

## **Mehrebenenanalysen - die Berücksichtigung von hierarchisch strukturierter Realität bei der Auswertung von Daten**

Erschien in: Mehrebenenanalysen. Beiträge zur Erfassung hierarchisch strukturierter Realität. Weinheim: Verlagsunion Psychologie, 1986

*Matthias v. Saldern*

Nach allgemeinem Konsens zwischen Forschern sollte empirische Forschung weitgehend theoriegeleitet sein. Gemeint ist damit, daß eine Annahme (gemeinhin als Hypothese bezeichnet) auf theoretischen Überlegungen basiert und anhand vorliegender Daten mit Hilfe der Statistik getestet werden sollte, um Zusammenhänge zwischen relevanten Merkmalen der Umwelt zu erklären. Ein großer Teil der sozialwissenschaftlichen Forschung testet Hypothesen, die sich im weitesten Sinne mit dem Menschen und dessen Einbeziehung in seine Umwelt beschäftigen. Die genaue Analyse der sozialwissenschaftlichen Literatur zeigt, daß nur ganz wenige Bereiche sich nicht mit dieser Aussage fassen lassen.

Während in der Hypothesen- und Modellbildung die Einbeziehung des Kontextes schon weitgehend berücksichtigt ist, wird dies im Bereich der Datenauswertung und der Statistik noch größtenteils vernachlässigt. Man kann sagen, daß die Modellbildung schon sehr viel weiter fortgeschritten ist als die Auswertungsmethodologie. Ein möglicher Ausweg aus dieser Kluft könnten die sogenannten Mehrebenenanalysen sein.

Mehrebenenanalysen resultieren aus der Überlegung, daß menschliches Handeln grundsätzlich in einem Kontext stattfindet, der bei der Auswertung berücksichtigt werden muß. Andernfalls kommt es nicht nur zu einer Schwächung der Erklärungskraft einzelner Variablen, viel schlimmer ist, daß man zu Fehlannahmen kommen kann. Mehrebenenanalysen unterstützen alle diejenigen Hypothesen, die in irgendeiner Weise annehmen, daß soziale Wirklichkeit besser durch ein Modell beschrieben werden kann, welches Hierarchien in institutionellen Gegebenheiten berücksichtigt.

Mehrebenenanalysen geben keine Antwort auf die Frage, wie es zu Hierarchien kommt und welche theoretischen Ansätze es dazu gibt. Man findet dazu aber in den strukturfunktionalistischen Theorien (T.Parsons, N.Luhmann) genügend Hinweise darauf, daß Hierarchien gewißermassen zwangsläufig für bestimmte Aufgabenstellungen prädestiniert zu sein scheinen. Ein den meisten Menschen zugängliches Beispiel ist der hierarchische Aufbau des Schulsystems (Die Beispiele in diesem Sammelband stammen deshalb aus diesem Bereich). Ein Schüler/eine Schülerin lebt und lernt ja nicht völlig von seiner Umwelt abgeschnitten, sondern er ist unterstes Element einer Hierarchie, wie sie beispielhaft in Abb. 1.1 dargestellt ist.

