als läge eine einfacher, nicht hierarchischer Datensatz vor. Die Unzulänglichkeiten, die daraus entstehen, sollen an dem genannten Beispiel aufgezeigt werden.

Zu untersuchen sei die Beziehung zwischen der sozialen Herkunft und den Schulleistungen der Schüler für einen Datensatz mit 40 Probanden aus zwei Schulklassen. Die für das Beispiel konstruierten Daten sind in Tabelle 1-1 am Ende des Kapitels angegeben.

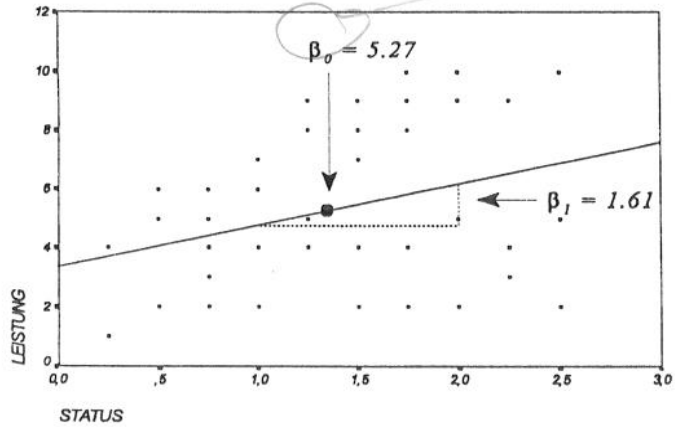

Abbildung 1-1: Beziehung zwischen sozialer Herkunft und Schulleistung in der Gesamtgruppe

Abbildung 1-1 zeigt die Beziehung beider Variablen für die Beispieldaten. In der Abbildung sind die 40 Wertepaare und die Regressionsgerade für die Gesamtgruppe zu sehen. Auf der Regressionsgeraden ist der Gesamtmittelwert (β_0=5.27) markiert und

außerdem ist die Steigung der Geraden (β_1=1.61) eingetragen. Die Korrelation zwischen sozialer Herkunft und Schulleistung beträgt für die Gesamtstichprobe .34 und ist auf dem 5%-Niveau signifikant (p=.0320, zweiseitiger Test). Zu diesem Ergebnis führt eine einfache Regressionsanalyse nach dem üblichen Modell:

$$Y_i = \beta_0 + \beta_1 X_i + r_i \qquad [1\text{-}1]$$

Das Ergebnis dieser einfachen Regressionsanalyse trifft die Wirklichkeit jedoch nicht, bzw. führt zu Informationen über den Datensatz, die zu falschen Schlußfolgerungen verleiten. Da es sich in dem Beispiel um die Daten aus *zwei Schulklassen* handelt, bleiben in der Analyse mit dem Gesamtdatensatz Fragen offen. Von Interesse ist vor allem, ob nicht bedeutsame Unterschiede zwischen den beiden Schulklassen bestehen. Hierbei gibt es mehrere Möglichkeiten.

Zum einen kann schon das Leistungsniveau in beiden Schulklassen unterschiedlich sein, die Schüler der einen Klasse könnten bedeutsam bessere Leistungen erzielt haben als die Schüler der anderen Klasse. Außerdem kann auch die Beziehung zwischen der sozialen Herkunft der Schüler und ihrem Leistungsstand in der einen Schulklasse enger und in der anderen weniger eng sein. Zu beiden Fragestellungen bietet die Gesamtanalyse keinerlei Informationen, vielmehr besteht die Gefahr, daß durch die gemeinsame Betrachtung wichtige Unterschiede verdeckt werden.

Damit kommen wir zu einer erneuten Betrachtung der einfachen Beispieldaten. In Abbildung 1-2 sind noch einmal die Wertepaare für die 40 Probanden eingetragen, vergleichbar wie in der Abbildung zuvor. Dabei sind aber nun die Probanden der beiden Schulklassen mit jeweils 20 Schülern durch unterschiedliche Symbole voneinander abgehoben. Schon auf den ersten Blick ist zu vermuten, daß sich die beiden Schulklassen sehr deutlich voneinander unterscheiden.

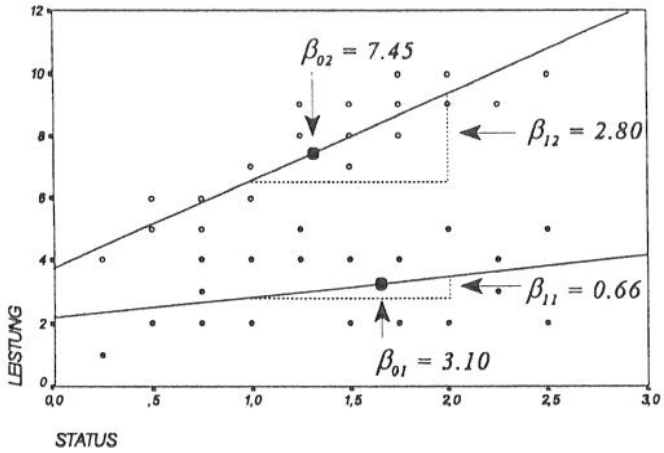

Abbildung 1-2: Beziehung zwischen sozialer Herkunft und Schulleistung getrennt für beide Gruppen

In der Abbildung ist zu erkennen, daß die Schüler aus der zweiten Schulklasse deutlich höhere Schulleistungen erzielen. Dieses Ergebnis läßt sich leicht durch einen Vergleich der Mittelwerte für die beiden Schulklassen bestätigen. Eine Varianzanalyse belegt, daß die Differenzen in den mittleren Schülerleistungen (3.10 vs. 7.45) signifikant und sehr bedeutsam sind.

Die Abbildung läßt aber noch einen weiteren Unterschied im Vergleich der beiden Schulklassen erkennen: In der ersten Schulklasse ist praktisch kein Zusammenhang zwischen dem sozialem Status der Schüler und ihren Schulleistungen festzustellen. Die durch den Punkteschwarm gezogene Regressionsgerade für die erste Schulklasse verläuft sehr flach, sie hat nur eine geringe Steigung (β_{11}=0.66). Anders sieht es in der zweiten Schulklasse aus. Der Zusammenhang zwischen sozialem Status und Schülerleistungen ist sehr ausgeprägt, wie die starke Steigung der Regressionsgeraden anzeigt (β_{12}=2.80).

Auf traditionellem Weg lassen sich diese Ergebnisse reproduzieren, indem statt einer Regressionsanalyse für die Gesamtstichprobe *zwei getrennte Regressionsanalysen für die beiden Schulklassen* gerechnet werden. Die Differenzen zwischen den beiden Schulklassen sind dann an den unterschiedlichen Regressionskoeffizienten aus beiden Analysen abzulesen: Die Differenzen in den mittleren Schülerleistungen zeigen sich in den voneinander abweichenden *Regressionskonstanten* (3.10 vs. 7.45), und die unterschiedliche Enge des Zusammenhanges zwischen sozialer Herkunft und Schülerleistungen ist an den unterschiedlichen *Steigungskoeffizienten* (0.66 vs. 2.80) abzulesen.

Getrennte Regressionsanalysen für beide Schulklassen wären ein erster Schritt, um nach Effekten der Aggregateinheit Schulklasse Ausschau zu halten - wenn auch, wie nachfolgend noch verdeutlicht werden soll, eine eher aufwendiger und nicht sehr zuverlässiger. Außerdem würde ein solches Vorgehen völlig unpraktikabel, wenn die Größe der

Stichprobe der Schulklassen zunimmt (z.B. 20 oder mehr). Immerhin würden aber getrennte Analysen Hinweise darauf geben, ob zwischen den Schulklassen relevante Unterschiede bestehen und es insofern unbefriedigend ist, bei dem Gesamtergebnis stehenzubleiben. Weiter zu untersuchen wäre dann immer noch die Frage, wodurch die Unterschiede erklärt werden können.

Denkbar wäre z.B., daß die Unterschiede im Leistungsniveau durch unterschiedliche kognitive Fähigkeiten der Schüler in beiden Schulklassen bedingt sind oder durch Unterschiede in deren Motivation. Die schulklassenspezifischen Effekte der sozialen Herkunft könnten sich verändern, wenn differenziertere Informationen zur häuslichen Umwelt der Schüler einbezogen werden. Solche Untersuchungen, die sich auf *individuell-spezifische* Erklärungsfaktoren beziehen, sind ohne Frage höchst bedeutsam.

Darüber hinaus könnten aber auch Merkmale der Schulklassen selbst, z.B. des Unterrichts oder der Lehrpersonen, mit verantwortlich für die Unterschiede zwischen den Schulklassen sein. Z.B. könnte der Unterricht in der zweiten Schulklasse besser sein als in der ersten. Bezogen auf die unterschiedliche Abhängigkeit von der sozialen Herkunft könnte es sein, daß in der einen Klasse durch Vorurteile der Lehrer Kinder der oberen Schichten bevorzugt und Kinder der unteren Schichten benachteiligt werden, während dies in der anderen Klasse nicht der Fall ist. Untersuchenswert sind insofern auch *Merkmale der Aggregateinheiten selbst*.

Damit gelangt man zu einer Fragestellung, die eine Mehrebenenanalyse erfordert und lauten kann: Gibt es über den Einfluß individueller Faktoren hinaus Faktoren der Bedingungen und Prozesse in den Schulklassen, die zu den Unterschieden im Leistungsniveau und zu einer größeren oder geringeren sozialen Selektivität beitragen?

