

Oliver Christ

Multivariate Verfahren und computergestützte Datenanalyse II

kultur- und
sozialwissenschaften

Der Inhalt dieses Dokumentes darf ohne vorherige schriftliche Erlaubnis durch die FernUniversität in Hagen nicht (ganz oder teilweise) reproduziert, benutzt oder veröffentlicht werden. Das Copyright gilt für alle Formen der Speicherung und Reproduktion, in denen die vorliegenden Informationen eingeflossen sind, einschließlich und zwar ohne Begrenzung Magnetspeicher, Computerausdrucke und visuelle Anzeigen. Alle in diesem Dokument genannten Gebrauchsnamen, Handelsnamen und Warenbezeichnungen sind zumeist eingetragene Warenzeichen und urheberrechtlich geschützt. Warenzeichen, Patente oder Copyrights gelten gleich ohne ausdrückliche Nennung. In dieser Publikation enthaltene Informationen können ohne vorherige Ankündigung geändert werden.

PD Dr. Oliver Christ, Diplom-Psychologe, vertritt seit Dezember 2013 das LG Psychologische Methodenlehre, Diagnostik und Evaluation an der FernUniversität in Hagen. Seit 2007 ist er Studienrat im Hochschuldienst für Psychologische Methodenlehre am Fachbereich Psychologie der Philipps-Universität Marburg.

Multivariate Verfahren und computergestützte Datenanalyse II

Stand: Mai 2014

Zusammenfassung

Der Kurs „Multivariate Verfahren und computergestützte Datenanalyse II“ (Kurs 3444) behandelt mit Strukturgleichungsmodellen und Hierarchisch Linearen Modellen weitere statistische Verfahren zur Auswertung multivariater Daten in der Psychologie. Darüber hinaus werden Methoden zum Umgang mit fehlenden Werten vorgestellt und mit dem Bayesianischen Ansatz Alternativen zu dem frequentistischen Ansatz der Hypothesentestung vorgestellt. Neben dem Verständnis der statistischen Grundlagen der Verfahren steht die praktische Anwendung im Zentrum, wobei die Umsetzung in AMOS (Strukturgleichungsmodelle) und SPSS (Hierarchische Lineare Modelle) anhand von Beispieldaten demonstriert wird.

Vorbemerkungen

Der Kurs 3444 wurde zusammen mit dem Kurs 3443 „Multivariate Verfahren und computergestützte Datenanalyse I“ für das Sommersemester 2014 neu konzipiert. Die Neukonzeption wurde in relativ kurzer Zeit vorgenommen, daher ist der vorliegende Studienbrief lediglich als „advanced organizer“ gedacht, der Ihnen einen Überblick über die im Kurs behandelten Inhalte und die dazugehörige Pflichtliteratur geben soll. Den didaktischen Schwerpunkt bilden die zugehörigen Online-Vorlesungen und Online-Tutorials, in denen, basierend auf der Pflichtliteratur, die zentralen statistischen Grundlagen, beispielhafte Anwendungen (Online-Vorlesung) sowie die praktische Anwendung der behandelten Verfahren (Online-Tutorials) vermittelt werden. Da der Kurs grundlegend neu konzipiert wurde, freue ich mich sehr über alle Hinweise und Anregungen zur weiteren Optimierung!

Ich wünsche Ihnen viel Erfolg bei der Bearbeitung der Studienmaterialien dieses Kurses und viel Spaß beim Studieren an der FernUniversität in Hagen!

Hagen, im Mai 2014

PD Dr. Oliver Christ

Lernziele

- ✓ Sie haben sich Wissen über die statistischen Grundlagen und Voraussetzungen von Strukturgleichungsmodellen und Hierarchischen Linearen Verfahren angeeignet.
- ✓ Sie haben eine allgemeine Kenntnis über das Vorgehen und die verwendeten Schätzverfahren der beiden Verfahren erworben.
- ✓ Sie wissen, bei welchen Fragestellungen die jeweiligen multivariaten Verfahren Antworten liefern und welche Aussagen auf Basis der Ergebnisse getroffen werden können.
- ✓ Sie sind in der Lage, die multivariaten Verfahren in SPSS bzw. AMOS durchzuführen und die Ergebnisausgabe zu interpretieren.
- ✓ Sie kennen unterschiedliche Methoden zum Umgang mit fehlenden Werten und sind in der Lage, diese Methoden in SPSS anzuwenden.
- ✓ Sie verfügen über Grundkenntnisse der Bayes-Statistik.