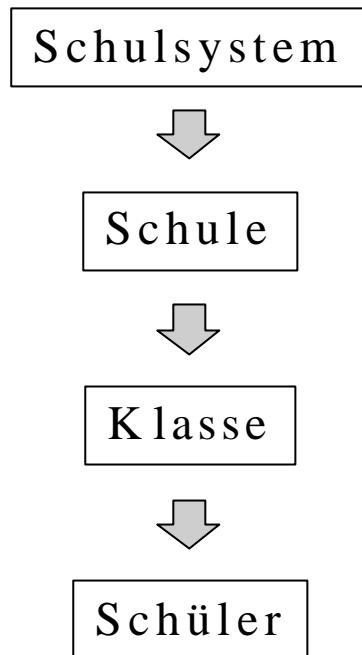


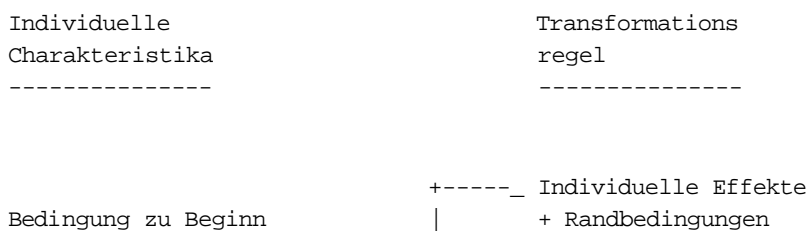
Abb. 1.1 Der hierarchische Aufbau des Schulsystems

Es ist durchaus plausibel, zu erwarten, daß das Schülerverhalten von der Klasse, der Schule etc. abhängig ist, in der der Schüler lebt. Dies kann man leicht durch eine einfache Formel verdeutlichen, in der die zwei additiven Komponenten eines Individualwertes auftauchen: der *Kontextwert* (=Mittelwert) und der *Abweichungswert* (= relative Position eines Schülers in seiner Klasse). Die Formel lautet:

\*Bitte Formel 1.1 einfügen\* (i=Individuum, j=Organisationseinheit)

Dieses Heranziehen des Kontextwertes muß theoretisch begründet werden, deshalb wäre nun zu klären, wie ein Kontexteffekt entsteht und warum er etwas Eigenständiges darstellt. Dies ist deshalb eine wichtige Frage, weil eine Reihe von Variablen zwar über das Individuum erhoben (Individuum als Untersuchungseinheit), aber auf einer hierarchisch höher liegenden Ebene (Gruppe als Analyseeinheit) ausgewertet werden. Es geht also um die Transformation von individuellen Effekten zu einem kollektiven Effekt. In Anlehnung an das Hempel-Oppenheim-Schema versuchte Lindenberg (1977) eine Lösung zu finden. Er führte die *Transformationsregel* ein, die graphisch wie folgt dargestellt werden kann (Abb. 1.2).

\*\* Bitte Abb. 1.2 einfügen \*\*



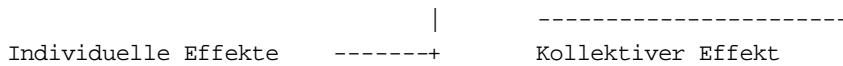


Abb.1.2: Die Transformation von individuellen Effekten zu einem kollektiven Effekt (nach dem H-O-Schema)

Die erklärten individuellen Effekte innerhalb eines Kollektivs werden ebenso wie die Randbedingungen und die Transformationsregel zum Explanans des kollektiven Effekts. Die Transformationsregel definiert die Art der Beziehungen zwischen den Bedingungen. Jeder kollektive Effekt ist eine Bedingungskonstellation individueller Effekte.

Was hier theoretisch recht einleuchtend erscheint, bereitet dem Statistiker Probleme. Er hat die schwierige Aufgabe herauszufinden, über welche Transformationsregel individuelle Rohwerte verarbeitet werden sollen, damit der resultierende Wert als ein Kollektivwert interpretiert werden kann. Gemeinhin verläuft ein richtiger Weg über die Indexbildung oder Aggregation von Daten, woraus man auch die Rechtfertigung für die Verwendung des Mittelwertes ableiten kann (siehe obige Formel).