Dieses einfache Beispiel soll einführend genügen, um aufzuzeigen, daß Analysen für hierarchisch strukturierte Daten auf nur einer Analyseebene zumindest unbefriedigend sind und oftmals irreführend bis hin zu grundlegend falsch sein können. Bei hierarchisch strukturierten Datensätzen müssen zumindest die folgenden möglichen Differenzen überprüft und ggf. in der Analyse berücksichtigt werden:

- Differenzen in den Mittelwerten (Regressionskonstanten)
- Differenzen in den Beziehungen zwischen den Variablen innerhalb der Aggregateinheiten (Regressionssteigungen)

Nur wenn in beiderlei Hinsicht keine signifikanten Differenzen vorliegen, ist es zulässig, die Daten der Einheiten gemeinsam zu analysieren, und umgekehrt: Sofern Differenzen bestehen, müssen diese in der Analyse berücksichtigt werden.

Die inzwischen verfügbaren Modelle und Programmpakete zur Mehrebenenanalyse erlauben es, die genannten Anforderungen zu erfüllen. In einem komplexen Design können Differenzen in den Regressionskoeffizienten zwischen Aggregateinheiten

ermittelt und außerdem Effekte von Aggregatmerkmalen auf die Regressionskoeffizienten innerhalb der Aggregateinheiten überprüft werden.

Ein gutes Beispiel für den Gewinn durch solche differenzierte Studien geben Analysen von BRYK und RAUDENBUSH mit den Daten aus der *High School and Beyond*-Studie von COLEMAN, HOFFER und KILGORE (1982; vgl. auch Kap. 5). COLEMAN u.a. hatten festgestellt, daß sich Schulen nach ihrem Leistungsniveau und ebenso hinsichtlich der Enge der Beziehung zwischen sozialer Herkunft und Schülerleistungen unterscheiden, in der Art, wie es im hier verwendeten vereinfachten Beispiel gezeigt wurde. Als bedeutsames Aggregatmerkmal stellte sich in den Untersuchungen der Schulträger heraus. Die katholischen Schulen in der US-Stichprobe erzielten ein höheres Leistungsniveau und waren zudem egalitärer als die Schulen in öffentlicher Trägerschaft. Dieses Ergebnis fordert unmittelbar zu der weiteren Frage heraus, warum das so ist und wie diese Effekte erklärt werden können. Die Katholizität des Schulträgers kann sicher nicht als eigenständig sinnvolles Erklärungsmerkmal gelten. In weiterführenden Analysen konnte aber gezeigt werden, daß Unterschiede in der Schul- und Unterrichtsorganisation und im sozial-normativen Umfeld der Schulen als Ursachen für die gefundenen Differenzen anzusehen waren. Die Anwendung von Verfahren der Mehrebenenanalyse hat somit zur Klärung inhaltlich relevanter Beziehungen beigetragen. Zugleich geben diese Studien anschauliche Belege für die Überlegenheit der Mehrebenenanalyse bei der Untersuchung komplexer Beziehungen in hierarchisch strukturierten Daten

gegenüber traditionellen Analysestrategien (vgl. z.B. BRYK und RAUDENBUSH 1989).

Den Kern der Kritik einer Vernachlässigung der Mehrebenenstruktur hat bereits CRONBACH (1976) formuliert: Die Koeffizienten des einfachen Regressionsmodells können mit CRONBACH als eine *uninterpretierbare Mischung aus Effekten innerhalb und zwischen den Aggregateinheiten* bezeichnet werden. In Kapitel 3.2 wird darauf im einzelnen eingegangen. Im Beispiel verdecken die Ergebnisse einer einfachen Regressionsanalyse über die Gesamtstichprobe die notwendige Unterscheidung nach Bedingungen und Prozessen im Vergleich *zwischen* den Schulen und *innerhalb* der Schulen. Sie sind somit für eine angemessene Erklärung der Beziehungen unbrauchbar.

Soweit läßt sich vorerst festhalten, daß Mehrebenenanalysen einen wichtigen Beitrag leisten können, um komplexen Beziehungsstrukturen gerecht zu werden. Eine Vielzahl von Fragestellungen in den Sozialwissenschaften werden durch Mehrebenenanalysen einer differenzierten Überprüfung zugänglich, indem Individual- und Aggregateffekte gemeinsam untersucht und in Beziehung zueinander geprüft werden. Die inzwischen verfügbaren Verfahren bieten die Grundlage, um den Besonderheiten der Datenstruktur im Hinblick auf die inhaltlichen und methodisch-statistischen Anforderungen gerecht zu werden.

Obwohl es sich damit im Grunde heute verbietet, hierarchisch strukturierte Daten mit nicht adäquaten Verfahren zu analysieren oder den Mehrebenencharakter der Daten zu ignorieren, ist ein Blick

auf die traditionell verwendeten Verfahren, mit denen Mehrebenendaten analysiert wurden, aufschlußreich. Vor allem lassen sich dabei die Besonderheiten von Mehrebenenanalysen weiter herausarbeiten und der Weg hin zu den aktuell ausgearbeiteten Verfahren kann nachgezeichnet werden.

Im Gegensatz zur gerade besprochenen Ignorierung des Mehrebenencharakters der Daten wurden in Analysen, bei denen versucht wurde, den hierarchischen Charakter der Daten zu beachten, üblicherweise zwei Wege beschritten (vgl. KEEVES und SELLIN 1990; KREFT und DE LEEUW 1991). Obwohl beide Wege ganz unterschiedlich sind, haben sie doch gemeinsam, daß die Mehrebenenstruktur nicht ignoriert, sondern vielmehr aufgehoben wird und die Daten so analysiert werden, als wären sie einer Ebene zuordenbar. Von daher erklären sich auch die langanhaltenden, im Kern aber unfruchtbaren Debatten dazu, was die "richtige" Analyseebene für Mehrebenendaten sei (die Individual- oder die Aggregatebene). Aus heutiger Sicht ist keine Ebene die "richtige", vielmehr ist es das Kennzeichen einer Mehrebenenanalyse, daß Daten von Einheiten, die unterschiedlichen Ebenen angehören, in eine simultane Analyse eingehen.

Aggregierung der Individualdaten

Bei der Aggregierung der Individualdaten werden (z.B. mit der SPSS-Prozedur *Aggregate*) aus den Daten der individuellen Einheiten zusammenfassende Statistiken erzeugt, die dann mit vorliegenden Aggregatdaten zusammengefügt werden kön-

nen. Es wird damit so getan, als würden nur Daten auf der Aggregatebene vorliegen.

Diese Analysestrategie ist in mehrfacher Hinsicht unergiebig. Vor allem ist problematisch, daß durch die Aggregierung der Daten (zu)viel an Information verlorengeht. Natürlich sind nach der Aggregierung nur noch Aussagen auf der Gruppenebene möglich, nicht mehr auf der Individualebene. Durch die Aggregierung gehen die Varianzen *innerhalb* der Aggregateinheiten verloren. In aller Regel ergeben sich denn auch in Analysen auf der Aggregatebene Ergebnisse, die von denen aus Individualanalysen abweichen, oftmals ganz erheblich (*aggregation-bias*, vgl. z.B. SELLIN 1990). Die Aggregierung der Individualdaten bewirkt, daß die Aggregate als in sich *homogen* betrachtet werden, was eine höchst fragwürdige Annahme ist. Eine Aggregierung wäre auch allenfalls dann sinnvoll, wenn die Varianzen der Variablen und die Beziehungen zwischen den Variablen innerhalb der Aggregateinheiten gleich sind. Andernfalls werden bedeutsame Unterschiede datentechnisch neutralisiert. Die Homogenität der Varianzen und Kovarianzen ist sicherlich nicht immer gegeben, und sie wird oftmals gar nicht überprüft.

Im Hinblick auf die Fragestellungen der Untersuchung macht es die Aggregierung unmöglich, das Zusammenwirken von Individual- und Aggregatmerkmalen zu untersuchen. Dieser oftmals besonders relevante Analyseaspekt wird durch die künstliche Vereinheitlichung der Aggregateinheiten ausgeschlossen. Nach der Aggregierung bleiben zwischen den Aggregaten nur Mittelwertdifferenzen der Variablen erhalten, aber keine Differenzen der

vidualebenenmodell daher auch als unbrauchbar ab (vgl. dazu BURSTEIN 1986). Es widerspricht der Vorstellung, den Einfluß von Individual- und Aggregatbedingungen sowie deren mögliche Wechselwirkungen zu analysieren und auseinanderzuhalten.

1.3 Modelle mit Zufallskoeffizienten als Verfahren der Mehrebenenanalyse

Mehrebenenanalysen hatten in den 80er Jahren eine gewisse Hochkonjunktur. Initiiert durch die Diskussion in den USA folgten in Deutschland - mit der üblichen zeitlichen Verzögerung- eine ganze Reihe von methodischen und anwendungsbezogenen Arbeiten (zusammenfassend: von SALDERN 1986). Dies führte jedoch keineswegs zu einer breiteren Anwendung von Mehrebenenanalysen, was nicht zuletzt auf den weitaus höheren Aufwand bei der Durchführung von Mehrebenenanalysen zurückzuführen sein dürfte. Daneben spielt sicherlich auch eine wichtige Rolle, daß wenig Einigkeit über die angemessenen statistischen Analyseverfahren bestand und explizit auf die Belange von Mehrebenenanalysen zugeschnittene Software nicht verfügbar war. Die teilweise ausgearbeiteten Vorschläge zur Analyse von Mehrebenendaten mit Standard-Statistikprogrammen waren schwer nachvollziehbar und erforderten eine außergewöhnlich aufwendige und kaum noch überschaubare Datenorganisation - bei einer weiterhin fragwürdigen Verläßlichkeit der Ergebnisse.