Didaktisches Konzept

Das didaktische Konzept des Kurses 03444 „Multivariate Verfahren und computergestützte Datenanalyse II“ basiert auf fünf Säulen:

Säule 1: Studienbrief. Der Studienbrief bildet den inhaltlichen Rahmen, der Sie durch den Kurs führt. Er skizziert die Inhalte der dem Kurs zugehörigen Online-Vorlesungen / -Tutorials und benennt die jeweils zugehörige Pflichtlektüre.

Säule 2: Online-Vorlesungen und -Tutorials. Den zentralen Baustein des Kurses bilden die Online-Vorlesungen, die in der Moodle-Umgebung bereitgestellt werden. Zu den verschiedenen in diesem Kurs behandelten multivariaten Verfahren gibt es eine oder mehrere dieser Vorlesungen und i.d.R. zusätzlich Online-Tutorials, in denen die Umsetzung in SPSS und AMOS demonstriert wird.

Säule 3: Pflichtliteratur. Zu jeder Online-Vorlesung gibt es begleitende Pflichtliteratur, anhand derer die Inhalte der Vorlesungen wiederholt und in Teilen vertieft werden können. Es handelt sich zumeist um Buchbeiträge, die meist in elektronischer Form über die UB der FernUniversität zur Verfügung stehen. Zum Teil wird auch auf Original-Artikel zurückgegriffen, die ebenfalls i.d.R. über die UB der FernUniversität, hier die elektronische Zeitschriftendatenbank, abgerufen werden können. Falls nicht, werden wir Ihnen entsprechende Links in Moodle zur Verfügung stellen. Weiterhin werden Ihnen Leitfragen zur Erarbeitung der Inhalte in Moodle zur Verfügung gestellt.

Säule 4: Moodle. Die Lernplattform Moodle ist in Ergänzung zu dem vorliegenden Studienbrief zentrales Kommunikationsinstrument dieses Kurses. Darin können Themen diskutiert und Fragen gestellt werden. Ferner werden in Moodle Links zu den Online-Vorlesungen und -Tutorials eingestellt.

Säule 5: Präsenzveranstaltung. Im Rahmen einer Präsenzveranstaltung für das gesamte Modul werden ausgewählte Inhalte vertieft und eingeübt. Bei der Präsenzveranstaltung handelt es sich nicht um ein Repetitorium.

Inhalt

Multivariate Verfahren und computergestützte Datenanalyse II	4
Zusammenfassung.....	4
Vorbemerkungen.....	4
Lernziele.....	5
Didaktisches Konzept	6
1 Einführung zum Kurs	8
1.1 Überblick über die behandelten Themen.....	8
2 Überblick über die Vorlesungsinhalte	11
2.1 Strukturgleichungsmodelle.....	11
2.2 Hierarchische Lineare Modelle	12
2.3 Umgang mit fehlenden Werten	12
2.4 Bayes-Statistik.....	12
3 Literatur	13
3.1 Pflichtliteratur.....	13
3.2 Weitere Literatur (nicht prüfungsrelevant).....	13

1 Einführung zum Kurs

1.1 Überblick über die behandelten Themen

Strukturgleichungsmodelle (englisch „Structural Equation Modeling“, SEM) stellen eine Familie von Verfahren dar, deren Gemeinsamkeit die Formulierung von Annahmen über Kovarianzstrukturen ist. Eine wichtige Unterscheidung in SEM ist die zwischen manifesten und latenten Variablen, die uns bereits in Kurs 3443 bei der explorativen Faktorenanalyse begegnet ist. Bei den manifesten Variablen handelt es sich um die beobachteten Variablen, bei den latenten Variablen um die den manifesten Variablen zugrunde liegenden, hypothetischen Merkmale. Die Verwendung von latenten Variablen ermöglicht im Rahmen von SEM die Berücksichtigung des Messfehlers, so dass die Beziehung zwischen Merkmalen messfehlerfrei geschätzt werden kann. Ein weiterer wichtiger Vorteil von SEM ist in der Überprüfbarkeit der Passung eines Modells auf die Daten (Modell-Fit) zu sehen, so dass ein konfirmatorisches Vorgehen möglich ist. SEM erlaubt die Prüfung komplexer Modelle mit einer Vielzahl von indirekten Effekten (s. hierzu die Vorlesung zum Thema Multiple Mediationsmodelle im Rahmen des Kurses 3443) und mehr als einer abhängigen Variablen. So haben Tausch, Hewstone, Kenworthy, Cairns und Christ (2007) mittels SEM den Einfluss von Kontaktquantität und -qualität auf Vorurteile untersucht und dabei simultan die Mediatoren Intergruppenangst und kollektive Bedrohung berücksichtigt.