Die Mehrebenenanalyse ist weniger durch statistische Probleme in die Diskussion geraten, als vielmehr durch theoretische, die sich immer dann auftun, wenn es gilt, individuelles Verhalten eines Einzelnen in einer Organisation zu erklären. Die Diskussion über diesen Problembereich ist zudem nicht neu, denn über Emergenz- und Aggregierungsansätze wird schon seit langem nachgedacht: Die Diskussion über die Notwendigkeit der Einbeziehung von Kontexteffekten wurde beispielsweise schon in der Antike geführt. Plato z.B. sprach von einem "übergeordneten Organismus", der die Menschheit schlechthin überlebensfähig macht. Auch Aristoteles definierte z.B. Kultur (sicherlich eine sehr weitgreifende Kontextvariable) als ein Ergebnis des Zusammenwirkens von Individuen (Bergius, 1976). Darauf aufbauend lassen sich etwas vereinfacht die gesamte Philosophiegeschichte zwei Denkrichtungen spezifizieren: der eher individualistisch orientierte Ansatz und der eher kollektivistisch oder sozial orientierte Ansatz. Die Psychologie ordnete man etwas grob nach dem Zweiten Weltkrieg gerne dem ersten Ansatz zu, die Soziologie dem zweiten. Diese Vereinfachung ist heute sicherlich nicht mehr zu halten. Denn Psychologen beschäftigen sich sehr wohl mit der Kontextbezogenheit menschlichen Verhaltens, Soziologen andererseits bauen Individualphänomene immer mehr in ihre Theoriengebäude ein.

Ein vorläufiger Höhepunkt in der Diskussion um diese beiden Denkrichtungen ist in den Veröffentlichungen von Hummel und Opp zu sehen, die kollektive Phänomene auf individuelle Phänomene durch Transformation zu reduzieren suchten. Diese *Reduktionismusthese* behauptete, daß sämtliche Begriffe der Soziologie durch Begriffe der Psychologie ersetzt werden können. Kollektivphänomene sollten gänzlich durch Individualphänomene erklärt werden können. Dieser Versuch von Hummel und Opp (1971) schlug fehl. Es stellte sich in der Diskussion immer mehr heraus, daß Kollektivbegriffe eine eigenständige Qualität haben. Grund genug, über Verfahren weiter nachzudenken, die diese Kollektiveffekte abzubilden versuchen.

Wenn man einen neuen Forschungsansatz wie die Mehrebenenanalyse propagiert, dann besteht einer der zentralen kritischen Rückfragen darin, ob eine neue Methodologie oder Auswertungsstrategie wirklich gegenüber den herkömmlichen so viele Vorteile besitzt, daß sich ein Einarbeiten auch lohnt.

Diese kritische Haltung ist nach einigen Modetrends in der Statistik (LISREL etc.) nachvollziehbar, besonders wenn man sich durch ein neues Verfahren eine bessere Annäherung von der Methode an die Theorie in der Hoffnung auf neue Erkenntnisse verspricht, eine Hoffnung, die nicht selten enttäuscht wurde.

Bevor nun genauer erklärt wird, was eigentlich das Besondere an Mehrebenenanalysen ist, soll geprüft werden, ob herkömmliche Verfahren nicht auch schon Kontexteffekte adäquat widerspiegeln können. Denn ein Weg, um den Vorteil einer neuen Analysestrategie zu zeigen, liegt sicher darin, die Nachteile herkömmlicher Verfahren aufzuzeigen, falls solche nachgewiesen werden können.

Versucht man, vorliegende statistische Verfahren grob zu klassifizieren, so lassen sich die Ansätze in regressionsanalytische und varianzanalytische aufteilen. Diese Trennung ist historisch begründet und würde sicherlich bei einem neuen Klassifikationsversuch über statistischen Methoden nicht mehr auftreten. Primär ist allerdings die Frage, ob und wie diese beiden Analysetechniken bisher angewendet wurden, um Kontexteffekte abzubilden. Es gibt in der Varianzanalyse die Möglichkeit, Individuen zu gruppieren. Ein Gruppierungsmerkmal könnte durchaus der Kontext sein. Theoretisch wird dabei aber vorausgesetzt, daß die Variable "Zugehörigkeit zu einer Gruppe" psychologisch relevant ist. Oft kann man feststellen, daß es keineswegs diese Variable ist, die psychologisch sinnvoll ist, sondern irgendwelche anderen Variablen, die zwischen den Gruppen variieren. Deshalb ist die einfache Varianzanalyse aus theoretischen Gründen kaum geeignet. Auch bei der Anwendung der regressionsanalytischen Verfahren läßt sich häufig zeigen, daß die formulierten Hypothesen zwar Kontexteffekte annehmen, aber das Regressionsmodell diese Kontexteffekte entweder nicht spezifiziert oder nur wiederum die Gruppenzugehörigkeit über Dummyvariablen einbezogen werden. Das Problem der Berücksichtigung von Kontexteffekten liegt also primär keineswegs in der Verwendung der Varianz- oder Regressionsanalyse, sondern in dem Erkennen der ebenenabhängigen, verhaltensrelevanten Erklärungsvariablen anstelle der Spezifizierung der eher vordergründigen Variablen Gruppenzugehörigkeit.