Ein wichtiger Vorschlag zur Analyse von Mehrebenendaten und grundlegender Vorstoß zu der Weiterentwicklung der Verfahren, ist unter dem Titel *slopes as outcomes* bzw. *systematic varying slopes* (*SVS*) bekannt geworden und geht auf Überlegungen von BURSTEIN und seinen Mitarbeitern zurück (BURSTEIN, LINN und CAPELL 1978). Dieser Vorstoß ist vor allem deshalb so bedeutsam, weil er eine genuin erziehungswissenschaftliche Sichtweise aufgreift und in eine Analysestrategie umzusetzen sucht. Die Idee ist, daß unterschiedliche Regressionsgeraden (besonders Unterschiede in den Steigungen / *slopes*) im Vergleich zwischen Schulklassen nicht als störende Effekte zu behandeln sind, sondern Ausdruck von Unterschieden in den ablaufenden Lernprozessen und/oder Lernbedingungen sein können. Die Varianz in den Regressionskoeffizienten wird damit als systematisch betrachtet und die Koeffizienten der Regressionen innerhalb der Schulklassen werden als abhängige Variablen (*outcomes*) behandelt. Die Analysestrategie ist damit folgende: In einem ersten Schritt werden die Beziehungen zwischen den Variablen innerhalb jeder Aggregateinheit (Schule bzw. Schulklasse) analysiert. Im zweiten Schritt wird dann versucht, die Varianz der Regressionskoeffizienten durch Merkmale der Aggregateinheiten zu erklären, d.h., die Koeffizienten werden als eine Funktion von Aggregatmerkmalen untersucht. In diesem Vorschlag von BURSTEIN u.a. stehen *Interaktionseffekte* im Mittelpunkt des Interesses, denn der Nachweis variierender Koeffizienten in Abhängigkeit von Aggregatmerkmalen bedeutet nichts anderes als eine Interaktion zwischen Individual- und Aggregatmerkmalen.

"Technically, this involves examination of the fixed effect estimates of macro-level variables on micro-level parameters. Practically speaking, then, we are still interested in 'slopes as outcomes', which implies different models for different classrooms and schools ... Regardless, the intent is to identify the antecedents of student performance and attitude and estimate the magnitude of their effects" (BURSTEIN u.a. 1989, S. 237).

Ohne Frage bedeutet die Idee der *systematic varying slopes* eine wichtige und auch unter theoretischem Aspekt höchst bedeutsame Weiterentwicklung mehrebenenanalytischer Verfahren. Der Umsetzung in ein angemessenes statistisches Analyseverfahren standen jedoch bislang eine Reihe gravierender Probleme entgegen (vgl. RAUDENBUSH und BRYK 1986; BRYK und RAUDENBUSH 1989; BRYK, RAUDENBUSH, SELTZER und CONGDON 1989; BRYK und RAUDENBUSH 1992; BURSTEIN, KIM und DELANDSHERE 1989):

- Die Koeffizienten der Regressionen innerhalb der Aggregateinheiten basieren in aller Regel auf einer geringen Stichprobengröße. Sie werden mit großem Stichprobenfehler geschätzt und ihre Reliabilität ist daher vergleichsweise gering. Schon wenige Ausreißer können eine erhebliche Pseudo-Varianz der geschätzten Koeffizienten bewirken (*bouncing betas*). Damit ist auch die Erklärung der vergleichsweise wenig reliablen Varianzen der Koeffizienten in einem zweiten Schritt problematisch. Verläßlich anwendbare Analyseverfahren müssen diesem Aspekt Rechnung tragen, indem die Schätzwerte der Koeffizienten (und deren Vertrauensintervalle) die Stichprobengrößen, die zudem sehr

häufig für die Aggregateinheiten unterschiedlich sind, in Rechnung stellen.

- Die Präzision der Innerhalb-Koeffizienten variiert somit in der Regel zwischen den Aggregateinheiten. Ein effizientes statistisches Schätzverfahren muß daher die Koeffizienten proportional zu ihrer (geschätzten) Präzision gewichten. Unterbleibt dies, so wirkt sich dies negativ auf die Analysen zur Erklärung der Differenzen aus.

- Die Variabilität der Koeffizienten besteht aus zwei Komponenten: Einerseits aus tatsächlichen Differenzen zwischen den Aggregateinheiten, dies ist die *Parametervarianz*. Zum anderen aber auch aus der genannten *Stichprobenvarianz* (*sampling variance*). Natürlich ist nur die Parametervarianz erklärbar, wohingegen die zweite Komponente durch Stichprobenbedingungen verursacht wird.

- Die Standard-Regression geht von der Unabhängigkeit der Beobachtungen aus; es werden unkorrelierte Fehlervarianzen angenommen. Aufgrund der üblichen und inhaltlich durchaus angebrachten Ziehung von Clustern statt individueller Einheiten im Stichprobenplan wird diese Bedingung in der Regel nicht erfüllt sein. Ein korrektes Analyseverfahren muß daher Fehlerstrukturen mit gruppenspezifischen Komponenten beinhalten.

- *Slopes as Outcomes* war bislang - schon wegen der Komplexität des Analyseverfahrens und des erforderlichen Datenmanagements - auf die

Analyse *einer* Innerhalb-Variablen begrenzt. Das widerspricht der Notwendigkeit, selbst wenn man nur an der Wirkung einer Variablen interessiert ist, den Effekt weiterer Variablen zu kontrollieren. In einem brauchbaren Analyseverfahren müssen konfundierende Variablen auf zwei Ebenen kontrolliert werden können, was ein statistisches Modell erfordert, das dieser komplexen Kovarianzstruktur angepaßt ist.

Obwohl die Idee der *systematic varying slopes* schon ursprünglich einen vielversprechenden Ansatz darstellte, erwies sich die Umsetzung in ein statistisches Analyseverfahren und die Entwicklung geeigneter Software als problematisch. Die Versuche, entsprechende Analysen über Standard-Verfahren zu realisieren, erwiesen sich als fragwürdig bzw. vom erforderlichen Ausführungsaufwand her als nicht vertretbar bis unpraktikabel. BURSTEIN u.a. kommentieren die frühen Versuche einer Umsetzung daher auch skeptisch:

"In summary, as applied in Burstein et.al., the slopes as outcomes approach is no basis for statistical inferences about the parameters from the general SVS model ... This approach can not be recommended except at perhaps exploratory stages as a means for describing potentially interesting associations of slopes with macro-level characteristics" (BURSTEIN, KIM und DELANDSHERE 1989, S. 252).

Inzwischen wurden jedoch in der Entwicklung mehrebenenanalytischer Analyseverfahren bedeutende Fortschritte erzielt (zusammenfassend BOCK 1989; CHEUNG, KEEVES, SELLIN und TSOI 1990), was bislang innerhalb der Bundesrepublik leider nur vereinzelt zur Kenntnis genommen wurde. Die

Weiter- bzw. Neuentwicklungen lassen sich unter dem Stichwort *Random Coefficient Models (RCM)* zusammenfassen (vgl. DE LEEUW und KREFT 1986). Damit liegen nunmehr anwendbare Analysemodelle vor, für die auch geeignete Software verfügbar ist. Den Modellen ist gemeinsam, daß sie auf der Idee der *systematic varying slopes* basieren und Verfahren entwickelt haben, um die methodischen und technischen Schwächen der frühen Ansätze zu überwinden. In allen Modellansätzen werden komplexe Daten- und Fehlerstrukturen berücksichtigt, und außerdem ist die Modellierung von Regressionskoeffizienten in Abhängigkeit von Faktoren übergeordneter Aggregatebenen möglich.

Im folgenden Text werden die Grundlagen und Anwendungen des *Hierarchical Linear Modelling* (HLM) von BRYK und RAUDENBUSH vorgestellt (BRYK und RAUDENBUSH 1989; BRYK, RAUDENBUSH, SELTZER und CONGDON 1989; RAUDENBUSH und BRYK 1986, 1989; RAUDENBUSH 1988; BRYK und RAUDENBUSH 1992). HLM ist der gegenwärtig vermutlich am weitesten verbreitete Modellansatz zur Mehrebenenanalyse, neben ML/3 (GOLDSTEIN 1987, 1989, 1995; GOLDSTEIN und MC DONALD 1988) und VARCL (AITKIN und LONGFORD 1986; LONGFORD 1986, 1989) - sowie einigen spezifischen und weniger bekannten Programmen zur Analyse von Mehrebenendaten.

Trotz der Gemeinsamkeiten der drei Ansätze und der entsprechend benannten Software-Pakete ist die Notation und Vorgehensweise in allen drei Verfahren unterschiedlich. Für den vorliegenden Text wurde der Bezug auf HLM zum einen wegen der weiten Verbreitung gewählt und zum anderen auch

deshalb, weil die Modellnotation einen gut nachvollziehbaren Überblick der Ebenenbeziehungen erlaubt

Mit dem Bezug auf HLM ist in keiner Weise eine Abwertung von ML/3 oder VARCL intendiert. Jeder der Ansätze und jedes der Software-Pakete hat seine spezifischen Vorzüge. So liegt für VARCL eine Programmvariante vor, die Analysen mit bis zu neun Ebenen ermöglicht. Eine spezifische Stärke von ML/3 ist ein umfangreicher Programmteil zur grafischen Darstellung der Daten. Aufgrund der - wenn auch nicht schon auf den ersten Blick erkennbaren - grundsätzlichen Gemeinsamkeiten der drei Ansätze und Programmpakete sollte es überdies möglich sein, sich aufgrund der Kenntnis des HLM-Ansatzes in den anderen Modellen bzw. Programmen zurecht zu finden.

