

Oliver Christ

Multivariate Verfahren und computergestützte Datenanalyse I

kultur- und
sozialwissenschaften



FernUniversität in Hagen

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 und Evaluation an der FernUniversität Hagen. Seit 2007 ist er Studienrat im Hochschuldienst für Psychologische Methodenlehre am Fachbereich Psychologie der Philipps-Universität Marburg.

Multivariate Verfahren und computergestützte Datenanalyse I

Stand: März 2014

Zusammenfassung

Der Kurs „Multivariate Verfahren und computergestützte Datenanalyse I“ (Kurs 3443) behandelt grundlegende statistische Verfahren zur Auswertung multivariater Daten in der Psychologie. Der Schwerpunkt liegt bei der multiplen Regression, wobei verschiedene Anwendungsmöglichkeiten zur Auswertung von Mediations- und Moderationshypothesen behandelt werden. Weiterhin werden die Kovarianzanalyse und die explorative Faktorenanalyse vorgestellt. Im Rahmen der Kovarianzanalyse wird auch auf die varianzanalytische Auswertung von sogenannten Split-Plot-Designs eingegangen, die in den Anwendungsfächern der Psychologie häufig zum Einsatz kommen. Neben dem Verständnis der statistischen Grundlagen der Verfahren steht die praktische Anwendung im Zentrum, wobei die Umsetzung in SPSS anhand von Beispieldaten aus unterschiedlichen Bereichen der wissenschaftlichen Psychologie demonstriert wird.

Vorbemerkungen

Der Kurs 3443 wurde zusammen mit dem Kurs 3444 „Multivariate Verfahren und computergestützte Datenanalyse II“ 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 Hagen!

Hagen, im März 2014

PD Dr. Oliver Christ

Lernziele

- ✓ Sie haben sich Wissen über die statistischen Grundlagen und Voraussetzungen wichtiger multivariater Verfahren angeeignet.
- ✓ Sie haben eine allgemeine Kenntnis über das Vorgehen bei und die verwendeten Schätzverfahren der verschiedenen multivariaten 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 durchzuführen und die Ergebnisausgabe zu interpretieren.

Didaktisches Konzept

Das didaktische Konzept des Kurses 03443 „Multivariate Verfahren und computergestützte Datenanalyse I“ 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 zusätzlich Online-Tutorials, in denen die Umsetzung in SPSS 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 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 I.....	4
Zusammenfassung.....	4
Vorbemerkungen.....	4
Lernziele.....	5
Didaktisches Konzept	6
1 Einführung zum Kurs	8
1.1 Warum ein weiterer Methodenkurs?	8
1.2 Überblick über die behandelten multivariaten Verfahren	8
2 Überblick über die Vorlesungsinhalte	11
2.1 Multiple Regression	11
2.2 Kovarianzanalyse und Zweifaktorielle Varianzanalyse mit Messwiederholung	12
2.3 Explorative Faktorenanalyse	12
3 Literatur	13
3.1 Pflichtliteratur.....	13
3.2 Weitere Literatur (nicht prüfungsrelevant).....	13

1 Einführung zum Kurs

1.1 Warum ein weiterer Methodenkurs?

In den Bachelorstudiengängen zur Psychologie - in Hagen, aber auch anderswo - stellt die Methodenausbildung einen wichtigen Schwerpunkt dar. Studierende müssen meist mehrere Module zu Forschungsmethoden und Statistik belegen und sich mit einer Vielzahl statistischer Verfahren auseinandersetzen. In der Regel handelt es sich bei den behandelten statistischen Verfahren um bivariate Methoden, die für die Analyse von Zusammenhängen zwischen zwei Variablen (z.B. Produkt-Moment-Korrelation) oder Unterschieden in einer Variable geeignet sind (z.B. t-Test für unabhängige Stichproben). Für viele Fragestellungen in der Psychologie ist die Berücksichtigung von lediglich zwei Variablen aber unzureichend. Entsprechend müssen Verfahren verwendet werden, die für den *multivariaten* Fall geeignet sind.

Die Auseinandersetzung mit multivariaten Verfahren ist zentral für jedes wissenschaftliche Studium der Psychologie. Ohne eine fundierte Methodenausbildung ist weder eine spätere wissenschaftliche Tätigkeit noch die Ausübung der Psychologie als Profession außerhalb der Universität auf hohem Niveau möglich. Auch Psychologinnen und Psychologen, die selbst keine wissenschaftliche Forschung betreiben, sollten in der Lage sein, wissenschaftliche Befunde kritisch zu reflektieren, z.B. wenn sie bewerten sollen, ob eine neue Behandlung für eine Aufmerksamkeitsstörung bei Kindern wirklich effektiv ist. Dazu gehört eben auch die Bewertung der statistischen Validität einer Untersuchung, also der Frage nach der Angemessenheit der verwendeten Verfahren in der Untersuchung und der korrekten Interpretation der Ergebnisse.

