

Christian Geiser

Datenanalyse mit Mplus

Eine anwendungsorientierte
Einführung

LEHRBUCH



CD-ROM

Christian Geiser

Datenanalyse mit Mplus

Christian Geiser

Datenanalyse mit Mplus

Eine anwendungsorientierte
Einführung



Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über
<<http://dnb.d-nb.de>> abrufbar.

1. Auflage 2010

Alle Rechte vorbehalten

© VS Verlag für Sozialwissenschaften | Springer Fachmedien Wiesbaden GmbH 2010

Lektorat: Kea S. Brahms

VS Verlag für Sozialwissenschaften ist eine Marke von Springer Fachmedien.

Springer Fachmedien ist Teil der Fachverlagsgruppe Springer Science+Business Media.

www.vs-verlag.de



Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Umschlaggestaltung: KünkelLopka Medienentwicklung, Heidelberg

Druck und buchbinderische Verarbeitung: Ten Brink, Meppel

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Printed in the Netherlands

ISBN 978-3-531-16393-2

Inhaltsverzeichnis

Vorwort	7
1 Datenaufbereitung in SPSS	11
1.1 Ländereinstellungen	12
1.2 Kodierung fehlender Werte	16
1.3 Abspeichern eines ASCII-Datensatzes für Mplus	21
2 Einlesen von Daten in Mplus	25
2.1 Einlesen von Individualdaten (Rohdaten)	25
2.1.1 Aufbau der basalen Mplus-Syntax und basic-Analyse	26
2.1.2 Mplus-Output zur basic-Analyse	29
2.2 Einlesen von Summary-Daten (Kovarianz- oder Korrelationsmatrizen)	37
3 Lineare Strukturgleichungsmodelle	41
3.1 Was sind lineare Strukturgleichungsmodelle?	41
3.2 Einfache Regressionsanalyse	44
3.3 Regressionsanalyse auf latenter Ebene	55
3.4 Konfirmatorische Faktorenanalyse (CFA)	65
3.4.1 Konfirmatorische Faktorenanalyse 1. Ordnung	65
3.4.2 Konfirmatorische Faktorenanalyse 2. Ordnung	71
3.5 Pfadmodelle und Mediatoranalyse	75
3.5.1 Einführung und manifeste Pfadanalyse	75
3.5.2 Manifeste Pfadanalyse in Mplus	78
3.5.3 Latente Pfadanalyse	85
3.5.4 Latente Pfadanalyse in Mplus	86
4 Strukturgleichungsmodelle der Veränderungsmessung	93
4.1 Latent-State-Analyse	93
4.1.1 LS- versus LST-Modelle	96
4.1.2 Analyse von LS-Modellen in Mplus	96
4.1.3 Modellierung indikatorsspezifischer Effekte	99
4.1.4 Überprüfung der Messinvarianz über die Zeit	107
4.2 Latent-State-Trait-Analyse	122
4.3 Autoregressive Modelle	131
4.3.1 Manifeste autoregressive Modelle	132
4.3.2 Latente autoregressive Modelle	138

4.4	Latent-Change-Modelle	151
4.5	Latente Wachstumskurvenmodelle	168
4.5.1	Latente Wachstumskurvenmodelle 1. Ordnung	169
4.5.2	Latente Wachstumskurvenmodelle 2. Ordnung	187
5	Mehrebenenregressionsmodelle	199
5.1	Einführung in die Mehrebenenanalyse	199
5.2	Spezifikation von Mehrebenenmodellen in Mplus	201
5.3	Option <code>twolevel basic</code>	203
5.4	Random-Intercept-Modelle	208
5.4.1	Nullmodell (Intercept-Only-Modell)	208
5.4.2	One-Way-Random-Effects-ANCOVA	212
5.4.3	Means-as-Outcomes-Modell	217
5.5	Random-Intercept-and-Slope-Modelle	222
5.5.1	Random-Coefficient-Regressionsanalyse	223
5.5.2	Intercepts-and-Slopes-as-Outcomes-Modell	227
6	Latent-Class-Analyse	235
6.1	Einführung in die Latent-Class-Analyse (LCA)	235
6.2	Spezifikation von LCA-Modellen in Mplus	238
6.3	Modellgütebeurteilung und Modellvergleiche	260
6.3.1	Absoluter Modellfit	260
6.3.2	Relativer Modellfit	265
6.3.3	Interpretierbarkeit	271
Anhang A: Zentrale Mplus-Befehle		273
Anhang B: Häufige Fehler bei der Mplus-Spezifikation und „Troubleshooting“		278
Anhang C: Literaturhinweise zu Mplus		281
Literatur		283
Register		289

Vorwort

Multivariate statistische Auswertungsmethoden – mit oder ohne Verwendung latenter Variablen – werden in den empirischen Sozialwissenschaften zunehmend häufiger eingesetzt. Gilt es doch in vielen Studien Fragestellungen bezüglich komplexer Zusammenhangsstrukturen zwischen einer Vielzahl von Variablen zu untersuchen. Da es hierbei erwünscht ist, zufällige Messfehlereinflüsse zu berücksichtigen, haben multivariate statistische Verfahren mit latenten Variablen in den letzten Jahrzehnten zunehmende Verbreitung in den Sozialwissenschaften gefunden. Statistische Modelle mit latenten Variablen ermöglichen es, Messfehlereinflüsse in der Analyse explizit zu berücksichtigen und die unerwünschten Effekte von Messfehlern auf statistische Ergebnisse – wie beispielsweise die verzerrte Schätzung von Korrelations- oder Regressionskoeffizienten – zu vermeiden.

Warum hat sich der Autor dazu entschlossen, ausgerechnet ein Buch zum Computerprogramm Mplus zu schreiben? Mplus ermöglicht die Analyse einer großen Vielzahl unterschiedlicher multivariater statistischer Modelle und Verfahren – mit und ohne latente Variablen – in einem einzigen Modellierungs-Framework (Muthén, 2002; Muthén & Muthén, 1998-2007; Muthén & Muthén, 2007). Eine Besonderheit ist dabei, dass in Mplus neben Modellen mit kontinuierlichen latenten Variablen („latenten Faktoren“) auch Modelle mit kategorialen latenten Variablen (latenten Klassenvariablen) analysiert werden können. Derartige *Latent-Class-Modelle* (siehe Kapitel 6) lassen sich in Mplus zudem mit dem Ansatz linearer Strukturgleichungsmodelle (siehe Kapitel 3) verbinden (z.B. bei der Analyse sogenannter *Growth-Mixture-Modelle*, bei denen unterschiedliche „Wachstumstypen“ Gegenstand der Analyse sind).

Daneben erweist sich Mplus als sehr flexibel, was die Behandlung von Datensituationen anbetrifft, welche die strengen Annahmen konventioneller Schätzmethoden verletzen. So macht beispielsweise das Maximum-Likelihood-Schätzverfahren zur Schätzung der Modellgrößen in linearen Strukturgleichungsmodellen die Annahme der multivariaten Normalverteilung der Variablen. Außerdem wird vorausgesetzt, dass die Beobachtungseinheiten (in der Regel die befragten oder getesteten Personen) eine Zufallsstichprobe darstellen und voneinander unabhängig sind. Die Normalverteilungsannahme ist in den empirischen Sozialwissenschaften in vielen Fällen nicht haltbar, insbesondere dann nicht, wenn einzelne Items mit lediglich ordinalem Skalenniveau als Indikatoren für latente Faktoren verwendet werden sollen (Finney & DiStefano, 2006). Auch die Annahme der Unabhängigkeit der Beobachtungseinheiten voneinander ist in vielen Fällen verletzt, etwa dann, wenn die gesammelten Daten eine hierarchische (Mehrebenen-)Struktur aufweisen (vgl. Kapitel 5). Mplus bietet für derartige Fälle eine Vielzahl spezieller robuster Schätzmethoden an, so dass beispielsweise auch ordinale und nicht-normalverteilte Variablen sowie geclusterte (Mehrebenen-)Daten

korrekt analysiert werden können (diese Themen werden im Fortsetzungsband von Geiser, Crayen & Enders [in Vorbereitung] behandelt).

Mplus bietet neben seiner Vielseitigkeit und Flexibilität gegenüber vielen anderen Programmen auch den Vorteil einer hohen Benutzerfreundlichkeit. Zwar handelt es sich bei Mplus um ein vorwiegend syntaxbasiertes Programm, das heißt Modelle und Analysen werden nicht über Menüfunktionen oder Pfaddiagramme spezifiziert, wie dies in anderen Programmen z.T. möglich ist, sondern mit Hilfe einer Befehlssprache. Die Mplus-Befehlssprache zeichnet sich dabei jedoch durch ein hohes Maß an Einfachheit und Klarheit aus. Komplexe Modelle lassen sich durch relativ wenige, leicht zu verstehende Befehle spezifizieren, ohne dass der Benutzer beispielsweise Experte in Matrixalgebra sein oder spezielle Programmiererfahrungen haben muss. Und nicht zuletzt sah sich der Autor auch deshalb motiviert, ein Buch zu Mplus zu schreiben, weil er in zahlreichen Seminaren und Workshops den steigenden Bedarf der Nutzer an einem grundlegenden Einführungsband zu diesem Programm regelmäßig zu spüren bekommen hat.

Das vorliegende Buch liefert eine Einführung in die Datenanalyse mit Mplus sowie in die Theorie und praktische Anwendung von vier in den empirischen Sozialwissenschaften besonders verbreiteten multivariaten Verfahren. Es werden „konventionelle“ lineare Strukturgleichungsmodelle (Regressions-, konfirmatorische Faktoren- und Pfadanalysen für Querschnittsdaten) sowie – aufgrund ihrer zunehmenden Verbreitung in den Sozialwissenschaften – auch Strukturgleichungsmodelle speziell für die Auswertung von Längsschnittsdaten (von identischen Merkmalsträgern zu mehreren Messgelegenheiten erhobene Daten, Messwiederholungsdaten) und die damit verbundenen zentralen Modellierungsaspekte behandelt. Hierbei wird auf Latent-State-Modelle, Latent-State-Trait-Modelle, manifeste und latente autoregressive Modelle, latente Differenzenmodelle und latente Wachstumskurvenmodelle eingegangen. Außerdem behandelt das Buch die Analyse von Mehrebenendaten in Mplus, wobei aus Platzgründen nur Modelle für manifeste Variablen (d.h. Mehrebenenregressionsmodelle) berücksichtigt werden konnten. Im letzten Kapitel wird die Durchführung der klassischen Latent-Class-Analyse in Mplus besprochen.

Ziel des vorliegenden Buches ist es, den Leserinnen und Lesern die Benutzung von Mplus von Grund auf in einer praxisorientierten und anwenderfreundlichen Weise nahe zu bringen. Hierzu wird durchgängig mit realen Datenbeispielen gearbeitet, die direkt auf vergleichbare eigene Problemstellungen übertragen werden können. Zahlreiche Screenshots und Outputexzerpte führen Schritt für Schritt durch die gezeigten Analysen und die Interpretation der Ergebnisse. Eine weitere Hilfe zur Übertragung auf eigene Datenanalysen liefert die Begleit-CD zum Buch, auf welcher sämtliche im Buch besprochenen Datenbeispiele, Mplus-Syntax-Dateien mit den entsprechenden Modellspezifikationen sowie die dazugehörigen Mplus-Output-Dateien enthalten sind. In die Syntaxdateien wurden zahlreiche Kommentare aufgenommen, um den Leserinnen und Lesern den Umgang mit der Syntax zu erleichtern. Viele der verwendeten Datenbeispiele sind auf eine geringe Anzahl von Variablen begrenzt und können daher auch mit der über die Mplus-Homepage frei verfügbaren Demoversion des Programms (siehe <http://statmodel.com/demo.shtml>) nachgerechnet werden.

Das Buch beginnt bei der Frage der Datenaufbereitung und des Datenimports für bzw. in das Programm Mplus. Aufgrund seiner großen Verbreitung (und nicht aufgrund einer etwaigen Präferenz des Autors) wird hierzu die Datenaufbereitung mit Hilfe des allgemeinen

Statistikprogramms SPSS gezeigt. Das Vorgehen kann jedoch problemlos auf alle gängigen statistischen Programme wie z.B. SAS, STATA, STATISTICA, Systat usw. übertragen werden.

Das vorliegende Buch versteht sich nicht als reiner Syntaxguide und stellt keinen Ersatz für, sondern eine Ergänzung zum Mplus User's Guide (Muthén & Muthén, 1998-2007; <http://statmodel.com/ugexcerpts.shtml>) dar. Es möchte dem Leser nicht nur zeigen, wie er welches statistische Modell mit Hilfe der Mplus-Syntax spezifizieren kann, sondern auch Hilfestellungen bei der Wahl geeigneter Modellierungsstrategien sowie bei der Interpretation der Schätzergebnisse liefern. Zudem soll das Buch dem Anwender helfen, typische, bei der Modellierung komplexer multivariater Daten auftretende Probleme (z.B. die Frage der Messinvarianztestung in Längsschnittuntersuchungen oder das Problem der Vermeidung lokaler Likelihood-Maxima in der Latent-Class-Analyse) zu erkennen und zu lösen. Aus diesem Grund werden die besprochenen statistischen Verfahren sowie aus Sicht des Autors sinnvolle und bewährte Modellierungsstrategien an geeigneten Stellen ebenfalls diskutiert. Bei besonders wichtigen Fragen geben Textboxen nochmals einen Überblick über wichtige Probleme und entsprechende Lösungsstrategien.

Trotz der Fokussierung des Bandes auf die praktische Anwendung der einzelnen Verfahren und Modelle in Mplus wird in der Einleitung zu jedem Kapitel auch noch einmal in knapper Form auf den theoretischen Hintergrund der statistischen Modelle und Verfahren eingegangen, wobei das Buch natürlich keinen Ersatz für ein Statistik-Lehrbuch sowie die jeweilige Spezialliteratur darstellt. Aus diesem Grund wird zu Beginn eines jeden Kapitels auf zentrale weiterführende Literatur zu dem jeweiligen Verfahren hingewiesen. Auch bei spezifischen Problemstellungen (wie z.B. der Frage nach der Testung von Mediatoreffekten in Kapitel 3.5) wird auf die entsprechende Spezialliteratur verwiesen. Als allgemeines Statistik-Lehrbuch, in welchem die Theorie der Mehrzahl der in diesem Buch beschriebenen Verfahren im Detail behandelt wird, sei dem Leser das Buch von Eid, Schmitt und Gollwitzer (2010) empfohlen.

Die Auswahl der in diesem Buch besprochenen statistischen Verfahren erfolgte subjektiv aufgrund der Erfahrungen des Autors aus seiner eigenen Forschungspraxis sowie aus zahlreichen von ihm gehaltenen Workshops, in welchen er einiges über die statistischen „Bedürfnisse“ von Sozialwissenschaftlern (insbesondere Psychologen) gelernt hat (wofür er ausgesprochen dankbar ist). Eine Einführung kann naturgemäß keine erschöpfende Darstellung aller relevanten Verfahren und Modellierungsaspekte liefern. Der Autor hofft jedoch, dass es ihm gelungen ist, eine für eine möglichst große Zahl von empirischen Sozialwissenschaftlern nützliche Auswahl an Modellen und Methoden getroffen zu haben. Komplexere multivariate Auswertungsmethoden und Modellierungsaspekte in Mplus werden im Fortsetzungsband von Geiser et al. (in Vorbereitung) behandelt.

Das vorliegende Buch ist entstanden aus den Lehrerfahrungen des Autors in zahlreichen Seminaren und Workshops zur angewandten multivariaten Statistik und Datenanalyse mit Mplus. Der Autor ist seinen Studierenden sowie den zahlreichen Workshopteilnehmerinnen und -teilnehmern sehr dankbar für die umfangreichen Erfahrungen, die er in diesen Veranstaltungen sammeln konnte und das Feedback der Teilnehmerinnen und -teilnehmer.

Der Autor dankt außerdem Martin Corth, David Cole und dem Deutschen Zentrum für Altersfragen (DZA) in Berlin für die Überlassung von Beispieldatensätzen, Christopher Marx für die Hilfe bei der Aufbereitung des Datenbeispiels zur Pfadanalyse in Kapitel 3.5

sowie die Durchführung von Analysen und das Korrekturlesen dieses Kapitels, Claudia Crayen für ihre Hilfe bei der Erstellung von Abbildungen und für Korrekturen sowie Hen- riette Hunold, Anne Janssen, Tanja Kutscher, Maike Luhmann, Natalie Mallach und Tabea Reuter für zahlreiche Korrekturen und hilfreiche Anregungen. Sie alle haben dazu beigetragen, das Buch verständlicher und anwenderfreundlicher zu machen und etliche Fehler auszumerzen. Für alle verbleibenden Fehler und Schwächen bin ich natürlich allein verantwortlich. Bitte kontaktieren Sie mich, wenn Sie Fehler entdecken sollten oder Vorschläge zur Verbesserung des Buches machen möchten. Ich wünsche Ihnen viel Erfolg bei der Durch- führung und Veröffentlichung Ihrer eigenen multivariaten Auswertungen mit Mplus.

Berlin im Oktober 2009

Christian Geiser

1 Datenaufbereitung in SPSS

In diesem Kapitel wird besprochen, was zu beachten ist, wenn man einen SPSS-Datensatz zum Einlesen in Mplus aufbereiten möchte. Das Programm SPSS wurde als Beispiel gewählt, da es in den Sozialwissenschaften zu den am weitesten verbreiteten Statistik-Software-Packages zu zählen ist. Die gezeigten Schritte lassen sich jedoch problemlos auch auf andere Statistikprogramme wie z.B. SAS, STATA, STATISTICA, Systat usw. übertragen.

Mplus kann keine Datensätze im SPSS-eigenen Datenformat (Dateiendung *.sav) korrekt verarbeiten. Man kann aber leicht mit Hilfe von SPSS (oder anderen gängigen Statistikprogrammen) Rohdatensätze (Individualdaten) als einfache Textdateien (sog. ASCII-Daten, *.txt oder *.dat) abspeichern lassen, die dann von Mplus gelesen werden können. Zu empfehlen ist hierbei insbesondere die Verwendung eines tabulator-getrennten („tab-delimited“) Formates in den ASCII-Dateien. Bei diesem Format sind die einzelnen Variablen Spalten mit Tabulatorzeichen voneinander getrennt. Aber auch andere Formate sind möglich (siehe dazu Mplus User’s Guide; Muthén & Muthén, 1998-2007), wie auch das Einlesen von Summary-Daten (keine Rohdaten, sondern z.B. Kovarianz- oder Korrelationsmatrizen; siehe Kapitel 2.2).

Die aus Sicht des Autors schnellste, praktischste und fehlerunanfällige Variante der Aufbereitung von Individualdaten in SPSS und Abspeicherung einer ASCII-Datendatei wird im Folgenden Schritt für Schritt anhand des Beispieldatensatzes **KFT.sav** beschrieben. Es sei darauf hingewiesen, dass viele Wege nach Rom führen, d.h. das hier vorgeschlagene Verfahren ist nur eines von mehreren möglichen. Die Datendatei **KFT.sav** sowie eine SPSS-Syntaxdatei mit den entsprechenden Syntaxbefehlen befinden sich auf der Begleit-CD zum Buch.

Box 1.1: Datenbeispiel KFT.sav

Der Datensatz **KFT.sav** enthält vollständige Daten von 455 Schülern, die mit dem Kognitiven Fähigkeitstest (KFT; Heller, Gaedicke & Weinläder, 1976) getestet wurden. Der KFT ist ein schulnäher Intelligenztest. Die Schüler mussten jeweils Aufgaben zum logisch-schlussfolgernden Denken bearbeiten, wobei verschiedene Untertests mit verbalem, numerischem und figuralem Stimulusmaterial eingesetzt wurden: Zwei verbale Untertests (Variablen **kft_v1** und **kft_v3**), zwei numerische („quantitative“) Untertests (Variablen **kft_q1** und **kft_q3**), und zwei figurale („nonverbale“) Untertests (Variablen **kft_n1** und **kft_n3**). Die erreichten Punkte der Schüler wurden für jeden Untertest jeweils zu einem Summenscore aggregiert.

1.1 Ländereinstellungen

Zu beachten ist, dass Mplus als Dezimaltrennzeichen keine Kommas (so wie es im Deutschen üblich ist), sondern nur Punkte akzeptiert. Entsprechend muss darauf geachtet werden,

*KFT.sav [DatenSet1] - SPSS Daten-Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Fenster Hilfe

1 : kft_v1 | 18 | Sichtbar: 6 von 6

	kft_v1	kft_v3	kft_q1	kft_q3	kft_n1	kft_n3	var	var	var
10	6,00	9,00	9,00	3,00	7,00	12,00			
11	13,00	16,00	12,00	6,00	7,00	10,00			
12	7,00	7,00	7,00	4,00	7,00	5,00			
13	10,00	5,00	7,00	2,00	5,00	4,00			
14	5,00	5,00	10,00	4,00	5,00	4,00			
15	8,00	9,00	7,00	2,00	6,00	5,00			
16	17,00	13,00	12,00	5,00	16,00	12,00			
17	9,00	9,00	10,00	6,00	9,00	14,00			
18	12,00	10,00	8,00	5,00	5,00	7,00			
19	16,00	5,00	12,00	12,00	21,00	11,00			
20	20,00	14,00	15,00	10,00	15,00	13,00			
21	20,00	10,00	14,00	13,00	22,00	11,00			
22	7,00	6,00	9,00	8,00	6,00	5,00			
23	2,00	4,00	10,00	5,00	6,00	5,00			
24	17,00	14,00	16,00	9,00	20,00	9,00			
25	12,00	6,00	12,00	6,00	8,00	9,00			
26	14,00	5,00	8,00	7,00	9,00	9,00			
27	11,00	5,00	7,00	4,00	6,00	11,00			
28	7,00	4,00	10,00	4,00	5,00	8,00			
29	14,00	6,00	9,00	8,00	9,00	7,00			
30	9,00	10,00	12,00	6,00	10,00	7,00			
31	20,00	4,00	16,00	9,00	5,00	4,00			
32	14,00	3,00	15,00	8,00	7,00	8,00			
33			
34			
35			
36	8,00	3,00	16,00	3,00	2,00	3,00			
37	11,00	15,00	13,00	6,00	4,00	12,00			
38	4,00	2,00	8,00	4,00	6,00	10,00			

Datenansicht / Variablenansicht SPSS Prozessor ist bereit

Abbildung 1.1. Ausschnitt aus dem SPSS-Beispieldatensatz mit Schülerdaten für 6 Subskalen aus dem Kognitiven Fähigkeitstest (KFT; Heller, Gaedicke & Weinläder, 1976). Nachkommastellen werden noch mit Kommas angezeigt. Fehlende Werte sind als leere Zellen („systemdefiniert fehlend“) dargestellt (siehe Zeilen 33-35). Beides würde in Mplus zu Dateneinlesefehlern führen. Das SPSS-Programm muss nun zunächst wieder geschlossen werden, und es müssen die Ländereinstellungen geändert werden (siehe Abb. 1.2 und 1.3).

dass in der zu erzeugenden ASCII-Datei Nachkommastellen durch Punkte abgetrennt werden. Die Anzeige (Komma vs. Punkt) hängt in älteren SPSS-Versionen von den Ländereinstellungen ab. Ist unter dem Windows-Betriebssystem in der Systemsteuerung unter „Län-

dereinstellungen“ beispielsweise „Deutsch (Deutschland)“ eingestellt, so zeigt SPSS Kommas als Dezimaltrennzeichen an. Auch in der mit SPSS zu erzeugenden ASCII-Datei (siehe Abschnitt 3.3) würden sich dann Kommas befinden, was zwangsläufig zu Datenlesefehlern in Mplus führen würde. Dies können Sie in älteren SPSS-Versionen leicht überprüfen, indem Sie den *.sav Datensatz, den Sie in Mplus einlesen möchte, in SPSS öffnen (siehe Abb. 1.1). Wenn Sie im SPSS-Dateneditor Kommas sehen (wie in Abb. 1.1), so müssen Sie zunächst das SPSS-Programm wieder schließen und die Ländereinstellungen ändern, z.B. auf „Englisch (USA)“. Wenn Sie danach SPSS erneut öffnen, sehen Sie Punkte anstelle der Kommas als Dezimaltrennzeichen (siehe Abb. 1.4). Die entsprechenden Schritte werden in den folgenden Screenshots verdeutlicht. Das in neueren SPSS-Versionen mögliche einfachere Vorgehen mittels Syntax wird im Anschluss gezeigt.

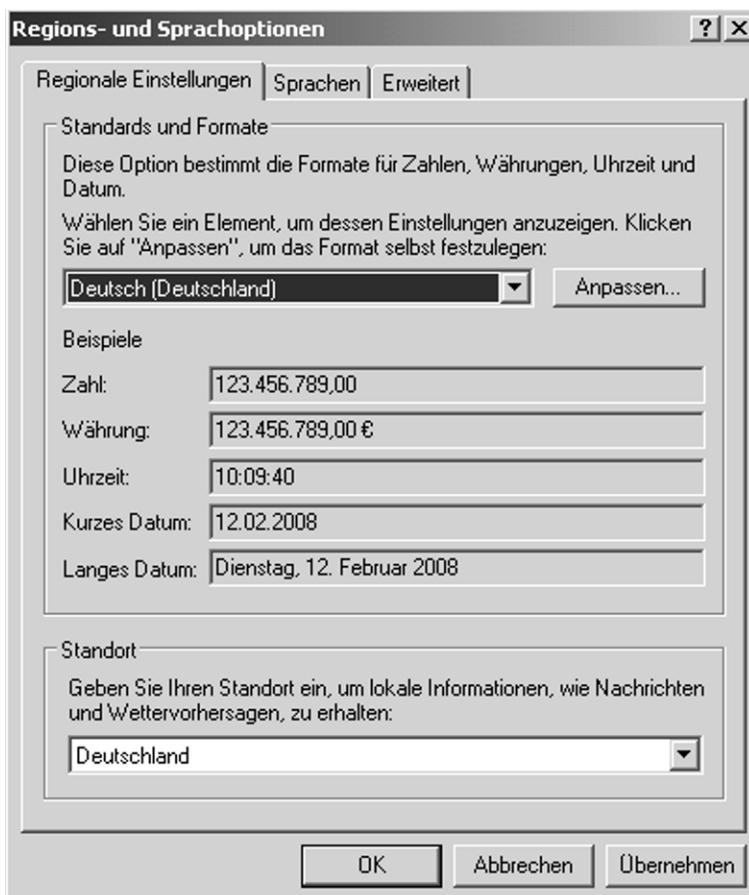


Abbildung 1.2. Im Betriebssystem Microsoft Windows gelangt man über die Option **Start → Einstellungen → Systemsteuerung → Regions- und Sprachoptionen** zu diesem Fenster. Man sieht, dass hier „Deutsch (Deutschland)“ eingestellt ist. Um Punkte als Dezimaltrennzeichen zu erhalten, muss diese Einstellung geändert werden, z.B. auf „Englisch (USA)“ (siehe Abb. 1.3).

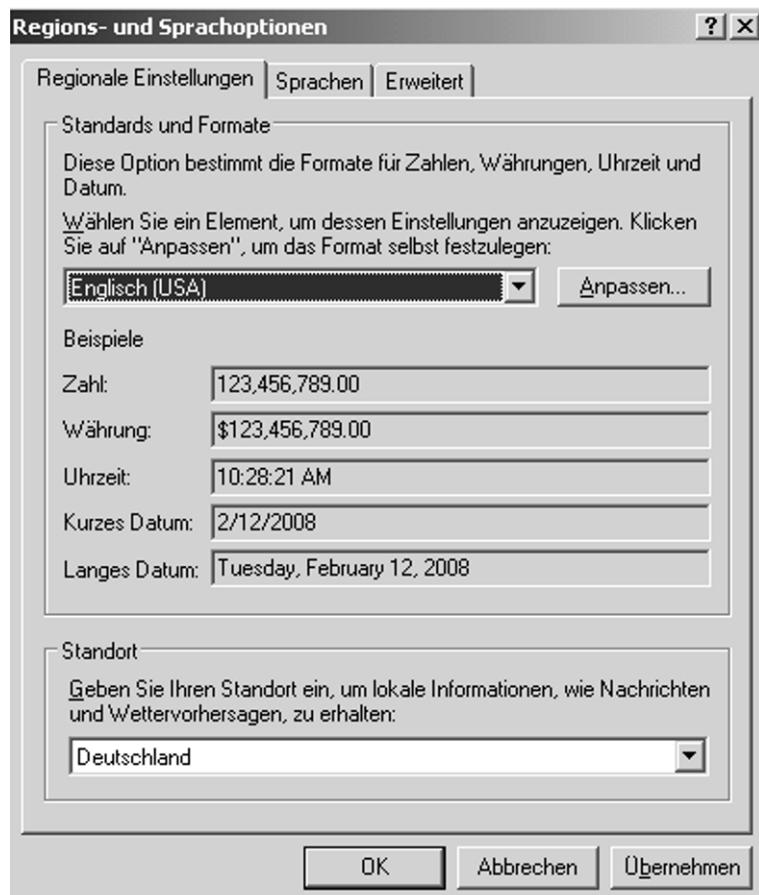


Abbildung 1.3. Geänderte regionale Einstellungen, damit Nachkommastellen mit Punkten abgetrennt werden. Danach muss der SPSS-Datensatz erneut geöffnet werden. Dezimaltrennzeichen werden nun in SPSS als Punkte dargestellt (siehe Abb. 1.4).

Dezimaltrennzeichen sind nun Punkte!

	kft_v1	kft_v3	kft_q1	kft_q3	kft_n1	kft_n3	var	var	var
10	6.00	9.00	9.00	3.00	7.00	12.00			
11	13.00	16.00	12.00	6.00	7.00	10.00			
12	7.00	7.00	7.00	4.00	7.00	5.00			
13	10.00	5.00	7.00	2.00	5.00	4.00			
14	5.00	5.00	10.00	4.00	5.00	4.00			
15	8.00	9.00	7.00	2.00	6.00	5.00			
16	17.00	13.00	12.00	5.00	16.00	12.00			
17	9.00	9.00	10.00	6.00	9.00	14.00			
18	12.00	10.00	8.00	5.00	5.00	7.00			
19	16.00	5.00	12.00	12.00	21.00	11.00			
20	20.00	14.00	15.00	10.00	15.00	13.00			
21	20.00	10.00	14.00	13.00	22.00	11.00			
22	7.00	6.00	9.00	8.00	6.00	5.00			
23	2.00	4.00	10.00	5.00	6.00	5.00			
24	17.00	14.00	16.00	9.00	20.00	9.00			
25	12.00	6.00	12.00	6.00	8.00	9.00			
26	14.00	5.00	8.00	7.00	9.00	9.00			
27	11.00	5.00	7.00	4.00	6.00	11.00			
28	7.00	4.00	10.00	4.00	5.00	8.00			
29	14.00	6.00	9.00	8.00	9.00	7.00			
30	9.00	10.00	12.00	6.00	10.00	7.00			
31	20.00	4.00	16.00	9.00	5.00	4.00			
32	14.00	3.00	15.00	8.00	7.00	8.00			
33			
34			
35			
36	8.00	3.00	16.00	3.00	2.00	3.00			
37	11.00	15.00	13.00	6.00	4.00	12.00			
38	4.00	2.00	8.00	4.00	6.00	10.00			

Abbildung 1.4. Darstellung der Dezimaltrennzeichen als Punkte in SPSS (KFT-Datensatz aus Abb. 1.1). Fehlende Werte sind noch nicht kodiert.

In neueren SPSS-Versionen lässt sich das Problem der Dezimaltrennzeichen leichter lösen. Hier kann bei der Speicherung der ASCII-Datei für Mplus mittels eines SPSS-Syntaxbefehls direkt ein zusätzlicher Unterbefehl eingefügt werden, der dafür sorgt, dass in der zu speichernden ASCII-Datei Dezimalstellen mit Punkten abtrennt werden. Der entsprechende vollständige SPSS-Syntax-Befehl zur Abspeicherung der KFT-Daten in einer für Mplus lesbaren tabulator-getrennten Textdatei mit dem Dateinamen **KFT.dat** wird im Folgenden gezeigt:

```
SAVE TRANSLATE OUTFILE='C:\KFT.dat'
/TYPE=TAB
/textoptions decimal = dot
/MAP
/REPLACE
/CELLS=VALUES.
```

Die ASCII-Datei **KFT.dat** wird hier direkt auf der Festplatte „C“ abgespeichert. Wichtig ist hierbei der zusätzlich eingefügte Unterbefehl /textoptions decimal = dot, welcher dafür sorgt, dass Punkte als Dezimaltrennzeichen verwendet werden. Nach Abspeicherung der ASCII-Datei empfiehlt es sich, diese zunächst zu öffnen (in Windows z.B. mit dem Editor- oder WordPad-Programm), um die Anzeige der Dezimaltrennzeichen zu kontrollieren.

1.2 Kodierung fehlender Werte

Es empfiehlt sich, systemdefiniert fehlende Werte (in SPSS: leere Zellen mit Punkt, siehe Abb. 1.1 bzw. Abb. 1.4) vor der Abspeicherung der ASCII-Datei in SPSS in „benutzerdefiniert fehlend“ umzukodieren und diese nicht als leere Zellen abzuspeichern, da dies ebenfalls zu Datenlesefehlern in Mplus führen kann. „Benutzerdefiniert fehlend“ bedeutet, dass man einen Missing-Value-Code für jeden fehlenden Wert vergibt. Der Code kennzeichnet die fehlenden Werte eindeutig als fehlend. Es sollte hierbei am besten ein numerischer Code verwendet werden, also eine Zahl. Diese Zahl muss natürlich so gewählt werden, dass sie in den Daten nicht als gültiger Wert vorkommen kann. Beliebte numerische Missing-Value-Codes sind z.B. „-9“ „-99“ oder „999“. Wir wählen für unseren KFT-Beispieldatensatz die Zahl „-99“ als Missing-Value-Code für alle 6 Variablen. Die Zahl „-99“ kann bei keiner der KFT-Skalen als gültiger Wert vorkommen.

Keinesfalls sollte man Merkmalsträger mit einzelnen fehlenden Werten komplett aus den Daten herauslöschen. Mplus bietet nützliche Optionen zur Behandlung fehlender Werte (u.a. *Full Information Maximum Likelihood Schätzung* mit fehlenden Werten, siehe z.B. Enders, 2010; Geiser et al., in Vorbereitung; Reinecke, 2005). Zudem kann auch *listenweiser Fallausschluss* in Mplus als Option gewählt werden. Bei listenweisem Fallausschluss werden nur solche Fälle in die Analyse einbezogen, die auf allen verwendeten Variablen gültige Werte haben.

Systemdefiniert fehlende Werte können in SPSS leicht mit Hilfe der Option **Transformieren → Umkodieren in dieselben Variablen** in benutzerdefiniert fehlende Werte umkodiert werden. Das entsprechende Vorgehen zeigen die Abbildungen 1.5-1.8 für die KFT-Daten. Die entsprechenden Syntaxbefehle (RECODE-Befehl) zur Kodierung fehlender Werte können der SPSS-Syntaxdatei auf der Begleit-CD entnommen werden.

Achtung: Nach dem Umkodieren von „systemdefiniert fehlend“ in „benutzerdefiniert fehlend“ muss man SPSS unbedingt noch mitteilen, dass nun ein benutzerdefinierter Missing-Value-Code vergeben wurde. Dies erfolgt entweder manuell über die Variablenansicht (siehe Abb. 1.9-1.12) oder mit Hilfe des SPSS-Syntax Commands MISSING VALUES (siehe Begleit-CD). Es ist wichtig, dies zu beachten, da SPSS ansonsten den Missing-Value-Code

(hier: -99) wie einen realen Wert behandeln und falsche statistische Berechnungen ausgeben würde. Die folgenden Screenshots zeigen das korrekte Vorgehen exemplarisch anhand der KFT-Daten.

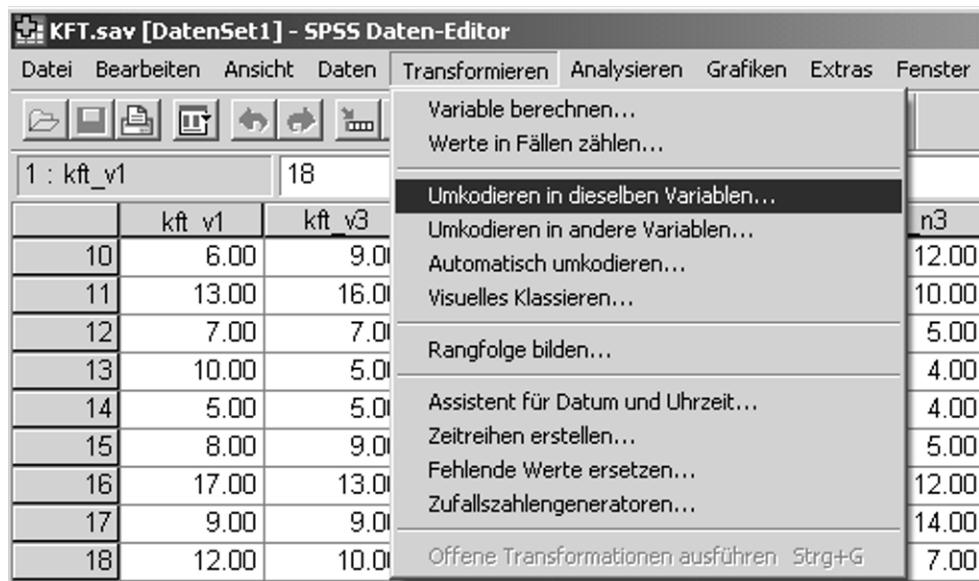


Abbildung 1.5: Option Transformieren → Umkodieren in dieselben Variablen in SPSS, hier benutzt zum Definieren eines Missing-Value-Codes.

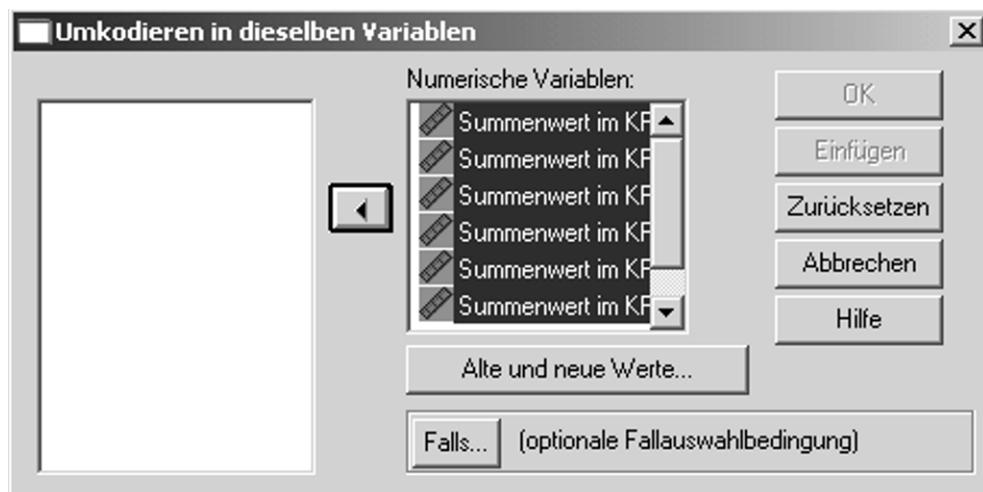


Abbildung 1.6: Alle Variablen werden zum Umkodieren in das rechte Fenster geklickt. Dann definiert man „Alte und neue Werte“ (siehe Abb. 1.7).