Tab. 1.1: Daten für das Beispiel in Kapitel 1.2

Nr.	Status	Leistung	Nr.	Status	Leistung
1	,25	1,00	2	,25	4,00
1	,50	2,00	2	,50	5,00
1	,50	2,00	2	,50	6,00
1	,75	3,00	2	,75	4,00
1	,75	2,00	2	,75	5,00
1	,75	4,00	2	,75	6,00
1	1,00	2,00	2	1,00	6,00
1	1,00	4,00	2	1,00	7,00
1	1,25	4,00	2	1,25	8,00
1	1,25	5,00	2	1,25	9,00
1	1,50	4,00	2	1,50	7,00
1	1,50	2,00	2	1,50	8,00
1	1,75	2,00	2	1,50	9,00
1	1,75	4,00	2	1,75	8,00
1	2,00	5,00	2	1,75	9,00
1	2,00	2,00	2	1,75	10,00
1	2,25	3,00	2	2,00	9,00
1	2,25	4,00	2	2,00	10,00
1	2,50	2,00	2	2,25	9,00
1	2,50	5,00	2	2,50	10,00

2 Grundlagen und Anwendungen des Hierarchisch Linearen Modells (HLM)

In diesem Kapitel werden die Grundlagen des Hierarchisch Linearen Modells von BRYK und RAUDENBUSH (1992) vorgestellt. Im Vordergrund steht nicht die Herleitung der statistisch-mathematischen Modellbeziehungen, sondern die Darstellung des Modells mit seinen vielfältigen Varianten und Anwendungsmöglichkeiten. Eine differenzierte Herleitung der Modellgrundlagen geben BRYK und RAUDENBUSH (1992). Die Ausführungen in diesem Kapitel sind auf den *Zwei-Ebenen-Fall* bezogen, was zur allgemeinen Grundlegung ausreichend ist. Anwendungen für mehr als zwei Ebenen sind als Ergänzung zu werten und aus der Logik des Zwei-Ebenen-Falles herleitbar. Eine Darstellung für den Drei-Ebenen-Fall gibt Kapitel 6.2.

2.1 Grundlagen des Hierarchisch Linearen Modells

Als Ausgangsbasis für das Hierarchisch Lineare Modell kann die übliche Regressionsgleichung angesehen werden, mit der in diesem Abschnitt begonnen wird. Der Bezugspunkt für die Darstellung

ist weiterhin das Beispiel aus dem ersten Kapitel zum Zusammenhang zwischen sozialer Herkunft und Schulleistung. Ausgegangen wird von der Beziehung der Variablen innerhalb einer Aggregateinheit, aus der dann die Verallgemeinerung für den Fall einer Stichprobe mit einer größeren Zahl von Aggregateinheiten vorgenommen wird. Schließlich interessiert besonders der Fall, daß Unterschiede in den Regressionskoeffizienten für eine größere Zahl von Aggregateinheiten durch deren spezifische Merkmale erklärt werden sollen.

2.1.1 Beziehungen in einer Aggregateinheit

Die Regressionsgleichung für die Beziehung zweier Variablen (im Beispiel: x, sozialer Status; y, Schulleistung) hat die bekannte Form:

$$Y_i = \beta_0 + \beta_1 X_i + r_i \qquad [2\text{-}1]$$

Die Schulleistung (y) eines Schülers (i) wird in der Gleichung in Abhängigkeit von der sozialen Herkunft (x) betrachtet. Die Koeffizienten des Modells haben eine entsprechende Bedeutung wie in jeder Standard-Regressionsanalyse:

β_0 Regressionskonstante, *intercept*: Entsprechend der Terminologie der Regression wird β_0 als die Regressionskonstante bzw. als *intercept* bezeichnet.

Sofern die Daten, wie in der obigen Gleichung, in der ursprünglichen Metrik analysiert werden, gibt

β_0 den erwarteten Wert in der abhängigen Variablen für den Fall an, daß die Prädiktorvariable den Wert Null annimmt. Im Beispiel wäre es die zu erwartende Leistung für einen Schüler mit dem Wert Null in der Variablen sozialer Status. Sofern dieser Bezugspunkt nicht sinnvoll interpretiert werden kann, ist eine andere Metrik der Prädiktorvariablen zu wählen (vgl. Kap. 3.1).

β_1 Regressionssteigung, *slope*: Sie gibt die erwartete Veränderung in Y an, wenn sich die Ausprägung von X um eine Einheit verändert.

Im Beispiel bezieht sich β_1 z.B. auf die Differenz in den Schulleistungen zwischen Schülern mit den Werten "1" und "2" in der Variable sozialer Status.

r_i Residuum oder Fehlerterm.

Das Residuum gibt die spezifische Abweichung der Individualwerte von der Regressionsgeraden an.

Das Regressionsmodell geht von der Annahme aus, daß die Residuen (r_i) normalverteilt sind, mit einem Mittelwert von Null und der Varianz σ^2, d.h., die Regressionsgerade ist so durch die Punktepaare gelegt, daß die Abweichungen der einzelnen Punkte von der Geraden im Mittel Null ergeben, bei einer Varianz von σ^2:

$$r_i \sim N(0, \sigma^2)$$

Eine wichtige Frage im Regressionsmodell betrifft die Regressionskonstante (β_0) und bezieht sich darauf, ob dieser Koeffizient interpretierbar ist. Oftmals ist das, wenn die Daten in der ursprünglichen

Metrik analysiert werden, nicht der Fall. Im hier verwendeten Beispiel macht die Angabe eines Koeffizienten für den Fall, daß der soziale Status eines Individuums den Wert Null annimmt, keinen Sinn, weil ein solcher Wert üblicherweise nicht definiert ist. In solchen Fällen ist es sinnvoll, eine Zentrierung (vgl. Kap. 3.1) vorzunehmen, um interpretierbare Koeffizienten für β_0 zu erhalten. Die Zentrierung bedeutet, daß Abweichungswerte der Individualwerte gebildet werden, z.B. als individuelle Abweichungswerte vom Gesamtmittelwert. Die Regressionskonstante β_0 gibt im Fall einer solchen Zentrierung den Leistungswert für Probanden mit einem mittleren sozialen Status an, und die Abweichungswerte werden in folgender Form gebildet:

$$X_i - \bar{X} \qquad [2\text{-}2]$$

Bezogen auf die Beispieldaten des ersten Kapitels könnte nun zu einer Analyse mit einer Stichprobe von zwei Aggregateinheiten weitergegangen werden, um deren Regressionskoeffizienten zu vergleichen. Allerdings ist eine Stichprobe von nur zwei Aggregateinheiten in aller Regel nicht sinnvoll und ausreichend, um zu gesicherten und verallgemeinerungsfähigen Aussagen über die vorliegenden Beziehungen zu kommen. Vergleichende Analysen für eine hinreichend große Zahl von Aggregateinheiten (z.B. Schulen) sind demgegenüber weitaus bedeutsamer und werden im nächsten Abschnitt behandelt.

2.1.2 Beziehungen in einer Stichprobe von Aggregateinheiten

Die Spezifik einer Mehrebenenanalyse wird erst dann deutlich, wenn eine größere Stichprobe von Aggregateinheiten aus der Population aller Aggregateinheiten analysiert werden soll; wenn etwa die Daten von Schülern aus einer größeren Zahl von Schulen oder Schulklassen zu untersuchen sind. Zur Darstellung der Modellbeziehungen wird dann ein weiterer Index (j= 1,...J) benötigt, der die Zugehörigkeit der individuellen Einheiten zu den Aggregateinheiten angibt. Die ergänzte *Modellgleichung für die Individualebene* (erste Ebene) lautet dann:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{.j}) + r_{ij} \qquad [2\text{-}3]$$

In der Gleichung wird von einer Zentrierung der Prädiktorvariablen (vgl. Kap. 3.1) ausgegangen, wobei im Fall einer größeren Anzahl von Aggregateinheiten zwei Möglichkeiten der Zentrierung gegeben sind: Zum einen die in der Gleichung oben angegebene Abweichung der individuellen Werte von den *Gruppenmittelwerten* der Aggregateinheiten, denen die Individuen angehören. Zum anderen könnten die individuellen Abweichungen vom *Gesamtmittelwert* verwendet werden. In der Mehrzahl der Fälle liefert die Zentrierung um den Gruppenmittelwert das intendierte Ergebnis für den Effekt der Prädiktorvariablen (siehe Kap. 3).