Zusammen mit dem Schwesterkurs „Multivariate Verfahren und computergestützte Verfahren II“ (Kurs 3444) werden die zentralen multivariaten Verfahren vorgestellt. In dem vorliegenden Kurs sind dies die multiple Regression, die Kovarianzanalyse, die zweifaktorielle Varianzanalyse mit einem messwiederholten Faktor (Auswertung von Split-Plot-Designs) und die explorative Faktorenanalyse. In Kurs 3444 werden schließlich noch Strukturgleichungsmodelle und Hierarchische Lineare Modelle als komplexere multivariate Verfahren behandelt (für weitere Inhalte siehe den Studienbrief zu Kurs 3444).

1.2 Überblick über die behandelten multivariaten Verfahren

Die *multiple Regression*, die Studierende mit einem B.Sc.-Abschluss in Hagen bereits kennengelernt haben, stellt eines der am häufigsten verwendeten multivariaten Verfahren dar. Die klassische Fragestellung, bei der die multiple Regression zum Einsatz kommt, ist, welche Kombination von Prädiktorvariablen (unabhängigen Variablen) in welchem Maße (wie gut) eine Kriteriumsvariable vorhersagen

kann. Beispielsweise hat Bernd Marcus, der das Lehrgebiet Arbeits- und Organisationspsychologie am hiesigen Psychologischen Institut leitet, in einer Publikation mit Heinz Schuler (Marcus & Schuler, 2004) die multiple Regression verwendet, um den Stellenwert von Selbstkontrolle in der Vorhersage von kontraproduktivem Verhalten im Vergleich zu anderen Prädiktoren zu untersuchen.

Die multiple Regression kann aber auch verwendet werden, um Mediationshypothesen, Moderationshypothesen und deren Kombination (Moderierte Mediation) zu überprüfen. In Mediationshypothesen wird der Frage nachgegangen, ob der Zusammenhang zwischen einer Variable X (unabhängige Variable bzw. Prädiktorvariable) und einer Variable Y (abhängige Variable bzw. Kriteriumsvariable) durch eine (M) oder mehrere Mediatorvariablen (M1, M2,...) vermittelt wird. So analysierten Christel Salewski und Manja Vollmann vom Lehrgebiet Gesundheitspsychologie zusammen mit Kollegen (Vollmann, Scharloo, Langguth, Kalkouskaya & Salewski, 2014) unterschiedliche Dimensionen des subjektiven Krankheitskonzeptes (z.B. wahrgenommene Konsequenzen der Erkrankung, wahrgenommene Kontrolle) als mögliche Mediatoren des Zusammenhangs zwischen dispositionellem Optimismus und psychischer Gesundheit.

In Moderationshypothesen wird geprüft, ob die Stärke (und evtl. auch die Richtung) des Zusammenhangs zwischen X und Y von der Ausprägung in einer Moderatorvariable Z abhängt. So konnten Stefan Stürmer und Bierte Siem vom Lehrgebiet Sozialpsychologie zusammen mit Kollegen (Stürmer, Snyder, Kropp & Siem, 2006) mittels einer moderierten Regression zeigen, dass Empathie insbesondere dann die Intention zur Hilfe beeinflusst, wenn die Person, der geholfen werden sollte, der Eigengruppe des Helfers/der Helferin und nicht einer fremden Gruppe angehörte.

Mittels einer *Kovarianzanalyse* (ANCOVA) kann der Einfluss einer Variable X auf das Kriterium Y untersucht werden bei gleichzeitiger Kontrolle von weiteren Prädiktorvariablen, die in der Regel Störvariablen darstellen. Anette Rohmann von Lehrgebiet Community Psychology und KollegInnen (Ramelli, Florack, Kosic & Rohmann, 2013) untersuchten den Zusammenhang zwischen der Kommunikationskompetenz und anfänglichen Kontakten mit Mitgliedern der Aufnahmegesellschaft von Immigranten auf deren Akkulturationsorientierungen. Die AutorInnen verwendeten die ANCOVA im regressionsanalytischen Kontext, um für den Einfluss von soziodemographischen Variablen (Alter, Geschlecht) zu kontrollieren.