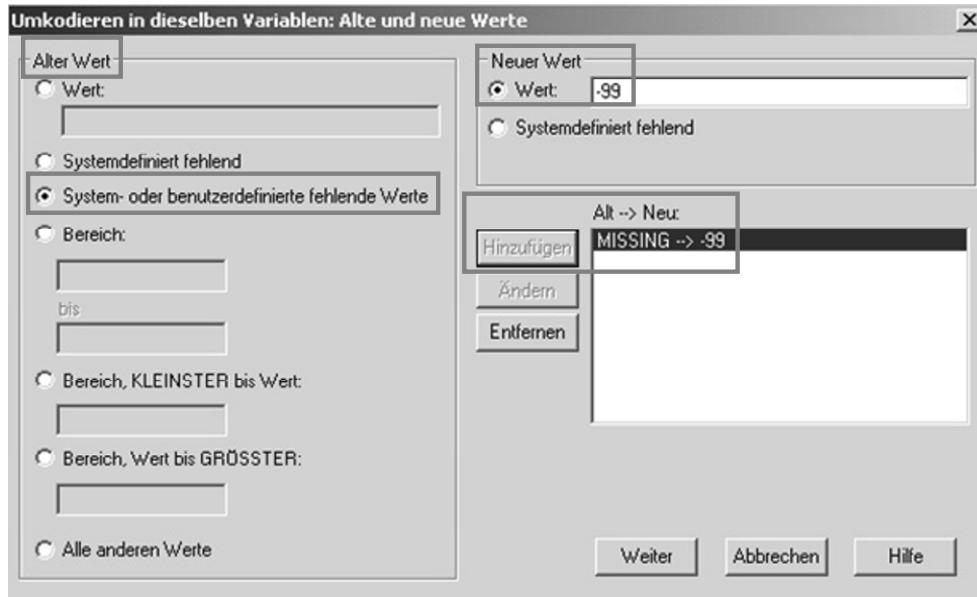


Abbildung 1.7: Option „Alte und neue Werte“ in der Umkodieroption von SPSS. Links werden alte (umzukodierende) Werte angegeben, rechts die neuen Werte (hier: Missing-Value-Code). Im gezeigten Beispiel wird die Zahl „-99“ als Missing-Value-Code für alle Werte vergeben, die bislang als systemdefiniert fehlend eingegeben waren oder einen anderen benutzerdefinierten Missing-Value-Code aufwiesen (z.B. -9). Dies dient der Vereinheitlichung zur Vermeidung von Fehlern. Nach dem Klicken auf Weiter und OK erhält man folgende veränderte Datenansicht (siehe Abb. 1.8).

*KFT.sav [DataSet1] - SPSS Daten-Editor

Datenansicht / Variablenansicht / SPSS Prozessor ist bereit

Abbildung 1.8: KFT-Beispieldatensatz. Alle fehlenden Werte sind nun bei allen 6 Variablen mit der Zahl „-99“ kodiert. Dass es sich bei den Werten „-99“ um fehlende Werte handelt, muss man SPSS nun unbedingt noch über die Variablenansicht oder mit Hilfe der Syntax mitteilen (siehe Abb. 1.9-1.12).

	Name	Typ	Spaltenformat	Dezimalstellen	Variablenlabel	Wertelabels	Fehlende Werte
1	kft_v1	Numerisch	8	2	Summenwert i	Kein	Kein
2	kft_v3	Numerisch	8	2	Summenwert i	Kein	Kein
3	kft_q1	Numerisch	8	2	Summenwert i	Kein	Kein
4	kft_q3	Numerisch	8	2	Summenwert i	Kein	Kein
5	kft_n1	Numerisch	8	2	Summenwert i	Kein	Kein
6	kft_n3	Numerisch	8	2	Summenwert i	Kein	Kein
7							

Abbildung 1.9: Die Zahl „-99“ ist in der Variablenansicht noch nicht als benutzerdefiniert fehlender Wert (Missing-Value-Code) eingetragen. Dies muss man unbedingt noch nachholen (siehe Abb. 1.10), da SPSS ansonsten die -99er Werte wie gültige Werte behandeln würde.

Um SPSS mitzuteilen, dass der Wert „-99“ nun bei allen Variablen fehlende Werte indiziert, kann man entweder den Syntax-Befehl MISSING VALUES (siehe SPSS-Syntax auf der Begleit-CD) verwenden, oder folgendermaßen über das SPSS-Menü vorgehen. Dazu klickt man in der SPSS-Variablenansicht bei einer beliebigen Variable in der Spalte „fehlende Werte“ rechts auf die drei kleinen Punkte, wie in Abbildung 1.10 gezeigt.

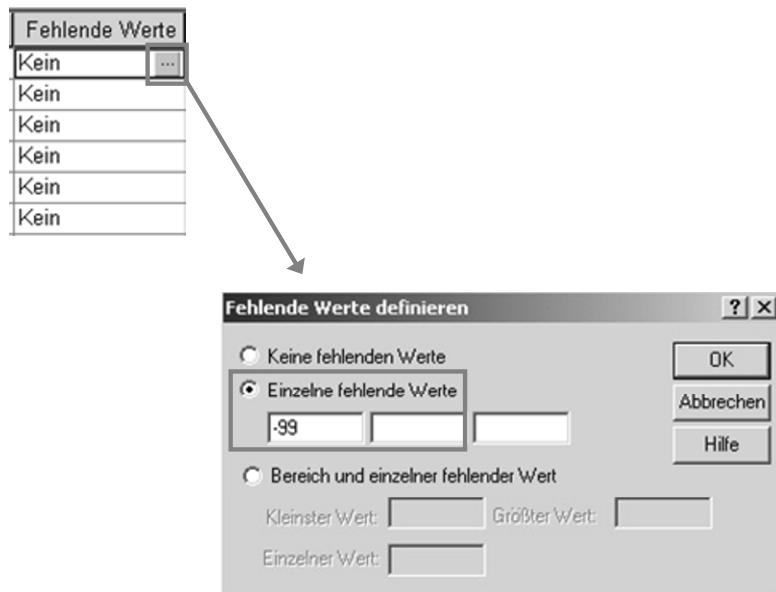


Abbildung 1.10: Der Wert -99 wird in SPSS als benutzerdefiniert fehlender Wert eingetragen. Diese Definition muss man praktischerweise nur für eine einzige Variable definieren. Dann kann man sie kopieren und für alle anderen Variablen ebenfalls einfügen (siehe Abb. 1.11).

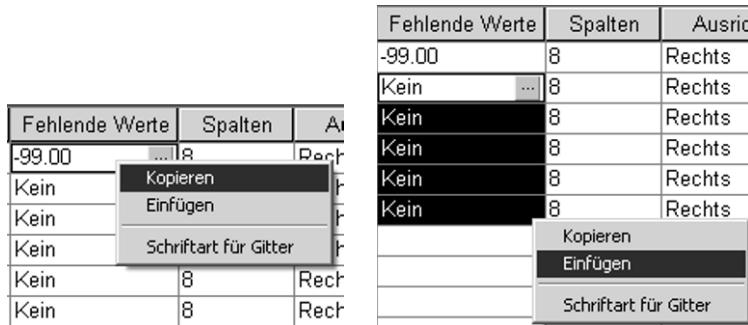


Abbildung 1.11: Kopieren und Einfügen der Missing-Value-Definition in der SPSS-Variablenansicht. Abbildung 1.12 zeigt die korrekt definierten Variablen.

*KFT.sav [DataSet1] - SPSS Daten-Editor							
	Name	Typ	Spaltenformat	Dezimalstellen	Variablenlabel	Wertelabels	Fehlende Werte
1	kft_v1	Numerisch	8	2	Summenwert i	Kein	-99.00
2	kft_v3	Numerisch	8	2	Summenwert i	Kein	-99.00
3	kft_q1	Numerisch	8	2	Summenwert i	Kein	-99.00
4	kft_q3	Numerisch	8	2	Summenwert i	Kein	-99.00
5	kft_n1	Numerisch	8	2	Summenwert i	Kein	-99.00
6	kft_n3	Numerisch	8	2	Summenwert i	Kein	-99.00
7							

Abbildung 1.12: Korrekt eingetragener Missing Value Code (-99) für alle 6 Variablen des KFT-Datensatzes in der SPSS-Variablenansicht.

1.3 Abspeichern eines ASCII-Datensatzes für Mplus

Nachdem Dezimaltrennzeichen als Punkte im Datensatz eingetragen sind und man fehlende Werte mit einem numerischen Missing-Value-Code (hier: -99) gekennzeichnet hat, kann man die Daten in dem für Mplus benötigten ASCII-Format abspeichern. Dies geschieht in der SPSS-Datenansicht über **Datei → Speichern unter** (siehe Abb. 1.13-1.15) oder mit Hilfe des Syntaxbefehls **SAVE TRANSLATE** (siehe Seite 16).

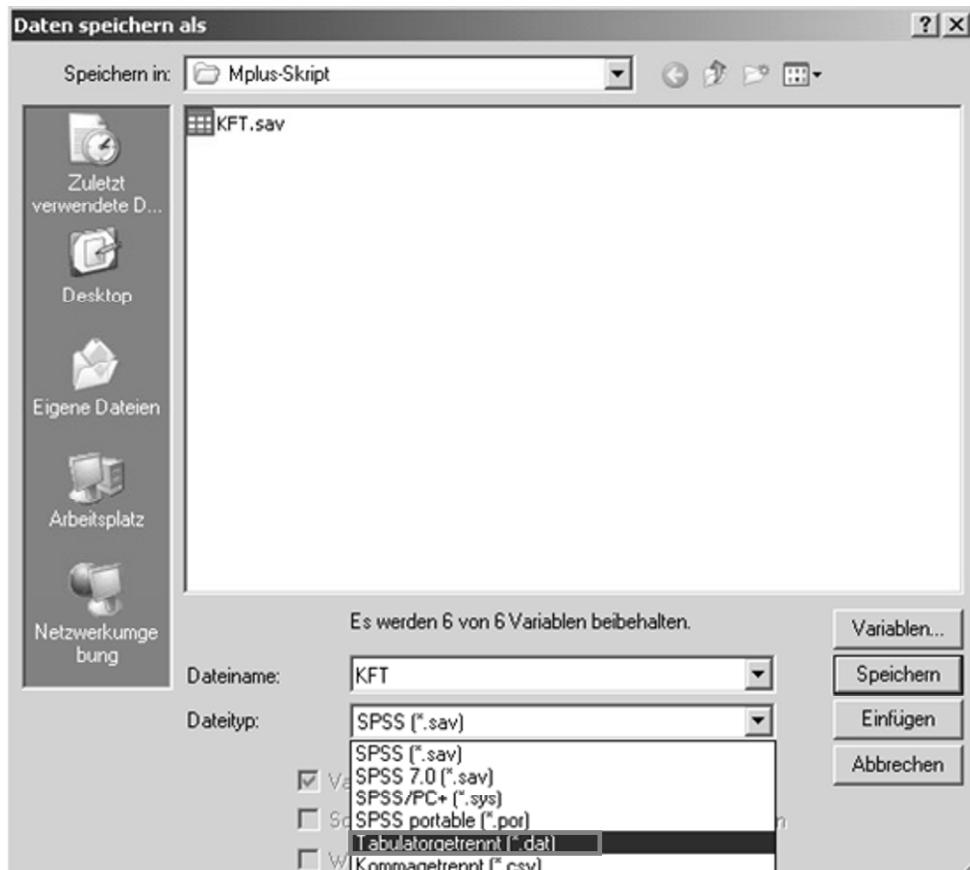


Abbildung 1.13: Abspeichern des KFT-Datensatzes im ASCII-Format. Der Autor empfiehlt hierbei die Verwendung der Variante „Tabulatorgetrennt (*.dat)“.

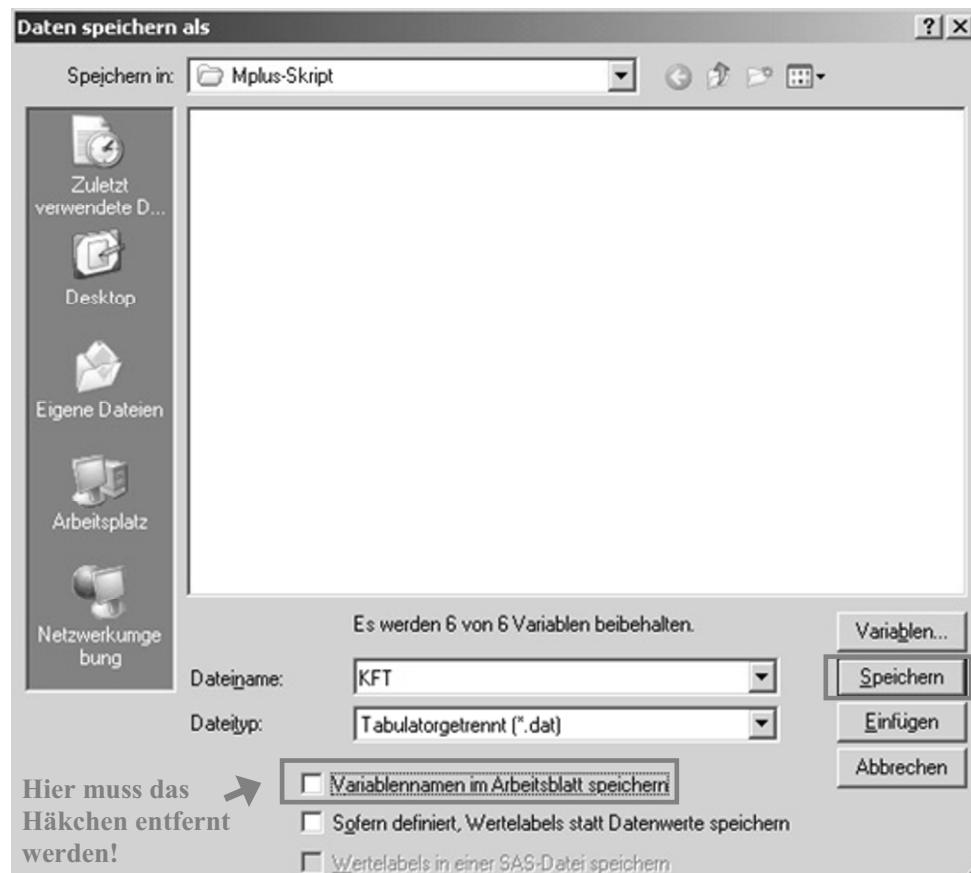


Abbildung 1.14: Abspeichern des KFT-Datensatzes im tabulator-getrennten ASCII-Format. Wichtig: Die Option **Variablennamen im Arbeitsblatt speichern** muss *deaktiviert* werden (Häkchen entfernen)! Dann auf **Speichern** klicken. Abbildung 1.15 zeigt einen Ausschnitt aus dem neu abgespeicherten ASCII-Datensatz mit dem Namen „KFT.dat“. Dieser lässt sich z.B. mit einem einfachen Textverarbeitungsprogramm (Editor, WordPad, Microsoft-Word etc.) öffnen.

18	13	14	7	9	5
16	7	4	2	12	10
17	13	8	8	9	10
20	17	13	9	20	12
13	6	19	9	10	10
6	5	8	6	3	5
16	7	19	9	9	10
12	5	10	4	13	9
7	10	9	7	9	9
6	9	9	3	7	12
13	16	12	6	7	10
7	7	7	4	7	5
10	5	7	2	5	4

Abbildung 1.15: Ausschnitt aus dem neu abgespeicherten ASCII-Datensatz (Variablen Spalten sind mit Tabulatorzeichen getrennt) mit dem Namen „KFT.dat“. Dieser wurde hier mit dem Windows-Programm „Editor“ geöffnet (In Windows: **Start** → **Programme** → **Zubehör**). Man beachte, dass Nachkommastellen hier nicht berücksichtigt wurden, da es sich bei allen Werten um ganzzahlige Scores handelt. Bei nicht-ganzzahligen Werten würden die Nachkommastellen nun mit Punkten angezeigt werden (hat man Kommas als Dezimaltrennzeichen, so würde das Einlesen dieser Daten in Mplus Fehlermeldungen verursachen, siehe Abschnitt 1.2).

2 Einlesen von Daten in Mplus

Mplus kann unterschiedliche Arten von Daten einlesen und verarbeiten. Die beiden wichtigsten Varianten sollen hier besprochen werden. Die erste Variante ist in der Praxis wohl die Häufigste. Hierbei werden Rohdaten verwendet, d.h. sogenannte *Individualdaten (individual data)*. In Individualdatensätzen hat jede Person auf jeder Variable einen Wert (oder ggf. einen fehlenden Wert), so wie dies z.B. im KFT-Datensatz (siehe Abb. 1.15) der Fall ist. Die zweite Variante bezieht sich auf das Einlesen sogenannter *Summary-Daten*. Dabei handelt es sich z.B. um eine Kovarianz- oder Korrelationsmatrix (und ggf. die Mittelwerte und Standardabweichungen der Variablen). Die Möglichkeit, mit Summary-Daten arbeiten zu können, ist beispielsweise dann nützlich, wenn man Daten aus einem Forschungsbericht (z.B. einem Fachzeitschriftenartikel) reanalysieren möchte. In vielen Publikationen werden Kovarianz- oder Korrelationsmatrizen, die die Grundlage für die berichteten Analysen (z.B. Pfad- oder Strukturgleichungsanalysen) darstellen, mit angegeben (wohingegen die Rohdaten nur selten verfügbar gemacht werden). Für viele Arten von linearen Strukturgleichungsmodellen ist es nicht unbedingt erforderlich, Individualdaten zu analysieren (obwohl Individualdaten in vielerlei Hinsicht vorteilhaft sind, z.B. wenn fehlende Werte oder besondere Schätzverfahren berücksichtigt werden sollen). Als Input genügt bei vielen Modellen die Kovarianzmatrix, ggf. ergänzt durch die Mittelwerte der zu analysierenden Variablen. Es wird zunächst das Einlesen von Individualdaten gezeigt, und zwar anhand des in Abschnitt 1.3 abgespeicherten ASCII-Datensatzes mit den KFT-Daten.

2.1 Einlesen von Individualdaten (Rohdaten)

Das Einlesen von Rohdaten verläuft erfahrungsgemäß problemlos, wenn man die in Kapitel 1 gezeigten Schritte korrekt ausgeführt hat. Trotzdem ist es unabdingbar, dass man sich in Mplus zunächst Deskriptivstatistiken (Mittelwerte, Varianzen, Kovarianzen und Korrelationen) der beobachteten Variablen ausgeben lässt und diese mit den entsprechenden Werten in SPSS vergleicht, bevor man erste weiterführende statistische Analysen durchführt. Der Vergleich von Deskriptivstatistiken in Mplus und SPSS dient dazu, zu überprüfen, ob die Daten von Mplus korrekt gelesen werden. Diesen Punkt kann man gar nicht genug betonen, da er allzu oft übersprungen und dadurch nicht (oder zu spät!) bemerkt wird, dass Mplus die verwendeten Daten nicht korrekt verarbeitet und deshalb inkorrekte Schätzergebnisse produziert. Aus diesem Grund sollte man sich die Zeit nehmen, in Mplus zunächst eine sogenannte **basic-Analyse** durchzuführen, um das korrekte Einlesen der Daten sicherzustellen, bevor man mit der ersten tatsächlichen Analyse beginnt.

2.1.1 Aufbau der basalen Mplus-Syntax und basic-Analyse

Mplus ist ein syntaxbasiertes Programm, d.h. alle Funktionen werden über eine Befehlsprache angefordert und sind nicht wie z.B. in SPSS über das Menü erhältlich. Zur Erstellung der basalen Mplus-Syntax steht jedoch eine menügesteuerte Funktion zur Verfügung. Dabei handelt es sich um den sogenannten *Mplus Language Generator* (im Mplus-Menü **Mplus → Language Generator**). Diese menügesteuerte Option ermöglicht es, die wichtigsten Grundbefehle vom Programm automatisch erstellen zu lassen. Die Befehle zur Spezifikation eines bestimmten Modells (`model`-Befehl) muss man jedoch trotzdem noch selbst einfügen. Auf die Darstellung der Variante der Syntaxgenerierung über den Mplus Language Generator wird hier verzichtet, da Benutzer in aller Regel nach Erlernen der Grundbefehle nicht mehr auf diese Funktion zurückgreifen.

Im Folgenden wird ein sinnvolles Vorgehen zum Dateneinlesen und zur Datenkontrolle mit der `basic`-Option anhand des Beispieldatensatzes KFT.dat demonstriert. Zunächst muss man den Mplus-Editor öffnen (in Windows: **Start → Programme → Mplus → Mplus Editor**). Es öffnet sich ein leeres Editor-Fenster. Man kann dann entweder direkt die benötigten Syntaxbefehle eingeben oder das Vorgehen über den Mplus Language Generator wählen. Die Minimalvariante benötigter Syntaxbefehle wird in Abbildung 2.1 gezeigt.

Sobald man begonnen hat, eine Mplus-Syntax zu erstellen, sollte man diese als Mplus-Input-Datei abspeichern (siehe Abb. 2.1). Mplus-Input-Dateien erhalten die Dateiendung `*.inp`. Auch regelmäßiges Zwischenspeichern neu eingefügter oder modifizierter Befehle (mittels Klicken auf das Diskettensymbol im Mplus-Menü oder Drücken der Tastenkombination Strg + S) sollte nicht vergessen werden, damit bei einem etwaigen Computerabsturz keine Arbeit verloren geht.

Der `title`-Befehl dient dazu, die Analyse zu benennen bzw. ausführlich zu kommentieren. Dieser Befehl wird nicht zwingend benötigt, allerdings ist es zur Dokumentation der eigenen Arbeit ratsam, stets einen ausführlichen Titel für die zu spezifizierende Analyse zu vergeben. Optional ist auch das Einfügen von Kommentarzeilen. Kommentarzeilen müssen mit einem Ausrufezeichen (!) beginnen (und brauchen nicht mit einem besonderen Zeichen zu enden). Mplus ignoriert alles, was in einer Kommentarzeile erscheint. Jede neue Kommentarzeile muss erneut mit einem „!“-Zeichen beginnen. Kommentarzeilen sind in Mplus praktischerweise in grüner Schriftfarbe abgehoben. Bei Kommentarzeilen ist die Länge übrigens nicht kritisch, da es sich nicht um relevante Befehle handelt. Allerdings schneidet Mplus zu lange Kommentarzeilen nach 80 bzw. 90 Zeichen ab, so dass die Kommentare u.U. unvollständig im Output erscheinen.

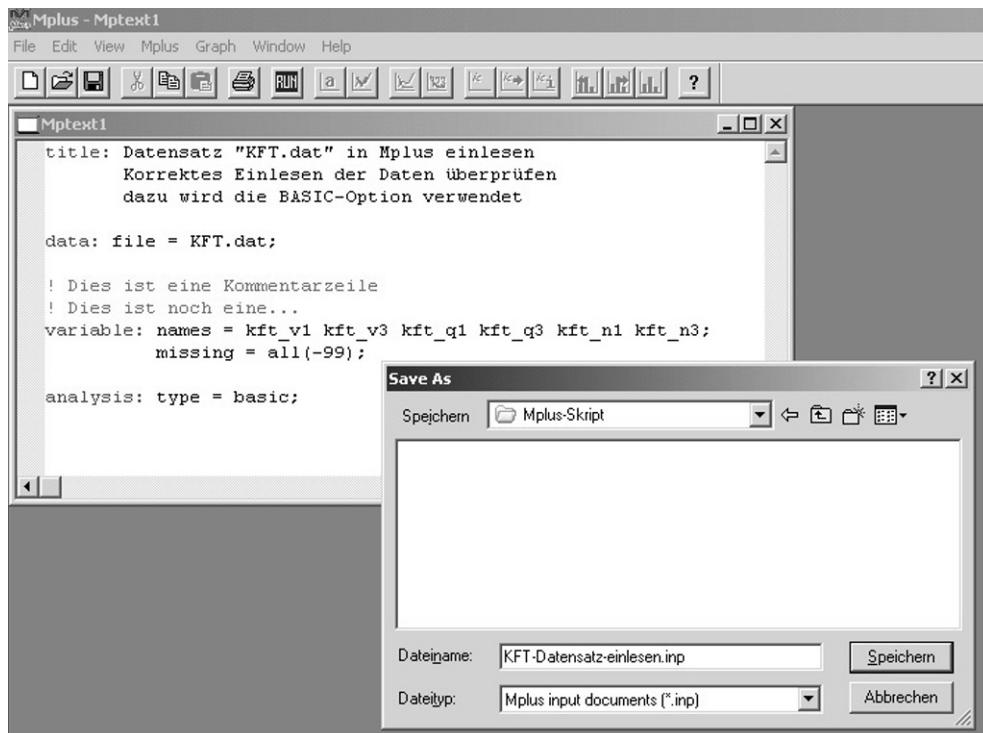


Abbildung 2.1: Syntax zur Überprüfung des korrekten Dateneinlesens des Datensatzes KFT.dat mittels basic-Analyse und Abspeichern des Mplus-Input-Files als *.inp-Datei. Hierbei wird der Input im selben Ordner abgespeichert, in welchem sich auch der Datensatz KFT.dat befindet. Deshalb ist eine genaue Spezifikation eines Pfades zu den Daten unter data: file = nicht erforderlich.

Der data-Befehl dient u.a. dazu, Mplus den Namen, Speicherort sowie die Art des zu verwendenden Datensatzes mitzuteilen. Wenn man den Datensatz auf dem Computer im gleichen Verzeichnis (Ordner) abspeichert wie die Mplus-Input-Datei, so braucht man den genauen Pfad zum Speicherort des Datensatzes nicht anzugeben (vgl. Abb. 2.1).

Der variable-Befehl dient u.a. zur Definition der im Datensatz enthaltenen und der zu analysierenden Variablen sowie zur Definition des verwendeten Missing-Value-Codes (vgl. dazu Abschnitt 1.2). Unter names = werden die Variablennamen vergeben. Wichtig ist hierbei die korrekte Reihenfolge der Variablennamen in der Liste hinter names =. Die Reihenfolge muss unbedingt der Reihenfolge der Variablen im einzulesenden ASCII-Datensatz entsprechen! Außerdem ist wichtig zu beachten, dass Variablennamen nicht länger als 8 Zeichen sein dürfen. Sehr bequem kann man die Variablennamen direkt aus SPSS mit Hilfe der SPSS-Menü-Option Extras → Variablen importieren. Diese Möglichkeit wird anhand von Abbildung 2.2 und 2.3 illustriert.

Ein weiterer Vorteil dieses Vorgehens ist, dass man die gleichen Variablennamen in beiden Programmen verwendet (Voraussetzung hierfür ist natürlich, dass man in SPSS auch

keine Variablennamen mit mehr als 8 Zeichen verwendet hat). Wichtig ist, hierbei darauf zu achten, dass SPSS die Variablennamen in der korrekten Reihenfolge in die Syntax einfügt, was leider nicht in jeder SPSS-Version der Fall ist.



Abbildung 2.2. Exportieren der Variablennamen aus SPSS. Links werden alle Variablen markiert. Durch das Klicken auf **Einfügen** werden die Variablennamen in ein Syntaxfenster eingefügt (siehe Abb. 2.3).



Abbildung 2.3. Exportieren der Variablennamen aus SPSS. Durch das Klicken auf **Einfügen** wurden die Variablennamen automatisch in das SPSS-Syntaxfenster eingefügt. Sie können nun aus dem Syntaxfenster herauskopiert und in Mplus-Editor im variable-Befehl hinter names = eingefügt werden. Hierbei muss man unbedingt auf die korrekte Reihenfolge der Variablennamen achten.

Weiterhin sollte beachtet werden, dass jeder Unterbefehl (mit Ausnahme des Titels) in Mplus mit einem Semikolon enden muss – das Fehlen eines Semikolons nimmt vermutlich in der Hitparade möglicher Fehlerquellen in Mplus den ersten Platz ein! Mit Hilfe des missing-Befehls teilt man Mplus mit, wie fehlende Werte kodiert sind („Missing-Value-Code“). Im vorliegenden Beispiel haben wir bei allen Variablen fehlende Werte mit -99 kodiert (siehe Abschnitt 1.2), also schreiben wir in Mplus unter variable: missing = all(-99);.

Mit Hilfe von analysis: type = basic; fordern wir Deskriptivstatistiken für alle Variablen an, die wir dann mit den entsprechenden Werten in SPSS vergleichen können.

Durch das Klicken auf RUN im Mplus-Menü (oder drücken der Tastenkombination **Alt + R**) wird die basic-Analyse ausgeführt. Box 2.1 gibt einen Überblick über einige grundlegende Regeln zur Mplus-Syntax. Im Anschluss werden die Ergebnisse der basic-Analyse besprochen.

Box 2.1: Einige wichtige Mplus-Syntaxregeln

- In der Mplus-Syntax wird nicht zwischen Groß- und Kleinschreibung unterschieden.
- Auch die Reihenfolge, mit welcher die einzelnen Befehle in der Syntax aufgelistet werden, ist weitgehend beliebig.
- Jede Befehlszeile muss mit einem Semikolon (;) abgeschlossen werden.
- Eine einzelne Befehlszeile darf nicht länger als 90 Zeichen sein (in älteren Mplus-Versionen nur 80 Zeichen!). Lange Befehlszeilen können zum Beispiel auftreten, wenn man viele Variablen einliest oder einen langen Pfad zum Datensatz angeben möchte. Die Folge ist, dass Mplus die Zeile nach 80 bzw. 90 Zeichen „abschneidet“, d.h. alle folgenden Zeichen ignoriert (was u.U. zu drastischen Fehlern führen kann!). Es erfolgt eine entsprechende Fehlermeldung im Mplus-Output, die auf dieses Problem hinweist. Man sollte sie keinesfalls ignorieren! Ein einfaches Mittel zum Beheben dieses Fehlers ist das simple Umbrechen der betreffenden zu langen Zeile mittels RETURN-Taste.
- Variablennamen dürfen nicht aus mehr als 8 Zeichen bestehen.
- Jede Kommentarzeile muss mit einem Ausrufezeichen (!) beginnen. Sie wird im Mplus-Input in grüner Farbe dargestellt. Kommentarzeilen müssen nicht mit einem besonderen Zeichen enden, jede neue Kommentarzeile muss jedoch wieder mit einem Ausrufezeichen beginnen.

2.1.2 Mplus-Output zur basic-Analyse

Nach dem Ausführen der Inputdatei zur basic-Analyse erscheint ein neues Fenster (sog. Outputfenster) mit dem entsprechenden Output. Der Output wird von Mplus automatisch als *.out-Datei mit demselben Dateinamen wie der Inputfile in dem Verzeichnis abgespeichert, in welchem sich der zugehörige Inputfile (*.inp) befindet. Der Output beginnt wie folgt:

```
Mplus VERSION 5
MUTHEN & MUTHEN
02/12/2008 3:40 PM

INPUT INSTRUCTIONS

title: Datensatz "KFT.dat" in Mplus einlesen
       Korrektes Einlesen der Daten überprüfen
       dazu wird die BASIC-Option verwendet

data: file = KFT.dat;

! Dies ist eine Kommentarzeile
! Dies ist noch eine...
variable: names = kft_v1 kft_v3 kft_q1 kft_q3 kft_n1 kft_n3;
           missing = all(-99);

analysis: type = basic;
```

Zunächst werden die Mplus-Input-Befehle reproduziert. Dies ist nützlich, da man so noch einmal kontrollieren kann, ob die Inputspezifikation korrekt war und diese auch auf einem möglichen Ausdruck des Outputs noch einmal dokumentiert ist. Dann erscheint eine Warnmeldung, da es im Datensatz **KFT.dat** insgesamt 131 Schüler gibt, die auf allen 6 Variablen fehlende Werte aufweisen:

```
*** WARNING
Data set contains cases with missing on all variables.
These cases were not included in the analysis.
Number of cases with missing on all variables: 131
1 WARNING(S) FOUND IN THE INPUT INSTRUCTIONS
```

Diese Meldung erscheint, da Mplus 5 standardmäßig (als „Default“) das Full Information Maximum Likelihood (FIML) Schätzverfahren zur Einbeziehung fehlender Werte verwendet (siehe z.B. Enders, 2010; Reinecke, 2005). Dieses Verfahren kann bei Personen, die auf *allen* verwendeten Variablen fehlende Werte aufweisen, nicht zum Einsatz gebracht werden, da für diese Personen in Bezug auf die zu analysierenden Variablen keinerlei Informationen zur Verfügung stehen.

Als nächstes erscheinen weitere technische Informationen zur Analyse, u.a. zur Stichprobengröße, zu den Variablen und zu den Daten allgemein:

```
Datensatz "KFT.dat" in Mplus einlesen
Korrektes Einlesen der Daten überprüfen
dazu wird die BASIC-Option verwendet
```

SUMMARY OF ANALYSIS

Number of groups	1
Number of observations	456

```

Number of dependent variables          6
Number of independent variables       0
Number of continuous latent variables 0

Observed dependent variables
  Continuous
    KFT_V1      KFT_V3      KFT_Q1      KFT_Q3      KFT_N1      KFT_N3

Estimator                           ML
Information matrix                 OBSERVED
Maximum number of iterations       1000
Convergence criterion              0.500D-04
Maximum number of steepest descent iterations 20
Maximum number of iterations for H1 2000
Convergence criterion for H1      0.100D-03

Input data file(s)
  KFT.dat

Input data format   FREE

```

Diese sind nützlich, um noch einmal zu kontrollieren, dass tatsächlich der korrekte Datensatz sowie die intendierten Variablen verwendet wurden.

Nachfolgend erhält man nützliche Informationen zu den in den Daten beobachteten Mustern fehlender Werte („missing data patterns“).

```

SUMMARY OF DATA

  Number of missing data patterns      2

SUMMARY OF MISSING DATA PATTERNS

  MISSING DATA PATTERNS (x = not missing)
    1  2
  KFT_V1   x  x
  KFT_V3   x  x
  KFT_Q1   x  x
  KFT_Q3   x  x
  KFT_N1   x  x
  KFT_N3   x

  MISSING DATA PATTERN FREQUENCIES

  Pattern  Frequency      Pattern  Frequency
    1        455           2        1

```

Weiter oben unter number of observations sehen wir, dass insgesamt 456 Schüler einen oder mehrere Werte beisteuern (hier sind diejenigen Fälle bereits ausgenommen, die fehlende Werte auf allen 6 Variablen aufweisen). Unter missing data patterns werden

alle beobachteten Ausfallmuster in einer eigenen Spalte dargestellt (mit Ausnahme des Musters, bei dem alle Werte fehlen). Die Spalte 1 (missing data pattern Nummer 1) enthält nur Kreuze, was anzeigt, dass es sich um dasjenige Muster handelt, bei welchem *keine* fehlenden Werte auftreten. In der zweiten Spalte (missing data pattern Nummer 2) sehen wir Kreuze bei allen Variablen, außer bei der letzten (**kft_n3**). Es handelt sich somit um das Ausfallmuster, bei welchem für alle Variablen gültige Werte vorliegen außer für die letzte.

Im Abschnitt missing data pattern frequencies zeigt sich, dass von den 456 Schülern fast alle vollständige Daten (Werte auf allen 6 Variablen) liefern: 455 Schüler haben das missing data pattern Nummer 1 (Frequency = 455). Es gibt nur einen Schüler mit dem missing data pattern Nummer 2 (Frequency = 1), also einem fehlenden Wert auf der Variable **kft_n3**.

Die covariance coverage gibt an, wie viel Prozent der Fälle Werte zu einer bestimmten Varianz oder Kovarianz beisteuern. In unserem Fall sind dies für alle Varianzen und Kovarianzen 100 % der Fälle, mit Ausnahme der Varianz und der Kovarianzen, die mit der Variable **kft_n3** assoziiert sind (dort nur 99.8 %, da hier eine Person einen fehlenden Wert aufweist). Der minimal akzeptable Wert für die covariance coverage ist in Mplus 10 % als Voreinstellung („Minimum covariance coverage value 0.100“).

COVARIANCE COVERAGE OF DATA					
Minimum covariance coverage value 0.100					
PROPORTION OF DATA PRESENT					
	Covariance Coverage				
	KFT_V1	KFT_V3	KFT_Q1	KFT_Q3	KFT_N1
KFT_V1	1.000				
KFT_V3	1.000	1.000			
KFT_Q1	1.000	1.000	1.000		
KFT_Q3	1.000	1.000	1.000	1.000	
KFT_N1	1.000	1.000	1.000	1.000	1.000
KFT_N3	0.998	0.998	0.998	0.998	0.998
Covariance Coverage					
KFT_N3					
KFT_N3		0.998			

Achtung: Die covariance coverage gibt nicht die tatsächlichen Werte der Kovarianzen zwischen den Variablen an, sondern gibt nur Auskunft über die Datengrundlage, die zu deren Berechnung verfügbar ist. Die tatsächlich geschätzten Kovarianzen erhalten wir weiter unten unter der Rubrik estimated sample statistics.

Als nächstes erhalten wir Deskriptivstatistiken für die sechs beobachteten KFT-Variablen (d.h. die geschätzten Mittelwerte, die Kovarianzmatrix und die Matrix der Pearson Produkt-Moment-Korrelationen der Variablen unter der Rubrik estimated sample

`statistics`). Diese Statistiken wurden mit Hilfe des FIML-Verfahrens geschätzt. Dadurch sind die Statistiken nicht eins-zu-eins mit in SPSS berechneten Statistiken vergleichbar. Das liegt daran, dass die SPSS-Statistiken entweder auf paarweisem Fallausschluss (pairwise deletion) oder auf listenweisem Fallausschluss (listwise deletion; Ausschluss aller Personen, die auf wenigstens einer der sechs Variablen einen fehlenden Wert aufweisen) beruhen und nicht auf FIML-Schätzungen.