Angenommen wird wiederum, daß die Residuen (r_{ij}) normalverteilt sind mit einem Mittelwert von Null und gleicher Varianz σ^2 in den Aggregateinheiten:

$$r_{ij} \sim N(0,\sigma^2)$$

Gegenüber der einfachen Regression in der Gesamtstichprobe resultieren nun als Ergebnis eine Vielzahl (J) von Regressionskonstanten und Steigungskoeffizienten. Das Modell geht von einer spezifischen Regressionskonstanten und einem spezifischen Steigungskoeffizienten für jede Aggregateinheit aus. Im Beispiel wird für jede Schule der für Sie spezifische Mittelwert der Schülerleistungen (Regressionskonstante) und der für die Schule spezifische Steigungskoeffizient für die Enge des Zusammenhanges zwischen dem sozialem Status und der Schulleistung ermittelt. Von daher ist es weiter erforderlich, Bezeichnungen für die Verteilungskennwerte dieser Koeffizienten zu vereinbaren. Im einzelnen werden Bezeichnungen für die Mittelwerte und die Varianzen bzw. Kovarianzen der Koeffizienten benötigt:

$$E(\beta_{0j}) = \gamma_0, \quad Var(\beta_{0j}) = \tau_{00}$$

$$E(\beta_{1j}) = \gamma_1, \quad Var(\beta_{1j}) = \tau_{11}$$

$$Cov(\beta_{0j}, \beta_{1j}) = \tau_{01}$$

Diese Koeffizienten haben folgende Bedeutung:

γ_0 Schulmittelwert für die Population der Schulen

γ_1 mittlerer Steigungskoeffizient über alle Schulen

τ_{00} Populationsvarianz der Schulmittelwerte

τ_{11} Populationsvarianz der Steigungskoeffizienten

τ_{01} Kovarianz zwischen den Mittelwerten und den Steigungskoeffizienten

Den hier aufgeführten Koeffizienten, bzw. - bei einer Erweiterung um mehrere Prädiktoren - den Mittelwert-Vektoren der β-Koeffizienten und deren Varianz-Kovarianz-Matrizen kommt im Hierarchisch Linearen Modell eine besondere Bedeutung zu. Dabei sind die wahren Populationsparameter nicht bekannt, sie müssen vielmehr auf der Basis der verfügbaren empirischen Daten mit Hilfe eines iterativen Verfahrens geschätzt werden.

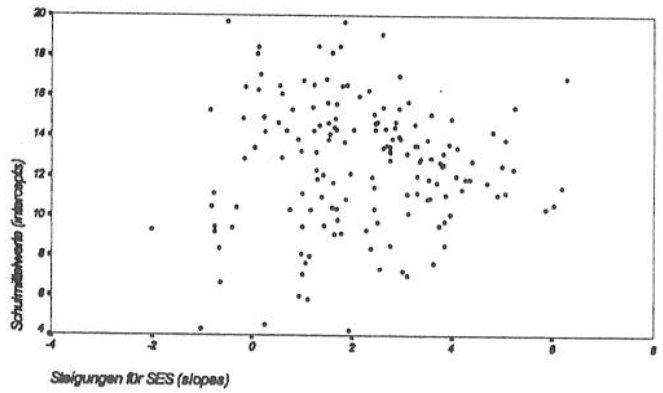

Abbildung 2-1: Verteilung der Intercepts und Slopes in 160 Schulen der *High School and Beyond*-Studie

Zur Veranschaulichung einer Verteilung von Regressionskoeffizienten zeigt Abbildung 2-1 die Re-

gressionskonstanten und Steigungskoeffizienten einer Regression der sozialen Herkunft (SES) auf die Schülerleistungen für den Datensatz der *High School and Beyond*-Studie (COLEMAN, HOFFER und KILGORE 1982; vgl. auch Kap. 5). Basierend auf einer Stichprobe von 160 Schulen lassen sich 160 Koeffizientenpaare ermitteln, die in der Abbildung als Scatterplot eingetragen sind.

Anhand dieser Abbildung läßt sich die Spezifik der Mehrebenenanalyse gegenüber dem üblicherweise verwendeten einfachen Regressionsmodell verdeutlichen: Im einfachen Regressionsmodell würde von einer einheitlichen Regressionskonstanten und einem einheitlichen Steigungskoeffizienten für alle 160 Schulen ausgegangen (es gäbe nur ein β_0 und nur ein β_1). Damit würde angenommen bzw. vorausgesetzt, daß beide Koeffizienten für alle Aggregateinheiten gleich sind. Wenn nun, wie in dem Beispiel, diese Koeffizienten zwischen den Schulen signifikant variieren, ist das einfache Regressionsmodell nicht angemessen und der Fehler bestünde darin, die Differenzen der Regressionskoeffizienten zwischen den Aggregateinheiten zu ignorieren. Folglich würde von nicht zutreffenden Annahmen ausgegangen und die tatsächlichen Beziehungen würden verfälscht wiedergegeben. Bezogen auf den Untersuchungsinhalt würde außer Acht gelassen, daß die mittleren Schülerleistungen zwischen den Schulen bedeutsam variieren und ebenso würde die unterschiedliche soziale Selektivität der Schulen nicht berücksichtigt.

Mit den in diesem Abschnitt eingeführten Modellbeziehungen ist soweit nur prüfbar, ob zwischen den Aggregateinheiten Differenzen in den Regressionskoeffizienten bestehen. Sofern dies der Fall ist, kann aber darüber hinaus nach Erklärungen für die Unterschiede in den Koeffizienten gesucht werden. Auf diese Fragestellung beziehen sich die folgenden Ausführungen.

2.1.3 Erklärung von variierenden Regressionskoeffizienten durch Aggregatmerkmale

Bezogen auf das oben angeführte Beispiel aus der *High School and Beyond*-Studie zeigen die Verteilungen der Regressionskonstanten und -steigungen folgendes: Zum einen gibt es hinsichtlich der Schülerleistungen bessere und schlechtere Schulen - dies zeigen die Differenzen der Regressionskonstanten für die Schulen. Außerdem sind die Schulen unterschiedlich sozial selektiv - das zeigen die Differenzen der Steigungskoeffizienten. Auf diesem Hintergrund kann nun versucht werden, die unterschiedliche Effektivität und soziale Selektivität der Schulen vorherzusagen, und zwar aufgrund von Merkmalen der Schulen (oder natürlich auch aufgrund von Merkmalen des Unterrichts bzw. der Lehrer in den Schulen bzw. allgemein: durch Merkmale der Aggregateinheiten).

Voraussetzung, um mögliche Erklärungen zu finden, ist, daß im Rahmen der Untersuchung auch Merkmale der Aggregateinheiten erhoben wurden. Auf das Beispiel bezogen könnten das Variablen sein zur Größe der Schulen, zum Alter und weiteren Merkmalen der Lehrerkollegien, zu den Anteilen einzelner Schülergruppen an der Schülerschaft usw. Diese Aggregatmerkmale werden mit

W_j (Merkmal W für Aggregateinheit j) bezeichnet. Nach den Ergebnissen von COLEMAN, HOFFER und KILGORE (1982) bestehen z.B. bezüglich beider Regressionsparameter Unterschiede zwischen katholischen und staatlichen Schulen (Variable: SECTOR). Ebenso könnte auch geprüft werden, ob Differenzen zwischen Stadt- und Landschulen bestehen oder ob sich die Lehrerfahrung und -kompetenz der Lehrkräfte an den Schulen auf das Leistungsniveau oder die Leistungsungleichheit unter den Schülern auswirkt.

Allgemein kann versucht werden, die Unterschiede in den β-Koeffizienten durch Merkmale der Aggregateinheiten (W_j) zu erklären. Damit werden nun die Regressionskoeffizienten selbst als abhängige Variablen betrachtet, deren Ausprägung durch Aggregatmerkmale vorhergesagt werden soll. Im Beispiel könnte W_j für die Variable SECTOR stehen, mit einem Wert von "0" für öffentliche Schulen und einem Wert von "1" für katholische Schulen:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \qquad [2\text{-}4]$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j} \qquad [2\text{-}5]$$

Im Beispiel haben die Koeffizienten der Gleichung folgende Bedeutung:

γ_{00} mittleres Leistungsniveau für öffentliche Schulen (SECTOR=0)

γ_{01} Differenz im Leistungsniveau zwischen öffentlichen und katholischen Schulen

γ_{10} mittlere Beziehung zwischen Status und Leistung in öffentlichen Schulen (SECTOR=0)

γ_{11} Differenz in der Beziehung zwischen Status und Leistung zwischen öffentlichen und katholischen Schulen

u_{0j} Residualanteil: spezifischer Effekt der Schule j auf die Leistung unter Konstanthaltung von W_j

u_{1j} Residualanteil: spezifischer Effekt der Schule j auf die Beziehung zwischen Status und Leistung unter Konstanthaltung von W_j

Die Berücksichtigung von Merkmalen der Aggregateinheiten (W_j) zur Vorhersage der variierenden Regressionskoeffizienten führt im Beispiel zu einem Kennwert für die Regressionskonstante (γ_{00}), der das mittlere Leistungsniveau der öffentlichen Schulen kennzeichnet und einem Abweichungswert (γ_{01}), um den die katholischen Schulen sich in den Schülerleistungen von den öffentlichen unterscheiden. Entsprechendes gilt für die Steigungskoeffizienten (γ_{10} und γ_{11}) zur Kennzeichnung der Beziehung zwischen dem sozialen Status der Schüler und ihren Leistungen in öffentlichen und katholischen Schulen. Außerdem ergeben sich auf der Aggregatebene Residualanteile für die Abweichungen zwischen den Regressionskonstanten (u_{0j}) und Steigungskoeffizienten (u_{1j}) zwischen den Schulen unter Kontrolle des Aggregatmerkmals W_j. Diese werden als Zufallsvariablen mit einem Mittelwert von Null und den Varianzen τ_{00} und τ_{11} sowie der Kovarianz τ_{01} betrachtet (vgl. Kap. 2.1.2).