Zu SEM zählt auch die *Konfirmatorische Faktorenanalyse* (englisch „confirmatory factor analysis“, CFA), mit der im Gegensatz zur explorativen Faktorenanalyse (s. Kurs 3443) eine angenommene faktorielle Struktur (ein sogenanntes Messmodell) überprüft werden kann.

Im Rahmen *multipler Gruppenvergleiche* ermöglicht SEM die Überprüfung von Messinvarianz und Populationsheterogenität. Grundlegende Frage bei der Prüfung von Messinvarianz ist, ob die den manifesten Variablen zugrunde liegenden latenten Variablen in den verschiedenen Gruppen wirklich gleich sind. Bei der Prüfung von Populationsheterogenität geht es um die Frage, ob die Beziehungen zwischen und das absolute Ausmaß in den latenten Variablen zwischen den Gruppen vergleichbar sind. Mittels eines multiplen Gruppenvergleichs können zwei oder mehr Gruppen in Hinblick auf die Parameter eines SEM verglichen werden, um entsprechend Aussagen über Messinvarianz und Populationsheterogenität treffen zu können. So haben Mewes et al. (2010) mittels eines multiplen Gruppenvergleichs die Messinvarianz zweier kurzer Screeningverfahren für depressive und somatoforme Störungen geprüft und hierzu die Messmodelle zwischen einer deutschen Stichprobe und einer Stichprobe von MigrantInnen verglichen.

Mittels unterschiedlicher Varianten von SEM lassen sich auch Längsschnittdaten auswerten. Ist die zeitliche Beziehung zwischen mehreren Merkmale von Interesse, können im Rahmen von SEM sogenannte *cross-lagged Modelle* geschätzt

werden. So haben Swart, Hewstone, Christ und Voci (2011) ein latentes cross-lagged Modell verwendet, um den Einfluss von Intergruppenkontakt auf unterschiedliche Indikatoren für Vorurteile über die Zeit zu untersuchen. Stehen dagegen die absoluten Veränderungen und interindividuelle Unterschiede in diesen Veränderungen im Mittelpunkt, kann dies mit *latenten Wachstumskurvenmodellen* oder „*Latent Change Score*“-Modellen untersucht werden.

Hierarchische Lineare Modelle (HLM) finden Anwendung bei sogenannten hierarchischen Datenstrukturen, die sich dadurch auszeichnen, dass auf mehreren Analyseebenen Beobachtungen vorgenommen wurden. Ein typisches Beispiel aus der Psychologie für eine solche hierarchische Datenstruktur wäre im Bereich der Bildungspsychologie die Untersuchung von SchülerInnen (Individualebene) aus unterschiedlichen Klassen (Klassenebene). In diesem Fall sind die SchülerInnen genestet in Klassen. Mittels HLM lassen sich nun die Daten auf diesen unterschiedlichen Analyseebenen simultan analysieren, um eine abhängige Variable, welche auf der untersten Analyseebene erhoben wurde, mit einer Reihe von unabhängigen Variablen, die auf allen Ebenen angesiedelt sein können, vorherzusagen. Jonkmann, Trautwein und Lüdtke (2009) haben den Effekt sozialer Dominanz von SchülerInnen auf unterschiedliche schulische Outcomes untersucht und sich hierbei besonders für die moderierende Wirkung der Normativität solchen Verhaltens innerhalb der unterschiedlichen Klassen auf diese Effekte interessiert.

Auch *Längsschnittdaten* können mit *HLM* ausgewertet werden. Im Vergleich zu klassischen Auswertungsverfahren für Längsschnittdaten wie z.B. die Varianzanalyse mit Messwiederholung (vgl. Kurs 3443) sind im Rahmen von HLM weniger restriktive Annahmen notwendig, so dass HLM eine wichtige Alternative zu diesen klassischen Auswertungsverfahren darstellen.