Um einen exakten Abgleich der Mplus- und SPSS-Ergebnisse vornehmen zu können, kann man in Mplus jedoch ebenfalls listenweisem Fallausschluss (anstelle der Voreinstellung FIML) anfordern. Listenweisem Fallausschluss erreicht man in Mplus, indem man im Input-Fenster unter dem `data`-Befehl den Unterbefehl `listwise = on` einfügt. Dies bewirkt, dass die FIML-Prozedur deaktiviert wird und alle Fälle, die auf wenigstens einer der verwendeten Variablen einen fehlenden Wert aufweisen, aus der Analyse ausgeschlossen werden (FIML ist erst seit der Mplus-Version 5 die Voreinstellung. Bei den Vorgängerversionen war listenweiser Fallausschluss die Voreinstellung und das FIML-Verfahren musste gesondert angefordert werden).

Die erweiterte Mplus-Syntax mit listenweisem Fallausschluss ist in Abbildung 2.4 dargestellt. Die resultierenden Deskriptivstatistiken sind darunter wiedergegeben. Zum Vergleich sind direkt unter dem Mplus-Output in Abbildung 2.5 die korrespondierenden SPSS-Statistiken gezeigt, die auf listenweisem Fallausschluss beruhen und mit Hilfe der SPSS-Option **Analysieren → Skalierung → Reliabilitätsanalyse → Statistik → Deskriptive Statistiken für Item/Zwischen Items Korrelationen/Kovarianzen** erzeugt wurden. Sie stimmen mit den Mplus-Ergebnissen unter listwise deletion (auf die dritte Nachkommastelle gerundet) überein, was uns zeigt, dass der Datensatz **KFT.dat** in Mplus korrekt gelesen wird. Auch kommt SPSS auf dieselbe Fallzahl ($N = 455$ listenweise gültige Fälle). Nach dieser Überprüfung können wir mit der Spezifikation des ersten Modells in Mplus beginnen (siehe Kapitel 3).

RESULTS FOR BASIC ANALYSIS [FIML-Schätzungen]					
ESTIMATED SAMPLE STATISTICS					
Means					
KFT_V1	KFT_V3	KFT_Q1	KFT_Q3	KFT_N1	KFT_N3
11.904	8.978	12.377	7.730	11.088	8.277
Covariances					
KFT_V1	21.592				
KFT_V3	10.761	17.938			
KFT_Q1	6.010	5.267	11.235		
KFT_Q3	4.645	4.406	4.027	6.754	
KFT_N1	11.072	11.644	6.870	6.111	29.826
KFT_N3	6.347	7.190	4.415	3.986	9.298
Covariances					
KFT_N3		KFT_N3			
KFT_N3		11.782			
Correlations					
KFT_V1	1.000				
KFT_V3	0.547	1.000			
KFT_Q1	0.386	0.371	1.000		
KFT_Q3	0.385	0.400	0.462	1.000	
KFT_N1	0.436	0.503	0.375	0.431	1.000
KFT_N3	0.398	0.495	0.384	0.447	0.496
Correlations					
KFT_N3		KFT_N3			
KFT_N3		1.000			
MAXIMUM LOG-LIKELIHOOD VALUE FOR THE UNRESTRICTED (H1) MODEL IS -7152.289					
Beginning Time: 15:40:25					
Ending Time: 15:40:25					
Elapsed Time: 00:00:00					
MUTHEN & MUTHEN					
3463 Stoner Ave.					
Los Angeles, CA 90066					
Tel: (310) 391-9971					
Fax: (310) 391-8971					
Web: www.StatModel.com					
Support: Support@StatModel.com					
Copyright (c) 1998-2007 Muthen & Muthen					

The screenshot shows the Mplus software interface with the title bar 'Mplus - [KFT-Datensatz-einlesen-listwise-deletion.inp]'. The menu bar includes File, Edit, View, Mplus, Graph, Window, and Help. Below the menu is a toolbar with various icons. The main window displays the Mplus input command language:

```

title: Datensatz "KFT.dat" in Mplus einlesen
      Korrektes Einlesen der Daten überprüfen
      dazu wird die BASIC-Option verwendet
      Abschalten der FIML-Prozedur, listenweiser Fallausschluss

data: file = KFT.dat;
      listwise = on; ! FIML-Prozedur wird abgeschaltet
                  ! Stattdessen listenweiser Fallausschluss

variable: names = kft_v1 kft_v3 kft_q1 kft_q3 kft_n1 kft_n3;
          missing = all(-99);

analysis: type = basic;

```

The status bar at the bottom left says 'Ready' and the bottom right says 'Ln 4, Col 65'.

Abbildung 2.4: Geänderter Mplus-basic-Input zum Abschalten der FIML-Prozedur mit Hilfe des Unterbefehls listwise = on;

SAMPLE STATISTICS [Berechnet gemäß listenweisem Fallausschluss]					
Means					
KFT_V1	KFT_V3	KFT_Q1	KFT_Q3	KFT_N1	KFT_N3
11.899	8.974	12.387	7.732	11.112	8.281

Covariances					
	KFT_V1	KFT_V3	KFT_Q1	KFT_Q3	KFT_N1
KFT_V1	21.677				
KFT_V3	10.799	18.008			
KFT_Q1	6.057	5.310	11.242		
KFT_Q3	4.669	4.429	4.038	6.783	
KFT_N1	11.172	11.745	6.794	6.120	29.686
KFT_N3	6.383	7.230	4.417	4.001	9.294

Covariances	
	KFT_N3
KFT_N3	11.811

Correlations

	KFT_V1	KFT_V3	KFT_Q1	KFT_Q3	KFT_N1
KFT_V1	1.000				
KFT_V3	0.547	1.000			
KFT_Q1	0.388	0.373	1.000		
KFT_Q3	0.385	0.401	0.462	1.000	
KFT_N1	0.440	0.508	0.372	0.431	1.000
KFT_N3	0.399	0.496	0.383	0.447	0.496

Correlations	KFT_N3
KFT_N3	1.000

Itemstatistiken

	Mittelwert	Std.-Abweichung	Anzahl
Summenwert im KFT-Subtest V1	11.8989	4.65585	455
Summenwert im KFT-Subtest V3	8.9736	4.24360	455
Summenwert im KFT-Subtest Q1	12.3868	3.35293	455
Summenwert im KFT-Subtest Q3	7.7319	2.60434	455
Summenwert im KFT-Subtest N1	11.1121	5.44845	455
Summenwert im KFT-Subtest N3	8.2813	3.43665	455

Inter-Item-Kovarianzmatrix

	Summenwert im KFT-Subtest V1	Summenwert im KFT-Subtest V3	Summenwert im KFT-Subtest Q1	Summenwert im KFT-Subtest Q3	Summenwert im KFT-Subtest N1	Summenwert im KFT-Subtest N3
Summenwert im KFT-Subtest V1	21.677	10.799	6.057	4.669	11.172	6.383
Summenwert im KFT-Subtest V3	10.799	18.008	5.310	4.429	11.745	7.230
Summenwert im KFT-Subtest Q1	6.057	5.310	11.242	4.038	6.794	4.417
Summenwert im KFT-Subtest Q3	4.669	4.429	4.038	6.783	6.120	4.001
Summenwert im KFT-Subtest N1	11.172	11.745	6.794	6.120	29.686	9.294
Summenwert im KFT-Subtest N3	6.383	7.230	4.417	4.001	9.294	11.811

Inter-Item-Korrelationsmatrix

	Summenwert im KFT-Subtest V1	Summenwert im KFT-Subtest V3	Summenwert im KFT-Subtest Q1	Summenwert im KFT-Subtest Q3	Summenwert im KFT-Subtest N1	Summenwert im KFT-Subtest N3
Summenwert im KFT-Subtest V1	1.000	.547	.388	.385	.440	.399
Summenwert im KFT-Subtest V3	.547	1.000	.373	.401	.508	.496
Summenwert im KFT-Subtest Q1	.388	.373	1.000	.462	.372	.383
Summenwert im KFT-Subtest Q3	.385	.401	.462	1.000	.431	.447
Summenwert im KFT-Subtest N1	.440	.508	.372	.431	1.000	.496
Summenwert im KFT-Subtest N3	.399	.496	.383	.447	.496	1.000

Abbildung 2.5: Deskriptivstatistiken für die 6 KFT-Skalen, erzeugt in SPSS über Analysieren → Skalierung → Reliabilitätsanalyse → Statistik → Deskriptive Statistiken für Item/Zwischen Items Korrelationen/Kovarianzen. Die Werte stimmen mit den Mplus-Berechnungen überein, wenn in Mplus ebenfalls listenweiser Fallausschluss (listwise = on;) angefordert wird.

2.2 Einlesen von Summary-Daten (Kovarianz- oder Korrelationsmatrizen)

Wie zu Beginn von Kapitel 2 bereits erwähnt, ist es in manchen Fällen praktisch, Summary-Daten anstelle von Individualdaten einzulesen. Summary-Daten sind z.B. Kovarianz- oder Korrelationsmatrizen, ggf. ergänzt durch Mittelwerte und/oder Standardabweichungen aller Variablen. Wir zeigen hier das Einlesen von Summary-Daten für das Beispiel mit den sechs KFT-Variablen. Die einfachste Möglichkeit zum Einlesen von Summary-Daten ist das Eingeben der Matrix in eine einfache Textdatei. Dazu kann man in Windows z.B. das einfache Editor-Programm (**Start → Programme → Zubehör → Editor**) oder WordPad benutzen.

Abbildung 2.6 zeigt ein Editorfenster (Dateiname: **KFT_Summary-Daten.txt**) mit den eingegebenen Mittelwerten (1. Zeile), Standardabweichungen (2. Zeile) und der Produkt-Moment-Korrelationsmatrix der sechs beobachteten KFT-Variablen (die Datei befindet sich ebenfalls auf der Begleit-CD). Als Dezimaltrennzeichen müssen wie bei Individualdaten Punkte (keine Kommas) verwendet werden. Die Statistiken in einer solchen Textdatei können problemlos von Mplus gelesen und als Grundlage für die Schätzung beispielsweise eines Strukturgleichungsmodells verwendet werden.

KFT_Summary-Daten.txt - Editor					
Datei	Bearbeiten	Format	Ansicht	?	
11.899	8.974	12.387	7.732	11.112	8.281
4.656	4.244	3.353	2.604	5.448	3.437
1.000					
.547	1.000				
.388	.373	1.000			
.385	.401	.462	1.000		
.440	.508	.372	.431	1.000	
.399	.496	.383	.447	.496	1.000

Abbildung 2.6: Editor-Datei mit Summary-Daten der sechs KFT-Variablen. Die erste Zeile enthält die Mittelwerte der Variablen. Die zweite Zeile enthält deren Standardabweichungen. Darunter ist die Produkt-Moment-Korrelationsmatrix dargestellt. Abb. 2.7 zeigt, wie diese Daten korrekt in Mplus eingelesen werden können.

Die Mplus-Syntax zum Einlesen der Daten in der Datei **KFT_Summary-Daten.txt** ist in Abbildung 2.7 dargestellt. Unter **data** wird der Name des Datensatzes angegeben. Außerdem muss festgelegt werden, was für eine Art von Summary-Statistiken eingelesen werden, wobei die Reihenfolge der Befehle wichtig ist. Die Zeile **type = means std corr;** bedeutet, dass im Datensatz **KFT_Summary-Daten.txt** zunächst die Mittelwerte, dann die Standardabweichungen und dann die Korrelationen der Variablen eingegeben sind (zum Einlesen einer Kovarianzmatrix würde man den Befehl **cova** anstelle von **corr** benutzen).

Der Befehl **nobservations = 455;** teilt Mplus mit, auf wie vielen Beobachtungseinheiten die einzulesenden Summary-Daten beruhen (hier: $N = 455$). Diese Information kann aus Summary-Daten (im Gegensatz zu Individualdaten) nicht vom Programm erschlossen werden. Die Option **variable: names =** dient wiederum dazu, die Variablennamen zu definieren, die ebenfalls nicht in der Datendatei enthalten sind.

Die **basic**-Option steht für Summary-Daten nicht zur Verfügung. Aus diesem Grund werden zur Datenkontrolle Deskriptivstatistiken über **output: sampstat;** angefordert. Die Option **sampstat** steht hierbei für „sample statistics“ und liefert uns Deskriptivstatistiken für die sechs Variablen (in diesem Fall die beobachteten Mittelwerte sowie die Kovarianzmatrix). Da wir kein Modell explizit spezifizieren, schätzt Mplus automatisch ein sogenanntes „Nullmodell“ (oder Unabhängigkeitsmodell). Das Nullmodell nimmt an, dass zwischen den Variablen keinerlei Zusammenhänge bestehen. Als Modellparameter werden daher nur die Mittelwerte und Varianzen der Variablen geschätzt. Der Output zu diesem Modell ist nicht gesondert dargestellt, kann jedoch der Begleit-CD entnommen werden.

The screenshot shows the Mplus software interface. The title bar reads "Mplus - KFT-Summary-Daten-einlesen.inp". The menu bar includes File, Edit, View, Mplus, Graph, Window, and Help. Below the menu is a toolbar with various icons. The main window displays the Mplus input command language. The code in the input file is:

```

title: KFT-Datensatz einlesen
      Hier: Summary-Daten aus Datei "KFT_Summary-Daten.txt"
      Datei enthält die Mittelwerte, Standardabweichungen
      und Produkt-Moment-Korrelationen der 6 KFT-Variablen
      Berechnung von Deskriptivstatistiken und Nullmodell
      Annahme im Nullmodell: alle Variablen sind unkorreliert

data: file = KFT_Summary-Daten.txt;
      type = means std corr;
      nobbservations = 455;

variable: names = kft_v1 kft_v3 kft_q1 kft_q3 kft_n1 kft_n3;

output: sampstat;

```

The status bar at the bottom left says "Ready" and at the bottom right says "Ln 1, Col 1".

Abbildung 2.7: Mplus-Input-Datei zum Einlesen der Summary-Daten aus Abb. 2.6 (Datei **KFT_Summary-Daten.txt**). Die basic-Option kann bei Summary-Daten nicht verwendet werden. Aus diesem Grund werden zur Datenkontrolle Deskriptivstatistiken über output: sampstat; angefordert. Da kein Modell explizit spezifiziert wird, schätzt Mplus automatisch ein „Nullmodell“ mit allen Variablen, die unter names = aufgelistet sind.

Im folgenden Kapitel wird anhand des Beispiels linearer Strukturgleichungsmodelle in die Modellspezifikation in Mplus eingeführt. Hierbei werden zunächst simple lineare Regressionsmodelle betrachtet und anschließend komplexere Strukturgleichungsmodelle wie konfirmatorische Faktorenanalysen und latente Pfadanalysen.

3 Lineare Strukturgleichungsmodelle

3.1 Was sind lineare Strukturgleichungsmodelle?

Lineare Strukturgleichungsmodelle mit latenten Variablen (häufig auch als *Kovarianzstrukturmödelle* bezeichnet) werden dazu verwendet, komplexe Zusammenhänge zwischen kontinuierlichen Variablen auf latenter (messfehlerbereinigter) Ebene zu untersuchen. Lineare Strukturgleichungsmodelle kann man als generalisierte (multivariate) Regressionsmodelle auffassen. In diesen Modellen können simultan gerichtete Zusammenhänge (Regressionen) zwischen multiplen abhängigen und multiplen unabhängigen Variablen untersucht werden. Variablen, die dabei ausschließlich als unabhängige Variablen fungieren, bezeichnet man als *exogene* Variablen, da sie im Modell nicht durch andere Variablen erklärt werden. Zwischen exogenen Variablen werden ggf. ungerichtete Zusammenhänge, d.h. Kovarianzen bzw. Korrelationen, modelliert.

Variablen, die im Modell durch eine oder mehrere andere Variablen vorhergesagt werden, nennt man *endogene* Variablen. Endogene Variablen, die sowohl als abhängige als auch als unabhängige Variable fungieren, heißen *intervenierende* Variablen oder *Mediatorvariablen*, da sie Zusammenhänge zwischen Variablen vermitteln. Mediatoreffekte werden häufig im Rahmen von manifesten oder latenten Pfadanalysen untersucht (siehe Kapitel 3.5).

Lineare Strukturgleichungsmodelle mit latenten Variablen lassen sich in zwei „Teilmödelle“ zergliedern, deren Parameter in einer Strukturgleichungsanalyse jedoch simultan geschätzt werden (siehe Abb. 3.1). Im *Messmodell* („measurement model“) wird spezifiziert, wie die latenten Variablen („Faktoren“, hier mit η_j bezeichnet) durch beobachtete („manifeste“) Variablen (Indikatoren, Y_i) gemessen werden. Hierbei wird meist angenommen, dass die latente Variable ursächlich für die Kovarianz zwischen verschiedenen Indikatoren eines Konstrukts verantwortlich ist (sog. *reflektives Messmodell*), d.h. Variation in der latenten Variable bedingt Variation in den Indikatoren und erklärt Zusammenhänge zwischen verschiedenen Indikatoren.

Die Verknüpfung von manifesten und latenten Variablen geschieht über die Spezifikation von Regressionen, wobei in reflektiven Messmodellen die Indikatoren als *abhängige* und die Faktoren als *unabhängige* Variablen fungieren (in sogenannten *formativen* Messmodellen ist dies umgekehrt, formative Messmodelle werden hier jedoch nicht besprochen). Die Regressionskoeffizienten λ_{ij} dieser Regressionen bezeichnet man häufig als Faktorladungen.

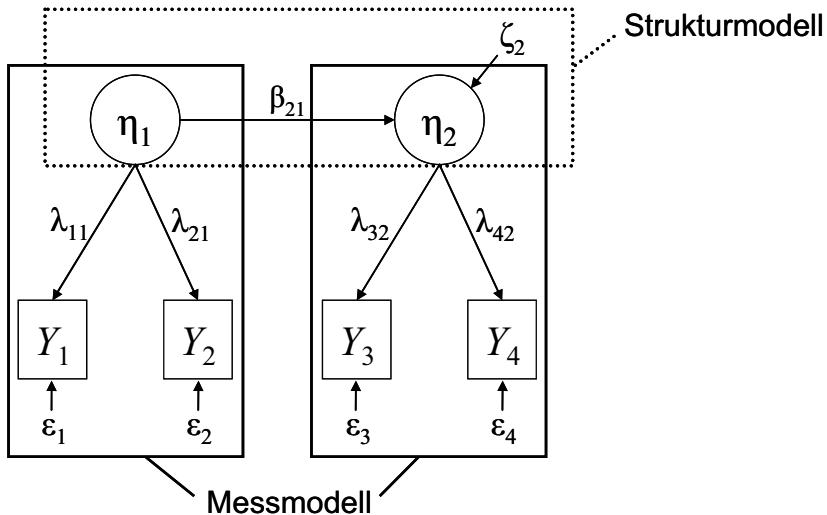


Abbildung 3.1: Beispiel für ein einfaches lineares Strukturgleichungsmodell mit zwei latenten Variablen (η_1 und η_2) welche jeweils durch zwei Indikatoren (Y_1 und Y_2 als Indikatoren für η_1 sowie Y_3 und Y_4 als Indikatoren für η_2) gemessen werden. Die Parameter λ_{11} - λ_{42} sind Faktorladungen. Die Variablen ε_1 - ε_4 bezeichnen Fehlervariablen für die Indikatoren. Im Strukturmodell wird eine latente Regression des Faktors η_2 auf η_1 geschätzt, wobei β_{21} den entsprechenden Regressionskoeffizienten und ζ_2 eine latente Residualvariable indizieren.

Im *Strukturmodell* („structural model“, „latent variable model“) werden Zusammenhangsstrukturen zwischen den latenten Variablen in Form von Kovarianzen oder Regressionen bzw. Pfadanalysen spezifiziert. Werden zwischen den latenten Variablen keine oder ausschließlich ungerichtete Zusammenhänge (Kovarianzen) angenommen, so spricht man von einer *konfirmatorischen Faktorenanalyse* (siehe Abschnitt 3.4). Werden hingegen gerichtete Zusammenhänge (Regressionen) spezifiziert, so bezeichnet man die Modelle häufig auch als *latente Regressionsanalysen* (siehe das Modell in Abb. 3.1 sowie Abschnitt 3.3) bzw. – sofern mehrere latente endogene Variablen betrachtet werden – als *latente Pfadanalysen* (siehe Abschnitt 3.5).

Der wichtigste Vorteil linearer Strukturgleichungsmodelle mit latenten Variablen besteht darin, dass durch die Verwendung latenter Variablen Messfehler in der Analyse explizit berücksichtigt werden. Dadurch lassen sich die Zusammenhänge im Strukturmodell korrekter schätzen als beispielsweise in Korrelations-, Regressions- oder Pfadanalysen, die ausschließlich auf der Betrachtung beobachteter (fehlerbehafteter) Variablen beruhen.

Weiterhin erlauben es Strukturgleichungsmodelle, komplexe Theorien über Mess- und Zusammenhangsstrukturen zwischen Variablen einer empirischen Prüfung (Modelltestung) zu unterziehen. Dabei können auch verschiedene konkurrierende Modelle statistisch miteinander verglichen werden. Strukturgleichungsmodelle sind sehr flexibel und eignen sich zur Auswertung einer Vielzahl komplexer Fragestellungen. So können Strukturgleichungsmodelle auch im Bereich der Veränderungsmessung nutzbringend eingesetzt werden, beispielsweise um stabile von situationsbedingten Einflüssen auf psychologische Messungen

zu trennen (siehe Kapitel 4.2 sowie Steyer, Schmitt & Eid, 1999; Steyer, Ferring & Schmitt, 1992), zur Messinvarianztestung (Kapitel 4.1.4) sowie zur Analyse latenter Veränderung (Kapitel 4.3-4.5) und latenter Wachstumskurven (Kapitel 4.5; siehe auch Bollen & Curran, 2006; Duncan, Duncan & Strycker, 2006; Steyer, Eid & Schwenkmezger, 1997). Ferner werden Strukturgleichungsmodelle häufig zur Auswertung von multimethodal erhobenen Daten (sog. *Multitrait-Multimethod-Daten*; siehe z.B. Eid, Nussbeck, Geiser, Cole, Gollwitzer & Lischetzke, 2008 sowie Eid, Nussbeck & Lischetzke, 2006) verwendet.

Zur graphischen Darstellung von Strukturgleichungsmodellen werden häufig Pfaddiagramme wie in Abbildung 3.1 eingesetzt. Auch im vorliegenden Text werden überwiegend Pfaddiagramme (anstelle mathematischer Gleichungen) verwendet, um die zu analysierenden Modelle zu illustrieren. In Pfaddiagrammen werden beobachtete Variablen (Synonyme: manifeste Variablen, Indikatoren) meistens in Kästchen dargestellt, während nicht direkt beobachtbare Variablen (Synonyme: latente Variablen, latente Faktoren) üblicherweise in Kreisen oder Ellipsen erscheinen. Gerichtete Zusammenhänge (Regressionsen) werden mit Pfeilen (\rightarrow) dargestellt, wobei die Variable, die den Pfeil aussendet in Bezug auf die Variable, auf die der Pfeil zeigt, eine unabhängige Variable (Prädiktor, Regressor) ist. Alle Variablen, auf die ein Pfeil zeigt, sind endogene (abhängige) Variablen. Variablen, die nur Pfeile aussenden, aber auf die keine einseitigen Pfeile zeigen werden als exogene („außenstehende“) Variablen bezeichnet, da ihre Varianz im Modell nicht durch andere Variablen erklärt wird.

Gebogene oder gerade Doppelpfeile (\leftrightarrow) indizieren – im Gegensatz zu einseitigen Pfeilen (\rightarrow) – ungerichtete Zusammenhänge, d.h. Kovarianzen bzw. Korrelationen. Manifeste Variablen werden im vorliegenden Text mit Y oder mit dem tatsächlichen Variablennamen bezeichnet. Latente Variablen werden mit η (Eta) bezeichnet oder erhalten eine Bezeichnung, welche am besten ihre Bedeutung charakterisiert. Fehlervariablen (Residualvariablen), die manifesten Variablen zugeordnet sind, werden mit ϵ (Epsilon) und latente Residualvariablen mit ζ (Zeta) symbolisiert.

Empfehlenswerte Einführungstexte zur Theorie linearer Strukturgleichungsmodelle sind z.B. die Bücher von Bollen (1989), Kaplan (2009), Kline (2004), Loehlin (1998), Raykov und Marcoulides (2006) sowie Schumacker und Lomax (1996). Ein deutschsprachiges Buch zu Strukturgleichungsmodellen hat Reinecke (2005) bereit gestellt, ein deutsches Einführungskapitel liefert Eid (1999).

Box 3.1: Die wichtigsten Parameter, die in Strukturgleichungsmodellen mit latenten Variablen geschätzt werden sowie gängige Notationen

Parameter des Messmodells (Verknüpfung zwischen manifesten Variablen Y_i und latenten Variablen η_j):

- Faktorladungen (Regressionsgewichte) λ_{ij} der beobachteten Variablen Y_i auf den latenten Variablen η_j
- Intercepts (Achsenabschnitte) α_i der beobachteten Variablen Y_i
- Residualvarianzen / Fehlervarianzen für beobachtete Variablen $Var(\varepsilon_i)$
- Ggf. Residualkovarianzen $Cov(\varepsilon_i, \varepsilon_{i'})$ bzw. -korrelationen $Corr(\varepsilon_i, \varepsilon_{i'})$

Parameter des Strukturmodells (Modell auf Ebene der latenten Variablen / Faktoren η_j):

- Faktorvarianzen $Var(\eta_j)$ (für exogene Faktoren)
- Faktormittelwerte $E(\eta_j)$ (für exogene Faktoren)
- Faktorkovarianzen $Cov(\eta_j, \eta_{j'})$ bzw. -korrelationen $Corr(\eta_j, \eta_{j'})$ (für exogene Faktoren)
- Latente Pfadkoeffizienten/Regressionskoeffizienten $\beta_{jj'}$ für die Regression eines endogenen Faktors η_j auf einen Faktor $\eta_{j'}$. Bei Faktorenanalysen höherer Ordnung wird die entsprechende Faktorladungen eines Faktors niedrigerer Ordnung η_j auf einem Faktor höherer Ordnung hier mit γ_j bezeichnet
- Latente Intercepts (Achsenabschnitte) für endogene Faktoren η_j werden im vorliegenden Buch in latenten Regressions- oder Pfadanalysen mit β_{0j} und in Faktorenanalysen höherer Ordnung mit γ_{0j} bezeichnet.
- Latente Residualvarianzen $Var(\zeta_j)$ (für endogene Faktoren)
- Ggf. latente Residualkovarianzen $Cov(\zeta_j, \zeta_{j'})$ bzw. -korrelationen $Corr(\zeta_j, \zeta_{j'})$ (für endogene Faktoren).

Korrelationen werden im vorliegenden Buch auch mit r bezeichnet.

3.2 Einfache Regressionsanalyse

Grundlage der linearen Strukturgleichungsmodelle ist die lineare Regressionsanalyse. Wir beginnen hier mit dieser einfachen Form der Analyse, um zunächst den basalen Aufbau der Modellspezifikation in Mplus kennenzulernen. Unser erstes Modell ist ein simples bivariate lineares Regressionsmodell mit einer manifesten unabhängigen Variable (UV) und einer manifesten abhängigen Variable (AV).

Box 3.2: Fallbeispiel einfache lineare Regression

In unserem konkreten Beispiel gehen wir davon aus, dass ein Forscher herausfinden will, in welchem Ausmaß sich die Leistung bei numerischen („quantitativen“ = q) Aufgaben (Aufgaben mit Zahlen) durch die Leistung bei figuralen („non-verbalen“ = n) Aufgaben vorhersagen lässt. Zur Vereinfachung ziehen wir in unserer Analyse nur die Leistung auf der Variable **kft_q1** als Maß für quantitative Fähigkeiten (entspricht der AV = Y) und die Variable **kft_n1** als Maß für non-verbale Fähigkeiten (entspricht der UV = X) heran. Das entsprechende Modell ist in Abbildung 3.2 als Pfaddiagramm dargestellt, wobei bereits die mit Mplus geschätzten Modellparameter eingetragen sind.

Formal lässt sich das einfache bivariate Regressionsmodell wie folgt aufschreiben (siehe z.B. Cohen, Cohen, West & Aiken, 2003):

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon,$$

wobei β_0 den Achsenabschnitt (Intercept), β_1 den Steigungskoeffizienten (Slope) für die Regression der AV (Kriteriumsvariable) Y auf die UV (Prädiktorvariable) X und ε eine Residualvariable bezeichnen.

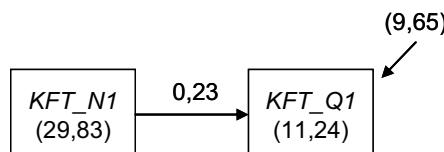


Abbildung 3.2: Pfaddiagramm des geschätzten bivariaten Regressionsmodells zur Vorhersage der beobachteten Variable **kft_q1** (quantitative Fähigkeiten) durch **kft_n1** (non-verbale Fähigkeiten). Die Werte geben das von Mplus geschätzte unstandardisierte Regressionsgewicht ($\hat{\beta}_1 = 0,23$) sowie die geschätzten Varianzen der Variablen (29,83 und 11,24) und die Residualvarianz ($\hat{\sigma}_\varepsilon^2 = 9,65$) an.

Die einfachste und schnellste Möglichkeit, ein Modell in Mplus zu spezifizieren ist, die Mplus-Input-Datei für die **basic**-Analyse erneut in Mplus zu öffnen und sie unter einem neuen Namen abzuspeichern (im Mplus-Menü **File → Save as**), z.B. mit dem Namen **Modell_1_Bivariate_Regression.inp** (siehe Abb. 3.3). Dann kann man in der neu abgespeicherten Datei die bestehenden Syntaxbefehle entsprechend ändern bzw. erweitern, um das Modell aus Abbildung 3.2 schätzen zu können, ohne die **basic**-Analyse zu überschreiben.

Box 3.3: Tipps für die erste Modellspezifikation

Nachdem man das korrekte Einlesen der Daten in Mplus mit Hilfe der basic-Option überprüft hat (vgl. Kapitel 2), kann man beginnen, das erste Modell zu spezifizieren. Grundsätzlich ist es sinnvoll, mit einem einfachen (Teil-)Modell zu beginnen, da man mögliche Fehler und Probleme in einem simplen Modell leichter lokalisieren kann als in einem komplexen. Möchte man ein komplexes Strukturgleichungsmodell mit vielen Variablen analysieren, so empfiehlt es sich, das Modell „langsam aufzubauen“, d.h. zunächst mit nur einigen Variablen oder Faktoren zu beginnen und dann die Komplexität des Modells schrittweise zu erhöhen. Auf diese Weise lassen sich mögliche Fehler und Probleme leichter einer bestimmten Konstellation von Variablen zuordnen. Entsprechendes „Troubleshooting“ ist dann einfacherer, als bei einem komplexen Modell mit vielen Variablen und Parametern.

Der Einfachheit halber wählen wir im Folgenden bei unseren Datenbeispielen stets listenweisen Fallausschluss anstelle der FIML-Prozedur. In realen Analysen sollten Nutzer jedoch die Benutzung der FIML-Prozedur in Erwägung ziehen (siehe dazu z.B. Enders, 2010; Geiser et al., in Vorbereitung; Reinecke, 2005, sowie Schafer & Graham, 2002).

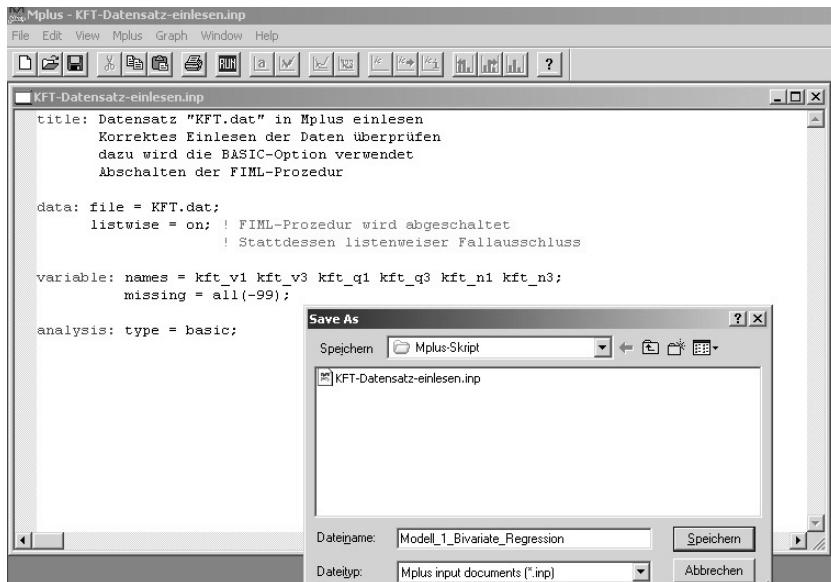


Abbildung 3.3: Abspeichern der Mplus-Input-Datei aus Abbildung 2.1 unter einem neuen Namen. Die Befehle können dann in der neuen Datei entsprechend abgeändert werden, um ein konkretes Modell zu schätzen (siehe Abb. 3.4).

Die Mplus-Spezifikation des bivariaten Regressionsmodells wird in Abbildung 3.4 gezeigt. Im variable-Befehl ist nun eine Zeile hinzugekommen, welche die im Modell zu benutzenen Variable spezifiziert.

zenden Variablen definiert. Es handelt sich um den usevar-Unterbefehl, der angibt, welche Variablen im Modell verwendet werden sollen. In unserem einfachen Regressionsmodell verwenden wir nur die Variablen kft_q1 und kft_n1, daher lautet der Befehl usevar = kft_q1 kft_n1;

The screenshot shows the Mplus software interface with the title bar 'Mplus - Modell_1_Bivariate_Regression.inp'. Below the menu bar, there is a toolbar with various icons. The main window displays the Mplus input command syntax:

```

title: Spezifikation eines einfachen bivariaten Regressionsmodells
      Vorhersage der Variable kft_q1 durch kft_n1
      kft_n1 --> kft_q1 <-- Error

data: file = KFT.dat;
      listwise = on; ! FIML-Prozedur wird abgeschaltet
                  ! Stattdessen listenweiser Fallausschluss

variable: names = kft_v1 kft_v3 kft_q1 kft_q3 kft_n1 kft_n3;
          usevar = kft_q1 kft_n1;
          missing = all(-99);

analysis: type = general;

model: kft_q1 on kft_n1; ! Regression von kft_q1 auf kft_n1

output: sampstat standardized stdyx;

```

Abbildung 3.4: Mplus-Spezifikation des bivariaten linearen Regressionsmodells aus Abb. 2.4

Zur korrekten Verwendung des usevar-Unterbefehls muss Folgendes unbedingt beachtet werden: Im usevar-Unterbefehl dürfen nur diejenigen Variablen aufgelistet werden, die im model-Befehl auch tatsächlich benutzt werden. Ansonsten kommt es zu einer u.U. schwerwiegenden Modellfehlspezifikation, da Mplus annimmt, dass diejenigen Variablen, die unter usevar zwar genannt sind, im model-Befehl jedoch nicht benutzt werden, mit allen anderen Variablen im Modell unkorreliert sind (s.u. sowie Anhang B). Hat man mehr Variablen unter usevar definiert als man tatsächlich im model-Befehl benutzt oder hat man vergessen, den usevar-Unterbefehl zu verwenden, so erscheinen im Mplus-Output Warnmeldungen, die wie folgt aussehen (hat man den usevar-Unterbefehl vergessen, so geht Mplus davon aus, dass alle Variablen, die unter names aufgelistet sind, die im Modell eine Rolle spielen):

```

*** WARNING in Model command
Variable is uncorrelated with all other variables:  KFT_N3
*** WARNING in Model command
All least one variable is uncorrelated with all other variables in
the model.
Check that this is what is intended.
2 WARNING(S) FOUND IN THE INPUT INSTRUCTIONS

```

In diesem Beispiel wurde (fälschlicherweise) die Variable kft_n3 unter usevar = aufgelistet, jedoch nicht in der Modellspezifikation verwendet. Sofern eine derartige Meldung auftritt, sollte man den usevar-Unterbefehl im Input entsprechend korrigieren bzw. hinzufügen und dann das Modell erneut mit dem korrigierten Input schätzen lassen.

Im analysis-Befehl wurde der Typ der Analyse von basic in general geändert, da Mplus unter der basic-Option nur deskriptive Statistiken berechnen, jedoch keine Modellparameter schätzen würde. Als zusätzliche Befehle wurden der model-Befehl und der output-Befehl aufgenommen. Im model-Befehl wird das zu schätzende Modell spezifiziert. Die Zeile

```
model: kft_q1 on kft_n1;
```

bedeutet, dass eine (in diesem Fall lineare) Regression von kft_q1 auf kft_n1 berechnet werden soll, wobei das Schlüsselwort „on“ für die englische Übersetzung für „Regression auf“ (engl. *regressed on*) steht. Hinweis: Eine andere Art von Regression (z.B. Probitregression oder logistische Regression) könnte dadurch angefordert werden, dass man die abhängige Variable im variable-Befehl als kategorial definiert (mit Hilfe des Unterbefehls categorical =) und eine spezielle Schätzmethode anfordert (siehe dazu den Geiser et al., in Vorbereitung sowie Muthén & Muthén, 1998-2007).

Mit Hilfe der Outputzeile werden zusätzlich Deskriptivstatistiken, Kovarianzen und Korrelationen für alle im Modell verwendeten beobachteten Variablen (sampstat steht für *sample statistics*) sowie die vollstandardisierte Lösung (standardized stdyx) angefordert. Im Folgenden ist der wichtigste Teil des Mplus-Outputs zu diesem Modell wiedergegeben. Hierbei ist wichtig, dass im Output im Anschluss an die Reproduktion der Inputspezifikationen von Mplus die folgende Meldung ausgegeben wird:

```
INPUT READING TERMINATED NORMALLY
```

Diese Meldung weist u.a. darauf hin, dass Mplus im Input keine Syntaxfehler feststellen konnte und dass Mplus keine Schwierigkeiten hatte, auf den angegebenen Datensatz zugreifen. Wichtig ist hierbei, dass diese Meldung dennoch keine Garantie dafür ist, dass beim Lesen der Daten und der Spezifikation des Inputs keine Fehler aufgetreten sind. Man sollte darum unbedingt zusätzlich die Plausibilität der ausgegebenen Stichprobengröße, der Deskriptivstatistiken sowie der Modellparameterschätzungen überprüfen, auch wenn bereits zuvor eine basic-Analyse (siehe Abschnitt 2.1.1) durchgeführt wurde. Auch ist diese Meldung keine Garantie dafür, dass das Modell konvergiert und die Schätzungen in Ordnung sind. Etwaige Probleme, die sich auf die Schätzung von Modellgrößen beziehen, würden von Mplus weiter unten im Output unter den sample statistics berichtet werden.