Die methodische Besonderheit dieses Mehrebenenansatzes gegenüber einer Standard-Regressionsanalyse läßt sich gut verdeutlichen, wenn die Gleichungen der Aggregatbeziehungen in die Individualgleichung eingesetzt und so zu einer kombinierten Gleichung zusammengezogen werden:

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}(X_{ij}-\bar{X}_{.j}) + \gamma_{11}W_j(X_{ij}-\bar{X}_{.j}) + u_{0j} + u_{1j}(X_{ij}-\bar{X}_{.j}) + r_{ij} \qquad [2\text{-}6]$$

Diese Modellbeziehung entspricht *nicht* dem üblichen Regressionsmodell, das unter Einbeziehung von Aggregatvariablen berechnet werden könnte. Das Schätzverfahren und die Hypothesentests in der Standard-Regressionsanalyse setzen unabhängige und normalverteilte Fehler mit konstanter Varianz voraus. Demgegenüber haben die Fehlerterme im Hierarchisch Linearen Modell eine komplexere Form, wie an der kombinierten Gleichung abzulesen ist ($u_{0j} + u_{1j}... + r_{ij}$). Berücksichtigt werden hierdurch *gruppenspezifische Fehlerkomponenten*, denn die Komponenten u_{0j} und u_{1j} sind für die Schüler innerhalb einer Schule gleich, sie variieren aber zwischen den Schulen. Der Gleichung ist damit auch zu entnehmen, in welchem Fall eine Standard-Regression angemessen wäre. Dies wäre dann der Fall, wenn u_{0j} und u_{1j} für alle Aggregateinheiten nicht von Null verschieden wären, die Modellgleichung also für alle Einheiten als identisch angenommen werden könnte. Sofern diese Bedingung nicht zutrifft, ist die Standard-Regression jedoch eine Fehlspezifikation. In jedem Fall ist also für

Mehrebenendaten zu prüfen, ob die gruppenspezifischen Komponenten vernachlässigbar gering sind oder nicht. Im ersten Fall wären auch Standardverfahren anwendbar, bei bedeutsamen gruppenspezifischen Komponenten dagegen nicht.

2.1.4 Zufalls- und feste Effekte in Modellgleichungen

Im Hierarchisch Linearen Modell können die Koeffizienten der Regressionsgleichungen als feste Effekte (invariant bzw. *fixed*) oder Zufallseffekte (variabel bzw. *random*) spezifiziert werden. Ebenso sind gemischte Modelle möglich, in denen ein Teil der Beziehungen im Sinne von festen Effekten und andere Beziehungen als Zufallseffekte behandelt werden. Im Beispiel mit den Schülerleistungen als abhängiger Variable könnte etwa ein variabler Effekt der sozialen Herkunft und ein invarianter Effekt der kognitiven Fähigkeiten der Schüler angenommen werden. Damit würde davon ausgegangen, daß für den Einfluß der kognitiven Fähigkeiten auf die Schülerleistungen ein einheitlicher Effekt in allen Schulen besteht, während für den Einfluß der sozialen Herkunft auf den Schulerfolg die Effekte schulspezifisch variieren.

Die entscheidende Frage bei der Modellspezifikation ist, ob die getesteten Beziehungen zwischen den Untersuchungsvariablen innerhalb der Aggregateinheiten als invariant oder variabel behandelt werden sollen. Es geht hierbei um die theoretisch begründete Plausibilität und empirische Haltbarkeit der zu prüfenden Annahmen. Bei der Spezifikation eines festen Effektes wird unterstellt, daß

ein einheitlicher Koeffizient für die gesamte Population der Aggregateinheiten gültig ist. Die Spezifikation eines Zufallseffektes besagt dagegen, es ist zu erwarten, daß in Abhängigkeit von Merkmalen der Aggregateinheiten unterschiedliche Koeffizienten, die eine Zufallsauswahl darstellen und ihrerseits erklärungsbedürftig sind, gefunden werden.

Die Annahme fester Effekte entspricht dem Vorgehen bei einer üblichen Regressionsanalyse, die variierende Koeffizienten zwischen Untersuchungseinheiten schon vom Ansatz her ausschließt. Ziemlich offensichtlich ist diese Annahme im Bereich der sozialwissenschaftlichen Forschung in vielen Fällen fragwürdig und würde sich bei einer Überprüfung - die mit dem Hierarchisch Linearen Modell möglich ist - als nicht haltbar erweisen. Als Beispiel dafür wurde bereits auf Ergebnisse der *High School and Beyond*-Studie hingewiesen.

Das konkrete Vorgehen sollte insofern so aussehen, daß die Modellbeziehungen auf vorhandene Zufallseffekte hin überprüft werden. Signifikante Varianzen der Koeffizienten können nicht gefixt werden, ohne eine Fehlspezifikation zu begehen. Sind die Zufallskomponenten dagegen nicht bedeutsam, ist es zulässig, die Beziehungen als feste Effekte zu behandeln.

2.1.5 Probleme und Grenzen des Hierarchisch Linearen Modells

Obwohl das Hierarchisch Lineare Modell - ebenso wie die anderen inzwischen verfügbaren Ansätze zur Mehrebenenanalyse - eine zuverlässige Analyse von hierarchisch strukturierten Datensätzen erlaubt, sollten die Probleme und Grenzen der bislang gegebenen Möglichkeiten nicht übersehen werden. Für eine Mehrebenenanalyse, die zu verläßlichen Ergebnissen führen soll, sind *hohe Anforderungen an die Daten und die Untersuchungsstichprobe zu erfüllen*. Die Grenzen des Modells sind, vor allem im Vergleich zu den Möglichkeiten linearer Strukturgleichungsmodelle, darin zu sehen, daß eine *Untersuchung von Kausalketten direkter und indirekter Beziehungen auf der Basis latenter Variablen nicht möglich* ist.

Bei der *Stichprobenziehung* muß ein komplexes Design realisiert werden, durch das sichergestellt ist, daß eine genügend große Zahl an Untersuchungseinheiten je Untersuchungsebene erreicht wird. Für eine Drei-Ebenen-Analyse zur Wirkung von Schulklassen- und Schulfaktoren auf individuelle Schülervariablen ist eine hinreichend große Zahl an Schülern, Schulklassen und Schulen erforderlich. Dazu, wie groß die Stichproben effektiv sein müssen, um die Modellvoraussetzungen zu erfüllen, ist bislang wenig bekannt (vgl. Kap. 4.2). Eher unklar sind ebenso die Folgen von Verletzungen der Modellannahmen. Problematisch ist in dieser Hinsicht weniger die Qualität der Schätzwerte für die Modellparameter als vielmehr die Brauchbarkeit der ermittelten Standardfehler. In der praktischen Anwendung empfiehlt sich von daher bei kleineren Stichproben eine Zurückhaltung in der Interpretation geringer Effekte, die das Signifikanzniveau nur knapp erreichen.

Hohe Anforderungen ergeben sich nicht nur bezüglich der Stichprobenziehung, sondern ebenso

hinsichtlich der *Auswahl der zu erhebenden Variablen auf allen Untersuchungsebenen*. Wenn Mehrebenenmodelle einen hohen Erklärungswert erzielen sollen, müssen für die Analyse aussagekräftige und theoretisch plausibel zueinander in Beziehung stehende Faktoren auf allen Ebenen erhoben worden sein.

Bei der *Spezifikation von Modellen* ist die Komplexität der impliziten Modellbeziehungen besonders zu beachten. Die Wirkungen der Variablen- und Ebenenbeziehungen sind nicht unabhängig voneinander, sondern bilden ein komplexes Beziehungsgefüge ab. Das Hinzufügen oder Weglassen einzelner Variablen beeinflußt auch die Beziehungen zwischen anderen Variablen und die Beziehungen auf anderen Modellebenen. Die Gesamtheit der Auswirkungen von Modellmodifikationen ist daher schwerer überschaubar als bei der Anwendung einfacherer Analyseverfahren. Auch von daher sollte die Modellentwicklung gut theoretisch fundiert und plausibel hergeleitet sein.

Die bisher verfügbaren Ansätze und Software-Pakete zur Mehrebenenanalyse haben auch ihre *Grenzen in der Plausibilität der in den Modellen abbildbaren Beziehungen*. Oftmals legt es die Theoriebildung nahe, von einem differenzierten Geflecht direkter und indirekter Beziehungen zwischen den Untersuchungsvariablen auszugehen. Die Wirkung der sozialen Herkunft von Schülern auf ihren Schulerfolg kann z.B. über eine Reihe von Vermittlungsgliedern (Einstellungen und Haltungen der Schüler, Wahrnehmung und Bewertung durch die Lehrer usw.) erfolgen, und es wäre von daher angebracht, statt der direkten Beziehung zwischen

beiden Variablen die entsprechende *Kausalkette* im einzelnen zu überprüfen, wie es in Pfadanalysen möglich ist. Gegenstand sozialwissenschaftlicher Forschung sind überdies in vielen Fällen keine direkt beobachtbaren oder meßbaren Einzelvariablen, sondern theoretische Konstrukte (wie etwa Intelligenz, Motivation usw.), die als *latente Variablen oder Faktoren* über mehrere beobachtbare Einzelvariablen (Indikatoren) erschlossen werden müssen. Beide Forschungsdesiderate - Berücksichtigung von Kausalketten und latenten Variablen - lassen sich mit den gegenwärtigen Mehrebenenmodellen nicht erfüllen. Statt dessen stehen dafür lineare Strukturgleichungsmodelle (*structural equation models*) zur Verfügung, in denen jedoch umgekehrt die spezifischen Anforderungen von Mehrebenenanalysen nicht oder allenfalls zum Teil erfüllt werden. Mehrebenenanalysen sollten insofern nicht als ein Allheilmittel zur Bewältigung aller relevanten Fragestellungen sozialwissenschaftlicher Forschung verstanden werden. Dennoch sind in vielen Untersuchungsbereichen durch Mehrebenenanalysen wichtige Einsichten in die Zusammenhänge von Individual- und Aggregatfaktoren zu erwarten, die durch andere Analyseverfahren nicht zuverlässig ermittelt werden können.