Ein einfaches Beispiel soll die Problematik um die Kontexteffekte illustrieren. Nehmen wir an, es soll eine Korrelation zwischen zwei Variablen berechnet werden, um den Zusammenhang zwischen Verhaltensauffälligkeiten von Schülern mit anderen Persönlichkeitsvariablen zu ermitteln, um pädagogische Maßnahmen wirksam zu gestalten. Ein - fiktiver - Forscher korreliert die Variablen X und Y über alle Schüler in einer Schule und kommt zu einem hoch positiven Zusammenhang. Aufgrund dieses Ergebnisses werden bestimmte Maßnahmen ergriffen, die sich nach einer Evaluation als Fehlschlag erweisen. Die Frage ist nun, woran dies gelegen haben kann. Der Grund liegt - in diesem vereinfachten Beispiel - darin, daß vom Forscher der Klassenverband nicht zu Genüge berücksichtigt worden ist. Denn macht man sich die Mühe, und berechnet die Regressionskoeffizienten für jede Klasse in dieser Schule, so wird man vielleicht zu genau entgegengesetzten Ergebnissen kommen: Der Zusammenhang ist nunmehr hoch negativ, die pädagogische Maßnahme hätte eine ganz andere sein müssen.

Dieser Sachverhalt ist graphisch anhand von vier Gruppen in Abbildung 1.3 (Stichprobe 2) dargestellt. Abgebildet sind dort Verteilungen von Individualwerten in vier Gruppen (A bis D). Die Punkte stellen den Gruppenmittelwert dar; um sie verlaufen die Umrandungen der Punktwolken. Eine Regressionsgerade über alle Schüler (oder durch die Klassenmittelwerte) würde leicht

ersichtlich anders verlaufen als die Regressionsgeraden in den Gruppen. Das Dilemma wäre bei Stichprobe 1 nicht aufgetreten.

\*\* Bitte Abb.1.3 einfügen \*\*

Dieses Problem ließe sich sicherlich sehr schnell lösen, würde man auf diese Fragestellung eine Kovarianzanalyse anwenden. Wenn aber nun die Beziehungen zwischen den beiden Variablen zwischen den Klassen schwanken, also die Regressionsgeraden mit unterschiedlicher Steigung verlaufen, dann darf man die Kovarianzanalyse nicht mehr verwenden. Der interessierte Forscher müßte aber gerade jetzt fragen, warum eventuell die Regressionsteigungen zwischen den Klassen schwanken (v.Saldern, 1982).