Bevor SEM oder HLM bzw. ganz allgemein statistische Verfahren zur Beantwortung von Fragestellungen herangezogen werden, sollte zunächst ein sogenanntes *Datenscreening* durchgeführt werden, im Rahmen dessen die Voraussetzungen der Verfahren geprüft und nach Ausreißern gesucht werden sollte (s. Vorlesung 4 im Kurs 3443). Ein weiterer wichtiger Aspekt betrifft die Frage des Umgangs mit *fehlenden Werten*. Es stehen unterschiedliche Methoden zur Verfügung, um mit fehlen Werten umzugehen. In der Praxis werden häufig solche Methoden verwendet, die zwar leicht einsetzbar, aber in der Regel nicht angebracht sind. Mittlerweile stehen der/dem AnwenderIn in vielen gängigen Statistikpaketen, u.a. auch SPSS und AMOS, bessere Alternativen zur Verfügung.

Schließlich wird mit dem *Bayesianischen Ansatz* eine alternative Möglichkeit für die Hypothesentestung vorgestellt. Die inferenzstatistische Absicherung bei allen in den Kursen 3443 und 3444 behandelten Verfahren beruht (üblicherweise) auf dem sogenannten frequentistischen Ansatz, der i.d.R. in den Statistik-Vorlesungen im Bachelor-Studiengang behandelt wird. Der frequentistische Ansatz liefert die Wahrscheinlichkeit der Daten gegeben einer Hypothese, die Bayes-Statistik dagegen die Wahrscheinlichkeit von Hypothesen gegeben den Daten. Trotz dieses of-

fenkundigen Vorteils des Bayesianischen Ansatzes ist die Bayes-Statistik noch wenig in der Psychologie verbreitet. Durch eine immer bessere Verfügbarkeit in Statistik-Paketen ist mit einer rasch steigenden Verwendung zu rechnen. Daher sollen im Rahmen der Vorlesung Grundkenntnisse vermittelt werden.

In den nächsten Abschnitten werden die Inhalte der Vorlesungen sowie die jeweils zugehörige Pflichtliteratur vorgestellt. Der Studienbrief selbst, wie bereits eingangs erwähnt, dient nicht zur Erarbeitung der Inhalte, sondern gibt lediglich einen Überblick.

2 Überblick über die Vorlesungsinhalte

2.1 Strukturgleichungsmodelle

Insgesamt gibt es zum Thema Strukturgleichungsmodelle drei Online-Vorlesungen mit jeweils einem zusätzlichen Online-Tutorial, in dem die Umsetzung ausgewählter Inhalte aus der Vorlesung in AMOS demonstriert wird. AMOS ist eine spezielle Software für Strukturgleichungsmodelle, die über einen grafischen Input verfügt, wodurch die Spezifikation von SEM relativ einfach ist.

In der ersten Online-Vorlesung V11 „*SEM I*“ werden die grundlegenden Begriffe vorgestellt, die Grundlagen von SEM behandelt und die zentralen Schritte bei der Modellspezifikation und –schätzung erörtert. Hierzu zählen die Modelidentifikation, die Modellschätzung sowie die Bewertung der Modellgültigkeit. Es wird auch auf die Prüfung alternativer Modelle eingegangen und die Problematik äquivalenter Modelle diskutiert.

In der zweiten Online-Vorlesung V12 „*SEM II*“ geht es um den Multiplen Gruppenvergleich. Es werden unterschiedliche Kriterien der Messinvarianz und Populationsheterogenität vorgestellt und das generelle Vorgehen bei einem multiplen Gruppenvergleich vorgestellt.

In der dritten Online-Vorlesung V13 „*SEM III*“ werden mit cross-lagged-Modellen, Latenten Wachstumskurvenmodellen und „Latent Change Score“-Modellen unterschiedliche Varianten von SEM für die Auswertung von Längsschnittdaten behandelt.

Die Pflichtlektüre für die Vorlesung V11 bilden die Kapitel 5 bis 10 aus dem Lehrbuch von Kline (2011). Grundlage für die Vorlesung V12 ist das Kapitel 7 aus dem Lehrbuch von Brown (2006). Für die Vorlesung V13 bilden relevante Abschnitte (für nähere Angaben, s. V13) aus drei englischsprachigen Artikel die Pflichtliteratur (Ferrer & McArdle, 2003; Hertzog & Nesselroade, 2003; Little, Preacher, Selig, & Card, 2007).