2.2 Anwendungen des Hierarchisch Linearen Modells

Das Hierarchisch Lineare Modell beinhaltet eine breite Klasse an spezifischen Teilmodellen, die mit ihm abgebildet werden können. U.a. lassen sich die Varianz- und Kovarianzanalyse mit Zufallseffekten durch eine Vereinfachung des allgemeinen Modells

spezifizieren. Die Darstellung in diesem Kapitel beginnt mit einfachen Anwendungsfällen und geht zu komplexeren Modellen über.

2.2.1 Einfaktorielle Varianzanalyse mit Zufallseffekten

Die einfaktorielle Varianzanalyse mit Zufallseffekten ist in der Notation des Hierarchisch Linearen Modells einfach zu spezifizieren und läßt sich für die Individualebene angeben als:

$$Y_{ij} = \beta_{0j} + r_{ij} \qquad [2\text{-}7]$$

Der relevante Parameter ist die Regressionskonstante (β_{0j}), die den Mittelwert für die Untersuchungseinheit j angibt. Der Meßwert einer Individualeinheit ergibt sich aus dem Mittelwert der Aggregateinheit und der individuellen Abweichung (r_{ij} ~$N(0,\sigma^2)$) von diesem Mittelwert.

Das Modell für die zweite Analyseebene lautet:

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad [2\text{-}8]$$

Dabei ist γ_{00} der Gesamtmittelwert der Population und u_{0j} ist der Abweichungswert oder Zufallseffekt für Einheit j, mit einem Mittelwert von 0 und der Varianz τ_{00}.

In kombinierter Form ergibt sich aus den beiden vorigen Gleichungen:

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \qquad [2\text{-}9]$$

Das Varianzanalysemodell beinhaltet den Gesamtmittelwert (γ_{00}), einen gruppenspezifischen Zufalls- (u_{0j}) und einen Individualeffekt (r_{ij}). Dieses einfache Modell ist ein bereits sehr informativer erster Schritt in einer Mehrebenenanalyse. Es kann als *vollständig unkonditioniertes Modell* bezeichnet werden, da es außer der Gruppenzugehörigkeit keine weiteren Variablen (weder auf Ebene 1 noch auf Ebene 2) beinhaltet. Dem Modell kommt von daher kein unmittelbarer Erklärungswert zu. Interessant ist das Modell aber dennoch, denn es wird geprüft, ob die abhängige Variablen überhaupt bedeutsam zwischen den Aggregateinheiten variiert, ob Unterschiede zwischen den Untersuchungsgruppen bestehen. Außerdem läßt sich ermitteln, wie groß die *Varianzanteile innerhalb und zwischen den Untersuchungsgruppen* sind.

Die Varianz in der abhängigen Variablen ergibt sich nach der Beziehung:

$$Var(Y_{ij}) = Var(u_{0j} + r_{ij}) = \tau_{00} + \sigma^2 \qquad [2\text{-}10]$$

D.h., die Varianz in der abhängigen Variablen (Y) kann zerlegt werden in einen Anteil zwischen den Untersuchungsgruppen (Varianz von $u_{0j} = \tau_{00}$) und einen Anteil innerhalb der Gruppen (Varianz von $r_{ij} = \sigma^2$).

Damit ist dann weiter die sogenannte *Intraklassen-Korrelation* (ρ) bestimmbar als:

vollständig unbedingtes Modell
- Wie Varianzanalyse
  1 Gesamtmittelwert
  1 gruppenspezif. Zufallsfaktor
  1 Individualeffekt

$Y_{ij} = \beta_{0j} + r_{ij}$

$\beta_{0j} = \gamma_{00} + u_{0j}$

$Var(Y_{ij}) = var(u_{0j} + r_{ij})$
$= \tau_{00} + \sigma^2$

Varianz von Y läßt sich zerlegen in Anteil zw. den Gruppen (τ_{00}) u. Anteil innerhalb der Gruppen (σ^2)

$$\rho = \tau_{00} \; / \; (\tau_{00} + \sigma^2) \qquad [2\text{-}11]$$

Die Intraklassen-Korrelation gibt den Anteil an Varianz in der abhängigen Variablen zwischen den Einheiten auf der zweiten Analyseebene an bzw. den auf die Gruppenzugehörigkeit rückführbaren Varianzanteil. Ein nur unbedeutender Varianzanteil würde besagen, daß keine gruppenspezifischen Effekte vorliegen. Auf das Standardbeispiel bezogen drückt die Intraklassen-Korrelation den Anteil an Varianz in den Schülerleistungen aus, der auf Unterschiede zwischen den Schulen rückführbar ist.

Trotz des geringen Erklärungswertes ist die einfache Varianzanalyse sehr häufig der erste Schritt in einer Mehrebenenanalyse, um Schätzungen der zu erwartenden Gruppeneffekte zu ermitteln. Das Vorgehen entspricht dem Vorschlag von FIREBAUGH (1979), zur Prüfung der vorliegenden Gruppendifferenzen die Effekte der Gruppenzugehörigkeit direkt zu ermitteln.

2.2.2 Regression mit Vorhersage der Mittelwerte

Eine zweite Modellvariante führt gegenüber dem Modell der Varianzanalyse weiter, indem versucht werden soll, Differenzen der Regressionskonstanten (β_{0j}) durch die Einbeziehung von Aggregatmerkmalen (W_j) zu erklären. Geprüft werden soll der Einfluß von Aggregatmerkmalen auf die Mittelwerte der Aggregateinheiten. Mit Blick auf das Standardbeispiel würde versucht, durch Merkmale

der Schulen - die natürlich erhoben worden sein müssen - ein unterschiedliches Niveau in den Schülerleistungen zu erklären.

Das Individualmodell entspricht dem der Varianzanalyse, es lautet also wiederum:

$$Y_{ij} = \beta_{0j} + r_{ij} \qquad [2\text{-}12]$$

Das Modell für die zweite Analyseebene wird um zumindest eine Prädiktorvariable erweitert, es lautet also im einfachsten Fall:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \qquad [2\text{-}13]$$

Damit wird nun versucht, Differenzen in den Mittelwerten der Aggregateinheiten (β_{0j}) durch die Berücksichtigung einer (oder auch mehrerer) Aggregatvariablen (W_j) vorherzusagen bzw. zu erklären. Im Beispiel könnte mit dieser Modellvariante geprüft werden, ob sich Schulmerkmale, wie z.B. eine mehr oder weniger enge Zusammenarbeit im Lehrekollegium oder die Unterrichtserfahrung der Lehrer, im erreichten Leistungsniveau der Schulen niederschlagen.

Kombiniert lautet das Modell:

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + u_{0j} + r_{ij} \qquad [2\text{-}14]$$

Hervorzuheben ist hierbei, daß in diesem *konditionierten Modell* u_{0j} nicht mehr die einfache Abweichung der Aggregateinheiten vom Gesamtmittel-

wert angibt, sondern den Residualanteil nach Kontrolle von W_j, d.h.:

$$u_{oj} = \beta_{0j} - \gamma_{00} - \gamma_{01} W_j \qquad [2\text{-}15]$$

Ebenso ist die Varianz von u_{0j} (τ_{00}) nun die Residualvarianz (in β_{0j}) nach Kontrolle von W_j.

2.2.3 Einfaktorielle Kovarianzanalyse mit Zufallseffekten

Bei Vergleichen zwischen Aggregateinheiten ist es sehr oft notwendig, die Effekte von Individualmerkmalen zu berücksichtigen bzw. zu kontrollieren, also eine Adjustierung der Gruppeneffekte um Individualeffekte vorzunehmen. Ein einfacher Vergleich der Mittelwerte für die Schülerleistungen zwischen Schulen wäre z.B. nicht angemessen, wenn bedeutsame Schülermerkmale in den Schulen unterschiedlich verteilt sind. In einem solchen Fall wären die Schulen mit der günstigeren Ausprägung der Schülermerkmale im Vorteil und diesen Schulen würden eventuell Wirkungen zugeschrieben, die allein schon durch die Schülermerkmale erklärbar sein könnten.

Ein Verfahren, um neben Gruppeneffekten auch Unterschiede in Individualmerkmalen berücksichtigen zu können, ist die Kovarianzanalyse. Die Anwendung einer traditionellen Kovarianzanalyse führt allerdings nur unter spezifischen Voraussetzungen zu verläßlichen Ergebnissen, wie noch zu zeigen sein wird.

In der Formulierung des Hierarchisch Linearen Modells lautet die Gleichung für die Individualebenen-Beziehung:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{.j}) + r_{ij} \qquad [2\text{-}16]$$

In dieser Gleichung wird von einer Zentrierung der Prädiktorvariablen um den jeweiligen Gruppenmittelwert ausgegangen (vgl. Kap. 3).