Wenn die oben gezeigte Meldung im Output nicht erscheint, so stehen an ihrer Stelle eine oder mehrere Fehlermeldungen, die auf Fehler bei der Inputspezifikation und/oder dem Datenelesevorgang hinweisen. Anhang B am Ende dieses Buches gibt einen Überblick über die „klassischen“ Probleme bei der Inputspezifikation sowie entsprechendes „Troubleshooting“.

Anschließend werden von Mplus unter summary of analysis technische Details zur durchgeführten Analyse wiedergegeben (Anzahl der Gruppen – hier nur eine, da es sich nicht um eine Multigruppenanalyse handelt; Anzahl der Versuchspersonen, Anzahl und

Namen der abhängigen und unabhängigen manifesten und latenten Variablen im Modell, Name und Art des Datensatzes). Diese Informationen sind nützlich, um noch einmal zu überprüfen, ob die gewünschte Analyse korrekt durchgeführt wurde.

```
SUMMARY OF ANALYSIS

Number of groups                                1
Number of observations                           456
Number of dependent variables                  1
Number of independent variables                1
Number of continuous latent variables          0

Observed dependent variables

  Continuous
    KFT_Q1

Observed independent variables

  KFT_N1

Input data file(s)
  KFT.dat

Input data format FREE
```

Nachfolgend werden Deskriptivstatistiken (sample statistics) für die verwendeten Variablen ausgegeben. Diese sollten zum einen nochmals gründlich auf ihre Plausibilität hin überprüft werden, falls beim Dateneinlesen doch noch etwas schief gegangen ist und sich dies häufig an dieser Stelle erkennen lässt. Zum anderen sind diese Statistiken häufig nützlich, um die Modellschätzergebnisse besser interpretieren zu können. Es zeigt sich hier beispielsweise, dass zwischen kft_q1 und kft_n1 eine moderate Korrelation von $r = 0,375$ besteht (Abschnitt Correlations).

```
SAMPLE STATISTICS

Means
      KFT_Q1        KFT_N1
  1   12.377     11.088

Covariances
      KFT_Q1        KFT_N1
KFT_Q1  11.235
KFT_N1  6.870     29.826

Correlations
      KFT_Q1        KFT_N1
KFT_Q1  1.000
KFT_N1  0.375     1.000
```

Im Anschluss an die deskriptiven Statistiken sollte der Benutzer ein weiteres Mal auf mögliche Fehlermeldungen achten. Wie bereits oben erwähnt, gibt Mplus an dieser Stelle ggf. Warnmeldungen aus, die sich speziell auf Probleme bei der Schätzung des spezifizierten Modells beziehen. Solche Meldungen können beispielsweise dann vorkommen, wenn ein Modell nicht identifiziert ist, wenn unzulässige Schätzergebnisse (z.B. negative Varianzen) aufraten oder wenn Konvergenzprobleme vorlagen. In unserem Beispiel erscheinen keine Warnmeldungen. Die Modellschätzung wurde problemlos durchgeführt, wie es bei einem einfachen Regressionsmodell auf manifestierter Ebene zu erwarten ist. Daher erhalten wir nur die (beruhigende) Meldung

THE MODEL ESTIMATION TERMINATED NORMALLY

Achtung: Wann immer an dieser Stelle zusätzlich zu dieser Meldung noch weitere Meldungen erscheinen, sollte man den weiteren Output (die Fitindices und die Modellschätzergebnisse) nur mit größter Vorsicht oder gar nicht interpretieren, da in einem solchen Fall möglicherweise gravierende Schätzprobleme vorliegen.

Nachfolgend erscheinen eine Reihe von Statistiken zur Beurteilung der Modellanpassung (TESTS OF MODEL FIT; siehe Box 3.7; für eine detaillierte Besprechung dieser Indizes siehe Bollen & Long, 1993; Schermelleh-Engel, Moosbrugger & Müller, 2003). Da wir in diesem Fall nur ein einfaches bivariate Regressionsmodell mit ausschließlich manifesten Variablen analysieren, sind diese Statistiken für uns nicht von Interesse. Ein einfaches Regressionsmodell schöpft die in den Daten vorhandenen Informationen voll aus (daher bezeichnet man es auch als ein *saturiertes* Modell). Aus diesem Grund sind alle globalen Fitindices für dieses Modell für uns wenig interessant (ein saturiertes Modell weist trivialerweise stets eine perfekte Anpassung an die Daten auf).

Einzig der Chi-Quadrat-Test für das Unabhängigkeitsmodell (Chi-Square Test of Model Fit for the Baseline Model) ist für uns aufschlussreich. Das Unabhängigkeitsmodell nimmt eine Null-Korrelation zwischen kft_q1 und kft_n1 an. Der Chi-Quadrat-Test für das Unabhängigkeitsmodell ist im vorliegenden Beispiel signifikant ($\chi^2 = 69,235, df = 1, p < 0,0001$) und weist somit darauf hin, dass das Unabhängigkeitsmodell verworfen werden muss. Daraus können wir schließen, dass die Korrelation zwischen kft_q1 und kft_n1 statistisch bedeutsam von Null verschieden ist.

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

Value	0.000
Degrees of Freedom	0
P-Value	0.0000

Chi-Square Test of Model Fit for the Baseline Model

Value	69.235
Degrees of Freedom	1
P-Value	0.0000

CFI / TLI

CFI	1.000
TLI	1.000

Loglikelihood

H0 Value	-2585.137
H1 Value	-2585.137

Information Criteria

Number of Free Parameters	3
Akaike (AIC)	5176.273
Bayesian (BIC)	5188.641
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	5179.120

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.000
90 Percent C.I.	0.000 0.000
Probability RMSEA <= .05	0.000

SRMR (Standardized Root Mean Square Residual)

Value	0.000
-------	-------

Im Anschluss an die Modellgütemaße erscheinen die Parameterschätzungen des Modells, und zwar zunächst in *unstandardisierter Form* (unter MODEL RESULTS).

		Two-Tailed		
	Estimate	S.E.	Est./S.E.	P-Value
KFT_Q1 ON KFT_N1	0.230	0.027	8.647	0.000
Intercepts KFT_Q1	9.823	0.329	29.833	0.000
Residual Variances KFT_Q1	9.652	0.639	15.100	0.000

Bei dem Wert 0,230 in der ersten Zeile (unter der Spalte Estimate) handelt es sich um den geschätzten unstandardisierten Regressionskoeffizienten $\hat{\beta}_1$ (Steigungskoeffizient, Slope) für die Regression von kft_q1 auf kft_n1. In der zweiten Spalte wird der für diesen Koeffizienten geschätzte Standardfehler (standard error, S.E.) von 0,027 mitgeteilt. Das Verhältnis Estimate/S.E. (dritte Spalte) ergibt eine empirische Prüfgröße (z-Wert), die zur Signifikanztestung herangezogen werden kann. Hier ergibt sich für den geschätzten Parameter $\hat{\beta}_1$ ein z-Wert von 8,647. Anhand der 4. Spalte (Two-tailed P-Value) erfahren wir, dass $\hat{\beta}_1$ bei einem zweiseitigen Test auf dem 0,1 % Niveau statistisch signifikant von Null verschieden ist (da der ausgegebene p-Wert kleiner ist als 0,001). Dies bedeutet, dass die Leistung im nonverbalen Untertest kft_n1 einen signifikanten Prädiktor für die Leistung im quantitativen Untertest kft_q1 darstellt.

In der nächsten Zeile wird der geschätzte Achsenabschnittsparameter (Intercept) ausgegeben. Er beträgt $\hat{\beta}_0 = 9,823$ und ist mit einem z-Wert von 29,833 ebenfalls statistisch bedeutsam von Null verschieden. Mit Hilfe der geschätzten Parameter $\hat{\beta}_0$ und $\hat{\beta}_1$ können wir die Regressionsgleichung aufstellen:

$$kft_q1 = 9,823 + 0,230 \cdot kft_n1 + \varepsilon.$$

Somit wird für Kinder, die einen kft_n1-Score von Null (keine Aufgabe richtig gelöst) erzielt haben ein kft_q1-Score von 9,823 vorhergesagt. Der erwartete Anstieg im kft_q1-Score für einen Anstieg des kft_n1-Scores um einen Punkt beträgt 0,23 Punkte.

Box 3.4: Exkurs Zentrierung

Der Intercept-Term ($\hat{\beta}_0$) in einer linearen Regressionsanalyse ist nur dann sinnvoll interpretierbar, wenn die unabhängige(n) Variable(n) einen bedeutsamen Nullpunkt haben. Dies ist jedoch bei vielen sozialwissenschaftlichen Variablen (z.B. Extraversion, Schulerfolg, Lebenszufriedenheit) – zumindest in der Rohmetrik – nicht gegeben. Daher sollte man in vielen Fällen in Erwägung ziehen, die unabhängige(n) Variable(n) vor der Analyse zu zentrieren. Zentrierung bedeutet, dass von jedem Rohwert eine bedeutsame Konstante (meistens der Mittelwert der Variablen) subtrahiert wird. Durch die Zentrierung wird der Nullpunkt der unabhängigen Variablen hin zum Mittelwert verschoben, wodurch der Nullpunkt der Variablen nach der Zentrierung einem bedeutsamen Wert entspricht, nämlich dem typischen (durchschnittlichen) Fall in der Stichprobe (Cohen et al., 2003).

Der Intercept-Term ist folglich bei zentrierten unabhängigen Variablen als vorhergesagter Kriteriumswert für Personen mit *mittleren Ausprägungen* auf der oder den Prädiktorvariablen sinnvoll interpretierbar. Eine besonders wichtige Rolle spielt die Zentrierung in der moderierten Regression (siehe Aiken & West, 1991) sowie bei Mehrebenenregressionsanalysen (vgl. Kapitel 5).

In Mplus kann eine Zentrierung sehr einfach durchgeführt werden, indem im variable-Befehl zusätzlich der Unterbefehl centering aufgenommen wird. In unserem Fall erweitert sich unser Mplus-Input dabei um die folgende Zeile (der vollständige Input mit zentrierter unabhängiger Variable befindet sich auf der Begleit-CD):

```
variable: centering = grandmean(kft_n1) ;
```

Diese Zeile bewirkt, dass die Prädiktorvariable kft_n1 vor Aufnahme in die Regression am Gesamtmittelwert (sogenannter *Grand Mean*) der Stichprobe zentriert wird. Unsere Regressionsgleichung lautet nach Zentrierung der unabhängigen Variable:

$$kft_q1 = 12,377 + 0,230 \cdot kft_n1_zentriert + \epsilon.$$

Wie man sieht, verändert sich in der einfachen Regression durch die Zentrierung nur der geschätzte Achsenabschnitt $\hat{\beta}_0$, nicht jedoch das geschätzte Regressionsgewicht $\hat{\beta}_1$. Der Intercept wird nun auf $\hat{\beta}_0 = 12,377$ geschätzt. Dieser Wert gibt den kft_q1-Wert an, den man für eine Person mit einem durchschnittlichen kft_n1-Wert erwartet (er entspricht genau dem kft_q1-Mittelwert).

Die dritte Zeile der MODEL RESULTS Tabelle liefert uns die unstandardisierte Residualvarianz (geschätzte Varianz der Fehlervariable ϵ). Diese beträgt hier 9,652. Inhaltlich bedeutsam wird dieser Wert, wenn wir ihn zu der beobachteten Varianz der abhängigen Variable (Varianz von kft_q1) in Beziehung setzen. Diese beobachtete Kriteriumsvarianz finden wir im oberen Teil des Outputs unter sample statistics (abzulesen aus der Hauptdiagonalen der Kovarianzmatrix). Die beobachtete Varianz von kft_q1 beträgt demnach 11,235.

Wir sehen hieran, dass die Residualvarianz (also die nicht durch die Regression aufgeklärte Varianz) im Verhältnis zur Gesamtvarianz recht groß ist. Daraus lässt sich schließen, dass mit dem non-verbalen Untertest kft_n1 nur ein geringer Teil der Varianz des quantitativen Untertests kft_q1 aufgeklärt werden kann. Durch die Berechnung des *Determinationskoeffizienten R²* kann man erkennen, dass nur ca. 14,1% der Varianz aufgeklärt werden kann:

$$R^2 = 1 - (\text{Residualvarianz} / \text{Gesamtvarianz}) = 1 - (9,652 / 11,235) = 0,141.$$

Im Anschluss an die unstandardisierten Modellschätzungen erhalten wir unter STANDARDIZED MODEL RESULTS (STDYX Standardization) die vollstandardisierten Parameterschätzwerte sowie unter R-SQUARE den berechneten Determinationskoeffizienten von $R^2 = 0,141$, den wir bereits oben anhand des Verhältnisses der unstandardisierten Residualvarianz zur Gesamtvarianz von Hand ausgerechnet hatten.

Der geschätzte standardisierte Regressionskoeffizient beträgt 0,375 und entspricht damit exakt der bivariaten Korrelation zwischen kft_n1 und kft_q1 (in einer bivariaten Regressionsanalyse entspricht das standardisierte Regressengewicht stets der bivariaten Korrelation zwischen dem Prädiktor und dem Kriterium).

STANDARDIZED MODEL RESULTS (STDYX Standardization)				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
KFT_Q1 ON KFT_N1	0.375	0.040	9.329	0.000
Intercepts KFT_Q1	2.931	0.166	17.678	0.000
Residual Variances KFT_Q1	0.859	0.030	28.448	0.000
R-SQUARE				
Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
KFT_Q1	0.141	0.030	4.664	0.000

Der standardisierte Intercept-Term (hier für die Lösung mit unzentrierter unabhängiger Variable) errechnet sich aus dem unstandardisierten Intercept (9,823) dividiert durch die modellimplizierte Standardabweichung der abhängigen Variable. Die Standardabweichung wird von Mplus nicht berichtet, kann jedoch als Wurzel aus der Varianz berechnet werden ($\sqrt{11,235} = 3,352$). Der standardisierte Intercept-Term beträgt also $9,823 / 3,352 = 2,931$. Dieser Wert ist jedoch in der Praxis eher selten von Interesse.

Die standardisierte Residualvarianz von 0,859 entspricht dem *Indeterminationskoeffizienten* ($1 - R^2$). Sie gibt den Anteil der Varianz von kft_q1 an, der nicht durch kft_n1 erklärt wird, hier ca. 85,9%.

3.3 Regressionsanalyse auf latenter Ebene

Das in Abschnitt 3.2 beschriebene bivariate Regressionsmodell betrachtete den Zusammenhang zwischen zwei manifesten Variablen (kft_q1 und kft_n1), das heißt auf der Basis von beobachteten (und somit noch messfehlerbehafteten) Variablen. Ein Nachteil von Regressionsanalysen auf manifester Ebene ist, dass diese von perfekt reliablen Messungen ausgehen und Messfehler in den unabhängigen Variablen nicht explizit berücksichtigen (Messfehlervarianz in der abhängigen Variable wird in manifesten Regressionsanalysen berücksichtigt, ist jedoch mit der Residualvarianz, d.h. Varianz von ϵ , konfundiert).

Strukturgleichungsmodelle mit latenten Variablen erlauben dagegen die Spezifikation von Regressionen auf latenter Ebene, d.h. zwischen messfehlerbereinigten, latenten Variablen (Faktoren). Dies hat den Vorteil, dass Messfehler sowohl in den unabhängigen als auch in den abhängigen Variablen explizit berücksichtigt werden und auch Schätzungen der Reliabilitäten der beobachteten Variablen erhältlich sind. Die Berücksichtigung von Messfehlern führt dazu, dass die Parameter der Regression (als Parameter des latenten Strukturmodells) präziser geschätzt werden als dies in manifesten Regressionsanalysen mit fehlerbehafteten Variablen möglich ist.

Für jede latente Variable benötigt man im Regelfall multiple Indikatoren (mindestens zwei, aus Identifikations- und schätztechnischen Gründen ist es jedoch empfehlenswert, mindestens drei Indikatoren pro Faktor zu verwenden). In diesem Abschnitt wird gezeigt, wie der Zusammenhang zwischen quantitativen und non-verbalen KFT-Fähigkeiten auf latenter – statt wie in Kapitel 3.2 auf manifester – Ebene untersucht werden kann.

Box 3.5: Fallbeispiel latente Regressionsanalyse

Wir verfügen in unserem Datensatz **KFT.dat** über zwei Indikatoren für jede Fähigkeitsfazette (quantitativ vs. non-verbal). Somit können wir quantitative und non-verbale Fähigkeiten jeweils als latenter Faktor modellieren. Die manifesten Variablen kft_q1 und kft_q3 dienen als Indikatoren für einen quantitativen Fähigkeitsfaktor (KFT_Q). Die Variablen kft_n1 und kft_n3 nutzen wir als Indikatoren für eine latente Variable „non-verbale Fähigkeiten“ (KFT_N). Die Regression quantitativer auf non-verbale Leistung wird auf Ebene der latenten Variablen (d.h. im Strukturmodell) spezifiziert. Das entsprechende Strukturgleichungsmodell kann man als latente Regressionsanalyse bezeichnen. Es ist in Abbildung 3.5 als Pfaddiagramm dargestellt.

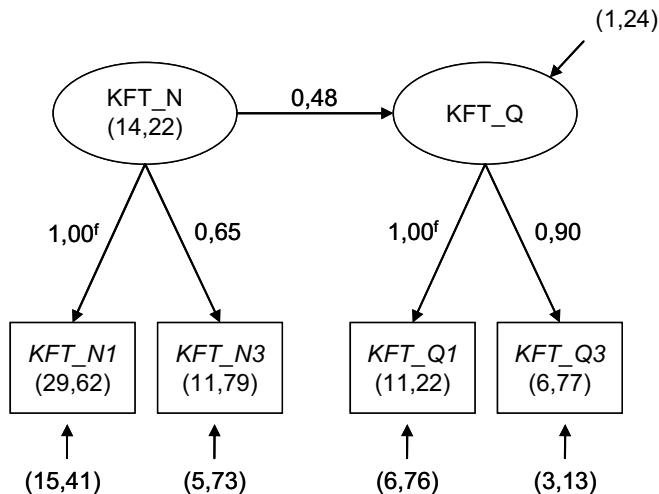


Abbildung 3.5: Latentes Regressionsmodell zur Vorhersage von quantitativen Fähigkeiten (Faktor KFT_Q) durch non-verbale Fähigkeiten (KFT_N). Jeder Faktor wird durch 2 Indikatoren (manifeste Variablen = KFT-Skalensummenwerte) gemessen. Die Werte repräsentieren die unstandardisierten Mplus-Parameterschätzungen (Ladungen, Pfadkoeffizienten und Varianzen bzw. Residualvarianzen). Die mit dem Superskript ^f bezeichneten Faktorladungen von KFT_N1 und KFT_Q1 wurden aus Identifikationsgründen auf 1 fixiert.

Die Mplus-Syntaxbefehle zur Spezifikation des Modells aus Abbildung 3.5 sind in Abbildung 3.6 zu finden. Zunächst werden wieder alle im Modell verwendeten manifesten Variablen unter `usevar` aufgeführt. Mit Hilfe des `by`-Statements im `model`-Befehl werden die Messmodelle spezifiziert. Hierbei wird angegeben, welche manifesten Variablen Ladungen auf welchen latenten Faktoren aufweisen. Das Wort `by` steht dabei für „gemessen durch“ (engl.: *measured by*). Der Faktor KFT_Q wird in unserem Beispiel durch die manifesten Variablen `kft_q1` und `kft_q3` gemessen (d.h. `kft_q1` und `kft_q3` „laden“ auf diesem Faktor). Der Faktor KFT_N wird durch die Variablen `kft_n1` und `kft_n3` gemessen.

Die Faktorladung des ersten Indikators hinter jedem `by`-Statement wird von Mplus standardmäßig auf 1 fixiert (hier also die Ladungen von `kft_q1` und `kft_n1`). Dies ist erforderlich, um die Metrik jedes latenten Faktors festzulegen (vgl. Box 3.6). Die Ladungen aller weiteren Indikatoren auf denselben Faktor werden standardmäßig als Modellparameter frei geschätzt.

The screenshot shows the Mplus software interface with the title bar "Mplus - [Modell_2_Bivariate_Regression_auf_latenter_Ebene.inp]". The menu bar includes File, Edit, View, Mplus, Graph, Window, and Help. Below the menu is a toolbar with various icons. The main window displays the Mplus input command language (INP) code for a bivariate regression model. The code specifies the title, data source, variables, analysis type, model specification (messmodelle and structural model), and output options. The status bar at the bottom indicates "Ready" and "Ln 24, Col 46".

```
title: Spezifikation eines bivariaten Regressionsmodells
      Vorhersage von kft_q durch kft_n auf latenter Ebene
      Spezifikation je eines latenten Faktors
      Faktor kft_q wird durch kft_q1 und kft_q3 gemessen
      Faktor kft_n wird durch kft_n1 und kft_n3 gemessen
      Faktor kft_q wird durch kft_n vorhergesagt

data: file = KFT.dat;
      listwise = on; ! FIML-Prozedur wird abgeschaltet
                  ! Stattdessen listenweiser Fallausschluss

variable: names = kft_v1 kft_v3 kft_q1 kft_q3 kft_n1 kft_n3;
           usevar = kft_q1 kft_q3 kft_n1 kft_n3;
           missing = all(-99);

analysis: type = general;

model: kft_q by kft_q1 kft_q3; ! Spezifikation der Messmodelle
       kft_n by kft_n1 kft_n3; ! "measured by"

       kft_q on kft_n; ! Spezifikation des Strukturmodells
                      ! "regressed on"

output: sampstat standardized stdyx residual;
```

Abbildung 3.6: Mplus-Spezifikation des latenten Regressionsmodells aus Abbildung 3.5.

Box 3.6: Alternative zur Voreinstellung „Fixierung der ersten Faktorladung auf 1“

Da die Metrik latenter Variablen zunächst beliebig ist, muss diese durch den Nutzer festgelegt werden (für Details siehe Bollen, 1989). Hierzu gibt es verschiedene Möglichkeiten. Der Standard in Mplus ist, dass die Faktorladung des ersten Indikators auf 1 fixiert wird und damit die Metrik des Faktors über die Metrik der ersten manifesten Variable definiert wird. Alternativ zur Fixierung einer Ladung pro Faktor kann man alle Faktorladungen frei schätzen lassen und zur Definition der Faktormetrik die Faktorvarianz bzw. -residualvarianz auf einen positiven Wert (in der Regel auf 1) fixieren. Fixierungen der Faktorvarianzen auf 1 (anstelle der Ladungen) sind in praktischen Anwendungen manchmal aus technischen Gründen hilfreich, z.B. dann, wenn Indikatoren eines Faktors sehr unterschiedliche Varianzen aufweisen und die Spezifikation mit einer fixierten Ladung aus diesem Grund nicht zu einer konvergierenden Lösung führt oder wenn man Parameter des Strukturmodells in der standardisierten Metrik (z.B. Faktorkorrelationen) auf bestimmte Werte fixieren möchte.

Andererseits können Fixierungen von Faktorvarianzen jedoch auch unerwünscht sein, etwa in Längsschnittmodellen, wenn Veränderungen in den Faktorvarianzen über die Zeit zu erwarten sind oder getestet werden sollen.

Als Beispiel betrachten wir den Faktor KFT_N. Um die Fixierung der Faktorladung des ersten Indikators (kft_n1) auf dem Faktor KFT_N aufzuheben, wird im by-Statement ein Sternchen (*) hinter dem Indikator eingefügt:

```
kft_n by kft_n1* kft_n3;
```

Damit ist diese Ladung nicht länger auf 1 fixiert, sondern wird nun vom Programm frei geschätzt. Die Faktorvarianz des Faktors KFT_N wird auf 1 fixiert, indem in einem zusätzlichen Befehl der Name des Faktors noch einmal aufgelistet und mit einem „@1“ versehen wird:

```
kft_n@1;
```

An der Modellgüte sowie den standardisierten Parameterschätzungen ändert sich im vorliegenden Beispiel nichts, wenn anstelle der Ladungen die Faktorvarianzen bzw. -residualvarianzen auf 1 fixiert werden (siehe Output auf der Begleit-CD).

Der Befehl kft_q on kft_n; ist analog zum on-Befehl, den wir für die manifeste Regressionsanalyse in Abschnitt 3.2 kennengelernt haben. Hier bedeutet er die Regression des Faktors KFT_Q auf den Faktor KFT_N, definiert also das Strukturmodell dieses einfachen Strukturgleichungsmodells. Residualvarianzen für die manifesten Variablen und den latenten Faktor KFT_Q werden von Mplus automatisch geschätzt. Alle übrigen Befehle wurden oben bereits eingeführt. Durch das Klicken auf **RUN** erhält man die Modellergebnisse (Maximum-Likelihood-Schätzungen), welche auszugsweise unten dargestellt sind.

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

Value	0.002
Degrees of Freedom	1
P-Value	0.9653

Chi-Square Test of Model Fit for the Baseline Model

Value	406.899
Degrees of Freedom	6
P-Value	0.0000

CFI/TLI

CFI	1.000
TLI	1.015

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.000
90 Percent C.I.	0.000 0.000
Probability RMSEA <= .05	0.980

SRMR (Standardized Root Mean Square Residual)

Value	0.000
-------	-------

Es zeigt sich, dass das Modell sehr gut auf die Daten passt. Der Chi-Quadrat-Wert für das Zielmodell ($\chi^2 = 0,002$, $df = 1$, $p = 0,9635$) weist darauf hin, dass das Modell nicht signifikant von den Daten abweicht, da $p > 0,05$. Der Chi-Quadrat-Wert für das Baselinemodell ($\chi^2 = 406,899$, $df = 6$) ist sehr groß und hochsignifikant ($p < 0,0001$), was zeigt, dass die Annahme der Unabhängigkeit (Nullkorrelation) der vier KFT-Variablen untereinander nicht haltbar ist.

Die übrigen Fitindices weisen ebenfalls auf einen guten Modellfit hin, da sie alle über bzw. unter den allgemein empfohlenen Richtwerten (siehe Box 3.7), liegen. Gemäß dieser Richtwerte sollten *CFI* und *TLI* für ein gutes Modell größer als 0,95 bzw. > 0,97 sowie *RMSEA* und *SRMR* beide kleiner als 0,05 ausfallen (Schermelleh-Engel, Moosbrugger, & Müller, 2003). Im vorliegenden Beispiel sind *CFI* und *TLI* > 0,97 sowie *RMSEA* und *SRMR* < 0,05. Auch der „Test of close fit“ fällt hier zugunsten des Modells aus. Mit diesem Test wird überprüft, ob der *RMSEA*-Koeffizient in der Population kleiner oder gleich 0,05 ist. Die Nullhypothese $RMSEA \leq 0,05$ muss hier nicht verworfen werden, da $p = 0,980$.

Box 3.7: Kriterien zur Beurteilung der Güte von Strukturgleichungsmodellen

In dieser Box wird ein Überblick über die wichtigsten in Mplus erhältlichen statistischen Tests bzw. Indices zur Beurteilung der Modellanpassung von Strukturgleichungsmodellen und gängige Cut-off-Richtwerte (siehe Hu & Bentler, 1999; Schermelleh-Engel et al., 2003) gegeben.

- **Chi-Quadrat-(χ^2) Test:** Testet die Nullhypothese, dass die Kovarianzmatrix in der Population gleich der vom Modell implizierten Kovarianzmatrix ist (Test auf exakten Modellfit). Ein signifikanter χ^2 -Wert führt zur Ablehnung der Nullhypothese, dass das Modell in der Population exakt passt. Die Freiheitsgrade (*degrees of freedom, df*) des Tests berechnen sich dabei aus der Differenz der Anzahl vorhandener Informationen (Varianzen, Kovarianzen und ggf. Mittelwerte der manifesten Variablen) minus der Anzahl geschätzter Modellparameter. In Mplus wird der χ^2 -Wert mit den entsprechenden *df* und dem *p*-Wert standardmäßig ausgegeben.
- **Chi-Quadrat-(χ^2) Differenzentest:** Mit Hilfe des χ^2 -Differenzentests kann man einen statistischen Modellvergleich zwischen zwei hierarchisch ineinander geschachtelten Modellen („nested models“) durchführen. Hierarchisch geschachtelt bedeutet, dass ein Modell durch zusätzliche Restriktionen als Spezialfall aus einem allgemeineren Modell hervorgeht. Dabei muss das allgemeinere Modell auf die Daten passen und durch die zusätzlichen Restriktionen darf kein Parameter auf die Grenze des zulässigen Parameterraumes (z.B. Varianz von 0 oder Korrelation von 1) festgelegt werden. Zur Durchführung des Tests wird die Differenz der χ^2 -Werte sowie die Differenz der *df* der beiden Modelle gebildet. Der χ^2 -Differenzwert wird anschließend bei den entsprechenden Differenz-*df* auf statistische Signifikanz geprüft. Ist der χ^2 -Differenzwert signifikant, so bedeutet dies, dass das stärker restriktive Modell signifikant schlechter auf die Daten passt als das allgemeinere Modell. χ^2 -Differenzentests werden z.B. häufig in Längsschnittanalysen bei der Messinvarianztestung (siehe Abschnitt 4.1.4) eingesetzt. Für konventionelle Maximum-Likelihood-Schätzung existiert in Mplus keine automatisierte Funktion zur Berechnung von χ^2 -Differenzentests. Man kann hierzu jedoch die von Crayen (2009) entwickelte Software *Mplus Model Comparison Gadget 1.0* benutzen, die in Kapitel 4.1.4 (Box 4.4) vor gestellt wird und auf der Begleit-CD enthalten ist.
- **Comparative-Fit-Index (CFI):** Zählt zur Gruppe der „Incremental Fit Indices“ und vergleicht den Fit des Zielmodells mit dem Fit eines Baseline-Modells. Das Baseline-Modell ist in Mplus das Unabhängigkeitsmodell. Das Unabhängigkeitsmodell nimmt an, dass die Kovarianzmatrix der beobachteten Variablen eine Diagonalmatrix ist, d.h. die beobachteten Variablen zwar Varianzen haben, jedoch keine (von Null verschiedenen) Kovarianzen (es werden keine Zusammenhänge zwischen den Variablen angenommen). Der CFI gibt an, um wie viel das Zielmodell besser auf die Daten passt als das (i.d.R. sparsamere) Unabhängigkeitsmodell. Für ein gutes Modell sollte der CFI-Wert über 0,95, besser jedoch über 0,97 liegen. In Mplus standardmäßig ausgegeben.

- **Tucker-Lewis-Index (TLI):** Zählt wie der CFI zur Gruppe der „Incremental Fit Indices“ und vergleicht den Fit des Zielmodells mit dem Fit eines Baseline-Modells. Es gelten dieselben Cut-off-Richtwerte wie für den CFI. In Mplus standardmäßig ausgegeben.
- **Root-Mean-Square-Error-of-Approximation (RMSEA):** Ein Maß für den approximativen Datenfit. Für ein gutes Modell sollte der RMSEA-Wert kleiner ausfallen als 0,05. In Mplus wird standardmäßig zusätzlich zum geschätzten RMSEA-Wert auch ein 90 %-Konfidenzintervall für diesen Index und ein *p*-Wert für den Test der Nullhypothese $\text{RMSEA} \leq 0,05$ ausgegeben.
- **Standardized-Root-Mean-Square-Residual (SRMR):** Der SRMR-Koeffizient ist ein standardisiertes Maß zur Gesamtbewertung der Residuen (Stichprobenkennwerte minus modellimplizierte Kennwerte). Werte < 0,05 werden als günstig angesehen. Sie deuten darauf hin, dass sich mit Hilfe des Modells die beobachteten Varianzen, Kovarianzen und ggf. Mittelwerte im Durchschnitt gut reproduzieren lassen. In Mplus standardmäßig ausgegeben.
- **Informationstheoretische Maße (AIC, BIC):** Diese Indices dienen zum Vergleich verschiedener Modelle auf deskriptiver Ebene. Dabei können, anders als beim χ^2 -Differenzentest, auch Modelle miteinander verglichen werden, die nicht in einer hierarchischen Beziehung zueinander stehen („non-nested models“). Als Entscheidungskriterium gilt, dass das Modell mit dem kleinsten Wert des betreffenden informationstheoretischen Maßes (z.B. AIC-Wert) ausgewählt wird. In Mplus standardmäßig ausgegeben.
- **Residuen:** Als Detailmaße des Modelfits können die Residuen betrachtet werden. Hierbei handelt es sich um die Differenz zwischen beobachteten (Stichproben-) Kennwerten (Varianzen, Kovarianzen und ggf. Mittelwerten) und den korrespondierenden modellimplizierten Kennwerten. Residuen geben an, welche der in den Daten beobachteten Varianzen, Kovarianzen und Mittelwerte durch das Modell gut versus nicht gut reproduziert werden können. Sie stellen daher ein wichtiges diagnostisches Instrument zur Detektion der Ursache eines schlechten globalen Modelfits dar. Nichtstandardisierte Residuen sind häufig schwer interpretierbar. Aus diesem Grund werden meistens die standardisierten Residuen (*z*-Werte) zur Modellgütebeurteilung herangezogen. Diese können absolut interpretiert werden. Ein großer absoluter *z*-Wert (> 1.96) spricht für eine signifikante Über- oder Unterschätzung eines Kennwertes durch das Modell. In Mplus erhält man ab Version 5 unstandardisierte und standardisierte Residuen über den Befehl `output: residual;`
- **Modifikationsindices:** Mit Hilfe von Modifikationsindices kann überprüft werden, welche Lockerung(en) von Restriktionen zu einer statistisch bedeutsamen Verbesserung des Modelfits führen. Modifikationsindices für eine einzige Modifikation (einen einzigen zusätzlich geschätzten Parameter) entsprechen in etwa einem χ^2 -Wert mit einem Freiheitsgrad. Es handelt sich somit ungefähr um die χ^2 -Differenz zwischen dem geschätzten Modell und dem weniger restriktiven Modell, in welchem die entsprechende Restriktion gelockert (ein zusätzlicher Parameter geschätzt) wurde. Große Modifikationsindices deuten auf einen durch eine bestimmte Restriktion verursachten Misfit hin. In Mplus erhält man Modifikationsindices über den Befehl

output: modindices; Mplus gibt zusätzlich zu den Modifikationsindices auch den „Expected Parameter Change“ aus, d.h. den Wert, der für den zusätzlich zu schätzenden Parameter im modifizierten Modell erwartet wird.

Ähnlich wie die Residuen liefern Modellmodifikationsindices häufig nützliche Anhaltspunkte zur Diagnose der Ursachen eines schlechten globalen Modelfits. Allerdings können sie z.T. auch verwirrende Ergebnisse liefern und lassen darüber hinaus in manchen Fällen keinen direkten Rückschluss auf die tatsächliche Ursache eines schlechten Modelfits zu. Modellmodifikationen sollten nie rein datengeleitet aufgrund von Modifikationsindices vorgenommen werden. Es sollte zudem nur eine (inhaltlich zu rechtfertigende) Modifikation zurzeit vorgenommen werden. Wichtig sind darüber hinaus die theoretische Begründung sowie der Nachweis der Replizierbarkeit von Modifikationen mit unabhängigen Daten.

Nachfolgend werden zunächst die unstandardisierten Mplus-Parameterschätzungen für das latente Regressionsmodell in Abbildung 3.5 gezeigt. Die unstandardisierten Parameterschätzungen erscheinen unter der Rubrik MODEL RESULTS. Die wichtigsten Schätzwerte sind auch in Abbildung 3.5 wiedergegeben.

MODEL RESULTS				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
KFT_Q BY				
KFT_Q1	1.000	0.000	999.000	999.000
KFT_Q3	0.903	0.094	9.588	0.000
KFT_N BY				
KFT_N1	1.000	0.000	999.000	999.000
KFT_N3	0.652	0.064	10.222	0.000
KFT_Q ON				
KFT_N	0.476	0.056	8.467	0.000
Intercepts				
KFT_Q1	12.387	0.157	78.890	0.000
KFT_Q3	7.732	0.122	63.398	0.000
KFT_N1	11.112	0.255	43.552	0.000
KFT_N3	8.281	0.161	51.457	0.000
Variances				
KFT_N	14.215	2.074	6.854	0.000
Residual Variances				
KFT_Q1	6.758	0.608	11.106	0.000
KFT_Q3	3.128	0.395	7.921	0.000
KFT_N1	15.405	1.591	9.683	0.000
KFT_N3	5.734	0.644	8.911	0.000
KFT_Q	1.239	0.425	2.916	0.004

Unter KFT_Q BY KFT_Q1 KFT_Q3 und KFT_N BY KFT_N1 KFT_N3 erscheinen die Schätzungen für die unstandardisierten Faktorladungen (Spalte Estimate). Die erste Faktorladung wurde – wie oben bereits erwähnt – jeweils von Mplus aus Identifikationsgründen auf 1 fixiert. Aus diesem Grund wird für diese Ladung kein Standardfehler (Spalte S.E.) berechnet. Der Wert „999.000“ in den Spalten Est./S.E. und Two-Tailed P-Value weist darauf hin, dass ein Signifikanztest für den betreffenden fixierten Parameter nicht erhältlich ist.

Die Ladungen der zweiten Indikatoren (KFT_Q3 und KFT_N3) wurden auf 0,903 bzw. 0,652 geschätzt und sind statistisch signifikant von Null verschieden (zweiseitige p -Werte < 0,000, siehe jeweils die letzte Spalte Two-Tailed P-Value). Der unstandardisierte Regressionskoeffizient für die Regression des latenten Faktors KFT_Q auf den Faktor KFT_N (Zeile KFT_Q ON KFT_N) beträgt 0,476 und ist ebenfalls statistisch bedeutsam von Null verschieden ($z = 8,467, p < 0,001$). Das zeigt, dass der Faktor KFT_N ein signifikanter Prädiktor für den Faktor KFT_Q ist.

Weiterhin gibt Mplus in der unstandardisierten Lösung die Intercepts für die beobachteten Variablen sowie die Varianzen bzw. Residualvarianzen der manifesten und latenten Variablen aus. Für alle exogenen Variablen (alle Variablen, die im Modell nicht durch andere Variablen erklärt vorhergesagt werden) schätzt Mplus deren Varianz als Modellparameter. Für alle endogenen Variablen (alle Variablen, die im Modell durch eine oder mehrere andere Variablen erklärt/vorhergesagt werden) ist die zugehörige *Residualvarianz* der im Modell geschätzte Parameter.

Die standardisierte Lösung [siehe Output-Excerpt unter STANDARDIZED MODEL RESULTS (STDRYX Standardization) weiter unten] enthält die Parameter in standardisierter Form. Hier wurde die vollstandardisierte Lösung stdyx gewählt, bei welcher sämtliche Variablen standardisiert werden. Die standardisierten Faktorladungen können in diesem Modell als Korrelationen zwischen den beobachteten Variablen und dem Faktor, auf dem sie laden, interpretiert werden (das ist deswegen hier möglich, da jede Variable nur auf einem Faktor lädt). Beispielsweise ist 0,631 die geschätzte Korrelation zwischen kft_q1 und dem Faktor KFT_Q.