Die beiden oben angeführten Regressionsanalysen sind also einmal eine für die gesamte Stichprobe:

\*Bitte Formel 1.2 einfügen\*

und jeweils eine für jede Gruppe j:

\*Bitte Formel 1.3 einfügen\*

Dieses kleine Beispiel berücksichtigt also neben der Individualebene auch die darüber liegende Gruppenebene. Damit liegt hier eine Parallelanalyse vor, ein untypischer Sonderfall der Mehrebenenanalyse. Denn meistens zeichnen sich Mehrebenenanalysen dadurch aus, daß Effekte *simultan* geschätzt werden, z.B. wie in dem folgenden Modell, wo Individual- und Kontextwert gleichzeitig mit eingehen:

\*Bitte Formel 1.4 einfügen\*

Nun könnte man natürlich noch weitere Ebenen mit einbeziehen: in unserem Beispiel die Schule, das Schulsystem etc. In diesem Falle würden allerdings die einzelnen statistischen Modelle sehr umfangreich sein. Aus diesem Grunde sind in der Literatur alle Aussagen und Vorschläge zur Mehrebenenanalyse weitestgehend auf die Zweiebenenanalyse beschränkt. Alle Aussagen gelten aber grundsätzlich auch für die Einbeziehung von mehr als zwei Ebenen in die statistische Auswertung.

Im folgenden soll gezeigt werden, daß die Fehlerquellen bei der Interpretation von Daten vielfältiger sind als es durch das o.g. Beispiel klar werden konnte. Die möglichen Fehleinschätzungen werden im folgenden klassifiziert, um zu verdeutlichen, daß hier eine besondere Quelle für die Produktion von Artefakten vorliegt. Man kann dazu zwei Gruppen von Fehlschlüssen grob klassifizieren: einmal die, die Aussagen über eine Einheit (Gruppe etc.) machen, wobei die Daten allerdings aus einer anderen Einheit stammen (*über-die-Einheiten*), und zum anderen die, die den Schluß von einer Hierarchieebene auf eine andere vollziehen (*über-die-Ebenen*). Beide Cluster lassen sich noch einmal differenzieren. In Anlehnung an Alker (1969) und Ziegler (1973) kann man folgende sieben Fehlschlüsse unterscheiden:

1. *Über-die-Ebenen* (ökologisch): Hierbei liegt ein Schluß von Kollektivindizes auf Individualeffekte vor. Ein Beispiel wäre der Schluß von einer Korrelation zwischen Mittelwerten auf eine Korrelation zwischen Individualwerten.

2. *Über-die-Ebenen* (individualistisch): Hier liegt genau der umgekehrte Weg des erstgenannten Fehlschlusses vor. Beispielsweise wird von einer Individualwertekorrelation auf eine Korrelation von Mittelwerten geschlossen.

3. *Über-die-Ebenen* (selektiv): Hier wird von den individuellen Beziehungen innerhalb eines Kollektivs oder einer Gruppe auf die Intergruppenbeziehungen geschlossen. Um beim Beispiel der Korrelation zu bleiben: Es ist der Schluß von einer Individualwertekorrelation innerhalb einer Gruppe auf die Mittelwertekorrelation.

4. *Über-die-Ebenen* (universalistisch): Hier liegt der umgekehrte Schluß von 3. vor. Es wird von einer Mittelwertkorrelation auf die Korrelation der Individualwerte innerhalb einer Gruppe geschlossen.

5. *Über-die-Einheiten* (individualistisch): Hierbei wird von einer individuellen Beziehung innerhalb einer Gruppe auf die Gesamtpopulation geschlossen. Es ist z.B. der Schluß von einer Korrelation von Individualwerten innerhalb einer Gruppe auf die Korrelation der Individualwerte der Gesamtpopulation.

6. *Über-die-Einheiten* (ökologisch): Hier liegt der Schluß einer Individualwertekorrelation auf Populationsbasis auf eine Individualwertekorrelation innerhalb einer Gruppe vor.

7. *Über-die-Einheiten* (gleichbleibend): Hier wird von den Beziehungen innerhalb einer Gruppe auf die Beziehungen innerhalb einer anderen geschlossen.