2.2 Hierarchische Lineare Modelle

Zum Themenblock Hierarchische Lineare Modelle gibt es ebenfalls drei Vorlesungen mit jeweils einem Tutorial. In der Vorlesung V14 „*HLM I*“ werden die Grundlagen von HLM dargestellt. Aufbauend hierauf geht es in der V15 „*HLM II*“ um die Frage der Zentrierung der unabhängigen Variablen in HLM sowie um das Thema Mediation. Schließlich wird in V16 „*HLM III*“ die Anwendung von HLM bei Längsschnittdaten diskutiert.

Pflichtliteratur für V11 sind die Kapitel 1, 2 und 4 aus dem Lehrbuch von Hox (2010). Grundlage und somit Pflichtliteratur für die V12 bilden zwei englischsprachige Artikel (Enders & Tofighi, 2007; Zhang, Zyphur, & Preacher, 2009), für V13 ein deutscher Artikel von Göllner, Gollwitzer, Heider, Zaby und Schröder (2010).

2.3 Umgang mit fehlenden Werten

In der Vorlesung V17 „*Fehlende Werte*“ werden unterschiedliche Methoden zum Umgang mit fehlenden Werten vorgestellt und diskutiert. In dem dazugehörigen Tutorial werden diese Methoden dann in SPSS an einem Datenbeispiel demonstriert.

Grundlage und für V17 somit Pflichtlektüre bildet ein deutscher Artikel von Lüdtke, Robitzsch, Trautwein und Köller (2007).

2.4 Bayes-Statistik

Die den Kurs 3444 abschließenden V18 „*Bayes-Statistik*“ gibt eine Einführung in den Bayesianischen Ansatz. Vor- und Nachteile im Vergleich zu dem frequentistischen Ansatz werden diskutiert. Zu dieser Vorlesung gibt es kein Tutorial.

Pflichtlektüre sind das Kapitel 19 aus dem Lehrbuch von Sedlmeier und Renkewitz (2013) und ein englischsprachiger Artikel von van de Schoot et al. (2013).

3 Literatur

3.1 Pflichtliteratur

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121-138.

Ferrer, E., & McArdle, J. J. (2003). Alternative structural models for multivariate longitudinal data analysis. *Structural Equation Modeling, 10*, 493-524.

Göllner, R., Gollwitzer, M., Heider, J., Zaby, A., & Schröder, A. (2010). Auswertung von Längsschnittdaten mit hierarchischen linearen Modellen. *Zeitschrift für Klinische Psychologie und Psychotherapie, 39*, 179-188.

Hertzog, C., & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology and Aging, 18*, 639-657.

Hox, J. J. (2010). *Multilevel analysis. Techniques and applications* (2nd ed.). New York, NY: Routledge. Kapitel 1, 2, 4.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development, 31*, 357-365.

Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen. *Psychologische Rundschau, 58*, 103-117.

Sedlmeier, P., & Renkewitz, F. (2013). *Forschungsmethoden und Statistik: Ein Lehrbuch für Psychologen und Sozialwissenschaftler* (2. überarbeitete und erweiterte Auflage). München, Deutschland: Pearson.

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods, 12*, 695-719.

Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2013). A gentle introduction to Bayesian Analysis: Applications to research in child development. *Child Development, 85*, 842-860.

3.2 Weitere Literatur (nicht prüfungsrelevant)

Jonkmann, K., Trautwein, U., & Lüdtke, O. (2009). Social dominance in adolescence – The moderating role of the classroom context and behavioral heterogeneity. *Child Development, 80*, 338-355.

Mewes, R., Christ, O., Rief, W., Glaesmer, A. Martin, E., & Brähler, E. (2010). Sind Vergleiche im Depressions- und Somatisierungsausmaß zwischen Migranten und Deutschen möglich?: Eine Überprüfung der Messinvarianz für den PHQ-9 und PHQ-15. *Diagnostica*, *56*, 230-239.

Swart, H., Hewstone, M., Christ, O., & Voci, A. (2011). Affective mediators of intergroup contact: A longitudinal analysis in South Africa. *Journal of Personality and Social Psychology*, *101*, 1221-1238.

Tausch, N., Hewstone, M., Kenworthy, J., Cairns, E., & Christ, O. (2007). Cross-community contact, intergroup attitudes and trust in Northern Ireland: The mediating roles of individual-level vs. group-level threats and the moderating role of social identification. *Political Psychology*, *28*, 53-68.