Unter KFT_Q ON KFT_N wird hier das standardisierte Regressionsgewicht für die Regression von KFT_Q auf KFT_N berichtet. Es beträgt 0,85 und kann als Korrelation zwischen den beiden latenten Faktoren KFT_Q und KFT_N interpretiert werden, da es sich um eine bivariate latente Regression handelt (für die multiple Regression mit mehreren Prädiktoren gilt dies nicht). Somit beträgt die geschätzte latente Korrelation zwischen KFT_Q und KFT_N $r = 0,85$, was als sehr hoher Zusammenhang angesehen werden kann.

STANDARDIZED MODEL RESULTS (STDRYX Standardization)				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
KFT_Q BY				
KFT_Q1	0.631	0.041	15.548	0.000
KFT_Q3	0.733	0.040	18.247	0.000
KFT_N BY				
KFT_N1	0.693	0.038	18.020	0.000
KFT_N3	0.717	0.038	18.717	0.000
KFT_Q ON				
KFT_N	0.850	0.049	17.244	0.000
Intercepts				
KFT_Q1	3.698	0.131	28.177	0.000
KFT_Q3	2.972	0.109	27.240	0.000
KFT_N1	2.042	0.082	24.799	0.000
KFT_N3	2.412	0.093	26.024	0.000
Variances				
KFT_N	1.000	0.000	999.000	999.000
Residual Variances				
KFT_Q1	0.602	0.051	11.780	0.000
KFT_Q3	0.462	0.059	7.840	0.000
KFT_N1	0.520	0.053	9.764	0.000
KFT_N3	0.487	0.055	8.869	0.000
KFT_Q	0.278	0.084	3.316	0.001
R-SQUARE				
Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
KFT_Q1	0.398	0.051	7.774	0.000
KFT_Q3	0.538	0.059	9.124	0.000
KFT_N1	0.480	0.053	9.010	0.000
KFT_N3	0.513	0.055	9.358	0.000
Latent Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
KFT_Q	0.722	0.084	8.622	0.000

Die standardisierten Fehlervarianzen (Residual Variances) entsprechen $1 - R^2$. Die R^2 -Werte für die manifesten Variablen (zu finden unter R-SQUARE - Observed Variable) geben den durch den jeweiligen Faktor erklärten Varianzanteil einer manifesten Variable an

(hier die quadrierte standardisierte Faktorladung). Diese Werte können als untere Grenze der Reliabilität einer manifesten Variable interpretiert werden. Es handelt sich um die untere Grenze der Reliabilität, da indikatorsspezifische Varianz in diesem speziellen Modell nicht von Messfehlereinflüssen trennbar ist. Die R^2 -Werte für die manifesten Variablen weisen auf geringe Reliabilitäten der beobachteten Variablen hin. Dies mag jedoch z.T. auch durch hohe spezifische Varianzanteile erklärbar sein.

Unter R-SQUARE – Latent Variable können wir erkennen, dass mit 72,2% ein beträchtlicher Anteil der Varianz des KFT_Q-Faktors durch den KFT_N-Faktor erklärt werden kann. Im Vergleich zur manifesten Regressionsanalyse (vgl. Abschnitt 3.2), in welcher nur kft_n1 und kft_q1 berücksichtigt wurden sowie keine Korrektur für Messfehler stattfand, ist die Varianzaufklärung im latenten Modell wesentlich höher. Dies ist zum einen dadurch zu erklären, dass im manifesten Modell Messfehlereinflüsse den Zusammenhang zwischen den Untertests kft_n1 und kft_q1 „verwässert“ haben. Zum anderen wurde im latenten Modell über die Verwendung multipler Indikatoren auch indikatorsspezifische Varianz als Teil des Messfehlers „ausgeschieden“, was als weitere Erklärung für den stark angestiegenen Zusammenhang angesehen werden kann.

3.4 Konfirmatorische Faktorenanalyse (CFA)

In diesem Abschnitt wird auf die Durchführung konfirmatorischer Faktorenanalysen (*confirmatory factor analysis*, CFA) in Mplus eingegangen. Zur Erinnerung: Von einer CFA spricht man, wenn das Strukturmodell ausschließlich ungerichtete Zusammenhänge (korrelierte Faktoren) enthält oder wenn zwischen den Faktoren überhaupt keine Zusammenhänge angenommen werden (CFA mit orthogonalen Faktoren).

Box 3.8: Fallbeispiel CFA

Als Beispiel betrachten wir die sechs KFT-Skalen im Datensatz **KFT.dat**. Folgende inhaltliche Fragestellung soll beantwortet werden: Messen die sechs KFT-Skalen kft_v1, kft_v3, kft_q1, kft_q3, kft_n1 und kft_n3 *drei verschiedene latente Dimensionen*, (verbale, quantitative und non-verbale/figurale Fähigkeiten) oder sind sie lediglich Indikatoren *einer einzigen* Merkmalsdimension, also eines „*g*-Faktors“ der Intelligenz? Dies wollen wir zunächst anhand der Analyse zweier CFA-Modelle erster Ordnung in Mplus überprüfen. Weiter unten wird auch noch die Spezifikation eines CFA-Modells mit einem Faktor zweiter Ordnung gezeigt.

3.4.1 Konfirmatorische Faktorenanalyse 1. Ordnung

Das erste Modell (siehe Abb. 3.7) ist ein 1-Faktormodell (so genanntes „Generalfaktor-“ oder kurz „*g*-Faktormodell“), in welchem alle sechs KFT-Skalen nur auf einem einzigen Faktor laden. Dieses Modell nimmt an, dass alle sechs Skalen nur eine einzige latente Merkmalsdimension messen und darüber hinaus nur zufällige Messfehler und indikatorsspezifische

Varianz. Würde das g-Faktormodell auf die Daten passen, so müsste man davon ausgehen, dass es nicht notwendig ist, zwischen verbalen, numerischen und figuralen Fähigkeiten zu unterscheiden bzw. dass eine solche Unterscheidung zumindest mit den vorliegenden Skalen nicht möglich ist (diese keine „diskriminante Validität“ aufweisen).

Das zweite Modell (siehe Abb. 3.8) beinhaltet drei latente Faktoren und ist im Einklang mit der Idee, dass die sechs KFT-Skalen drei voneinander abgrenzbare Subdimensionen messen (kft_v1 und kft_v3 als Indikatoren für einen Faktor, der verbale Fähigkeiten repräsentiert, kft_n1 und kft_n3 als Indikatoren für einen quantitativ-numerischen Faktor sowie kft_n1 und kft_n3 als Indikatoren für einen non-verbalen/figuralen Faktor).

$$\chi^2 = 33,939, df = 9, p = 0,0001, AIC = 14343$$

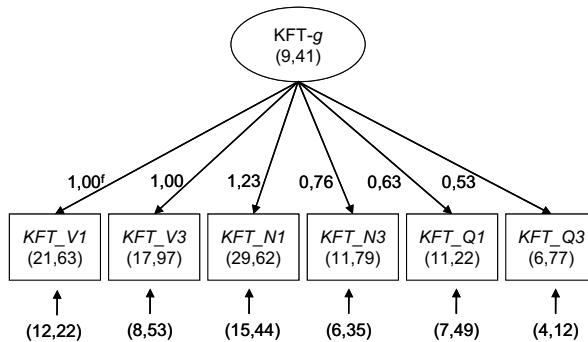


Abbildung 3.7: 1-Faktormodell (g-Faktormodell) für die sechs KFT-Skalen (Datei **KFT.dat**) mit Modellfitindices und unstandardisierten Parameterschätzwerten. Die mit dem Superskript f bezeichnete Ladung wurde von Mplus aus Identifikationsgründen automatisch auf 1 fixiert.

$$\chi^2 = 7,093, df = 6, p = 0,3123, AIC = 14323$$

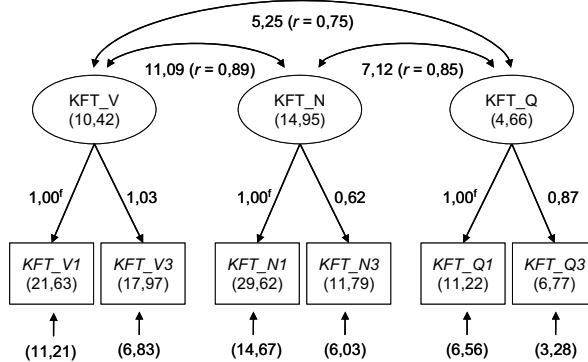
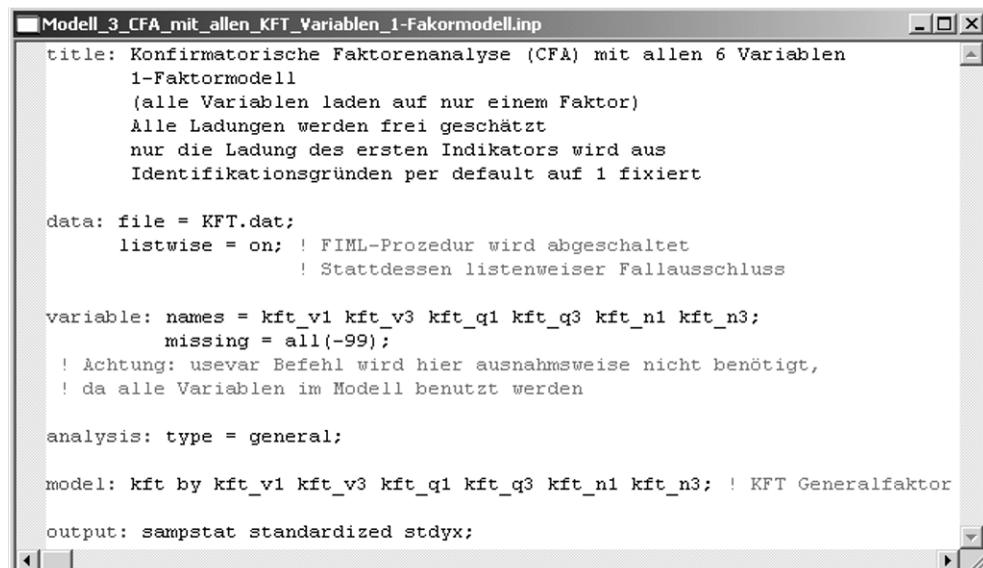


Abbildung 3.8: 3-Faktormodell für die sechs KFT-Skalen (Datei **KFT.dat**) mit Modellfitindices und unstandardisierten Parameterschätzwerten sowie latenten Faktorkorrelationen (r). Die mit dem Superskript f bezeichneten Ladungen wurden von Mplus aus Identifikationsgründen automatisch auf 1 fixiert.

Statistisch kann das 1-Faktormodell auch als ein (restringiertes) 3-Faktormodell aufgefasst werden, in welchem perfekte Zusammenhänge ($r = 1$) zwischen allen drei Faktoren vorliegen. Ein Vergleich der Modelle erlaubt es uns herauszufinden, ob eine „1-Faktortheorie“ oder eine „3-Faktortheorie“ für die KFT-Skalen angemessener ist.

Der Mplus-Input für das 1-Faktormodell wird in Abbildung 3.9 gezeigt, während der Input für das 3-Faktormodell in Abbildung 3.10 wiedergegeben ist. Man beachte, dass der usevar-Befehl hier ausnahmsweise nicht benötigt wird, da alle Variablen im Datensatz (alle Variablen, die unter names = aufgelistet sind) im model-Befehl benutzt werden. In allen anderen Fällen wird der usevar-Befehl benötigt (vgl. Abschnitt 3.2).

Weiterhin ist von Interesse, dass die zu schätzenden Kovarianzen zwischen den drei Faktoren im 3-Faktormodell im Mplus-Input nicht explizit spezifiziert werden müssen. Mplus schätzt per Voreinstellung alle Kovarianzen zwischen exogenen Faktoren automatisch (da die drei Faktoren im Modell nicht durch andere Variablen vorhergesagt werden, handelt es sich um exogene Variablen). Unerwünschte Korrelationen können mit dem with-Befehl in Verbindung mit @ auf Null fixiert werden (z.B. F1 with F2@0, um die Kovarianz/Korrelation zwischen zwei Faktoren F1 und F2 auf Null zu setzen). Mit Hilfe des tech4-Befehls in der Zeile output: kann zusätzlich die Kovarianz- und die Korrelationsmatrix der latenten Faktoren angefordert werden.



```

Modell_3_CFA_mit_allen_KFT_Variablen_1-Fakormodell.inp

title: Konfirmatorische Faktorenanalyse (CFA) mit allen 6 Variablen
       1-Faktormodell
       (alle Variablen laden auf nur einem Faktor)
       Alle Ladungen werden frei geschätzt
       nur die Ladung des ersten Indikators wird aus
       Identifikationsgründen per default auf 1 fixiert

data: file = KFT.dat;
       listwise = on; ! FIML-Prozedur wird abgeschaltet
                   ! Stattdessen listenweiser Fallausschluss

variable: names = kft_v1 kft_v3 kft_q1 kft_q3 kft_n1 kft_n3;
           missing = all(-99);
       ! Achtung: usevar Befehl wird hier ausnahmsweise nicht benötigt,
       ! da alle Variablen im Modell benutzt werden

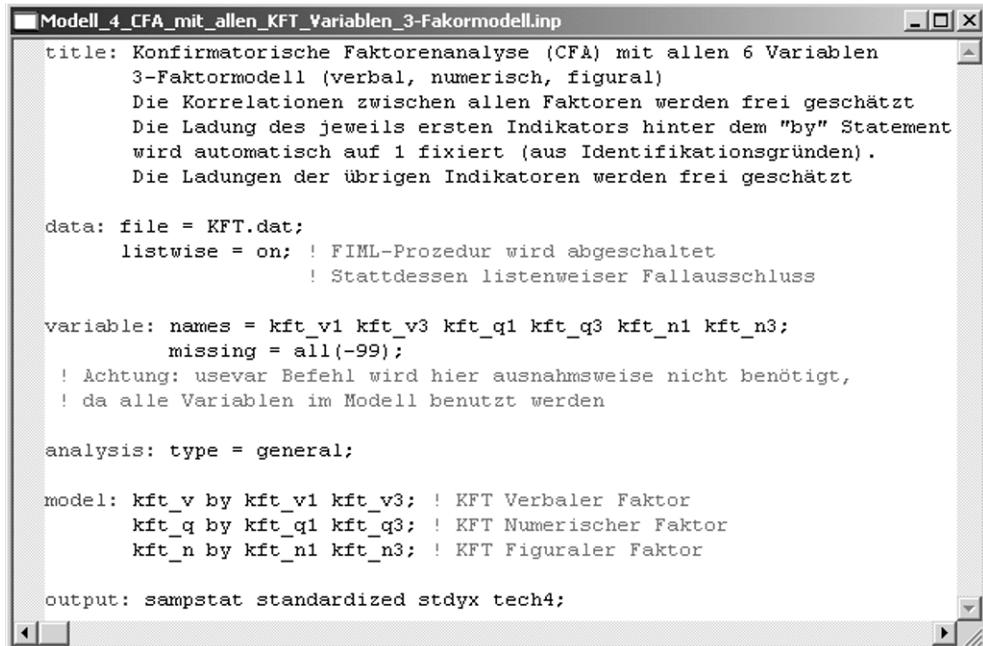
analysis: type = general;

model: kft by kft_v1 kft_v3 kft_q1 kft_q3 kft_n1 kft_n3; ! KFT Generalfaktor

output: sampstat standardized stdyx;

```

Abbildung 3.9: Mplus-Input-Datei zur Spezifikation des KFT-g-Faktormodells aus Abbildung 3.7 anhand der Daten im Datensatz **KFT.dat**.



```

Modell_4_CFA_mit_allen_KFT_Variablen_3-Fakormodell.inp

title: Konfirmatorische Faktorenanalyse (CFA) mit allen 6 Variablen
       3-Faktormodell (verbal, numerisch, figural)
       Die Korrelationen zwischen allen Faktoren werden frei geschätzt
       Die Ladung des jeweils ersten Indikators hinter dem "by" Statement
       wird automatisch auf 1 fixiert (aus Identifikationsgründen).
       Die Ladungen der übrigen Indikatoren werden frei geschätzt

data: file = KFT.dat;
      listwise = on; ! FIML-Prozedur wird abgeschaltet
                  ! Stattdessen listenweiser Fallausschluss

variable: names = kft_v1 kft_v3 kft_q1 kft_q3 kft_n1 kft_n3;
           missing = all(-99);
           ! Achtung: usevar Befehl wird hier ausnahmsweise nicht benötigt,
           ! da alle Variablen im Modell benutzt werden

analysis: type = general;

model: kft_v by kft_v1 kft_v3; ! KFT Verbaler Faktor
       kft_q by kft_q1 kft_q3; ! KFT Numerischer Faktor
       kft_n by kft_n1 kft_n3; ! KFT Figuraler Faktor

output: sampstat standardized stdyx tech4;

```

Abbildung 3.10: Mplus-Input-Datei zur Spezifikation des KFT-3-Faktormodells aus Abbildung 3.8 anhand der Daten im Datensatz **KFT.dat**.

Die von Mplus berechneten Fitstatistiken für beide Modelle sind in Tabelle 3.1 dargestellt. Es zeigt sich, dass das 1-Faktormodell im Gegensatz zum 3-Faktormodell einen signifikanten χ^2 -Wert aufweist. Nach diesem Kriterium müsste das 1-Faktormodell verworfen werden, während das 3-Faktormodell nicht signifikant von den Daten abweicht. Auch liegt der RMSEA-Wert für das 1-Faktormodell deutlich über dem Richtwert von 0,05. Gemäß der CFI-, TLI- und SRMR-Werte wäre sowohl das 1-Faktormodell als auch das 3-Faktormodell akzeptabel. Der AIC-Wert für das 1-Faktormodell fällt etwas größer aus als der entsprechende Wert für das 3-Faktormodell, so dass nach diesem Kriterium das 3-Faktormodell zu bevorzugen ist. Insgesamt scheint aufgrund der Fitstatistiken die 3-Faktorlösung angemessener zu sein.

Die Mplus-Parameterschätzungen für das 3-Faktormodell sind nachfolgend dargestellt. Zunächst wird wiederum die unstandardisierte Lösung (Tabelle MODEL RESULTS) ausgegeben. Es zeigt sich, dass alle Ladungen (zu finden unter BY) mit Ausnahme der jeweils ersten, welche zur Identifikation der jeweiligen Faktoremetrik automatisch auf 1 fixiert wurde, signifikant von Null verschieden sind. Alle frei geschätzten Faktorkovarianzen (zu finden unter WITH) sind ebenfalls bedeutsam von Null verschieden. Die Höhe der Zusammenhänge der Faktoren ist aufgrund der Kovarianzen nicht leicht abzuschätzen, es sei denn, man setzt diese in Beziehung zu den jeweiligen Faktorvarianzen. Hier hilft uns die standardisierte Lösung, welcher wir die Korrelationen (standardisierte Kovarianzen) der Faktoren entnehmen können (s.u.).

Tabelle 3.1: Modellfitstatistiken zum Vergleich des 1-Faktormodells mit dem 3-Faktormodell für die KFT-Daten

	1-Faktormodell	3-Faktormodell
$\chi^2 (df)$	33,939 (9)	7,093 (6)
$p(\chi^2)$	0,0001	0,3123
CFI / TLI	0,969 / 0,948	0,999 / 0,997
RMSEA (90 % Konfidenzintervall)	0,078 (0,051, 0,107)	0,020 (0,000, 0,066)
$p(RMSEA \leq 0,05)$	0,043	0,820
SRMR	0,031	0,014
AIC	14343,362	14322,516

Anmerkung. Die Bedeutung der Fitstatistiken sowie gängige Richtwerte für akzeptablen Fit werden in Box 3.7 besprochen.

Der unstandardisierten Lösung können wir außerdem noch die geschätzten Intercepts der manifesten Variablen (diese entsprechen in diesem Fall den beobachteten Mittelwerten), die geschätzten Faktorvarianzen sowie die Residualvarianzen der manifesten Variablen entnehmen.

MODEL RESULTS				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
KFT_V BY				
KFT_V1	1.000	0.000	999.000	999.000
KFT_V3	1.034	0.083	12.435	0.000
KFT_Q BY				
KFT_Q1	1.000	0.000	999.000	999.000
KFT_Q3	0.865	0.084	10.338	0.000
KFT_N BY				
KFT_N1	1.000	0.000	999.000	999.000
KFT_N3	0.620	0.050	12.394	0.000
KFT_Q WITH				
KFT_V	5.247	0.696	7.534	0.000
KFT_N WITH				
KFT_V	11.088	1.222	9.074	0.000
KFT_Q	7.122	0.839	8.488	0.000
Intercepts				
KFT_V1	11.899	0.218	54.574	0.000
KFT_V3	8.974	0.199	45.156	0.000
KFT_Q1	12.387	0.157	78.889	0.000
KFT_Q3	7.732	0.122	63.400	0.000
KFT_N1	11.112	0.255	43.551	0.000
KFT_N3	8.281	0.161	51.457	0.000

Variances				
KFT_V	10.424	1.401	7.443	0.000
KFT_Q	4.659	0.720	6.472	0.000
KFT_N	14.953	1.969	7.596	0.000
Residual Variances				
KFT_V1	11.206	1.004	11.156	0.000
KFT_V3	6.830	0.853	8.011	0.000
KFT_Q1	6.559	0.586	11.193	0.000
KFT_Q3	3.283	0.366	8.980	0.000
KFT_N1	14.668	1.382	10.614	0.000
KFT_N3	6.033	0.550	10.967	0.000

Unter STANDARDIZED MODEL RESULTS (STDRYX Standardization) wird die voll-standardisierte Lösung ausgegeben. Die standardisierten Faktorladungen erscheinen wiederum unter BY, die latenten Faktorkorrelationen unter WITH. Es zeigt sich, dass alle drei Faktoren sehr hoch miteinander korreliert sind ($0,75 < r < 0,89$), was auf eine geringe diskriminante Validität der KFT-Untertests hindeutet.

Obwohl die drei Faktoren hoch miteinander korreliert sind, zeigte sich anhand der Fitstatistiken, dass eine g -Faktorlösung (Annahme perfekt korrelierter Faktoren) die Daten nicht zufriedenstellend abbilden kann. Somit kann man annehmen, dass den drei Unterdimensionen zwar ein starker gemeinsamer Faktor zu Grunde liegt, darüber hinaus jedoch noch systematische skalenspezifische (verbale, numerische und figurale) Varianz vorliegt.

STANDARDIZED MODEL RESULTS (STDRYX Standardization)				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
KFT_V BY				
KFT_V1	0.694	0.033	20.964	0.000
KFT_V3	0.787	0.031	25.288	0.000
KFT_Q BY				
KFT_Q1	0.644	0.039	16.701	0.000
KFT_Q3	0.718	0.038	18.957	0.000
KFT_N BY				
KFT_N1	0.711	0.033	21.590	0.000
KFT_N3	0.699	0.033	21.046	0.000
KFT_Q WITH				
KFT_V	0.753	0.051	14.764	0.000
KFT_N WITH				
KFT_V	0.888	0.041	21.589	0.000
KFT_Q	0.853	0.049	17.458	0.000

Intercepts				
KFT_V1	2.558	0.097	26.401	0.000
KFT_V3	2.117	0.084	25.084	0.000
KFT_Q1	3.698	0.131	28.178	0.000
KFT_Q3	2.972	0.109	27.241	0.000
KFT_N1	2.042	0.082	24.798	0.000
KFT_N3	2.412	0.093	26.024	0.000
Variances				
KFT_V	1.000	0.000	999.000	999.000
KFT_Q	1.000	0.000	999.000	999.000
KFT_N	1.000	0.000	999.000	999.000
Residual Variances				
KFT_V1	0.518	0.046	11.268	0.000
KFT_V3	0.380	0.049	7.753	0.000
KFT_Q1	0.585	0.050	11.756	0.000
KFT_Q3	0.485	0.054	8.932	0.000
KFT_N1	0.495	0.047	10.589	0.000
KFT_N3	0.512	0.046	11.037	0.000
R-SQUARE				
Observed Variable				
		Estimate	S.E.	Two-Tailed P-Value
KFT_V1	0.482	0.046	10.482	0.000
KFT_V3	0.620	0.049	12.644	0.000
KFT_Q1	0.415	0.050	8.350	0.000
KFT_Q3	0.515	0.054	9.479	0.000
KFT_N1	0.505	0.047	10.795	0.000
KFT_N3	0.488	0.046	10.523	0.000

3.4.2 Konfirmatorische Faktorenanalyse 2. Ordnung

Die hohen (jedoch nicht perfekten) Korrelationen der 3 KFT-Faktoren gaben uns Anlass zu der Vermutung, dass allen sechs Skalen ein gemeinsamer Fähigkeitsfaktor (g -Faktor der Intelligenz?) zugrunde liegt und darüber hinaus noch ein (relativ geringer) Anteil an spezifischer verbaler, numerischer und figuraler Fähigkeitsvarianz vorliegt. Diese Annahme kann man mit Hilfe der Spezifikation eines Faktormodells 2. Ordnung ausdrücken (siehe Abb. 3.11), in welchem die KFT-Faktoren 1. Ordnung (KFT_V, KFT_Q und KFT_N) auf einem Faktor 2. Ordnung (KFT_g) laden.

$$\chi^2 = 7,093, df = 6, p = 0,3123, AIC = 14323$$

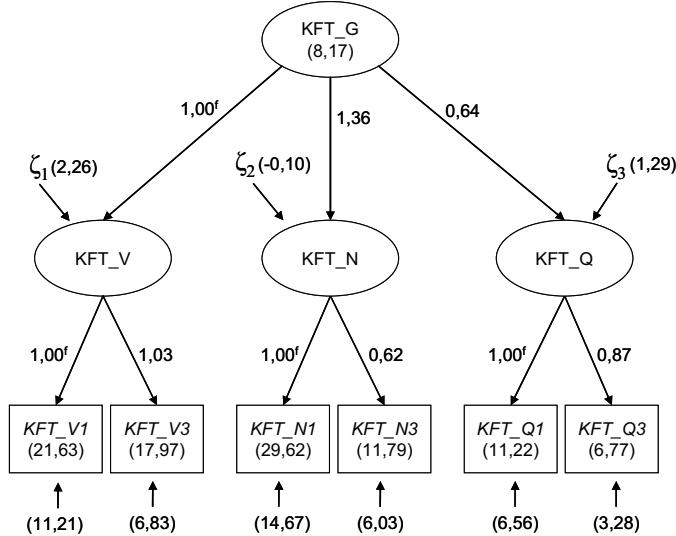


Abbildung 3.11: Faktormodell 2. Ordnung für die sechs KFT-Skalen. Die KFT-Faktoren 1. Ordnung (KFT_V, KFT_Q und KFT_N) laden auf einem Faktor 2. Ordnung (g-Faktor KFT_G). Die angegebenen Werte sind die unstandardisierten Mplus-Parameterschätzungen. Die mit dem Superskript ^f bezeichneten Ladungen wurden von Mplus aus Identifikationsgründen automatisch auf 1 fixiert. Die in Klammern angegebenen Werte sind die geschätzten Varianzen bzw. Residualvarianzen. Man beachte die negativ geschätzte Residualvarianz (sogenannter „Heywood Case“) für den Faktor KFT_N (-0,10).

Die entsprechende Mplus-Spezifikation für das Modell in Abbildung 3.11 wird in Abbildung 3.12 gezeigt. Es sollte beachtet werden, dass die Annahme eines zugrundeliegenden g-Faktors 2. Ordnung in diesem Fall nicht empirisch testbar ist, da ein Modell mit drei korrelierten Faktoren 1. Ordnung mathematisch identisch mit einem Modell 2. Ordnung ist und daher auch denselben Modellfit aufweist wie dieses Modell. Man spricht in einem solchen Fall von *statistischer Äquivalenz* der Modelle. Erst bei vier oder mehr Faktoren 1. Ordnung (oder bei Einführung zusätzlicher Restriktionen wie z.B. der Restriktion gleicher Ladungen der Faktoren 1. Ordnung auf dem Faktor 2. Ordnung) ist das Strukturmodell eines Faktormodells 2. Ordnung restriktiver als das (saturierte) Strukturmodell einer CFA 1. Ordnung mit korrelierten Faktoren und wäre dann auch testbar.

```

title: Konfirmatorische Faktorenanalyse (CFA) mit allen 6 Variablen
      3-Faktormodell (verbal, numerisch, figural)
      Die 3 Faktoren 1. Ordnung laden
      auf einen Faktor 2. Ordnung (g-Faktor)
      Die Ladung des jeweils ersten Indikators hinter dem "by" Statement
      wird automatisch auf 1 fixiert (aus Identifikationsgründen).
      Die Ladungen der übrigen Indikatoren werden frei geschätzt.

data: file = KFT.dat;
      listwise = on; ! FIML-Prozedur wird abgeschaltet
                  ! Stattdessen listenweiser Fallausschluss

variable: names = kft_v1 kft_v3 kft_q1 kft_q3 kft_n1 kft_n3;
           missing = all(-99);

analysis: type = general;

model: kft_v by kft_v1 kft_v3; ! KFT Verbaler Faktor
       kft_q by kft_q1 kft_q3; ! KFT Numerischer Faktor
       kft_n by kft_n1 kft_n3; ! KFT Figuraler Faktor

       kft_g by kft_v kft_q kft_n; ! KFT-Faktor 2. Ordnung (g-Faktor)

output: sampstat standardized stdyx tech4 residual;

```

Abbildung 3.12: Mplus-Input-Datei für ein CFA-Modell 2. Ordnung. Die KFT-Faktoren 1. Ordnung (KFT_V, KFT_Q und KFT_N) laden auf einem Faktor höherer Ordnung (KFT_G).

Allerdings weist die Anwendung in diesem Fall dennoch auf eine Missspezifikation hin, da bei der Schätzung der Modellparameter eine unzulässige Parameterschätzung auftritt (*Heywood Case*; Chen, Bollen, Paxton, Curran & Kirby, 2001; siehe auch Box 3.9). In diesem Fall wird die Residualvarianz des Faktors KFT_N auf einen negativen Wert geschätzt (-0,099), was nicht akzeptabel ist, da Varianzen per Definition nicht negativ sein können. Mplus weist mit folgender Warnmeldung auf dieses Problem hin:

```

WARNING: THE LATENT VARIABLE COVARIANCE MATRIX (PSI) IS NOT POSITIVE
DEFINITE. THIS COULD INDICATE A NEGATIVE VARIANCE/RESIDUAL VARIANCE
FOR A LATENT VARIABLE, A CORRELATION GREATER OR EQUAL TO ONE BETWEEN
TWO LATENT VARIABLES, OR A LINEAR DEPENDENCY AMONG MORE THAN TWO
LATENT VARIABLES. CHECK THE TECH4 OUTPUT FOR MORE INFORMATION. PROB-
LEM INVOLVING VARIABLE KFT_N.

```

Die geschätzte negative Residualvarianz könnte ein Hinweis darauf sein, dass die Faktoren erster Ordnung ein inhomogenes Korrelationsmuster aufweisen, welches der Idee eines gemeinsamen Faktors 2. Ordnung widerspricht. In diesem Fall empfiehlt es sich, das unproblematischere Modell mit 3 korrelierten Faktoren 1. Ordnung dem Faktormodell 2. Ordnung vorzuziehen und die Parameter des Modells 2. Ordnung allenfalls mit größter Vorsicht zu interpretieren.

Box 3.9: Unzulässige Schätzergebnisse („Heywood Cases“) in Strukturgleichungsmodellen

Als *Heywood Case* bezeichnet man eine Lösung, welche unzulässige Parameterschätzungen (sog. *out-of-range parameter estimates*) wie beispielsweise negative Varianzen oder Korrelationen > 1 aufweist. Derartige unzulässige Schätzungen können vielfältige Ursachen haben, welche auch in Kombination miteinander auftreten können. Zu den typischen Ursachen zählen:

- Modellmisspezifikation, z.B. durch inhomogene Indikatoren im Messmodell oder die Spezifikation zu vieler oder zu weniger Faktoren
- Verwendung einer zu kleinen Stichprobe
- zu wenige Indikatoren pro Faktor
- Stichprobenfehler
- Ausreißer- oder Extremwerte

Gemäß der Erfahrung des Autors sind Modellmisspezifikationen die häufigste Ursache für das Auftreten unzulässiger Schätzungen. Aus diesem Grund sollten unzulässige Schätzungen nicht einfach durch „kosmetische Maßnahmen“ (z.B. Fixierung der entsprechenden negativen Varianzen auf 0) entfernt, sondern die möglichen Ursachen gründlich untersucht werden. Nähere Hinweise zum Umgang mit Heywood Cases findet man bei Chen et al. (2001).

Im Folgenden sind die Mplus-Schätzungen für die Residualvarianzen des CFA-Modells 2. Ordnung aus Abbildung 3.11 abgedruckt. Man beachte die auf einen negativen Wert geschätzte Residualvarianz für den Faktor KFT_N:

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Residual Variances					
KFT_V1	11.206	1.004	11.156	0.000	
KFT_V3	6.829	0.853	8.010	0.000	
KFT_Q1	6.559	0.586	11.192	0.000	
KFT_Q3	3.283	0.366	8.979	0.000	
KFT_N1	14.667	1.382	10.614	0.000	
KFT_N3	6.033	0.550	10.967	0.000	
KFT_V	2.256	0.754	2.992	0.003	
KFT_Q	1.289	0.402	3.209	0.001	
KFT_N	-0.099	1.206	-0.082	0.935	



Negative Residualvarianz
(*„Heywood Case“*)!

3.5 Pfadmodelle und Mediatoranalyse

3.5.1 Einführung und manifeste Pfadanalyse

Pfadanalysen zählen neben konfirmatorischen Faktorenanalysen zu den häufigsten Modellen, die im Rahmen linearer Strukturgleichungsmodelle analysiert werden. Unter einer Pfadanalyse versteht man im Allgemeinen ein *multivariates* Regressionsmodell, also eine Regressionsanalyse, in welcher simultan *mehrere abhängige* Variablen betrachtet werden. Im Gegensatz dazu beschränken sich konventionelle Regressionsanalysen (vgl. Abschnitt 3.2) auf die Vorhersage einer einzelnen abhängigen Variable.

Exogene Variablen können in einem Pfadmodell untereinander korreliert sein. Pfadanalysen können entweder auf der Ebene manifester Variablen (wie in Abb. 3.13 dargestellt) oder auf latenter Ebene (latente Pfadanalysen, siehe Abschnitt 3.5.3) durchgeführt werden. Wir besprechen hier zunächst den einfacheren Fall (die Pfadanalyse auf der Ebene manifester Variablen).

Box 3.10: Fallbeispiel Pfadanalyse

Als Beispiel dient ein einfaches pfadanalytisches Modell mit drei Konstrukten aus der Gesundheitspsychologie. In diesem von Whitelaw und Liang (1991) vorgeschlagenen Modell geht es um den Zusammenhang der drei Konstrukte *physische Gesundheit*, *funktionelle Gesundheit* und *subjektive Gesundheit*. Unter physischer Gesundheit wird hierbei der faktische, physiologische Gesundheitsstatus (z.B. Vorhandensein versus Abwesenheit von Beschwerden bzw. Krankheiten) verstanden. Funktionelle Gesundheit bezieht sich auf die Fähigkeit zur Ausführung alltäglicher Aktivitäten und Handlungen (z.B. Laufen, Treppen steigen). Subjektive Gesundheit bedeutet die individuelle, global gefasste subjektive Einschätzung der eigenen Gesundheit (zu einer genaueren Beschreibung der drei Konstrukte siehe Pinquart, 2001).

Die Beispieldaten stammen von einer zufällig ausgewählten Teilstichprobe ($N = 1648$) aus dem Alterssurvey des Deutschen Zentrums für Altersfragen (DZA)¹ und befinden sich in der Datei **gesundheit.dat**. Der Individualdatensatz enthält sieben Variablen: Die Anzahl an Krankheiten (Variable **KRANK**) sowie die Anzahl der Arztbesuche in den letzten 12 Monaten (Variable **ARZT**) als Indikatoren für die physische Gesundheit, die Subskala zur funktionellen Gesundheit aus dem SF-36-Fragebogen (Ware & Sherborne, 1992) – einmal als Gesamtsummenwert **FG** und einmal aufgeteilt in zwei Skalenhälfte **FG1** und **FG2** – zur Messung der funktionellen Gesundheit und zwei Items zur Erfassung der subjektiven Gesundheit in der Gegenwart (**SGG**) sowie Veränderungen der subjektiven Gesundheit in den letzten sechs Jahren (**SGV**).

¹ Der Autor dankt Christopher Marx sowie Susanne Wurm vom DZA ganz herzlich für ihre Unterstützung bei der Bereitstellung, Aufbereitung und Analyse dieses Datenbeispiels.

Abbildung 3.13 zeigt das Pfadmodell nach Whitelaw und Liang (1991) auf der Basis von drei manifesten Variablen. In diesem Modell wird angenommen, dass die exogene Variable „Physische Gesundheit“ (X ; hier operationalisiert über die Anzahl der Krankheiten in den letzten 12 Monaten) die endogene Variable „Funktionelle Gesundheit“ (Y_1 ; Summenscore aus dem SF-36-Fragebogen zur funktionellen Gesundheit) beeinflusst, welche ihrerseits wiederum einen Effekt auf die zweite endogene Variable „Subjektive Gesundheit“ (Y_2 ; gemessen über eine Selbsteinschätzung der gegenwärtigen subjektiven Gesundheit) hat. Außerdem wird angenommen, dass das Konstrukt physische Gesundheit die subjektive Gesundheit auch direkt beeinflusst. Die Koeffizienten β_1 , β_2 und β_3 bezeichnen Regressions- oder Pfadkoeffizienten und kennzeichnen die Stärke des Einflusses einer Variable auf eine andere.

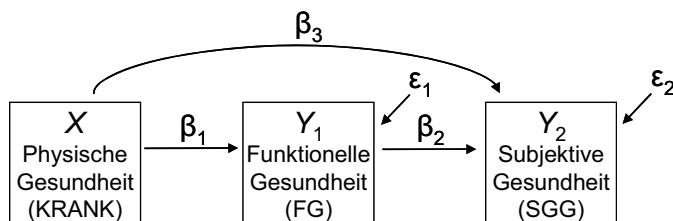


Abbildung 3.13: Manifestes Pfadmodell nach Whitelaw und Liang (1991) mit drei beobachteten Gesundheitsvariablen X , Y_1 und Y_2 . Die Konstanten β_1 , β_2 und β_3 bezeichnen Regressions- bzw. Pfadkoeffizienten. ε_1 und ε_2 sind Residualvariablen. Das Modell enthält den indirekten Effekt $\beta_1 \cdot \beta_2$. In Klammern sind die in der Mplus-Analyse verwendeten Variablennamen dargestellt.