In vielen Anwendungsfächern der Psychologie, vor allen Dingen, wenn es um die Evaluation von Maßnahmen und Interventionen geht (vgl. Kurs 3445) werden sogenannte Split-Plot-Designs angewendet. Bei einem Split-Plot-Design handelt es sich um einen zweifaktoriellen Versuchsplan mit einem messwiederholten Faktor. Entsprechend wird üblicherweise eine *zweifaktorielle Varianzanalyse mit einer messwiederholten Faktor* zur Auswertung herangezogen. Anette Rohmann

und Kolleginnen (Klinge, Rohmann & Piontkowski, 2009) haben einen Online-Kurs zur interkulturellen Sensibilisierung evaluiert und dabei in einer Studie (Studie 2) diesen Kurs mit einer alternativen Intervention verglichen. Dabei erfasste ein Faktor (der between-subject Faktor) die Zugehörigkeit zu einer der beiden Interventionen. Der andere, messwiederholte Faktor kodiert die Messung der kulturellen Sensibilität vor und nach Durchführung der Intervention. Die Ergebnisse einer ANOVA mit einem messwiederholten Faktor zeigten, dass sich die beiden Interventionen in ihrer Effektivität nicht unterschieden. Alternativ, und unter bestimmten Umständen der Varianzanalyse vorzuziehen, können Hierarchisch Lineare Modelle für die Auswertung von solchen Versuchsplänen verwendet werden, die im Kurs 3444 behandelt werden.

Schließlich werden wir mit der *explorativen Faktorenanalyse* ein weiteres, insbesondere für die Theorienbildung in der Psychologie wichtiges multivariates Verfahren kennenlernen. Ziel der explorativen Faktorenanalyse ist es, die latente, d.h. nicht direkt beobachtbare dimensionale Struktur aufzudecken, die dem Zusammenhangsmuster mehrerer Variablen (z.B. Fragebogenitems) zugrunde liegt. Bernd Marcus und Kolleginnen (Schütz, Marcus & Sellin, 2004) haben eine Faktorenanalyse durchgeführt, um die dimensionale Struktur einer deutschen Übersetzung des Narcissistic Personality Inventory zu überprüfen. Dabei handelt es sich um ein Messinstrument, mittels dessen Narzissmus als Persönlichkeitsmerkmal erfasst werden soll. Es geht also um die üblicherweise theoriefreie Bestimmung der Anzahl von Faktoren bzw. Dimensionen, die nötig sind, um das Zusammenhangsmuster mehrerer Variablen hinreichend zu beschreiben. Im Kurs 3444 wird die konfirmatorische Faktorenanalyse behandelt, bei der es im Gegensatz zur explorativen Faktorenanalyse um die Bestätigung einer theoretisch angenommenen latenten Struktur geht.

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 Multiple Regression

Insgesamt gibt es zum Thema Multiple Regression vier Online-Vorlesungen mit jeweils einem zusätzlichen Online-Tutorial, in dem die Umsetzung ausgewählter Inhalte aus der Vorlesung in SPSS demonstriert wird.

In der ersten Online-Vorlesung VI1 „*Multiple Regression I*“ wird nach einer kurzen Wiederholung der einfachen linearen Regression das Modell der multiplen Regression vorgestellt und auf die Berechnung und Interpretation der Regressionskoeffizienten eingegangen. Weiterhin werden das Problem der Multikollinearität und die Rolle von Suppressorvariablen angesprochen.

Die zweite Online-Vorlesung VI2 „*Multiple Regression II*“ behandelt unterschiedliche Möglichkeiten der Auswahl der unabhängigen Variablen und geht auf die Prüfung von Mediatorhypothesen mittels der multiplen Regression ein.

In der dritten Online-Vorlesung VI3 „*Multiple Regression III*“ wird die Prüfung von Moderatorhypothesen vorgestellt, wobei hierbei sowohl kontinuierliche, aber auch kategoriale Moderatoren berücksichtigt werden. Es wird auch auf die Prüfung von konditionalen indirekten Effekten eingegangen, bei der simultan Mediator- und Moderatorhypothesen geprüft werden.

Schließlich wird in der letzten Vorlesung VI4 „*Multiple Regression IV*“ zum Thema multiple Regression auf Annahmen und Voraussetzungen der multiplen Regression und deren Überprüfung eingegangen.