Es ist oft nicht leicht, in vorliegenden Untersuchungsberichten solche Fehlerquellen ausfindig zu machen. Dies mag zum einen daran liegen, daß diesbezüglich die Kriterien für eine Berichtlegung noch nicht genügend ausdifferenziert sind, zum anderen ist sich der Forscher oft nicht über die Tragweite seiner Entscheidungen im klaren. Auch heute noch kann man Berichten entnehmen, daß z.B. eine Korrelation mit Mittelwerten berechnet wurde, weil die ökonomischer sei (*weniger Fälle!*).

Wie ist diesem Problem beizukommen? Eine Hilfe ist, sich den Forschungsprozeß an denen für diese Problematik wichtigen Stellen wiederholt bewußt zu machen:

- a) die Formulierung der Hypothesen,
- b) die Hypothesentestung und
- c) die Interpretation der Ergebnisse.

Jede Phase ist bestimmt durch die vorherige(n). Die Folge daraus ist, daß man im Forschungsprozeß einmal eingeschlagene Wege nicht wechseln sollte, also beispielsweise Hypothesen über das Verhältnis zwischen Individualebenenphänomenen zu machen, bei der Hypothesentestung oder erst später bei der Interpretation dann über Kontextphänomene zu reden.

Die Mehrebenenanalysen kommen vorwiegend in der mittleren Phase b) zum Tragen. Hinzu kommt allerdings, daß neue Methoden auch andere Hypothesen provozieren. Dies steht zwar im Widerspruch zum Ideal des Forschungsprozesses, aber in der Realität sind Entscheidungen in diesem

Prozeß selten unidirektional, eher schon zirkulär. Ein Forscher formuliert also nicht selten Hypothesen so, wie er sie mit dem ihm zur Verfügung stehenden Methodeninventar analysieren zu können glaubt. Die Beschäftigung mit Mehrebenenanalysen wird also eine - vermutlich fruchtbare - Rückkoppelung auf die Hypothesenformulierung bewirken.

Durch die Verwendung von Mehrebenenanalysen soll aber nicht nur verhindert werden, daß der theoretische Einbezug von Kontexteffekten überhaupt nicht berücksichtigt wird, sondern auch, daß eventuell vorhandene Kontextauswertungen falsch interpretiert werden. Die Beiträge in diesem Buch zeigen auf, in welch vielfältigen Bereichen es möglich und sinnvoll ist, den Mehrebenencharakter von Daten zu berücksichtigen.

In dem Beitrag von Burstein u. Linn (Kap. 2) wird eine leicht verständliche Einführung in das grundlegende Gedankengebäude der Mehrebenenanalyse für die Konzeption eines Forschungsprojektes versucht. Dabei ist es eine Hilfe, daß dieser Beitrag (wie andere auch) an einem Beispiel aufgehängt ist, um die Notwendigkeit und Relevanz dieser Methode zu verdeutlichen. Dieser Beitrag verzichtet weitgehend auf formelle Darstellung.

Wie in Kap. 2 wird auch im 3. Kapitel (L. Burstein) auf das SIMS-Projekt (*Second International Mathematics Study*) zurückgegriffen, um einige herkömmliche und neuere Regressionsansätze auf ihre Fruchtbarkeit hin zu überprüfen. Dabei wird die Individualwerteregression mit mehrebenenanalytischen Modellen verglichen. Auch werden Unterschiede und Gemeinsamkeiten in den Schätzwerten aufgezeigt und erklärt. Ein Teil ist der Frage gewidmet, ob Gruppenmittelwerte nach ihrer Mitgliederanzahl gewichtet werden müssen oder nicht und ob standardisierte oder unstandardisierte Regressionskoeffizienten Grundlage der Interpretation werden sollen. Der Leser sollte mit einfacher und komplexer Regression vertraut sein.

In Kap. 4 (W.H. Schmidt und R.T. Houang) werden drei mehrebenenanalytische Verfahren miteinander verglichen: die getrennte Zwischengruppen- und Innergruppenanalyse nach Cronbach (1976), die Zwei-Phasen-Analyse nach Keesling und Wiley (1974) und die Gesamtmodell-Analyse nach Keesling (1977). Der Vergleich dieser zentralen Modelle (die bereits in Kap. 3 angesprochen wurden) ist wichtig. Der Leser möge sich durch die Matrizennotation nicht abschrecken lassen.