Pfadanalysen sind insbesondere deshalb beliebt, weil viele sozialwissenschaftliche Theorien Hypothesen über *direkte* versus *indirekte Effekte* von Variablen implizieren (für Beispiele siehe MacKinnon, 2008). Indirekte Effekte sind Effekte, die über andere Variablen vermittelt sind. Sie werden daher häufig auch als *Mediatoreffekte* bezeichnet (Baron & Kenny, 1986). Mediatoreffekte lassen sich mit Hilfe von (manifesten oder latenten) Pfadanalysen untersuchen. Mediatoreffekte liegen in einem Pfadmodell dann vor, wenn das Modell eine oder mehrere Variablen enthält, die abhängige und unabhängige Variable zugleich sind. Wie oben bereits erwähnt bezeichnet man derartige „zwischengeschaltete“ Variablen als intermittierende Variablen oder Mediatorvariablen.

Im Modell in Abbildung 3.13 handelt es sich bei der Variablen Y_1 (funktionelle Gesundheit) um eine Mediatorvariable, da diese in Bezug auf X (physische Gesundheit) eine abhängige, in Bezug auf Y_2 (subjektive Gesundheit) jedoch eine unabhängige Variable ist. Anders ausgedrückt: Y_1 „empfängt“ einen Effekt von X (Pfad β_1) und hat ihrerseits einen Effekt auf Y_2 (Pfad β_2). Somit vermittelt Y_1 einen Teil des Effekts von X auf Y_2 , was den Begriff Mediatorvariable erklärt. Der indirekte (oder medierte) Effekt von X auf Y_2 kann als das Produkt der Regressions- bzw. Pfadkoeffizienten $\beta_1 \cdot \beta_2$ quantifiziert werden (Eid, 1999; MacKinnon, 2008). Die Variable X hat in diesem Beispiel außerdem noch einen direkten Effekt auf Y_2 , der sich über den Pfadkoeffizienten β_3 ausdrücken lässt.

Inhaltlich bedeutet dies, dass im Modell der „objektive“ Gesundheitsstatus (die physische Gesundheit) sowohl direkt als auch indirekt (über den durch Krankheit verursachten

Verlust an funktionellen Kompetenzen) auf die subjektiv erlebte Gesundheit wirkt. Mit dem Modell kann unter anderem getestet werden, ob die physische Gesundheit einen direkten Effekt auf die subjektive Gesundheit hat oder ob der Effekt vollständig durch die funktionelle Gesundheit mediiert wird.

Box 3.11: Mediator- versus Moderatoreffekte

Mediatoreffekte sollten nicht mit Moderatoreffekten verwechselt werden (siehe dazu ausführlich Baron & Kenny, 1986). Im Gegensatz zu einem Mediatoreffekt liegt ein Moderatoreffekt dann vor, wenn eine *Interaktion* zwischen zwei Variablen besteht. Eine Variable wird als Moderatorvariable bezeichnet, wenn sie die Stärke des direkten Einflusses einer Variable auf ein Kriterium beeinflusst. Wenn also beispielsweise das Regressionsgewicht β_3 (welches den direkten Effekt von X auf Y_2 in Abb. 3.13 kennzeichnet) für unterschiedliche Ausprägungen von Y_1 unterschiedliche Werte annehmen würde (der direkte Einfluss von X auf Y_2 also durch Y_1 moderiert würde), so könnte man Y_1 als Moderatorvariable bezeichnen. Die Analyse von Moderatoreffekten wird im vorliegenden Buch nicht besprochen. Aiken und West (1991) sowie Cohen et al. (2003, Kapitel 7-9) gehen ausführlich auf die Analyse von Moderatoreffekten in manifesten Regressionsmodellen (sog. *moderierte Regression*) ein. Latente Moderatormodelle werden u.a. von Klein (2000) sowie Marsh, Wen und Hau (2006) diskutiert.

In Pfadanalysen mit einem oder mehreren indirekten Effekten wird häufig eine Effektzerlegung in direkten Effekt plus indirekte(n) Effekt(e) vorgenommen und außerdem der *totale Effekt* einer Variablen bestimmt. Der totale Effekt einer Variablen berechnet sich dabei aus der Summe des direkten Effektes und aller indirekten Effekte. In unserem Beispiel beträgt der direkte Effekt von X auf Y_2 wie bereits erwähnt β_3 und der indirekte Effekt $\beta_1 \cdot \beta_2$. Somit kann der totale Effekt in diesem Fall als $\beta_3 + (\beta_1 \cdot \beta_2)$ ermittelt werden.

Box 3.12: Signifikanztestung von indirekten Effekten

Häufig ist die Überprüfung der statistischen Signifikanz von indirekten Effekten eine zentrale Fragestellung in Studien, in denen Mediatoreffekte untersucht werden sollen. Eine Schwierigkeit dabei ist, dass der indirekte Effekt das Produkt von zwei oder mehr Regressionskoeffizienten darstellt (in unserem Beispiel das Produkt $\beta_1 \cdot \beta_2$). Die Annahme der Normalverteilung des Effektes in der Population ist aus diesem Grund häufig verletzt. Daher führen konventionelle Methoden der Signifikanztestung von indirekten Effekten unter Umständen zu verzerrten Ergebnissen (MacKinnon, Lockwood & Williams, 2004). MacKinnon und Kollegen haben sich ausführlich mit verschiedenen Methoden zur Signifikanztestung indirekter Effekte befasst (MacKinnon, 2008; MacKinnon et al., 2004; MacKinnon, Lockwood, Hoffman, West & Sheets, 2002). Sie schlagen unter anderem die Verwendung sogenannter Bootstrap-Methoden zur Signifikanzprüfung mittels asymmetrischer Konfidenzintervalle um indirekte Effekte vor. Der theoretische Hintergrund sol-

cher Verfahren wird ausführlich von MacKinnon (2008) besprochen. Ein indirekter Effekt ist nach dieser Methode auf dem 5%-Niveau signifikant, wenn der Wert 0 nicht in das geschätzte 95%-Bootstrap-Konfidenzintervall um diesen Effekt fällt. MacKinnon et al. (2004) empfehlen u.a. den sogenannten Bias-korrigierten Bootstrap als Methode der Signifikanzprüfung von indirekten Effekten über Konfidenzintervalle. Bootstrap-Methoden sowie Konfidenzintervalle sind in Mplus implementiert und leicht zu handhaben. Ihre Anwendung zur Testung von Mediatoreffekten wird in Abschnitt 3.5.2 anhand des Datenbeispiels gezeigt.

3.5.2 Manifeste Pfadanalyse in Mplus

Abbildung 3.14 zeigt den Mplus-Input zur Spezifikation des manifesten Pfadmodells aus Abbildung 3.13. In diesem Modell wird die Krankheitshäufigkeit (Variable **KRANK**) zur Messung des Konstrukt physische Gesundheit verwendet. Funktionelle Gesundheit wird über den Gesamtscore der SF-36-Subskala zur funktionellen Gesundheit (Variable **FG**) operationalisiert. Subjektive Gesundheit wird durch das Item zur subjektiven Gesundheit in der Gegenwart (Variable **SGG**) repräsentiert.

```

1_Pfadanalyse_manifest.inp
title: Manifeste Pfadanalyse zum Zusammenhang von
       physischer Gesundheit, funktioneller Gesundheit und subjektiver Gesundheit
       Daten des Deutschen Zentrums für Altersfragen (DZA) Berlin (Alterssurvey)

data: file = gesundheit.dat;

variable: names = SGG ! Subjektive Gesundheit Gegenwart
           SGV ! Veränderungen der subjektiven Gesundheit vergangene 6 Jahre
           KRANK ! Anzahl an Krankheiten in den letzten 12 Monaten
           ARZT ! mittlere Anzahl Arztbesuche in den letzten 12 Monaten
           FG ! Funktionelle Gesundheit Gesamtscore
           FG1 FG2; ! Funktionelle Gesundheit Teilscore 1 & 2

usevar = SGG KRANK FG;

analysis: type = general;

model: SGG on KRANK FG; ! Pfadanalyse auf Ebene beobachteter Variablen
       FG on KRANK;

model indirect: SGG ind KRANK; ! Gesonderte Ausgabe des indirekten Effektes

output: sampstat stdyx;

```

Abbildung 3.14: Mplus-Input-Datei zur Spezifikation des manifesten Pfadmodells aus Abbildung 3.13 anhand der Daten im Datensatz **gesundheit.dat**.

Im Unterschied zur Regressionsanalyse in Abschnitt 3.2 kommt in der vorliegenden Pfadanalyse mehr als nur ein **on**-Befehl zum Einsatz, da simultan zwei lineare Regressionen geschätzt werden. Es gibt in diesem Fall zwei abhängige Variablen (SGG und FG). Somit sind auch zwei **on**-Befehle notwendig. Das Modell besteht aus einer multiplen Regression

mit zwei Prädiktoren (SGG on KRANK FG;) und einer bivariaten Regression (FG on KRANK;):

SGG on KRANK FG;

FG on KRANK;

Mit Hilfe des zusätzlichen `model indirect`-Befehls kann man in Mplus (sowohl bei manifesten als auch bei latenten Pfadanalysen) eine gesonderte Aufstellung aller direkten, indirekten und totalen Effekte anfordern. Dies ist insbesondere dann von Nutzen, wenn bestimmte indirekte Effekte auf Signifikanz geprüft werden sollen (s.u.). Zu diesem Zweck wird im manifesten Modell folgender Befehl zusätzlich in das `model`-Statement aufgenommen:

```
model indirect: SGG ind KRANK;
```

Das Schlüsselwort `ind` zwischen den Variablennamen `SGG` und `KRANK` bewirkt, dass u.a. alle indirekten Effekte der Variable `KRANK` auf die Variable `SGG` mit entsprechenden Standardfehlern und Signifikanztests im Output ausgegeben werden.

Im Mplus-Output erhalten wir unter `SAMPLE STATISTICS` die Mittelwerte, Kovarianzen und Korrelationen der drei manifesten Variablen `SGG`, `KRANK` und `FG` (nur die Korrelationen werden hier gezeigt). Wie zu erwarten korreliert die subjektive Gesundheit positiv mit der funktionellen Gesundheit und negativ mit der Anzahl an Krankheiten in den letzten 12 Monaten. Die Krankheitshäufigkeit ist mit funktioneller Gesundheit ebenfalls negativ korreliert.

Correlations			
	SGG	FG	KRANK
SGG	1.000		
FG	0.594	1.000	
KRANK	-0.458	-0.447	1.000

Unter der Rubrik `TESTS OF MODEL FIT` können wir erkennen, dass das von uns geschätzte manifeste Pfadmodell Null Freiheitsgrade aufweist. Dies liegt daran, dass alle Varianzen und Kovarianzen der drei beobachteten Variablen in diesem Modell zur Schätzung von Modellparametern verbraucht werden (man spricht in einem solchen Fall von einem *saturierten* oder *gerade identifizierten Modell*). Eine Konsequenz ist, dass das Modell keine testbaren Restriktionen für die Kovarianzstruktur der Variablen enthält und somit (trivialerweise) „perfekt passt“ (zu den Details siehe Bollen, 1989 oder Eid, 1999). Dies kann man insbesondere am ausgegebenen χ^2 -Modelltest erkennen:

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

Value	0.000
Degrees of Freedom	0
P-Value	0.0000

Der χ^2 -Wert beträgt Null und hat Null Freiheitsgrade und ist somit zur Bewertung der Modellgüte hier nicht von Nutzen (der p -Wert wird in einem solchen Fall in Mplus fälschlicherweise als $p = 0,0000$ berichtet; statistisch korrekt müsste der p -Wert hier als $p = 1$ angegeben werden). Die Beurteilung der Modellgüte eines saturierten Modells konzentriert sich daher eher auf die geschätzten Modellparameter (insbesondere die geschätzten Pfadkoeffizienten) und die erzielte Varianzaufklärung (gemessen über die R^2 -Werte) als auf die von Mplus unter TESTS OF MODEL FIT ausgegebenen Modellfitindices.

Box 3.13: Exkurs saturierte versus nichtsaturierte Pfadmodelle

Man beachte, dass nicht jedes manifeste Pfadmodell saturiert ist. Entfernt man beispielsweise im vorliegenden Modell den direkten Pfad von physischer Gesundheit (Variable KRANK) auf subjektive Gesundheit (Variable SGG), d.h. fixiert man den Pfadkoeffizienten β_3 implizit auf 0, so ist das Modell nicht mehr saturiert. Es weist stattdessen nun einen Freiheitsgrad auf und enthält somit eine testbare Restriktion. Anhand des Modelltests bzw. der Modellfitindices kann man in diesem Fall die Annahme überprüfen, dass der Effekt von physischer Gesundheit auf subjektive Gesundheit vollständig durch die funktionelle Gesundheit mediert wird, d.h. die physische Gesundheit nur einen indirekten (aber keinen direkten) Effekt auf die subjektive Gesundheit hat.

In Mplus muss man dazu lediglich in der Modellspezifikation den on-Befehl wie folgt ändern (der vollständige Input zu diesem Alternativmodell befindet sich auf der Begleit-CD):

SGG on FG;

FG on KRANK;

Das resultierende restriktivere Modell ohne direkten Effekt von KRANK auf SGG weist einen hochsignifikanten χ^2 -Wert von 122,11 bei einem Freiheitsgrad ($p < .0001$) auf. Die übrigen Fit-Indices (insbesondere der RMSEA-Index) deuten ebenfalls auf eine eher schlechte Anpassung dieses Modells an die Daten hin ($CFI = 0,90$, $RMSEA = 0,27$; $SRMR = 0,06$). Somit muss die Hypothese vollständiger Mediation verworfen werden (physische Gesundheit hat auch einen signifikanten direkten Effekt auf subjektive Gesundheit).

Die Tabelle mit den MODEL RESULTS liefert uns die Schätzungen der unstandardisierten Pfadkoeffizienten. Die standardisierten Pfadkoeffizienten finden wir unter STANDARDIZED MODEL RESULTS (STDX Standardization). Es ist erkennbar, dass die subjekti-

ve Gesundheit von der Krankheitshäufigkeit direkt und signifikant negativ beeinflusst wird (direkter Effekt $\hat{\beta}_3 = -0,094$, $z = -11,258$, $p < 0,001$; $\hat{\beta}_3^{\text{standardisiert}} = -0,24$). Eine gute funktionelle Gesundheit wirkt sich dagegen positiv auf die subjektiv erlebte Gesundheit aus (direkter Effekt $\hat{\beta}_2 = 0,885$, $z = 22,765$, $p < 0,001$; $\hat{\beta}_2^{\text{standardisiert}} = 0,486$). Physische Beinträchtigungen haben zudem einen signifikanten negativen Einfluss auf die funktionelle Gesundheit ($\hat{\beta}_1 = -0,096$, $z = -20,273$, $p < 0,001$; $\hat{\beta}_1^{\text{standardisiert}} = -0,447$). Je mehr Krankheiten berichtet werden, desto schlechter wird die eigene funktionelle Gesundheit eingeschätzt. Der indirekte Effekt der Krankheitshäufigkeit auf die subjektive Gesundheit wird weiter unten diskutiert.

Mit dem vorliegenden Modell kann ca. 39,9% der Varianz in der subjektiven Gesundheit ($R^2 = 0,399$) und ca. 20% der Varianz in der funktionellen Gesundheit ($R^2 = 0,20$) aufgeklärt werden. Dies können wir unter R-SQUARE (unter den standardisierten Parameterschätzungen) ablesen.

MODEL RESULTS

		Two-Tailed			
		Estimate	S.E.	Est./S.E.	P-Value
SGG	ON				
KRANK		-0.094	0.008	-11.258	0.000
FG		0.885	0.039	22.765	0.000
FG	ON				
KRANK		-0.096	0.005	-20.273	0.000
Intercepts					
SGG		1.408	0.117	12.034	0.000
FG		2.936	0.016	179.357	0.000
Residual Variances					
SGG		0.432	0.015	28.705	0.000
FG		0.174	0.006	28.705	0.000
STANDARDIZED MODEL RESULTS (STDYX Standardization)					
		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SGG	ON				
KRANK		-0.240	0.021	-11.441	0.000
FG		0.486	0.019	25.057	0.000
FG	ON				
KRANK		-0.447	0.020	-22.661	0.000
Intercepts					
SGG		1.661	0.151	10.973	0.000
FG		6.304	0.099	63.378	0.000

Residual Variances				
SGG	0.601	0.019	32.153	0.000
FG	0.800	0.018	45.431	0.000
R-SQUARE				
Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SGG	0.399	0.019	21.304	0.000
FG	0.200	0.018	11.330	0.000

Der über den model indirect-Befehl zusätzlich angeforderte Output liefert uns Informationen über die Größe und statistische Signifikanz des im Modell geschätzten indirekten Effektes ($\hat{\beta}_1 \cdot \hat{\beta}_2$) von physischer Gesundheit (PG) auf subjektive Gesundheit (SG). Zunächst werden die Effekte mit den entsprechenden Standardfehlern und Signifikanztests wiederum in unstandardisierter Form ausgegeben:

TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Effects from KRANK to SGG				
Total	-0.178	0.009	-20.892	0.000
Total indirect	-0.085	0.006	-15.140	0.000
Specific indirect				
SGG				
FG				
KRANK	-0.085	0.006	-15.140	0.000
Direct				
SGG				
KRANK	-0.094	0.008	-11.258	0.000

Der *totale Effekt*, also die Summe aus dem direkten Effekt $\hat{\beta}_3$ und dem indirekten Effekt $\hat{\beta}_1 \cdot \hat{\beta}_2$, beträgt (in unstandardisierter Form):

$$\hat{\beta}_3 + (\hat{\beta}_1 \cdot \hat{\beta}_2) = -0,094 + (-0,096 \cdot 0,885) = -0,178.$$

Der *totale indirekte Effekt* errechnet sich als

$$\hat{\beta}_1 \cdot \hat{\beta}_2 = -0,096 \cdot 0,885 = -0,085.$$

In diesem Beispiel gibt es lediglich einen indirekten Effekt von physischer Gesundheit auf subjektive Gesundheit (vermittelt über funktionelle Gesundheit), so dass der *totale indirekte Effekt* (Total indirect = -0,085) gleich dem *spezifischen indirekten Effekt* (Rubrik Specific indirect) ist. Der von Mplus geschätzte konventionelle Standardfehler für den indirekten Effekt beträgt 0,006 ($z = -15,14$). Somit ist der indirekte Effekt nach diesem Verfahren (dem so genannten *Sobel-Test*; Sobel, 1982) signifikant von Null verschieden ($p < 0,001$). Das bedeutet, dass ein signifikanter Mediatoreffekt vorliegt.

Darunter wird auch noch einmal der bereits weiter oben unter MODEL RESULTS berichtete direkte Effekt $\hat{\beta}_3 = -0,094$ ausgegeben. Nach dem gleichen Prinzip erfolgt danach die Auflistung der einzelnen Effekte bezogen auf die standardisierte Lösung:

STANDARDIZED TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

STDX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Effects from KRANK to SGG				
Total	-0.458	0.019	-23.496	0.000
Total indirect	-0.217	0.013	-16.600	0.000
Specific indirect				
SGG				
FG				
KRANK	-0.217	0.013	-16.600	0.000
Direct				
SGG				
KRANK	-0.240	0.021	-11.441	0.000

Die Verwendung konventioneller Signifikanztests zur Signifikanzprüfung indirekter Effekte kann in der Praxis problematisch sein, da das Produkt zweier oder mehrerer Regressionskoeffizienten hier Gegenstand der Signifikanzprüfung ist. Dieses Produkt ist häufig nicht normalverteilt, was die Verlässlichkeit herkömmlicher Signifikanztests einschränkt (MacKinnon, 2008; vgl. Box 3.12). MacKinnon et al. (MacKinnon, 2008; MacKinnon et al., 2004) empfehlen daher zur Signifikanztestung indirekter Effekte alternativ die Betrachtung von asymmetrischen Konfidenzintervallen, welche mit Hilfe von Bootstrapping-Methoden geschätzt werden können. Hierbei wird von diesen Autoren insbesondere der so genannte Bias-korrigierte (bias-corrected) Bootstrap empfohlen. In Mplus kann man Konfidenzintervalle um Modellparameterschätzungen allgemein (d.h. unabhängig von Bootstrap-Methoden) über den Befehl

output: cinterval;

anfordern. Wenn man Konfidenzintervalle auf der Basis des Bias-korrigierten Bootstraps schätzen lassen möchte, so ist zusätzlich folgende Spezifikation im analysis sowie im output-Befehl erforderlich:

```
analysis: bootstrap = <Anzahl der Bootstrap-Stichproben>;
output: cinterval(bcbootstrap);
```

Der Unterbefehl `bootstrap` = dient zur Wahl der Anzahl der Bootstrap-Stichproben. Um eine hinreichende Genauigkeit zu erzielen, sollten möglichst viele Bootstrap-Stichproben gewählt werden (z.B. 10.000). Die Art des Bootstrappings wird im `output`-Befehl hinter `cinterval` spezifiziert. Der Befehl `cinterval(bcbootstrap)` fordert hierbei die bias-korrigierte Variante des Bootstraps an. Normale Bootstrap-Konfidenzintervalle erhält man über `output: cinterval(bootstrap)`;

Die vollständigen Input- und Output-Dateien zur Spezifikation des Pfadmodells mit Bias-korrigierten Bootstrap-Konfidenzintervallen befinden sich auf der Begleit-CD. Nachfolgend ist der Teil des Mplus-Outputs dargestellt, der die entsprechenden, auf der Basis von 10.000 Bootstrap-Stichproben geschätzten, Konfidenzintervalle mit Bias-Korrektur enthält. Unter den Spalten `Lower .5%` sowie `Upper .5%` kann man die Unter- bzw. Obergrenze von 99%-Konfidenzintervallen um die Parameterschätzungen ablesen, wogegen unter den Spalten `Lower 2.5%` und `Upper 2.5%` die Unter- bzw. Obergrenzen von 95%-Konfidenzintervallen erscheinen.

```
CONFIDENCE INTERVALS OF TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT,
AND DIRECT EFFECTS
```

	Lower .5%	Lower 2.5%	Estimate	Upper 2.5%	Upper .5%
--	-----------	------------	----------	------------	-----------

Effects from KRANK to SGG

Total	-0.203	-0.197	-0.178	-0.159	-0.153
Total indirect	-0.102	-0.098	-0.085	-0.072	-0.068

Specific indirect

SGG					
FG					
KRANK	-0.102	-0.098	-0.085	-0.072	-0.068

Direct

SGG					
KRANK	-0.117	-0.111	-0.094	-0.076	-0.071

```
CONFIDENCE INTERVALS OF STANDARDIZED TOTAL, TOTAL INDIRECT, SPECIFIC
INDIRECT, AND DIRECT EFFECTS
```

STDX Standardization

	Lower .5%	Lower 2.5%	Estimate	Upper 2.5%	Upper .5%
--	-----------	------------	----------	------------	-----------

Effects from KRANK to SGG					
Total	-0.513	-0.500	-0.458	-0.415	-0.402
Total indirect	-0.257	-0.247	-0.217	-0.187	-0.178
Specific indirect					
SGG					
FG					
KRANK	-0.257	-0.247	-0.217	-0.187	-0.178
Direct					
SGG					
KRANK	-0.299	-0.285	-0.240	-0.196	-0.182

Es zeigt sich, dass der indirekte Effekt auch gemäß Bias-korrigierten Bootstrap-Konfidenzintervallen auf dem 1%-Niveau signifikant von Null verschieden ist. Dies ist daran zu erkennen, dass weder das 95%-Konfidenzintervall [-0.098, -0.072] noch das 99%-Konfidenzintervall [-0.102, -0.068] um den indirekten Effekt den Wert Null einschließt. Gleiches ist auch für den standardisierten indirekten Effekt der Fall (relevante Output-Zeilen sind unten jeweils fett gedruckt). Die Ergebnisse zeigen also, dass die physische Gesundheit neben einem signifikanten direkten auch einen signifikanten indirekten Effekt auf die subjektive Gesundheit hat.

3.5.3 Latente Pfadanalyse

In latenten Pfadanalysen wird die Idee der CFA mit der Idee der manifesten Pfadanalyse verbunden. So wird einerseits wie in der CFA für jedes zu untersuchende Konstrukt ein Messmodell mit einer latenten Variable (Faktor) und zumeist mehreren Indikatoren spezifiziert. Gleichzeitig sind die latenten Faktoren wie in einer manifesten Pfadanalyse über gerichtete (Regressions-)Pfade miteinander verknüpft (im Unterschied zur CFA, in welcher zumeist lediglich ungerichtete Zusammenhänge zwischen den Faktoren angenommen werden; Ausnahme: Faktorenanalysen höherer Ordnung, siehe Abschnitt 3.4.2).

Abbildung 3.15 zeigt eine latente Pfadanalyse mit drei latenten Variablen (η_1 , η_2 und η_3), welche jeweils mit zwei Indikatoren (Y_{11} - Y_{23}) gemessen wurden. Diese Pfadanalyse entspricht dem in Abbildung 3.13 gezeigten theoretischen Modell nach Whitelaw und Liang (1991), nur dass das Mediatormodell nun auf latenter Ebene (auf Ebene der η -Variablen) spezifiziert wird. Latente Pfadanalysen haben gegenüber manifesten Pfadanalysen den Vorteil, dass Messfehlereinflüsse bei der Schätzung von direkten, indirekten und sonstigen Effekten berücksichtigt werden.

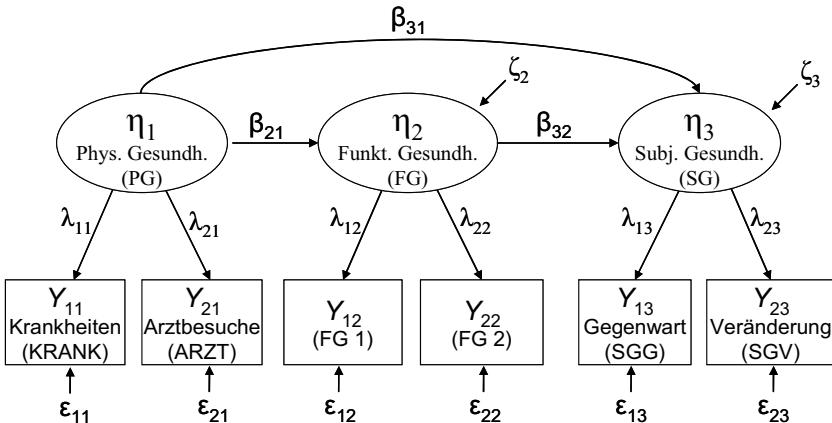


Abbildung 3.15: Latentes Pfadmodell mit drei latenten Variablen η_1 , η_2 und η_3 , welche jeweils mit zwei Indikatoren (Y_{11} - Y_{23}) gemessen werden. In Klammern erscheinen die in Mplus verwendeten Variablennamen der Indikatoren. Die Konstanten λ_{11} - λ_{23} sind Faktorladungen. ε_{11} - ε_{23} bezeichnen Messfehlervariablen. Die Konstanten β_{21} , β_{32} und β_{31} bezeichnen latente Regressions- bzw. Pfadkoeffizienten. ζ_2 und ζ_3 sind latente Residualvariablen. Das Modell enthält auf latenter Ebene den indirekten Effekt $\beta_{21} \cdot \beta_{32}$.

3.5.4 Latente Pfadanalyse in Mplus

Jedes der drei Konstrukte wurde in unserem Mplus-Beispiel über zwei Indikatoren (manifeste Variablen) operationalisiert: Physische Gesundheit über die Anzahl der Krankheiten (Variable KRANK) sowie die Anzahl der Arztbesuche in den letzten 12 Monaten (ARZT), funktionelle Gesundheit über die in zwei Testhälften (FG1 und FG2) aufgeteilte Subskala zur funktionellen Gesundheit aus dem SF-36-Fragebogen und subjektive Gesundheit über zwei Einzelitems, wobei sich das erste Item auf die subjektiv erlebte Gesundheit in der Gegenwart (SGG) und das zweite Item auf Veränderungen der subjektiven Gesundheit in den letzten sechs Jahren (SGV) bezieht.

Die drei η -Variablen repräsentieren die um Messfehler sowie item- bzw. skalenspezifische Einflüsse bereinigten Scores hinsichtlich der drei Konstrukte physische Gesundheit, funktionelle Gesundheit und subjektive Gesundheit. Die von Whitelaw und Liang (1991) postulierte Zusammenhangsstruktur (das Pfadmodell) zwischen den drei Konstrukten wird nun also auf Ebene der um Messfehler und spezifische Einflüsse bereinigten latenten Variablen untersucht. Abbildung 3.16 zeigt den Mplus-Input zur Schätzung der Parameter der latenten Pfadanalyse aus Abbildung 3.15.

```

title: Latente Pfadanalyse zum Zusammenhang von
      physischer Gesundheit, funktioneller Gesundheit und subjektiver Gesundheit
      Daten des Deutschen Zentrums für Altersfragen (DZA) Berlin (Alterssurvey)

data: file = gesundheit.dat;

variable: names = SGG ! Subjektive Gesundheit Gegenwart
           SGV ! Veränderungen der subjektiven Gesundheit vergangene 6 Jahre
           KRANK ! Anzahl an Krankheiten in den letzten 12 Monaten
           ARZT ! mittlere Anzahl Arztbesuche in den letzten 12 Monaten
           FG ! Funktionelle Gesundheit Gesamtscore
           FG1 FG2; ! Funktionelle Gesundheit Teilscore 1 & 2

usevar = SGG SGV KRANK ARZT FG1 FG2;

analysis: type = general;

model: SG by SGG SGV; ! Messmodell subjektive Gesundheit
       PG by KRANK ARZT; ! Messmodell physische Gesundheit
       FG by FG1 FG2; ! Messmodell funktionelle Gesundheit

       FG on PG; ! Pfadanalyse auf Ebene latenter Variablen
       SG on PG FG;

       model indirect: SG ind PG; ! Ausgabe des indirekten Effektes

output: sampstat stdyx;

```

Abbildung 3.16: Mplus-Input-Datei zur Spezifikation des latenten Pfadmodells aus Abbildung 3.15 anhand der Daten im Datensatz gesundheit.dat.

Die Besonderheit im latenten Pfadmodell ist, dass (ähnlich wie im latenten Regressionsmodell in Abschnitt 3.3) zunächst die Messmodelle für die latenten Faktoren (hier SG, PG und FG) spezifiziert werden müssen. Dies geschieht wiederum über by-Befehle:

```

SG by SGG SGV;
PG by KRANK ARZT;
FG by FG1 FG2;

```

Dieser Teil der Modellspezifikation entspricht somit der Spezifikation einer CFA in Mplus (vgl. Abschnitt 3.4). Der latente Faktor SG repräsentiert subjektive Gesundheit, der Faktor PG physische Gesundheit und der Faktor FG funktionelle Gesundheit. Im nächsten Schritt wird das Strukturmodell spezifiziert, welches hier der Pfadanalyse auf latenter Ebene entspricht:

```

FG on PG;
SG on PG FG;

```

Man beachte, dass SG, PG und FG in diesem Input latente Variablen bezeichnen, die nicht direkt im Datensatz enthalten sind (und darum auch nicht unter variable: names gelistet werden). Im latenten Pfadmodell ist der Befehl zur Ausgabe der indirekten Effekte

äquivalent zum entsprechenden Befehl im manifesten Modell, nur dass er sich auf die latenten Faktoren bezieht:

```
model indirect: SG ind PG;
```

Im Output sehen wir unter TESTS OF MODEL FIT, dass das latente Pfadmodell im Gegensatz zum manifesten Pfadmodell nicht saturiert ist, sondern sechs Freiheitsgrade aufweist. Die Freiheitsgrade entstammen allerdings in diesem Beispiel allesamt dem Messmodell. Das Strukturmodell (die Pfadanalyse auf Ebene der latenten Variablen) ist auch in diesem Modell saturiert, d.h. es erwachsen keine testbaren Restriktionen aus dem Strukturmodell. Die Modellgüteindices weisen darauf hin, dass das Gesamtmodell sehr gut auf die Daten passt:

```
TESTS OF MODEL FIT

Chi-Square Test of Model Fit

      Value           5.509
      Degrees of Freedom          6
      P-Value          0.4804

CFI/TLI

      CFI           1.000
      TLI           1.000

RMSEA (Root Mean Square Error Of Approximation)

      Estimate        0.000
      90 Percent C.I.    0.000  0.031
      Probability RMSEA <= .05     1.000

SRMR (Standardized Root Mean Square Residual)

      Value          0.007
```

Bezüglich der Parameterschätzungen (Rubrik MODEL RESULTS) können wir zunächst erkennen, dass alle frei geschätzten Faktorladungen signifikant von Null verschieden sind (die jeweils erste unstandardisierte Ladung pro Faktor ist wie bei der CFA standardmäßig auf 1 fixiert, um jedem Faktor eine Metrik zu verleihen). Auch sind alle standardisierten Ladungen [siehe STANDARDIZED MODEL RESULTS (STDX Standardization)] substantiell (Range: 0,494-0,972). Insbesondere die Indikatoren des Faktors FG (funktionelle Gesundheit) weisen sehr hohe standardisierte Ladungen auf ($\hat{\lambda}_{12}^{\text{standardisiert}} = 0,972$; $\hat{\lambda}_{22}^{\text{standardisiert}} = 0,955$), was auf eine starke Homogenität (die Variablen FG1 und FG2 messen etwas sehr ähnliches) bzw. hohe Reliabilität der verwendeten Skalenhälften der SF-36-Subskala zur funktionellen Gesundheit hinweist.

Beim Konstrukt physische Gesundheit (Faktor PG) sind die standardisierten Ladungen am geringsten. Dies kann vor allem dadurch erklärt werden, dass die Krankheitshäufigkeit mit der Anzahl der Arztbesuche zwar statistisch signifikant, aber nur moderat hoch korre-

liert ist ($r = 0,308$, $z = 13,799$, $p < 0,001$; Output dazu ist nicht gesondert dargestellt, jedoch auf der CD erhältlich). Somit sind die beiden Indikatoren für physische Gesundheit als eher heterogen zu betrachten (sie erfassen unterschiedliche Facetten des Konstruktions physische Gesundheit). Die Variable Krankheitshäufigkeit lädt dabei standardisiert etwas höher auf dem Faktor als die Anzahl der Arztbesuche, ist also eher als Markiervariable für den Faktor physische Gesundheit zu betrachten als die ARZT-Variable (als Markiervariablen werden solche Indikatoren bezeichnet, die besonders hoch auf einem Faktor laden und denen daher besonderes Gewicht bei der inhaltlichen Interpretation eines Faktors zukommt).

Hinsichtlich des Strukturmodells finden wir (wie zu erwarten) von der Richtung her dieselben Zusammenhänge wie im manifesten Modell: Physische Leiden (physische Gesundheit ist wiederum negativ gepolt mit den Indikatoren Häufigkeit von Krankheiten bzw. Arztbesuchen) beeinflussen sowohl die funktionelle als auch die subjektive Gesundheit negativ ($\hat{\beta}_{21} = -0,246$, $z = -13,189$, $p < 0,001$; $\hat{\beta}_{21}^{\text{standardisiert}} = -0,722$; $\hat{\beta}_{31} = -0,379$, $z = -6,944$, $p < 0,001$; $\hat{\beta}_{31}^{\text{standardisiert}} = -0,71$). Die funktionelle Gesundheit wirkt sich dagegen positiv auf die subjektive Gesundheit aus ($\hat{\beta}_{32} = 0,314$, $z = 2,631$, $p = 0,009$; $\hat{\beta}_{32}^{\text{standardisiert}} = 0,20$).

MODEL RESULTS

		Estimate	S.E.	Two-Tailed	
				Est. / S.E.	P-Value
SG	BY				
SGG		1.000	0.000	999.000	999.000
SGV		0.710	0.031	23.145	0.000
PG	BY				
KRANK		1.000	0.000	999.000	999.000
ARZT		5.547	0.363	15.265	0.000
FG	BY				
FG1		1.000	0.000	999.000	999.000
FG2		0.981	0.014	68.107	0.000
FG	ON				
PG		-0.246	0.019	-13.189	0.000
SG	ON				
PG		-0.379	0.055	-6.944	0.000
FG		0.314	0.119	2.631	0.009
Intercepts					
SGG		3.525	0.021	168.751	0.000
SGV		2.521	0.020	128.763	0.000
KRANK		2.706	0.054	50.435	0.000
ARZT		13.082	0.375	34.892	0.000
FG1		2.646	0.012	226.224	0.000
FG2		2.709	0.012	232.059	0.000

Variances				
PG		1.839	0.179	10.276
Residual Variances				
SGG		0.194	0.019	10.360
SGV		0.367	0.016	23.627
KRANK		2.906	0.159	18.304
ARZT		175.062	7.166	24.429
FG1		0.012	0.002	5.135
FG2		0.020	0.002	8.231
SG		0.132	0.026	5.114
FG		0.102	0.009	11.015
STANDARDIZED MODEL RESULTS (STDYX Standardization)				
		Estimate	S.E.	Two-Tailed P-Value
SG	BY			
SGG		0.854	0.016	54.333
SGV		0.647	0.018	36.105
PG	BY			
KRANK		0.623	0.026	24.309
ARZT		0.494	0.025	19.906
FG	BY			
FG1		0.972	0.006	173.398
FG2		0.955	0.006	165.700
FG	ON			
PG		-0.722	0.030	-24.378
SG	ON			
PG		-0.710	0.080	-8.836
FG		0.200	0.076	2.630
Intercepts				
SGG		4.157	0.076	54.351
SGV		3.172	0.060	52.435
KRANK		1.242	0.033	37.891
ARZT		0.860	0.029	29.817
FG1		5.573	0.100	55.647
FG2		5.716	0.103	55.731
Variances				
PG		1.000	0.000	999.000
				999.000

Residual Variances				
SGG	0.270	0.027	10.063	0.000
SGV	0.581	0.023	25.042	0.000
KRANK	0.612	0.032	19.202	0.000
ARZT	0.756	0.025	30.787	0.000
FG1	0.055	0.011	5.063	0.000
FG2	0.088	0.011	7.966	0.000
SG	0.251	0.045	5.549	0.000
FG	0.479	0.043	11.219	0.000

R-SQUARE				
Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SGG	0.730	0.027	27.166	0.000
SGV	0.419	0.023	18.052	0.000
KRANK	0.388	0.032	12.154	0.000
ARZT	0.244	0.025	9.953	0.000
FG1	0.945	0.011	86.699	0.000
FG2	0.912	0.011	82.850	0.000
Latent Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SG	0.749	0.045	16.517	0.000
FG	0.521	0.043	12.189	0.000

Es fällt auf, dass im latenten Modell deutlich stärkere Zusammenhänge zwischen den Konstrukten vorliegen als im manifesten Pfadmodell. Dies wird insbesondere anhand der R^2 -Werte für die latenten Variablen deutlich (R-SQUARE Latent Variable). So können im latenten Modell etwa 74,9% der Varianz in der subjektiven Gesundheit und 52,1% der Varianz in der funktionellen Gesundheit aufgeklärt werden.