Für die zweite Analyseebene lauten die getesteten Beziehungen:

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad [2\text{-}17]$$

und

$$\beta_{1j} = \gamma_{10} \qquad [2\text{-}18]$$

Wie zu sehen ist, entspricht die obere der beiden Gleichungen dem zuvor angegebenen Modell der Varianzanalyse mit der Annahme variierender Mittelwerte zwischen den Aggregateinheiten.

An der zweiten Gleichung ($\beta_{1j} = \gamma_{10}$) ist außerdem die zentrale Annahme der traditionellen Kovarianzanalyse abzulesen: Es wird davon ausgegangen, daß der Effekt von X auf Y für alle Einheiten auf der zweiten Analyseebene gleich ist, daß es sich also um einen festen Effekt handelt, der nicht zwischen den Aggregateinheiten variiert. Das würde im Beispiel bedeuten, daß ein identischer Koeffizient für die Beziehung zwischen sozialem

Status und Schülerleistung in allen Untersuchungseinheiten angenommen werden kann. Sofern diese Bedingung (wie im Beispiel) nicht erfüllt ist, führt eine traditionelle Kovarianzanalyse zu verfälschten Ergebnissen. Die Adjustierung des Individualeffektes ist dann nämlich nicht korrekt, sie wird für die Aggregateinheiten in einheitlicher Art und Weise vorgenommen, statt spezifisch - wie es erforderlich wäre.

In kombinierter Form lautet das Modell der Kovarianzanalyse:

$$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij}-\bar{X}_{.j}) + u_{0j} + r_{ij} \quad [2\text{-}19]$$

Der Unterschied zur Standard-Kovarianzanalyse besteht darin, daß der Gruppeneffekt (u_{0j}) hier als Zufalls- und nicht als fester Effekt spezifiziert ist. Dagegen ist γ_{10}, wie in der Standard-Kovarianzanalyse, ein gemeinsamer Regressionskoeffizient für die Beziehung innerhalb der Gruppen (*pooled within-groups*). Außerdem sind die β_{0j} die adjustierten Mittelwerte bezüglich der Variablen X und σ^2 ist die Residualvarianz nach Kontrolle von X.

Die nicht unproblematische Annahme eines für alle Aggregateinheiten gleichermaßen zutreffenden Regressionskoeffizienten γ_{10} muß im Hierarchisch Linearen Modell *nicht* zwangsläufig gemacht werden. Darauf gehen die beiden folgenden Abschnitte ein.

2.2.4 Regression mit Zufallskoeffizienten

Eine gravierende Einschränkung und nicht immer erfüllte Voraussetzung der Kovarianzanalyse ist die Annahme eines für alle Aggregateinheiten gültigen Steigungskoeffizienten (γ_{10}) für die Regression von Y auf X. Diese Annahme wird im Hierarchisch Linearen Modell bei einer Regression mit Zufallskoeffizienten aufgegeben. Die Steigungskoeffizienten dürfen also für die Aggregateinheiten unterschiedlich sein. Damit werden nicht mehr nur, wie in den Modellen zuvor, die *Regressionskonstanten*, sondern auch die *Steigungen* (*slopes*) als variable Zufallskoeffizienten behandelt. Diese spezifische Erweiterung ermöglicht es, im Hierarchisch Linearen Modell *systematic varying slopes* im Sinne von BURSTEIN u.a. (1978; vgl. Kap. 1.3) zu untersuchen.

Das Modell für die Individualebene entspricht dem der Kovarianzanalyse:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij}-\bar{X}_{.j}) + r_{ij} \quad [2\text{-}20]$$

Die Gleichungen für die zweite Analyseebene lauten jetzt aber:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad [2\text{-}21]$$

und

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad [2\text{-}22]$$

Die zweite Gleichung drückt die unterschiedlichen Steigungen innerhalb der Aggregateinheiten als Abweichungswerte von der mittleren Steigung in der Gesamtstichprobe (*pooled within*) aus. Zugleich ergibt sich damit eine *Varianz der Steigungskoeffizienten*, die mit τ_{11} bezeichnet wird.

Dieses Modell kann als *unkonditioniertes Modell der Beziehungen auf der zweiten Analyseebene* bezeichnet werden. Hierbei werden zwar die Regressionskonstanten *und* -steigungen als Zufallskoeffizienten behandelt, es werden aber keine Prädiktoren aufgenommen, um die Variabilität dieser Koeffizienten vorherzusagen.

Insofern kommt auch dieser Modellvariante - wie dem vollständig unkonditioniertem Modell (Kap. 2.2.1) - kein eigentlicher Erklärungswert zu, denn es bleibt unbeantwortet, wodurch die Variabilität der Koeffizienten bedingt ist. Dennoch ist die Regression mit Zufallskoeffizienten in aller Regel ein sinnvoller Ausgangspunkt für weitere Analysen. Auf der Grundlage der gefundenen Ergebnisse läßt sich entscheiden, für welche Koeffizienten im Modell aufgrund unbedeutender Varianzanteile die Annahme fester Effekte zulässig ist und welche Koeffizienten aufgrund bedeutsamer Unterschiede zwischen den Aggregateinheiten als Zufallseffekte spezifiziert sein müssen.

2.2.5 Regressionskoeffizienten als abhängige Variablen

Die Behandlung der Regressionskoeffizienten als abhängige Variablen stellt den allgemeinen Fall

der Anwendung des Hierarchisch Linearen Modells dar. Die Erweiterung im Vergleich zum vorhergehenden Modell besteht darin, daß durch die Einbeziehung geeigneter Prädiktorvariablen versucht werden soll, die *Unterschiede der Regressionskoeffizienten in den Aggregateinheiten erklären* zu können. Dies setzt natürlich voraus, daß entsprechend geeignete Variablen auf der Aggregatebene für die Analyse zur Verfügung stehen. Im Vergleich zu dieser allgemeinen Form des Modells stellen die zuvor aufgeführten Varianten Vereinfachungen dar. Zugleich kann in der allgemeinen Form des Modells davon ausgegangen werden, daß sowohl auf der ersten als auch auf der zweiten Analyseebene die Wirkung von mehr als nur einer einzelnen Prädiktorvariablen zu überprüfen ist.

Bezeichnet man die abhängige Variable für Person i in der Aggregateinheit j wie zuvor mit Y_{ij}, dann lautet das Modell auf der *Individualebene* für einen Satz unabhängiger Variablen X_{qij} (q= 1,...Q) und den Residualanteil r_{ij} wie folgt:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + \ldots + \beta_{Qj} X_{Qij} + r_{ij} \qquad [2\text{-}23]$$

Dabei wird wieder die Annahme getroffen, daß die Residuen normalverteilt sind, mit einem Mittelwert von Null und Varianz σ^2:

$$r_{ij} \sim N(0,\sigma^2)$$

Das *Modell für die Beziehungen auf der zweiten Analyseebene* beschreibt in allgemeiner Form die

Vorhersage der zwischen den Aggregateinheiten variierenden β-Koeffizienten. Relevant sind damit Q+1 Vorhersagegleichungen, nämlich Q Gleichungen für die β_{qj}-Koeffizienten und eine weitere Gleichung für β_{0j}. Alle β_{qj} sowie $ß_{0j}$ können als Variablen behandelt werden, die in Abhängigkeit von einem Satz von Merkmalen der Aggregate (W_{sj}) variieren und außerdem durch die Residualanteile u_{qj} charakterisiert sind. Für jeden Koeffizienten kann ein eigener Satz von Variablen W_{sj} ($s=1,...,S_q$) analysiert werden:

$$\beta_{qj} = \gamma_{q0} + \gamma_{q1}W_{1j} + \gamma_{q2}W_{2j} + \ldots + \gamma_{qS_q}W_{S_qj} + u_{qj} \qquad [2\text{-}24]$$

Hierbei wird von der Annahme* ausgegangen, daß die Residuen auf der zweiten Analyseebene multivariat normalverteilt sind, mit einem Mittelwert von Null und einer Varianz-Kovarianzmatrix T:

$$u_{qj} \sim N(0,T)$$

Mit diesen Gleichungen ist das Hierarchisch Lineare Modell in seiner allgemeinen Form für Anwendungsfälle mit zwei Untersuchungsebenen angegeben. Auf der *Individualebene* wird die abhängige Variable durch eine oder mehrere Individualvariablen vorhergesagt und auf der *Aggregatebene* werden die Regressionskoeffizienten selbst zu abhängigen Variablen, deren Varianz zwischen den Aggregateinheiten durch ein Aggregatmerkmal oder mehrere Aggregatmerkmale erklärt werden soll. Die zuvor aufgeführten Varianten sind Vereinfachungen des allgemeinen Modells, indem einzelne Parameter auf den Wert Null gesetzt werden.

Der spezifische Anwendungsfall für Analysen von *Längsschnittdaten* und die Erweiterung für Anwendungen mit *drei Untersuchungsebenen* werden in Kapitel 6 besprochen. Auf Strategien der Modellbildung, die Modellvoraussetzungen und Besonderheiten des verwendeten Schätzverfahrens geht das vierte Kapitel ein. Zunächst wird in Kapitel 3 die Metrik der Variablen und die Adjustierung von Effekten in Mehrebenenmodellen behandelt.

* Von dieser Annahme wird in aller Regel ausgegangen und sie ist für eine breite Klasse von Anwendungen angemessen; allerdings sind auch andere Spezifikationen der Fehlerstrukturen möglich (vgl. BRYK und RAUDENBUSH 1992, 24ff.)