Die Pflichtliteratur für die Vorlesungen VI1, VI2 und VI4 ist das Kapitel 18 aus dem Lehrbuch von Eid, Gollwitzer und Schmitt (2013). Für die Vorlesung VI3 bilden zwei englischsprachige Artikel die Pflichtliteratur (Preacher & Hayes, 2008; Preacher, Rucker & Hayes, 2007). In VI2 wird auch auf einen kurzen Artikel von Fiedler, Schott und Meiser (2011) verwiesen, der ebenfalls zur Pflichtliteratur gehört. Studierende, die in ihrem Bachelor-Studium nicht die einfache lineare Regression kennengelernt haben, bekommen mit Kapitel 16 aus dem Lehrbuch von Eid und Kollegen (2013) eine gute didaktische Einführung, deren Lektüre ich daher zusätzlich empfehle.

2.2 Kovarianzanalyse und Zweifaktorielle Varianzanalyse mit Messwiederholung

In der Vorlesung V15 „*Kovarianzanalyse und Zweifaktorielle Varianzanalyse mit Messwiederholung*“ wird zum einen die Kovarianzanalyse im regressionsanalytischen Kontext dargestellt und auf Probleme im Falle von quasi-experimentellen Designs eingegangen. Weiterhin wird mit der zweifaktoriellen Varianzanalyse mit Messwiederholung auf einem Faktor ein multivariates Verfahren vorgestellt, welches für die Auswertung von Split-Plot-Designs geeignet ist. Split-Plot-Designs finden häufig in der Evaluationsforschung Anwendung (s. Kurs 3445).

Pflichtliteratur bildet wieder das Lehrbuch von Eid, Gollwitzer und Schmitt (2013): Für die Kovarianzanalyse Kapitel 18.12.1 und 18.12.2 und für die zweifaktorielle Varianzanalyse mit Messwiederholung auf einem Faktor Kapitel 14.2.2.

2.3 Explorative Faktorenanalyse

Die explorative/exploratorische Faktorenanalyse (EFA) ist Gegenstand von insgesamt zwei Online-Vorlesungen. In Vorlesung V16 „*Explorative Faktorenanalyse I*“ wird der Grundgedanke einer EFA erörtert und verschiedene Methoden der EFA vorgestellt. Zusätzlich werden die zentralen Kennwerte der EFA erläutert und das generelle Vorgehen skizziert.

Vorlesung V17 „*Explorative Faktorenanalyse II*“ vertieft das Vorgehen bei der EFA, wobei verschiedene Extraktionskriterien und Rotationsmethoden vorgestellt werden und wichtige Aspekte für die Anwendung diskutiert. Schließlich werden die Voraussetzungen für eine EFA erläutert.

Für die EFA sind das Kapitel 6 aus dem Lehrbuch von Bühner (2010) und ein Originalartikel von Fabrigar, Wegener, MacCallum & Strahan (1999) Pflichtlektüre.

3 Literatur

3.1 Pflichtliteratur

Bühner, M. (2010). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson-Education.

Eid, M., Gollwitzer, M. & Schmitt, M. (2013). *Statistik und Forschungsmethoden* (3. Aufl.). Weinheim: Beltz.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272-299.

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, 47, 1231-1236.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40, 879–891.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, 42, 185-227.

3.2 Weitere Literatur (nicht prüfungsrelevant)

Marcus, B., & Schuler, H. (2004). Antecedents of counterproductive behavior at work: A general perspective. *Journal of Applied Psychology*, 89, 647-660.

Klinge, K., Rohmann, A., & Piontkowski, U. (2009). Intercultural sensitization with synthetic cultures: Evaluation of a computer based multimedia learning tool. *International Journal of Intercultural Relations*, 33, 507-515.

Hayes, A. F. (2012). *PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling* [White paper]. Retrieved from <http://www.afhayes.com/public/process2012.pdf>

Ramelli, M., Florack, A., Kosic, A., & Rohmann, A. (2013). Being prepared for acculturation: On the importance of the first months after immigrants enter a new culture. *International Journal of Psychology*, 48, 363-373.

Schütz, A., Marcus, B., & Sellin, I. (2004). Die Messung von Narzissmus als Persönlichkeitskonstrukt. Psychometrische Eigenschaften einer Lang- und einer Kurzform des Deutschen NPI (Narcissistic Personality Inventory). *Diagnostica*, 50, 202-218.

Stürmer, S., Snyder, M., Kropp, A., & Siem, B. (2006). Empathy-motivated helping: The moderating role of group membership. *Personality and Social Psychology Bulletin*, 32, 943-956.

Vollmann, M., Scharloo, M., Langguth, B., Kalkouskaya, N. & Salewski, C. (2014). Illness representations as mediators of the relationship between dispositional optimism and depression in patients with chronic tinnitus: A cross-sectional study. *Psychology and Health, 29*, 81-93.