Diese dramatischen Unterschiede im Vergleich zum manifesten Modell (hier betrug die Varianzaufklärung lediglich 39,9% bzw. 20%) lassen sich folgendermaßen erklären: Zum einen werden in der Analyse auf latenter Ebene Messfehler berücksichtigt, was im manifesten Pfadmodell nicht der Fall ist. Die Berücksichtigung von Messfehlern führt dazu, dass die Zusammenhänge korrekter geschätzt werden (sie sind auf latenter Ebene nicht durch Messfehler „verwässert“). Zusätzlich spielen vermutlich auch spezifische Varianzanteile der Indikatoren (sog. indikatorspezifische Varianz) eine Rolle. Wie oben gesehen sind die Indikatoren für physische Gesundheit (Variablen KRANK und ARZT) als eher heterogen einzustufen. Auch die Items SGG (subjektive Gesundheit Gegenwart) und SGV (Veränderungen der subjektiven Gesundheit) sind nur moderat hoch miteinander korreliert ($r = 0,553$, $z = 32,325$, $p < 0,001$; entsprechender Output dazu befindet sich auf der CD). Diese Indikatoren weisen somit jeweils relativ hohe spezifische Varianzanteile auf.

Im manifesten Pfadmodell hatten wir die Variable SGV als Indikator für subjektive Gesundheit und die Variable ARZT als Indikator für physische Gesundheit nicht berücksich-

tigt und konnten somit bei beiden Konstrukten spezifische Varianz der Indikatoren nicht „herausrechnen“ (bei funktioneller Gesundheit stellt sich dieses Problem nicht, da die Indikatoren FG1 und FG2 aus derselben Subskala stammen, sowie inhaltlich sehr homogen und auch sehr hoch korreliert sind, $r = 0,928$, $z = 273,013$, $p < 0,001$).

Die Ergebnisse im latenten Modell basieren dagegen auf multiplen, teilweise heterogenen Indikatoren und berücksichtigen somit neben Messfehlereinflüssen auch indikatorsspezifische Varianzanteile. Diese indikatorsspezifische Varianz wird hier wie Fehlervarianz behandelt und über die Messmodelle von gemeinsamer Konstruktvarianz separiert.

Im vorliegenden Fall liegt bei den Indikatoren der Konstrukte subjektive Gesundheit und physische Gesundheit jeweils ein hohes Maß an Indikatorsspezifität vor. Diese ist – neben der Berücksichtigung von Messfehlereinflüssen – mitverantwortlich dafür, dass auf latenter Ebene deutlich höhere Zusammenhänge ermittelt werden als im manifesten Modell.

Bei der Interpretation der Ergebnisse muss in einem solchen Fall sehr sorgfältig überlegt werden, ob die latenten Variablen – so wie sie durch die Messmodelle operationalisiert sind – die theoretisch interessierenden Konstrukte adäquat abbilden. Dies ist vor allem ein theoretisch-inhaltliches und weniger ein statistisches Problem, das der Anwender anhand von inhaltlichen Überlegungen lösen muss.

Der über den `model indirect`-Befehl zusätzlich angeforderte Output (nicht gesondert dargestellt) liefert uns auch im latenten Pfadmodell detaillierte Informationen über die Größe und statistische Signifikanz des in unserem Modell enthaltenen indirekten Effektes von physischer Gesundheit auf subjektive Gesundheit (Effekt $\hat{\beta}_{21} \cdot \hat{\beta}_{32}$). Auch können wir analog zum manifesten Pfadmodell Bootstrap-Konfidenzintervalle zur Signifikanzprüfung dieses indirekten Effektes anfordern. Da das Prinzip exakt dasselbe ist wie beim manifesten Pfadmodell, wird hier auf eine ausführliche Darstellung verzichtet. Die entsprechenden Input- und Outputdateien können der Begleit-CD entnommen werden.

4 Strukturgleichungsmodelle der Veränderungsmessung

Lineare Strukturgleichungsmodelle werden nicht nur häufig zur Auswertung querschnittlich erhobener Daten eingesetzt. Sie bieten sich aufgrund ihrer hohen Flexibilität bei der Modellierung komplexer Zusammenhänge und durch die Möglichkeit, Messfehlereinflüsse zu berücksichtigen, auch hervorragend zur Auswertung von Längsschnittdaten (Messwiederholungsdesigns) an. In den folgenden Abschnitten wird die Spezifikation und Analyse fünf verschiedener Klassen von Längsschnittstrukturgleichungsmodellen in Mplus besprochen. Grob kann hierbei zwischen Modellen zur Analyse von Variabilität (*Variabilitätsmodelle*) und Modellen zur Messung von Trait-Veränderungen (*Trait-Veränderungsmodelle*) unterschieden werden (vgl. Eid, 2003). Variabilitätsmodelle befassen sich eher mit der Modellierung situationsspezifischer Schwankungen von Messwerten um einen zeitstabilen Eigenschaftswert („Trait“). Trait-Veränderungsmodelle werden dagegen vor allem dann eingesetzt, wenn es darum geht, tatsächliche, überdauernde Veränderungen von Merkmalen über die Zeit zu analysieren.

In Hinblick auf Variabilitätsmodelle werden wir *Latent-State-* sowie *Latent-State-Trait-Modelle* (Steyer et al., 1992, 1999) besprechen (siehe Kapitel 4.1 und 4.2). In Kapitel 4.1 werden wir auch auf allgemeine Aspekte von Längsschnittanalysen wie beispielsweise die Modellierung indikatorsspezifischer Effekte über die Zeit und die Testung von Messinvarianzannahmen eingehen. Im Anschluss betrachten wir drei besonders gängige Typen von Trait-Veränderungsmodellen: (1) *Manifeste* sowie *latente autoregressive Modelle* (Jöreskog, 1979a, 1979b; Hertzog & Nesselroade, 1987; siehe Kapitel 4.3), (2) *Latent-Change-Modelle* (McArdle, 1988; Steyer et al., 1997; Steyer, Partchev, & Shanahan, 2000; siehe Kapitel 4.4) und (3) *Latente Wachstumskurvenmodelle* (*latent growth curve models*; Bollen & Curran, 2006; Duncan et al., 2006; siehe Kapitel 4.5).

4.1 Latent-State-Analyse

Hauptziel der Analyse von Variabilitätsmodellen ist es, den Grad der Situations- bzw. Messgelegenheitsabhängigkeit von Messwerten zu bestimmen. Hierbei wird zwischen einer *Trait-Komponente*, einer *State-Residual-Komponente* und einer *Messfehler-Komponente* unterschieden, weswegen man von *Latent-State-Trait-(LST-)Modellen* spricht (Steyer et al., 1992, 1999, siehe Abschnitt 4.2). Beispielsweise ist es häufig von Interesse, in welchem Ausmaß Messinstrumente (z.B. Fragebogenskalen) situationspezifische Aspekte (z.B. Zustandsangst = Angst in einer bestimmten Situation), stabile Persönlichkeitsdispositionen („Traits“; z.B. Eigenschaftsangst = generelle Ängstlichkeit einer Person über Situationen hinweg) und zufällige Messfehlereinflüsse reflektieren. Diese Fragen betreffen die zeitliche Stabilität, Gültigkeit (Validität) sowie die Genauigkeit (Reliabilität) der Messung interindividueller

Unterschiede zwischen Merkmalsträgern. LST-Modelle erlauben es, den Grad der Situationspezifität (*occasion-specificity*), der Konsistenz (*consistency*) und die Höhe der Reliabilität (*reliability*) über die Berechnung von Varianzkomponenten zu quantifizieren (Steyer et al., 1992, 1999).

Zur Einführung betrachten wir zunächst verschiedene Varianten der (weniger komplexen) sogenannten *Latent-State (LS-) Modelle* (siehe Abb. 4.1; vgl. Steyer et al., 1992), die eine „Vorstufe“ von LST-Modellen darstellen. LS-Modelle stellen allgemeine CFA-Modelle zur Auswertung von Längsschnittdaten dar (zur CFA für Querschnittsdaten siehe Abschnitt 3.4), in welchen noch keine Trennung stabiler von messgelegenheitsspezifischen Varianzanteilen vorgenommen wird. Da sie recht einfach aufgebaut und zudem eher unrestriktiv sind, werden LS-Modelle in der Praxis häufig zur Auswertung von Längsschnittdaten eingesetzt.

Box 4.1: Fallbeispiel Längsschnittanalyse

Grundlage unserer Beispielanalysen ist ein Längsschnittdatensatz, in welchem u.a. die Depressivität und Ängstlichkeit von $N = 620$ amerikanischen Kindern mit Hilfe von Selbsteinschätzungen zu mehreren Messgelegenheiten untersucht wurden (für Details siehe Cole, Martin, Powers & Truglio, 1996; Cole, Martin & Powers, 1997). Für unsere Illustration verwenden wir einen Teil dieser Daten, welcher sechs Variablen umfasst (die ersten drei Messzeitpunkte mit je zwei beobachteten Variablen als Indikatoren für selbstberichtete Depressivität). Die Daten liegen in Summary-Form vor (Mittelwerte, Standardabweichungen sowie Korrelationsmatrix; vgl. Abschnitt 2.2) und befinden sich in der Datei **depression.dat** auf der Begleit-CD.

Die beobachteten Variablen Y_{ik} ($i = \text{Indikator}$, $k = \text{Messgelegenheit}$) werden im LS-Modell zerlegt in einen konstanten Intercept-Term (Achsenabschnitt) α_{ik} , eine messgelegenheitsspezifische latente State-Variable und eine messgelegenheitsspezifische Messfehlervariable ε_{ik} :

$$Y_{ik} = \alpha_{ik} + \lambda_{ik} \cdot (\text{State } k) + \varepsilon_{ik},$$

wobei λ_{ik} die Faktorladung bezeichnet. Abbildung 4.1 zeigt ein LS-Modell für das Konstrukt Depression, gemessen mit zwei beobachteten Variablen Y_{1k} und Y_{2k} , welche zu drei Messgelegenheiten ($k = 1, 2, 3$) erhoben wurden.

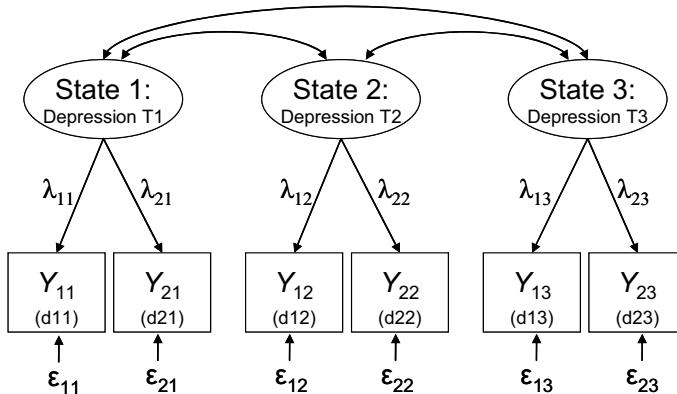


Abbildung 4.1: Latent State-Modell für Depression, gemessen zu drei Messgelegenheiten (T1-T3) mit jeweils zwei beobachteten Variablen Y_{ik} ($i = \text{Indikator}$, $k = \text{Messgelegenheit}$). ε_{ik} : Messfehlervariable; λ_{ik} : Faktorladung. In Klammern erscheinen die im Mplus-Beispiel verwendeten Variablennamen der Indikatoren.

Wie aus Abbildung 4.1 erkennbar, können die latenten State-Variablen im LS-Modell alle untereinander korreliert sein. Diese Korrelationen reflektieren die Stabilität interindividueller Unterschiede über die Zeit auf messfehlerbereinigter Ebene. Man kann an diesen Korrelationen somit erkennen, ob es sich bei einem Merkmal um eine zeitstabile Persönlichkeiteigenschaft („Trait“) handelt oder eher um ein Merkmal, welches starken messgelegenheitspezifischen Schwankungen unterworfen ist. Wenn die Korrelationen der State-Variablen sehr hoch sind (d.h. nahe 1 liegen), spricht dies für eine hohe Stabilität interindividueller Unterschiede und damit für ein Trait-artiges Konstrukt (z.B. Intelligenz). Bei geringen State-Korrelationen handelt es sich dagegen eher um ein eher situationsbezogenes Merkmal (z.B. Ärger).

Zusätzlich kann man an den Korrelationen der State-Faktoren beispielsweise auch erkennen, ob eine sogenannte autoregressive Struktur vorliegt. Dies wäre der Fall, wenn die State-Korrelationen mit zunehmendem zeitlichen Abstand der States voneinander immer weiter abnehmen. Latente autoregressive Modelle, welche Spezialfälle von LS-Modellen darstellen, werden in Abschnitt 4.3 besprochen.

In LS-Modellen werden meist keine besonderen Annahmen hinsichtlich der Kovarianz- / Korrelationsstruktur der latenten State-Variablen getroffen (wie dies z.B. in LST-Modellen sowie in latenten autoregressiven Modellen der Fall ist, s.u.). Das heißt, alle möglichen Korrelationen zwischen den latenten State-Variablen werden üblicherweise frei geschätzt. Somit ist das Strukturmodell in LS-Modellen in der Regel *saturiert*, was bedeutet, dass keine testbaren Restriktionen aus dem Strukturmodell ableitbar sind und somit auch keine Freiheitsgrade aus diesem Teil des Modells hervorgehen. Dieser Aspekt macht LS-Modelle zu eher unrestriktiven Modellen, weswegen sie sich gut als Ausgangspunkt einer Längsschnittanalyse auf latenter Ebene eignen. Ursachen eines möglichen Modell-Misfits können leichter identifiziert werden, da sie sich nur aus dem Messmodell, nicht aber aus dem Strukturmodell ergeben können.

4.1.1 LS- versus LST-Modelle

In LS-Modellen werden – im Gegensatz zu LST-Modellen (siehe Abschnitt 4.2) – nur latente Zustände (States) und Messfehlereinflüsse modelliert. Die latenten State-Variablen werden in LS-Modellen nicht weiter in eine Trait- und eine State-Residualkomponente zerlegt (Steyer et al., 1992). Somit erlauben LS-Modelle keine Trennung stabiler von messgelegenthspezifischen Varianzkomponenten.

LS-Modelle sind trotz dieser Einschränkung im Allgemeinen sehr nützliche Modelle für Längsschnittdatenauswertungen. Sie eignen sich nicht nur bei LST-Analysen gut als Ausgangsmodelle („baseline models“), da viele spezielle Längsschnittmodelle auf ihnen basieren (z.B. latente autoregressive Modelle, Latent-Change-Modelle und latente Wachstumskurvenmodelle 2. Ordnung, siehe Kapitel 4.3, 4.4 und 4.5) und sie im Vergleich zu diesen Erweiterungen weniger restriktive Annahmen machen.

LS-Modelle ermöglichen es somit, zunächst grundlegende Aspekte longitudinaler Messmodelle zu testen (beispielsweise die Frage der Homogenität der Indikatoren, der Invarianz der Faktorenstruktur sowie der Invarianz spezifischer Parameter über die Zeit), bevor zusätzliche restriktive Annahmen in Bezug auf die latenten State-Faktoren eingeführt werden (z.B. hinsichtlich einer State-Trait-Zerlegung, autoregressiver Strukturen oder der Form latenter Wachstumskurven). Möglichkeiten zur Modellierung indikatorsspezifischer Effekte werden in Abschnitt 4.1.3 besprochen. Auf die Frage der Messinvarianztestung wird in Abschnitt 4.1.4 eingegangen. In Abschnitt 4.2 wird dann die Erweiterung des LS-Modells zu einem LST-Modell gezeigt.

4.1.2 Analyse von LS-Modellen in Mplus

Im ersten Schritt einer Längsschnittanalyse wird häufig ein möglichst unrestrictives LS-Modell spezifiziert, d.h. ein Modell, in welchem beispielsweise noch keine Annahmen über zeitlich invariante Faktorladungen etc. getroffen werden. Eine wichtige Frage, die zunächst geklärt werden sollte, ist die Frage nach der Homogenität der Indikatoren bzw. die Frage, ob *indikatorspezifische Effekte* auftreten.

Bei indikatorspezifischen Effekten handelt es sich um spezifische (Methoden-)Effekte von Indikatoren (Indikatorspezifität), die dadurch sichtbar werden, dass dieselben Indikatoren wiederholt gemessen werden (vgl. Raffalovich & Bohrnstedt, 1987). Teilt ein Indikator spezifische Varianz (Varianz, die dieser nicht mit den übrigen beobachteten Variablen gemeinsam hat) mit sich selbst über die Zeit, so verletzt das die Homogenitätsannahme im konventionellen LS-Modell mit unkorrelierten Fehlervariablen. Der Indikator wäre dann mit sich selbst über die Zeit höher korreliert, als es ein LS-Modell mit unkorrelierten Fehlervariablen vorhersagen würde. Mit anderen Worten, ein LS-Modell mit unkorrelierten Fehlervariablen wäre in diesem Fall zu restriktiv und es müsste überlegt werden, wie diese indikatorspezifischen Effekte adäquat modelliert werden können.

Im Folgenden wird zunächst die Schätzung eines einfachen LS-Modells *ohne* Berücksichtigung indikatorspezifischer Effekte in Mplus gezeigt. Wie wir sehen werden, weist dieses einfache Modell im vorliegenden Beispiel aufgrund vorhandener indikatorspezifischer Effekte eine verbesserungswürdige Datenanpassung auf (s.u.). Anschließend wird

demonstriert, wie das einfache LS-Modell erweitert werden kann, um indikatorsspezifische Effekte zu berücksichtigen. Abbildung 4.2 zeigt den Mplus-Input zur Schätzung des einfachen LS-Modells ohne Berücksichtigung indikatorsspezifischer Effekte aus Abbildung 4.1 für die sechs Depressions-Indikatoren d11 bis d13.

The screenshot shows the Mplus software window with the title bar 'Mplus - [1_LS-Modell.inp]'. The menu bar includes File, Edit, View, Mplus, Graph, Window, and Help. Below the menu is a toolbar with various icons. The main area contains the Mplus input code:

```

title: Latent-State-Modell ohne Invarianzrestriktionen
       und ohne Berücksichtigung indikatorsspezifischer Effekte
       Summary-Daten aus Datei "depression.dat"
       (Mittelwerte, Standardabweichungen und Korrelationen)
       Depression-Selbstbericht gemessen mit je 2 Indikatoren
       zu 3 Messgelegenheiten

data: file = depression.dat;
      type = means std corr; ! Angabe der Art der Summary-Daten
                           ! Wichtig hier: Beachtung der Reihenfolge!

      nobs = 620; ! Angabe der Stichprobengröße

variable: names = d11 d21 ! entspricht Y11 und Y21 in Abb. 34
           d12 d22 ! entspricht Y12 und Y22 in Abb. 34
           d13 d23; ! entspricht Y13 und Y23 in Abb. 34

model: state1 by d11 d21; ! Spezifikation der Messmodelle
       state2 by d12 d22; ! für die Latent-State-Variablen
       state3 by d13 d23;

output: sampstat standardized stdyx;

```

The status bar at the bottom shows 'Ready', 'Ln 2, Col 63', and 'NUM'.

Abbildung 4.2: Mplus-Input-Datei zur Spezifikation eines einfachen LS-Modells ohne spezifische Messinvarianzrestriktionen für die Depressionsdaten (drei Messgelegenheiten, vgl. Abb. 4.1).

Mit Hilfe der Unterbefehle

```

state1 by d11 d21;
state2 by d12 d22;
state3 by d13 d23;

```

werden im model-Befehl die Indikatoren wie in einem konventionellen CFA-Modell (vgl. Abschnitt 3.4) den jeweiligen State-Faktoren zugeordnet.

Box 4.2: Mplus-Default für die Mittelwertsstruktur

In unserem Ausgangsmodell schätzen wir zunächst noch keine Mittelwerte für die latenten State-Faktoren, sondern verwenden die Mplus-Standardeinstellung, gemäß welcher die Mittelwerte aller latenten Variablen auf Null fixiert sind, während die Intercepts aller manifesten Variablen (α_{ik}) als Modellparameter frei geschätzt werden. Sofern keine Restriktionen auf die Intercepts der beobachteten Variablen gelegt werden, sind die Intercepts gleich den Mittelwerten [Erwartungswerten $E(Y_{ik})$] der beobachteten Variablen. Dies folgt aus der Zerlegung der Erwartungswerte der beobachteten Variablen im LS-Modell, da die Mittelwerte der latenten States per Mplus-Voreinstellung auf Null fixiert sind und die Residuen ε_{ik} per Definition Mittelwerte von Null haben (vgl. Steyer et al., 1992):

$$\begin{aligned} E(Y_{ik}) &= E(\alpha_{ik}) + E(\lambda_{ik} \cdot \text{State } k) + E(\varepsilon_{ik}) \\ &= \alpha_{ik}. \end{aligned}$$

In Abschnitt 4.1.4 wird demonstriert, wie die Mittelwerte der latenten State-Variablen (an denen in Längsschnittstudien häufig Interesse besteht) identifiziert und in Mplus geschätzt werden können.

Die wichtigsten im Mplus-Output berichteten Modellgütemaße zum einfachen LS-Modell sind im Folgenden dargestellt:

Chi-Square Test of Model Fit

Value	43.480
Degrees of Freedom	6
P-Value	0.0000

CFI/TLI

CFI	0.988
TLI	0.970

Information Criteria

Number of Free Parameters	21
Akaike (AIC)	-1222.299
Bayesian (BIC)	-1129.275
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	-1195.947

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.100
90 Percent C.I.	0.074 0.129
Probability RMSEA <= .05	0.001

SRMR (Standardized Root Mean Square Residual)

Value	0.009
-------	-------

Es zeigt sich, dass das Modell einen verbessерungsbedürftigen Fit aufweist, was insbesondere aufgrund des Chi-Quadrat-Tests ($\chi^2 = 43,48, df = 6, p < 0,0001$) und des RMSEA-Koeffizienten [$RMSEA = 0,10; p(RMSEA \leq 0,05) = 0,001$] diagnostiziert werden kann (CFI und $SRMR$ zeigen mit $CFI = 0,988$ und $SRMR = 0,009$ akzeptable Werte). Ein schlechter Modellfit kann in Längsschnittanalysen häufig durch das Vorliegen von indikatorsspezifischen Effekten erklärt werden, die im einfachen LS-Modell nicht hinreichend berücksichtigt werden.

4.1.3 Modellierung indikatorsspezifischer Effekte

Eine Möglichkeit zur Modellierung indikatorsspezifischer Effekte besteht darin, Korrelationen zwischen denjenigen Fehlervariablen zuzulassen, die demselben Indikator zugeordnet sind (Sörbom, 1975; siehe Abbildung 37). Man spricht dann häufig von einem LS-Modell mit *autokorrelierten Fehlervariablen* (oder im Multitrait-Multimethod-Kontext von *Correlated-Uniqueness-Modellen*, siehe z.B. Lance, Noble & Scullen, 2002; Cole & Maxwell, 2003).

Da die Schätzung von Fehlerkorrelationen eine relativ unkomplizierte Art der Modellierung darstellt, erfreut sie sich großer Beliebtheit. Ein Nachteil dieser Methode besteht jedoch darin, dass indikatorsspezifische Effekte auf diese Weise mit Fehlervarianz konfundiert bleiben und nicht als Quelle systematischer (reliabler) Varianz von der Fehlervarianz separiert werden können. Somit führt diese Methode im Allgemeinen zu einer Unterschätzung der Reliabilitäten der Indikatoren. Aus diesem Grund werden häufig statt autokorrelierter Fehlervariablen sogenannte *indikatorsspezifische Faktoren* (ähnlich wie Methodenfaktoren in Multitrait-Multimethod-Modellen) mit in das Modell aufgenommen. Wir besprechen hier zunächst die Spezifikation autokorrelierter Fehlervariablen und anschließend eine Modellvariante mit einem indikatorsspezifischen Faktor.

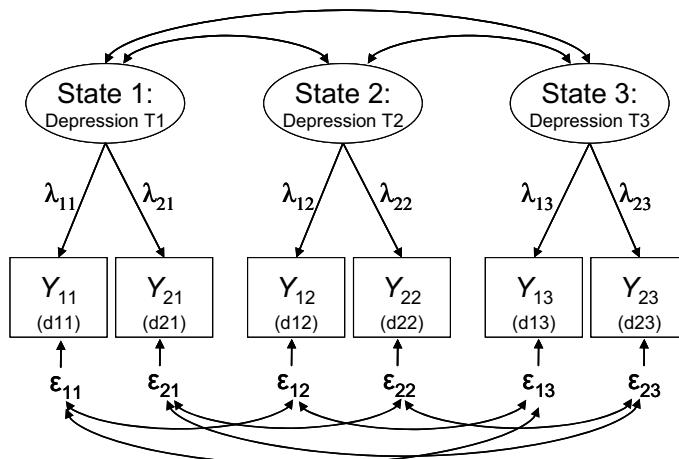


Abbildung 4.3: Latent-State-Modell mit autokorrelierten Fehlervariablen zur Modellierung indikatorsspezifischer Effekte über die Zeit. Das Modell ist im vorliegenden Beispiel aufgrund einer zu geringen Zahl von Indikatoren unteridentifiziert.

Für unser vorliegendes Beispiel mit lediglich zwei Indikatoren pro State-Faktor sowie nur drei Messzeitpunkten ist ein LS-Modell, in welchem sämtliche möglichen Fehlerautokorrelationen frei geschätzt werden, aufgrund der zu geringen Zahl vorliegender Informationen unteridentifiziert und somit nicht schätzbar (zu den Grundlagen des Identifikationsproblems bei Strukturgleichungsmodellen sowie bezüglich allgemeiner Regeln für die Identifizierbarkeit von Modellen siehe Bollen, 1989). Um ein Modell mit allen Fehlerautokorrelationen schätzen zu können, wären mehr Indikatoren pro Messgelegenheit bzw. mehr als drei Messgelegenheiten nötig.

Um das Prinzip der Spezifikation autokorrelierter Fehlervariablen in Mplus dennoch anhand unseres Beispiels illustrieren zu können, greifen wir auf eine restriktivere Modellvariante zurück, in welcher nur die Korrelationen zwischen Fehlervariablen benachbarter Messzeitpunkte zugelassen werden (siehe Abb. 4.4). Diese Modellvariante ist auch für die vorliegende Datensituation identifiziert und schätzbar. Der relevante Ausschnitt aus der Mplus-Syntax ist in Abbildung 4.5 dargestellt.

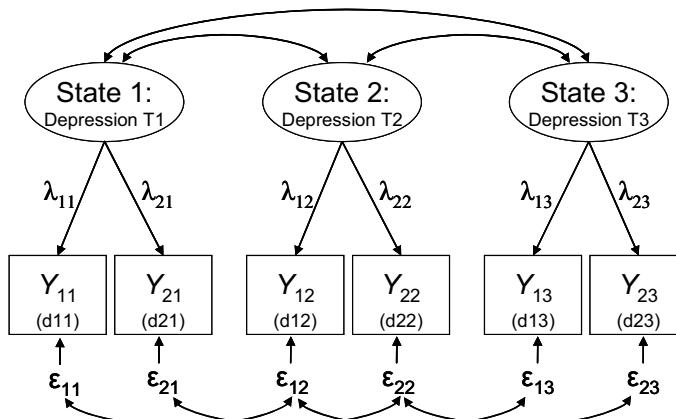


Abbildung 4.4: Latent-State-Modell mit autokorrelierten Fehlervariablen zur Modellierung indiktorspezifischer Effekte über die Zeit. Im Gegensatz zum Modell in Abbildung 4.3 werden hier nur die Fehlervariablen benachbarter Messgelegenheiten korreliert.

```
model: state1 by d11 d21; ! Spezifikation der Messmodelle
       state2 by d12 d22; ! für die Latent-State-Variablen
       state3 by d13 d23;

       ! Spezifikation autokorrelierter Fehlervariablen
       ! (nur für benachbarte Messzeitpunkte)
       d11 with d12;
       d12 with d13;
       d21 with d22;
       d22 with d23;
```

Abbildung 4.5: Mplus-model1-Befehl zur Spezifikation des LS-Modells mit autokorrelierten Fehlervariablen zu benachbarten Messzeitpunkten.

Die Modellgütemaße für dieses Modell ($\chi^2 = 16.182, df = 2, p < 0,001; CFI = 0,995; RMSEA = 0,107; p(RMSEA \leq 0,05) = 0,019; SRMR = 0,006$) fallen günstiger aus als für das einfache LS-Modell ohne Fehlerkorrelationen [Fit-Maße für das einfache LS-Modell zur Erinnerung: $\chi^2 = 43,48, df = 6, p < 0,0001; RMSEA = 0,10; p(RMSEA \leq 0,05) = 0,001; CFI = 0,988; SRMR = 0,009$]. Die geschätzten Fehlerkorrelationen liegen zwischen $-0,120$ (n.s.) und $0,277$ ($p < 0,001$).

Allerdings zeigt der χ^2 -Wert auch für das Modell mit Autokorrelationen eine signifikante Abweichung zwischen Modell und Daten an. Der $RMSEA$ -Wert liegt zudem nach wie vor erheblich über dem empfohlenen Richtwert von 0,05. Dies weist darauf hin, dass indikatorspezifische Effekte über mehr als nur benachbarte Messzeitpunkte hinweg generalisieren und diese somit mit dem vorliegenden Modell, in welchem nur die Fehlervariablen benachbarter Messgelegenheiten korreliert sind, nicht hinreichend abgebildet werden.

Eid, Schneider und Schwenkmezger (1999, siehe auch Eid, 2000) haben eine Modellierungsart für indikatorspezifische Effekte vorgeschlagen, die ohne die Schätzung von Fehlerautokorrelationen auskommt und auch den Nachteil der Unterschätzung der Reliabilitäten von Indikatoren vermeidet. Bei der von ihnen empfohlenen Art der Modellierung wird zunächst ein Indikator als *Referenzindikator* ausgewählt. Hierbei kann es sich z.B. um eine Markiervariable handeln, die das zu messende Konstrukt, hier Depression, besonders gut repräsentiert. Der Einfachheit halber nehmen wir zur Illustration im Folgenden an, dass der erste Indikator (Y_{1k}) als Referenzindikator gewählt wurde. Rein mathematisch könnte jedoch auch jeder andere Indikator ausgewählt werden.

Der Referenzindikator Y_{1k} lädt nur auf dem State-Faktor, während alle anderen (Nicht-referenz-)Indikatoren $Y_{ik}, i \neq 1$, zusätzlich auf einem sogenannten indikatorspezifischen (Methoden-)Faktor laden (siehe Abbildung 4.6). Der indikatorspezifische Faktor bildet die Spezifität eines Nichtreferenzindikators $Y_{ik}, i \neq 1$, in Kontrast zum Referenzindikator Y_{1k} ab. Im vorliegenden Fall (zwei Indikatoren) wird somit nur für Y_{2k} ein zusätzlicher Faktor benötigt, allgemein für alle Indikatoren $Y_{ik}, i \neq 1$.

Der indikatorspezifische Faktor ist als *Residualfaktor* bezüglich der State-Faktoren definiert, d.h. er repräsentiert den Teil der reliablen Varianz in den Nicht-Referenzindikatoren, welcher nicht durch die reliable Varianz des Referenzindikators erklärt werden kann. Daraus folgt, dass der indikatorspezifische Faktor in jedem Fall einen Mittelwert von Null hat und mit allen State-Faktoren unkorreliert ist (vgl. Eid et al., 1999; Eid, 2000). Es ist wichtig, diesen Aspekt zu betonen, da die Unkorreliertheit des indikatorspezifischen Faktors mit den State-Faktoren im Mplus-Input explizit spezifiziert werden muss (der Mittelwert des Faktors wird dagegen bereits per Voreinstellung auf Null fixiert, vgl. Box 4.2).

Die Bedeutung des Faktors *State 1* in Abbildung 4.6 wird durch die Bedeutung des Indikators Y_{11} (d11) bestimmt. Aus diesem Grund sollte als Referenzindikator wie oben bereits erwähnt stets eine Variable gewählt werden, die das zu untersuchende Konstrukt theoretisch bzw. inhaltlich besonders gut abbildet (sogenannte Markiervariable). Dies ist insbesondere dann wichtig, wenn man eher heterogene Indikatoren zur Verfügung hat, die unterschiedliche Facetten eines Konstrukttes repräsentieren. Auf diesen Punkt werden wir bei der Besprechung der Schätzergebnisse nochmals zurück kommen.

Ein LS-Modell mit einem indikatorspezifischen Faktor (IS_2) für den zweiten Indikator wird in Abbildung 4.6 gezeigt. Die Mplus-Syntax zur Spezifikation dieses Modells kann Abbildung 4.7 entnommen werden.

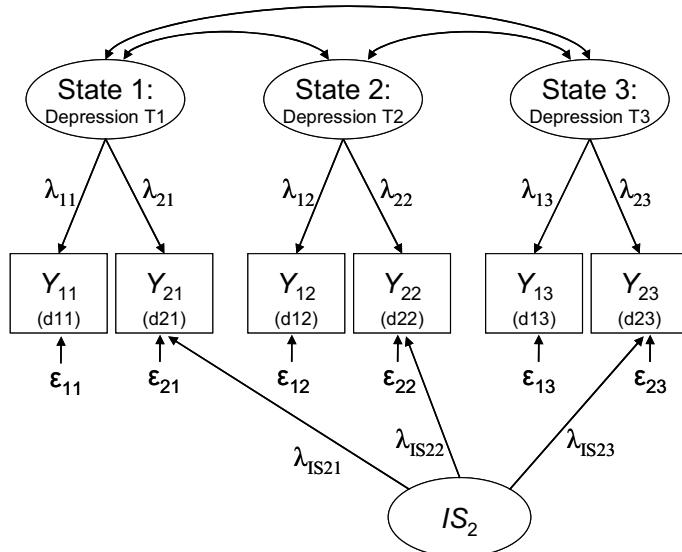


Abbildung 4.6: Latent-State-Modell mit einem indikatorspezifischen Faktor (IS_2) für den zweiten Indikator (Y_{2k}) zur Modellierung indikatorspezifischer Effekte über die Zeit. Die Parameter λ_{IS2k} bezeichnen die Faktorladungen auf dem indikatorspezifischen Faktor IS_2 . Der indikatorspezifische Faktor ist als Residualfaktor mit den State-Faktoren unkorreliert.

```
model: state1 by d11 d21; ! Spezifikation der Messmodelle
       state2 by d12 d22; ! für die Latent-State-Variablen
       state3 by d13 d23;

       is2 by d21 d22 d23; ! Indikator spezifischer Faktor
                           ! für den 2. Indikator

       is2 with state1-state3@0;
```

Abbildung 4.7: Mplus-model-Befehl zur Spezifikation des LS-Modells mit einem indikatorspezifischen Faktor ($is2$) für den zweiten Indikator aus Abbildung 39.

Der indikatorspezifische Faktor $is2$ wird in der Mplus-Syntax analog zu den State-Faktoren mit einem `by`-Statement eingeführt. Wichtig ist, dass Korrelationen zwischen dem indikatorspezifischen Faktor und den State-Faktoren desselben Konstrukts nicht zugelassen werden dürfen, da es sich bei $is2$ um einen Residualfaktor handelt. Mit Hilfe des Befehls

```
is with state1-state3@0;
```

werden diese unzulässigen Korrelationen auf Null fixiert.

Es zeigt sich, dass das LS-Modell mit indikatorspezifischem Faktor für den zweiten Indikator einen deutlich besseren Fit aufweist als das einfache LS-Modell ohne Berücksichtigung indikatorspezifischer Effekte. Es passt zudem besser als das Modell mit autokorrelier-

ten Fehlern zu benachbarten Messzeitpunkten. Der χ^2 -Wert für das LS-Modell mit indikatorspezifischem Faktor ist wesentlich geringer und statistisch nicht signifikant ($\chi^2 = 1,24, df = 3, p = 0,74$). Auch CFI ($CFI = 1,000$), RMSEA [$RMSEA = 0,000; p(RMSEA \leq 0,05) = 0,96$] und SRMR ($SRMR = 0,003$) zeigen einen wesentlich verbesserten Fit an. Zudem ist der AIC-Wert für das Modell mit indikatorspezifischem Faktor kleiner ($AIC = -1258,54$) als für das Modell ohne diesen Faktor ($AIC = -1222,30$) sowie kleiner als für das Modell mit autokorrelierten Fehlervariablen ($AIC = -1241,598$).

Aus Platzgründen werden daher nachfolgend nur die Parameterschätzungen des gut passenden Modells mit einem indikatorspezifischen Faktor für den zweiten Indikator gezeigt. Die vollständigen Outputs für die beiden anderen Modelle können der Begleit-CD entnommen werden.

Mplus gibt wie üblich zunächst die unstandardisierten Parameterschätzungen aus. Es lässt sich erkennen, dass die State-Faktorladungen des zweiten Indikators ($\lambda_{21}, \lambda_{22}$ und λ_{23}) auf ähnlich hohe Werte geschätzt wurden, wie die des ersten Indikators (letztere sind aus Identifikationsgründen zu jedem der drei Messzeitpunkte auf 1 fixiert). Dies spricht dafür, dass die beiden Indikatoren ähnliche Trennschärfen aufweisen.

Weiterhin sehen wir, dass sich die State-Ladungen des zweiten Indikators über die Zeit kaum verändert haben. Somit könnte die Annahme über die Zeit gleichbleibender („zeitlich invariant“) Ladungen gerechtfertigt sein (für einen formalen Test dieser Annahme siehe Abschnitt 4.1.4). Außerdem können wir erkennen, dass die beiden frei geschätzten unstandardisierten Ladungen der Variablen d22 und d23 auf dem indikator spezifischen Faktor (is2; Schätzungen: $\hat{\lambda}_{IS22} = 1,040$ und $\hat{\lambda}_{IS23} = 1,735$) signifikant von Null verschieden sind (z-Werte: 3,828 und 2,588). Dies bedeutet, dass statistisch signifikante indikator spezifische Effekte vorliegen, welche im einfachen LS-Modell nicht berücksichtigt wurden. Die standardisierten Ladungen auf dem indikator spezifischen Faktor liegen zwischen 0,176 und 0,308.