M. Miller diskutiert in Kap. 5 eine Erweiterung der Itemanalyse für hierarchisch strukturierte Daten. Dieser Abschnitt ist leicht verständlich dargestellt und verweist auf die Notwendigkeit, die hierarchische Struktur von Wirklichkeit nicht erst bei der Datenauswertung zu berücksichtigen, sondern auch schon bei der Fragebogenkonstruktion.

T.R. Knapp greift im 6. Kapitel Probleme der Varianzanalyse auf. Anhand von drei veröffentlichten Artikeln wird die Wahl der Analyseeinheit und des angemessenen Fehlerterms diskutiert. Dieser Abschnitt schließt mit Ratschlägen für den Forscher ab.

In Kap. 7 wird deutlich, daß die Ansichten über die Wahl der geeigneten Varianzanalyse durchaus zu verschiedenen Schlußfolgerungen führen kann. Anschaulich wird von K.D.Hopkins das Einbeziehen oder Ausschließen von Faktoren (unabhängigen Variablen) beschrieben und die Folgen davon an Beispielen problematisiert.

Es ist kaum verwunderlich, daß auch Strukturgleichungsansätze (= regressionsanalytische Ansätze mit latenten Variablen) Eingang in den Bereich der Mehrebenenanalysen gefunden haben. J.M. Wisenbaker und H. Schmidt stellen im 8. Kapitel eine Erweiterung des allgemeinen LISREL-Modells dar. Der Ansatz hat große Tragweite, ist allerdings nur für den Leser leicht verständlich, der auch das LISREL-Modell gut kennt.

Eines der in der Mehrebenenanalyse diskutierten Probleme besteht darin, wie man durch Aggregatdaten Beziehungen auf Individualebene schätzen kann. Individualdaten liegen also in diesem Falle nicht vor, sondern nur aggregierte Werte, wie Anteile in Prozent etc. E. Haertel und D.E. Wiley widmen sich diesem Problem im 9. Kapitel und stellen ein Modell vor, mit dem man Individualbeziehungen aus Aggregatdaten gewissermaßen "rückgewinnen" kann.

Ein ähnlich gelagertes Problem wird im 10. Kapitel angerissen (G. Firebaugh). Es geht darum, wie man die verschiedenen Größen von Subpopulationen am besten berücksichtigen kann. Dazu werden zwei Modelle kritisch u.a. durch eine Simulationsstudie verglichen.

Auch innerhalb der Vertreter mehrebenenanalytischer Denkstrategien gibt es unterschiedliche Meinungen über den Wert von verschiedenen Regressionsmodellen. C.-P. Chou kritisiert in Kap. 11 insbesondere die Verwendung des Mittelwertes und stellt eine Alternative zu ihm vor: Die Beziehungen zwischen den Gruppenmitgliedern werden nicht durch den Mittelwert, sondern durch eine Interaktionsmatrix (Interaktionen zwischen den Mitgliedern einer Gruppe) repräsentiert.

Im vorletzten Kapitel (12) stellen W. Schneider und A. Helmke ebenfalls nach einer kritischen Diskussion vorliegender Modelle ein Differenzierungskriterium zwischen Gruppen vor: die Steigung der Innerklassenregressionsgeraden. Am Beispiel eines Projektes über Schulleistungen wird die Fruchtbarkeit dieses Ansatzes verdeutlicht.

Der Reader schließt ab mit Hinweisen des Herausgebers über Möglichkeiten, vorhandene Software für Mehrebenenanalysen zu verwenden. Dies erscheint notwendig, weil oft als angemessen erkannte Auswertungsstrategien deshalb nicht angewandt werden, weil die dazugehörige Software vermeintlich fehlt.