In der Praxis ist es sehr wichtig, den Grad der Heterogenität der Indikatoren (Stärke der indikator spezifischen Effekte) genau zu untersuchen. Liegen hohe standardisierte Ladungen der Indikatoren auf den indikator spezifischen Faktoren vor (etwa um 0,50 oder höher), so zeigt dies, dass ein relativ großer Prozentsatz der Varianz der Indikatoren durch die indikator spezifischen Faktoren erklärt wird (die quadrierten standardisierten Ladungen geben den Anteil der beobachteten Varianz an, der durch den indikator spezifischen Faktor erklärt wird). Praktisch würde das bedeuten, dass die Indikatoren eher heterogen sind und unterschiedliche Facetten des Konstrukt abilden.

In einem solchen Fall sollte man sehr sorgfältig überlegen, welchen Indikator man als Referenzindikator (ohne indikator spezifischen Faktor) verwendet, da die Interpretation der State-Faktoren vom Referenzindikator abhängt. Sind Indikatoren sehr heterogen, so können beispielsweise die Korrelationen der State-Faktoren sowie andere Parameterschätzungen des Modells stark variieren, je nach dem, welcher Indikator als Referenzindikator verwendet wird.

Wenn in einer Anwendung ein hohes Maß an Indikator spezifität vorliegt, so sollte man vor allem aufgrund inhaltlicher Überlegungen entscheiden, welcher der Indikatoren das Konstrukt am besten repräsentiert und diesen Indikator als Referenzindikator verwenden

(vgl. dazu die Diskussion in Geiser, Eid und Nussbeck, 2008). Zudem sollten gründliche Sensitivitätsanalysen durchgeführt werden, d.h. es sollte überprüft werden, in welcher Weise sich die Wahl unterschiedlicher Referenzindikatoren auf die Ergebnisse des Strukturmodells auswirkt.

In unserem Beispiel liegen nur geringe bis moderate indikatorsspezifische Effekte vor (standardisierte Ladungen zwischen 0,176 und 0,308). Somit gehen höchstens 9,4 % (0,308²) der beobachteten Varianz auf Indikatorsspezifität zurück, was als unbedenklich angesehen werden kann. Inhaltlich ist daraus zu schließen, dass die Indikatoren größtenteils homogen sind (dasselbe Konstrukt erfassen) und nur geringe spezifische Anteile haben.

Wir sehen weiterhin, dass wie gewünscht alle Kovarianzen des indikatorsspezifischen Faktors mit den drei State-Faktoren auf Null gesetzt wurden (wodurch der indikatorsspezifische Faktor als Residualfaktor interpretiert werden kann). Die Kovarianzen der State-Faktoren untereinander wurden dagegen wie gewünscht von Mplus frei geschätzt. Diese Kovarianzen sind alle signifikant von Null verschieden ($11,659 \leq z \leq 13,547$). Somit liegt eine statistisch bedeutsame Stabilität interindividueller Unterschiede über die Zeit vor. Das Ausmaß der Stabilität ist jedoch anhand der Kovarianzen schwer zu beurteilen, da es sich bei Kovarianzen um nichtstandardisierte Zusammenhangsmaße handelt. Aus diesem Grund betrachtet man meist eher die Faktorkorrelationen (siehe STANDARDIZED MODEL RESULTS) zur Beurteilung der Stabilitäten.

Weiter unten werden noch die geschätzten Intercepts, Faktorvarianzen sowie Residualvarianzen angegeben. Die unstandardisierten Intercepts entsprechen hier den beobachteten Mittelwerten der Indikatoren, da wir in diesem Modell keine Restriktionen auf die Mittelwertsstruktur der beobachteten Variablen gelegt haben (vgl. Box 4.2).

MODEL RESULTS

				Two-Tailed	
		Estimate	S.E.	Est./S.E.	P-Value
STATE1	BY				
D11		1.000	0.000	999.000	999.000
D21		0.950	0.034	28.154	0.000
STATE2	BY				
D12		1.000	0.000	999.000	999.000
D22		0.918	0.030	30.365	0.000
STATE3	BY				
D13		1.000	0.000	999.000	999.000
D23		0.913	0.031	29.496	0.000
IS2	BY				
D21		1.000	0.000	999.000	999.000
D22		1.040	0.272	3.828	0.000
D23		1.735	0.670	2.588	0.010

IS2 WITH				
STATE1	0.000	0.000	999.000	999.000
STATE2	0.000	0.000	999.000	999.000
STATE3	0.000	0.000	999.000	999.000
 STATE2 WITH				
STATE1	0.063	0.005	13.097	0.000
 STATE3 WITH				
STATE1	0.052	0.004	11.659	0.000
STATE2	0.064	0.005	13.547	0.000
 Intercepts				
D11	0.314	0.013	24.301	0.000
D21	0.281	0.012	23.386	0.000
D12	0.280	0.013	21.669	0.000
D22	0.259	0.012	21.155	0.000
D13	0.286	0.012	22.917	0.000
D23	0.260	0.012	21.799	0.000
 Variances				
STATE1	0.085	0.006	13.718	0.000
STATE2	0.089	0.006	14.449	0.000
STATE3	0.086	0.006	14.703	0.000
IS2	0.003	0.001	2.158	0.031
 Residual Variances				
D11	0.018	0.002	7.431	0.000
D21	0.010	0.002	4.763	0.000
D12	0.015	0.002	6.960	0.000
D22	0.015	0.002	7.981	0.000
D13	0.010	0.002	4.711	0.000
D23	0.008	0.003	2.348	0.019

Anhand der vollstandardisierten Modellschätzungen [STANDARDIZED MODEL RESULTS (STDYX Standardization)] zeigt sich, dass die Variablen sehr hohe standardisierte Ladungen auf den State-Faktoren ($\geq 0,897$) und vergleichsweise geringe standardisierte Ladungen auf dem indikatorsspezifischen Faktor (größte standardisierte Ladung = 0,308) aufweisen (die standardisierten Ladungen können hier wiederum als geschätzte Korrelationen zwischen den Indikatoren und den Faktoren interpretiert werden, da die State-Faktoren mit dem indikatorsspezifischen Faktor unkorreliert sind). Anhand der hohen State- und der geringen indikatorsspezifischen Ladungen kann ersehen werden, dass die Indikatoren reliabel und eher homogen sind (d.h. dass indikatorsspezifische Effekte, wie oben bereits erwähnt, hier zwar statistisch signifikant sind, von der Größe her jedoch eine eher untergeordnete Rolle spielen).

Der standardisierten Lösung können wir darüber hinaus die Korrelationen der State-Faktoren untereinander entnehmen. Es zeigt sich, dass die drei State-Variablen recht hoch miteinander korreliert sind ($0,61 < \hat{r} < 0,74$), was auf eine vergleichsweise hohe zeitliche

Stabilität interindividueller Unterschiede hinsichtlich der Depressivität schließen lässt. Zeitlich „benachbarte“ State-Faktoren (*State 1* und *State 2* sowie *State 2* und *State 3*) sind dabei höher miteinander korreliert als die zeitlich weiter voneinander entfernten Faktoren *State 1* und *State 3*. Dies weist auf das Vorhandensein einer autoregressiven Struktur hin (vgl. Kapitel 4.3).

Von Interesse sind außerdem die geschätzten R^2 -Werte für die beobachteten Variablen, die als letztes unter der standardisierten Lösung berichtet werden. Diese Werte können als Reliabilitäts-Schätzungen für die beobachteten Variablen verwendet werden. Wie aufgrund der hohen standardisierten Ladungen bereits erkennbar war, weisen alle sechs Variablen hohe Reliabilitäten (*Rel*) auf ($0,82 < \text{Rel} < 0,91$). Das zeigt, dass die Messgenauigkeit der verwendeten Depressions-Indikatoren (hier: Fragebogenskalen) hoch ist.

STANDARDIZED MODEL RESULTS (STDYX Standardization)					
		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
STATE1	BY				
D11		0.907	0.014	65.166	0.000
D21		0.926	0.014	68.362	0.000
STATE2	BY				
D12		0.926	0.012	79.024	0.000
D22		0.897	0.013	70.943	0.000
STATE3	BY				
D13		0.944	0.013	75.501	0.000
D23		0.902	0.014	66.504	0.000
IS2	BY				
D21		0.176	0.041	4.315	0.000
D22		0.180	0.042	4.273	0.000
D23		0.308	0.064	4.847	0.000
IS2	WITH				
STATE1		0.000	0.000	999.000	999.000
STATE2		0.000	0.000	999.000	999.000
STATE3		0.000	0.000	999.000	999.000
STATE2	WITH				
STATE1		0.729	0.023	31.693	0.000
STATE3	WITH				
STATE1		0.606	0.029	20.985	0.000
STATE2		0.737	0.022	33.422	0.000

Intercepts				
D11	0.976	0.049	20.001	0.000
D21	0.939	0.048	19.493	0.000
D12	0.870	0.047	18.455	0.000
D22	0.850	0.047	18.134	0.000
D13	0.920	0.048	19.207	0.000
D23	0.875	0.047	18.546	0.000
Variances				
STATE1	1.000	0.000	999.000	999.000
STATE2	1.000	0.000	999.000	999.000
STATE3	1.000	0.000	999.000	999.000
IS2	1.000	0.000	999.000	999.000
Residual Variances				
D11	0.178	0.025	7.037	0.000
D21	0.112	0.024	4.613	0.000
D12	0.143	0.022	6.601	0.000
D22	0.163	0.022	7.466	0.000
D13	0.108	0.024	4.584	0.000
D23	0.092	0.040	2.320	0.020
R-SQUARE				
Observed Variable		Estimate	S.E.	Two-Tailed P-Value
D11		0.822	0.025	32.583 0.000
D21		0.888	0.024	36.742 0.000
D12		0.857	0.022	39.512 0.000
D22		0.837	0.022	38.450 0.000
D13		0.892	0.024	37.751 0.000
D23		0.908	0.040	22.806 0.000

Das von uns analysierte LS-Modell mit einem indikatorspezifischen Faktor für den zweiten Indikator zeigt – im Gegensatz zu dem einfacheren Modell ohne indikatorspezifischen Faktor bzw. dem Modell mit korrelierten Fehlervariablen – einen sehr guten Datenfit. Es bietet sich daher im vorliegenden Fall an, dieses Modell als Basismodell für Messinvarianztests über die Zeit und sonstige weitergehende Modellierungen heranzuziehen (wären keine signifikanten indikatorspezifischen Effekte nachweisbar gewesen, so hätte man das sparsamere Modell ohne indikatorspezifischen Faktor bevorzugt).

4.1.4 Überprüfung der Messinvarianz über die Zeit

In den meisten Längsschnittstudien kommt der Frage nach der Vergleichbarkeit der einzelnen Messungen über die Zeit (Messkonstanz, Messäquivalenz) eine entscheidende Bedeutung zu. Es muss sichergestellt werden, dass sich die psychometrischen Eigenschaften der wiederholt erhobenen Messinstrumente (Indikatoren) nicht verändert haben, damit die gemessenen latenten Variablen über die Zeit hinweg vergleichbar sind. Insbesondere müs-

sen wir in unserem LS-Modell für Depression sicherstellen, dass unsere State-Faktoren zu jedem Messzeitpunkt eine vergleichbare Messstruktur aufweisen und damit vergleichbare Konstrukte repräsentieren. Andernfalls wäre es fragwürdig, ob man die Korrelationen der State-Faktoren als Indikatoren für die Stabilität interindividueller Unterschiede auf ein- und demselben Konstrukt (Depressivität) interpretieren kann.

Strukturgleichungsmodelle, die auf mehreren Indikatoren pro Messzeitpunkt basieren (wie das LS-Modell), haben den Vorteil, dass Annahmen bezüglich der Messinvarianz über die Zeit mit Hilfe von Modelltests bzw. Modellfitindices empirisch überprüft werden können. Konkret bezieht sich die Frage nach der Messinvarianz in Strukturgleichungsmodellen auf die Verknüpfung der Indikatoren mit den latenten Variablen, also auf die zeitliche Konstanz (Gleichheit) von Parametern des Messmodells (Faktorladungen, Intercepts und Messfehlervarianzen).

Folglich kann das Vorliegen bzw. der Grad an Messinvarianz mit Hilfe von *Gleichheitsrestriktionen* getestet werden, die man auf diese Parameter legen kann. Eine Gleichheitsrestriktion ist eine Restriktion, mit der man erzwingt, dass ein Parameter mit einem anderen Parameter gleichgesetzt wird, so dass beide Parameter auf exakt denselben Wert geschätzt werden müssen. Solche Restriktionen sind in Mplus einfach zu implementieren (s.u.).

Das Vorliegen eines Mindestgrades an Messinvarianz ist in der Regel erwünscht, da ansonsten bestimmte Vergleiche über die Zeit (z.B. latente Mittelwertsvergleiche) nicht gerechtfertigt (oder zumindest schwer interpretierbar) sein können.

Box 4.3: Messinvarianzbedingungen nach Widaman & Reise (1997)

Widaman und Reise (1997; siehe auch Meredith, 1993; Meredith & Horn, 2001) unterscheiden folgende Grade an Messinvarianz:

- *Konfigurale Invarianz (configural invariance)* ist die schwächste Form von Invarianz. Diese liegt bereits dann vor, wenn allein die Faktorenstruktur (Anzahl der Faktoren und das Ladungsmuster) über die Zeit hinweg unverändert bleibt. Es müssen keine spezifischen Parameter über die Zeit hinweg exakt gleich bleiben, um dieses Kriterium zu erfüllen.
- Man spricht dagegen von *schwacher faktorieller Invarianz (weak factorial invariance)*, wenn zusätzlich zu konfiguraler Invarianz die Faktorladungen λ_{ik} über die Zeit konstant bleiben (d.h. zu jeder Messgelegenheit gleiche Werte annehmen).
- Die Bedingung der *starken faktoriellen Invarianz (strong factorial invariance)* ist erfüllt, wenn darüber hinaus auch die Intercepts α_{ik} über die Zeit gleich bleiben.
- *Strikte faktorielle Invarianz (strict factorial invariance)* bedeutet, dass zusätzlich zur Gleichheit der Ladungen und Intercepts auch die Messfehlervarianzen $Var(\varepsilon_{ik})$ invariant sind.

Weitere, noch striktere Invarianzbedingungen wie beispielsweise die Konstanz der State-Faktorvarianzen und -kovarianzen über die Zeit können ebenfalls überprüft werden. Neben der Varianz-/Kovarianzstabilität kann zudem auch die Mittelwertsstabilität der State-Faktoren getestet werden, welche in vielen Studien von Interesse ist. Eine Voraussetzung

für die sinnvolle Überprüfung latenter Mittelwertsveränderungen ist, dass bezüglich der Messstruktur wenigstens starke faktorielle Invarianz angenommen werden kann.

Die Annahme der Messinvarianz ist in der Praxis manchmal nur für einen Teil der Indikatoren haltbar. Man spricht in einem solchen Fall von *partieller Messinvarianz*. Byrne, Shavelson und Muthén (1989) diskutieren die damit verbundenen Probleme im Detail.

Der vorliegende Grad an Messinvarianz (vgl. Box 4.3) wird häufig über Modellvergleiche ermittelt. Dabei werden Modelle mit unterschiedlichen Gleichheitsrestriktionen geschätzt und hinsichtlich ihrer Modellanpassung verglichen. Beispielsweise kann ein konfigurales Invarianzmodell (Modell ohne Gleichheitsrestriktionen) gegen ein Modell mit über die Zeit gleichgesetzten Ladungen (Modell schwacher Invarianz) und/oder gegen ein Modell mit gleichen Ladungen und gleichen Intercepts (Modell starker Invarianz) getestet werden. Anschließend wird das restaktivste Modell (Modell mit dem höchsten Grad an Invarianz) gewählt, das noch einen akzeptablen Datenfit aufweist.

Eine Strategie zur Bestimmung des vorliegenden Grades an Messinvarianz besteht darin, ein LS-Modell schrittweise durch Gleichsetzung von Parametern über die Zeit immer weiter zu restringieren, bis durch eine signifikante Verschlechterung des Fits (z.B. des Chi-Quadrat-Wertes) angezeigt wird, dass bestimmte Invarianzrestriktionen nicht mit den Daten in Einklang stehen. Das Modell mit der größten Anzahl an Invarianzrestriktionen, welches noch einen akzeptablen Datenfit aufweist, wird dann als endgültiges Modell angenommen, da ein solches restringiertes Modell eine bessere Interpretierbarkeit und Vergleichbarkeit bestimmter Parameter über die Zeit ermöglicht und es zudem sparsamer ist (mit weniger freigeschätzten Parametern auskommt) als Modelle mit weniger Restriktionen.

In unserer Anwendung hatten wir in Abschnitt 4.1.3 bereits ein LS-Modell mit konfiguraler Invarianz geschätzt. Wir beginnen unsere Messinvarianzanalyse daher mit dem nächst höheren Grad an Invarianz, dem Modell schwacher faktorieller Invarianz. In diesem Modell werden nur die Faktorladungen, nicht jedoch die Intercepts und Fehlervarianzen über die Zeit hinweg gleichgesetzt.

Allgemein können in Mplus Parameter gleichgesetzt werden, indem man im model-Befehl hinter dem entsprechenden Parameter dieselbe Zahl oder dasselbe Label in Klammern vergibt. Welche Zahl hierbei gewählt wird, ist im Prinzip egal. Es muss jedoch darauf geachtet werden, dass für alle gleichzusetzenden Parameter dieselbe Zahl/dasselbe Label verwendet wird und dass für unterschiedliche Parametertypen unterschiedliche Zahlen/Labels verwendet werden. In unserem Fall müssen wir nur die State-Faktorladungen des zweiten Indikators über die Zeit gleichsetzen, da die Ladungen des ersten Indikators standardmäßig zur Fixierung der Metrik der latenten Variablen zu jedem Messzeitpunkt auf 1 fixiert und somit „automatisch“ invariant gesetzt sind (sie sind zu jedem Zeitpunkt gleich 1).

Wir wählen die Zahl (1), um die State-Ladungen des zweiten Indikators über die Zeit hinweg gleichzusetzen. Wichtig ist zu verstehen, dass die Spezifikation (1) nicht bewirkt, dass der entsprechenden Parameter auf einen Wert nahe 1 geschätzt wird oder dieser gar auf 1 fixiert ist (Fixierungen von Parametern werden in MPlus mit dem @-Befehl vorgenommen):

```

state1 by d11
    d21 (1);
state2 by d12
    d22 (1);
state3 by d13
    d23 (1);

```

Die Zahl (1) bedeutet hier, dass die Ladungen λ_{21} , λ_{22} und λ_{23} der Indikatoren d21, d22 und d23 auf dem jeweiligen State-Faktor auf *denselben* Wert geschätzt werden sollen. Welcher Wert das genau ist, wird von Mplus im Rahmen der Parameterschätzung ermittelt.

Für den indikator spezifischen Faktor (is2) spezifizieren wir ebenfalls invariante Ladungen über die Zeit. Die entsprechenden Restriktionen fallen für diesen Faktor jedoch anders aus, da hier die Ladungen eines einzelnen wiederholt gemessenen Indikators (Variablen d21, d22 und d23) auf einem *zeitunspezifischen* Faktor betroffen sind. Da die erste Ladung auf dem is2-Faktor, also die Ladung der Variable Y_{21} (λ_{IS21}), von Mplus wiederum per Voreinstellung auf 1 fixiert wird, müssen wir zur Gleichsetzung *aller drei* Ladungen die Ladungsparameter λ_{IS22} und λ_{IS23} ebenfalls auf 1 fixieren (bei den State-Ladungen von d21, d22 und d23 war dies nicht notwendig, da keine der State-Ladungen dieser Variable per Voreinstellung auf einen bestimmten Wert fixiert ist). Die Ladungen λ_{IS22} und λ_{IS23} werden somit im model-Befehl beide mit @ auf 1 fixiert:

```

is2 by d21
    d22@1
    d23@1;

```

Der vollständige Mplus-model-Befehl zur Spezifikation des Modells schwacher faktorieller Invarianz ist in Abbildung 4.8 dargestellt.

```

model: state1 by d11
    d21 (1); ! Gleichsetzung der State-Ladungen des 2. Indikators
state2 by d12
    d22 (1);
state3 by d13
    d23 (1);

is2 by d21
    d22@1 ! Fixierung aller Ladungen auf
    d23@1; ! dem indikator spezifischen Faktor auf 1

is2 with state1-state3@0;

```

Abbildung 4.8: Mplus-model-Befehl zur Spezifikation des LS-Modells mit einem indikator spezifischen Faktor (is2) für den zweiten Indikator sowie invarianten Ladungen über die Zeit (Modell schwacher faktorieller Invarianz).

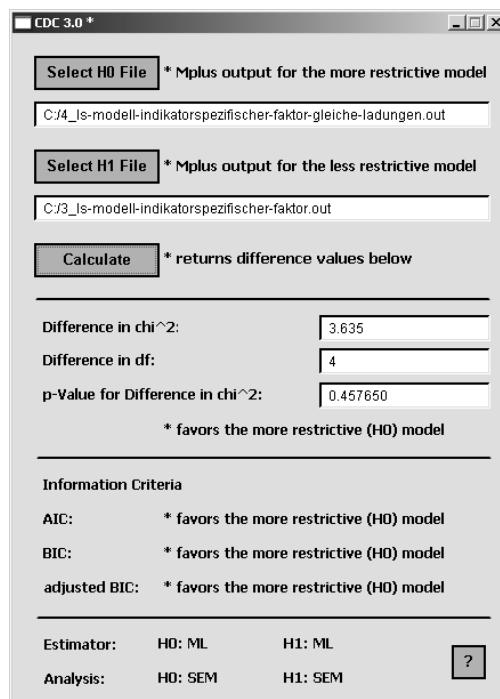
Es zeigt sich anhand der von Mplus berechneten Fitmaße, dass das Modell schwacher faktorieller Invarianz – ebenso wie das Modell konfiguraler Invarianz (vgl. Abschnitt 4.1.3) –

einen sehr guten Fit aufweist [$\chi^2 = 4,87, df = 7, p = 0,68; CFI = 1,00; RMSEA = 0,00; p(RMSEA \leq 0,05) = 0,99; SRMR = 0,005$]. Um zu überprüfen, ob das restriktivere Modell schwächer faktorieller Invarianz signifikant schlechter auf die Daten passt als das Ausgangsmodell mit unrestringierten Faktorladungen, kann ein χ^2 -Differenzentest durchgeführt werden, da die Modelle hierarchisch ineinander geschachtelt sind (vgl. Box 3.7). Man spricht in diesem Zusammenhang auch von „nested models“, da das Modell mit invarianten Ladungen ein Spezialfall des Ausgangsmodells mit unrestringierten Ladungen darstellt.

χ^2 -Differenzentests zwischen verschiedenen geschätzten Modellen werden von Mplus nicht direkt ausgegeben. Sie können jedoch von Hand oder mit einer entsprechenden Spezialsoftware berechnet werden. In Box 4.4 wird beschrieben, wie χ^2 -Differenzentests mit Hilfe der von Crayen (2010) entwickelten Software *Chi-Square Difference Calculator* (CDC) automatisiert durchgeführt werden können. Der CDC (Version 3.0) sowie ein entsprechendes Manual befinden sich auf der Begleit-CD zum Buch.

Box 4.4: χ^2 -Differenzentest mit Crayens (2010) Programm CDC

Mit dem Programm *Mplus Model Comparison Gadget* kann ein Vergleich in Mplus geschätzter Modelle mittels χ^2 -Differenzentest direkt auf Basis der Mplus-Output-Dateien zweier geschachtelter Modelle durchgeführt werden. Hierzu müssen im ersten Schritt die beiden zu vergleichenden Modelle in Mplus geschätzt werden. Anschließend wird die Datei „CDC3.exe“ (siehe Begleit-CD) ausgeführt.



Anschließend teilen wir dem Programm mit, in welchem Verzeichnis auf dem Computer die Mplus-Outputs für das restriktivere Modell (sogenanntes *H0-Modell*) sowie das weniger restriktive (*H1-Modell*) abgelegt sind. Die entsprechenden Datei-Pfade können entweder direkt in den entsprechenden Feldern eingegeben oder aber bequem über die Schaltflächen „Select H0 File“ bzw. „Select H1 File“ ausgewählt werden. Durch Klicken auf „Calculate“ ermittelt das Programm dann den χ^2 -Differenzwert ($\Delta\chi^2$ = „Difference in chi²“) und die Differenz der Freiheitsgrade (Δdf = „Difference in df“) zwischen den beiden Modellen sowie den entsprechenden *p*-Wert für den Differenzentest. Außerdem teilt das Programm dem Benutzer mit, für welches der beiden Modelle die von Mplus berichteten informationstheoretischen Maße *AIC*, *BIC* und sample size adjusted *BIC* jeweils günstiger ausfallen (d.h. für welches Modell der *AIC*-, *BIC*- und sample size adjusted *BIC*-Wert kleiner ist, vgl. Box 3.7). In diesem Beispiel ist das restriktivere Modell nach allen Kriterien zu bevorzugen (vgl. Diskussion im Text).

Wichtig ist, dass das Programm nicht überprüft, ob die Voraussetzungen zur Durchführung des χ^2 -Differenzentests (vgl. Box 3.7) erfüllt sind. Z.B. beinhaltet das Programm keinen Test dafür, ob die vom Benutzer angegebenen Modelle tatsächlich ineinander geschachtelt sind. Dies muss man vor Benutzung des Programms selbst klären. Nicht-geschachtelte Modelle können zwar ggf. mit Hilfe von *AIC*, *BIC* und sample size adjusted *BIC* verglichen werden, nicht jedoch über den χ^2 -Differenzentest.

In unserem Beispiel ergibt sich für den Vergleich des Modells schwacher Invarianz mit dem Modell konfiguraler Invarianz ein χ^2 -Differenzwert von $\Delta\chi^2 = 4,872 - 1,237 = 3,635$. Dieser Wert ist bei $\Delta df = 7 - 3 = 4$ statistisch nicht signifikant ($p = 0,46$; vgl. Box 4.4). Daraus können wir schließen, dass das restringierte Modell (mit gleichen Ladungen über die Zeit) nicht signifikant schlechter auf die Daten passt als das Ausgangsmodell mit unrestringierten Ladungen. Somit muss die Annahme schwacher faktorieller Invarianz über die Zeit aufgrund der vorliegenden Daten nicht verworfen werden. Auch der *AIC*-Index favorisiert das Modell mit gleichen Ladungen (*AIC* = -1262,91) gegenüber dem weniger sparsamen Ausgangsmodell (-1258,54; allerdings ist der Unterschied in den *AIC*-Werten nicht besonders groß).

Box 4.5: Kontrolle der korrekten Umsetzung von Restriktionen

Bei Modellen mit Gleichheits- oder sonstigen spezifischen Restriktionen empfiehlt es sich, die korrekte Umsetzung der gewünschten Restriktionen in Mplus stets sehr gründlich im Output zu überprüfen. Hier können sich – insbesondere bei komplexen Modellen mit vielen Restriktionen – sehr leicht Fehler einschleichen. Man sollte daher für alle Parameter genau kontrollieren, ob diese in der gewünschten Weise gleichgesetzt bzw. nicht gleichgesetzt wurden. Ein unerwartet schlechter Modellfit kann ein Zeichen dafür sein, dass man in der Spezifikation Fehler gemacht hat. Beispielsweise kann es vorkommen, dass man versehentlich dieselbe Zahl in Klammern für Gleichheitsrestriktionen verschiedener Parametergruppen (z.B. Ladungen und Intercepts) vergeben hat, so dass unbeab-

sichtigt völlig unsinnige Gleichheitsrestriktionen (Ladungen = Intercepts) vom Programm vorgenommen wurden, die zu einer sehr schlechten Modellanpassung führen. Aus diesem Grund sollte man niemals auf eine gründliche Kontrolle der korrekten Umsetzung dieser Restriktionen verzichten!

Anhand der unstandardisierten Parameterschätzungen (Rubrik MODEL RESULTS, s.u.) können wir sehen, dass die von uns intendierten Invarianzrestriktionen von Mplus korrekt umgesetzt wurden, da die unstandardisierten State-Ladungen des zweiten Indikators alle auf denselben Wert geschätzt wurden ($\hat{\lambda}_{21} = \hat{\lambda}_{22} = \hat{\lambda}_{23} = 0,924$) und diese auch alle exakt denselben Standardfehler aufweisen ($SE = 0,021$). Sämtliche Ladungen auf dem indikator-spezifischen Faktor wurden wie gewünscht auf 1 fixiert. Dass es sich tatsächlich um fixierte Parameter handelt, können wir daran erkennen, dass für keine der drei unstandardisierten Ladungen auf dem IS-Faktor ein Standardfehler geschätzt wurde.

MODEL RESULTS

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
STATE1	BY				
D11		1.000	0.000	999.000	999.000
D21		0.924	0.021	44.753	0.000
STATE2	BY				
D12		1.000	0.000	999.000	999.000
D22		0.924	0.021	44.753	0.000
STATE3	BY				
D13		1.000	0.000	999.000	999.000
D23		0.924	0.021	44.753	0.000
IS2	BY				
D21		1.000	0.000	999.000	999.000
D22		1.000	0.000	999.000	999.000
D23		1.000	0.000	999.000	999.000
IS2	WITH				
STATE1		0.000	0.000	999.000	999.000
STATE2		0.000	0.000	999.000	999.000
STATE3		0.000	0.000	999.000	999.000
STATE2	WITH				
STATE1		0.064	0.005	13.389	0.000
STATE3	WITH				
STATE1		0.053	0.004	11.939	0.000
STATE2		0.064	0.005	13.679	0.000

Intercepts				
D11	0.314	0.013	24.185	0.000
D21	0.281	0.012	23.430	0.000
D12	0.280	0.013	21.722	0.000
D22	0.259	0.012	21.089	0.000
D13	0.286	0.012	22.951	0.000
D23	0.260	0.012	21.828	0.000
Variances				
STATE1	0.088	0.006	15.380	0.000
STATE2	0.088	0.006	15.447	0.000
STATE3	0.085	0.005	15.651	0.000
IS2	0.004	0.001	5.660	0.000
Residual Variances				
D11	0.017	0.002	7.850	0.000
D21	0.010	0.002	5.872	0.000
D12	0.015	0.002	7.826	0.000
D22	0.014	0.002	8.575	0.000
D13	0.011	0.002	6.095	0.000
D23	0.011	0.002	6.970	0.000

Man beachte, dass die *standardisierten* Ladungen jeweils *nicht* invariant sind, sondern über die Zeit leicht variieren. Dies liegt daran, dass wir keine Restriktionen auf die State-Faktor- und Fehlervarianzen gelegt haben:

STANDARDIZED MODEL RESULTS (STDYX Standardization)					
		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
STATE1	BY				
	D11	0.916	0.011	81.902	0.000
	D21	0.916	0.012	79.304	0.000
STATE2	BY				
	D12	0.925	0.010	90.745	0.000
	D22	0.897	0.011	80.620	0.000
STATE3	BY				
	D13	0.939	0.010	90.725	0.000
	D23	0.908	0.011	80.840	0.000
IS2	BY				
	D21	0.220	0.020	10.940	0.000
	D22	0.215	0.020	10.930	0.000
	D23	0.221	0.020	10.949	0.000

IS2 WITH				
STATE1	0.000	0.000	999.000	999.000
STATE2	0.000	0.000	999.000	999.000
STATE3	0.000	0.000	999.000	999.000
STATE2 WITH				
STATE1	0.726	0.023	31.694	0.000
STATE3 WITH				
STATE1	0.609	0.029	21.211	0.000
STATE2	0.739	0.022	33.715	0.000
Intercepts				
D11	0.971	0.048	20.115	0.000
D21	0.941	0.048	19.608	0.000
D12	0.872	0.047	18.608	0.000
D22	0.847	0.047	18.198	0.000
D13	0.922	0.048	19.320	0.000
D23	0.877	0.047	18.665	0.000
Variances				
STATE1	1.000	0.000	999.000	999.000
STATE2	1.000	0.000	999.000	999.000
STATE3	1.000	0.000	999.000	999.000
IS2	1.000	0.000	999.000	999.000
Residual Variances				
D11	0.161	0.020	7.851	0.000
D21	0.112	0.019	5.824	0.000
D12	0.145	0.019	7.672	0.000
D22	0.150	0.018	8.349	0.000
D13	0.118	0.019	6.047	0.000
D23	0.127	0.018	6.887	0.000
R-SQUARE				
Observed Variable		Estimate	S.E.	Two-Tailed P-Value
D11		0.839	0.020	40.951 0.000
D21		0.888	0.019	46.063 0.000
D12		0.855	0.019	45.373 0.000
D22		0.850	0.018	47.472 0.000
D13		0.882	0.019	45.362 0.000
D23		0.873	0.018	47.483 0.000

Im nächsten Schritt testen wir ein Modell starker faktorieller Invarianz. In diesem Modell werden zusätzlich zu den Ladungen auch die Intercepts aller Indikatoren über die Zeit hinweg gleichgesetzt. Da wir Unterschiede in den latenten State-Mittelwerten zulassen und

analysieren wollen, müssen wir dabei zunächst die Mplus-Standardeinstellung umgehen, gemäß welcher die Mittelwerte aller latenten Variablen auf Null fixiert werden (vgl. Box 4.2). Ansonsten wäre unsere Spezifikation mit gleichgesetzten Intercepts der Indikatoren in Verbindung mit jeweils auf Null fixierten latenten Mittelwerten sehr restriktiv, da sie gleichbedeutend mit der Annahme ist, dass keinerlei Mittelwertsunterschiede über die Zeit hinweg zu verzeichnen sind.

Zudem müssen wir uns überlegen, wie wir die Mittelwerte der latenten State-Faktoren identifizieren können. Hierzu ist es nützlich, sich noch einmal die Formel anzuschauen, gemäß welcher die Mittelwerte der beobachteten Variablen, $E(Y_{ik})$, im Modell vorhergesagt werden (vgl. Box 4.2):

$$\begin{aligned} E(Y_{ik}) &= E(\alpha_{ik}) + E(\lambda_{ik} \cdot \text{State } k) \\ &= \alpha_{ik} + \lambda_{ik} \cdot E(\text{State } k). \end{aligned}$$

Aus dieser Formel ist ersichtlich, dass die Mittelwerte der latenten State-Faktoren $E(\text{State } k)$ relativ einfach über die Mittelwerte der ersten beobachteten Variablen Y_{1k} identifizierbar sind. Wenn wir die Intercept-Terme α_{1k} dieser Variablen auf Null fixieren, entspricht der latente Mittelwert dem Mittelwert der ersten beobachteten Variable:

$$\begin{aligned} E(Y_{1k}) &= \alpha_{1k} + \lambda_{1k} \cdot E(\text{State } k) \\ &= 0 + 1 \cdot E(\text{State } k) \\ &= E(\text{State } k). \end{aligned}$$

Somit sind die latenten State-Mittelwerte $E(\text{State } k)$ unter diesen Bedingungen direkt über die Mittelwerte der beobachteten Variablen $E(Y_{1k})$ identifiziert und schätzbar. Die Intercepts der restlichen Indikatoren werden nach dem gleichen Prinzip wie die Faktorladungen über die Zeit hinweg gleichgesetzt.

Die entsprechende Mplus-Spezifikation (nur das model1-Statement, der restliche Input ist unverändert) wird in Abbildung 4.9 demonstriert. Zu beachten ist hier, dass in Spezifikationen, die sich auf Intercepts bzw. Mittelwerte beziehen, die Variablennamen in eckigen Klammern erscheinen. Ob sich ein Befehl in eckigen Klammern auf Intercepts oder auf Mittelwerte bezieht, hängt davon ab, ob die Variable im Modell exogen (d.h. im Modell ausschließlich unabhängige Variable) oder endogen (d.h. im Modell durch andere Variablen beeinflusst) ist.

Abbildung 4.9 zeigt, wie die Intercepts der ersten Indikatoren auf Null fixiert, die State-Mittelwerte angefordert und die restlichen Intercepts α_{2k} (der zweiten Indikatoren Y_{2k}) über die Zeit hinweg gleichgesetzt werden. Man beachte, dass für die Gleichsetzung der Intercepts die Zahl (2) verwendet wurde. Die Zahl (1) dient nach wie vor zur Gleichsetzung der Faktorladungen des zweiten Indikators. Der Mittelwert des indikatorspezifischen Faktors wird nicht gesondert aufgelistet. Dies hat zur Folge, dass er gemäß der Mplus-Voreinstellung automatisch auf Null fixiert wird (was nach wie vor erwünscht und korrekt ist).

```

model: state1 by d11
        d21 (1); ! Gleichsetzung der State-Ladungen des 2. Indikators
state2 by d12
        d22 (1);
state3 by d13
        d23 (1);

is2 by d21
        d22@1 ! Fixierung aller Ladungen auf
        d23@1; ! dem indikatorspezifischen Faktor auf 1

is2 with state1-state3@0;

[d11@0 d12@0 d13@0]; ! Fixierung der Intercepts des ersten Indikators auf 0

[state1-state3]; ! Schätzung der Mittelwerte der latenten State-Faktoren

[d21 d22 d23] (2); ! Gleichsetzung der Intercepts des 2. Indikators

```

Abbildung 4.9: Mplus-model-Befehl zur Spezifikation des LS-Modells mit einem indikatorspezifischen Faktor (is2) für den zweiten Indikator sowie invarianten Ladungen und Intercepts über die Zeit (Modell starker faktorieller Invarianz).

Der entsprechende Mplus-Output zeigt, dass auch das Modell starker faktorieller Invarianz sehr gut auf die Daten passt ($\chi^2 = 5,95, df = 9, p = 0,75; CFI = 1,00; RMSEA = 0,00; SRMR = 0,006$). Der χ^2 -Differenzwert zwischen diesem Modell und dem Modell schwacher faktorieller Invarianz fällt nicht signifikant aus ($\Delta\chi^2 = 1,073, \Delta df = 2, p = 0,58$). Auch beim Vergleich der AIC-Werte schneidet das Modell starker Invarianz ($AIC = -1265,84$) besser ab als das Modell schwacher Invarianz ($AIC = -1262,90$). Daraus können wir schließen, dass auch die Annahme starker faktorieller Invarianz der Depressionsskalen für die vorliegenden Beispieldaten nicht verworfen werden muss.

Anhand der unstandardisierten Parameterschätzwerte (MODEL RESULTS, s.u.) können wir erkennen, dass unsere Gleichheitsrestriktionen wiederum korrekt umgesetzt wurden ($\hat{\lambda}_{21} = \hat{\lambda}_{22} = \hat{\lambda}_{23} = 0,923; \hat{\alpha}_{21} = \hat{\alpha}_{22} = \hat{\alpha}_{23} = -0,004$). Auch die Restriktion zur Fixierung der Intercepts des ersten Indikators auf Null wurde von Mplus implementiert (man erkennt, dass für diese drei Intercepts jetzt keine Standardfehler mehr berichtet werden, da es sich nunmehr um fixierte Parameter handelt).

Zusätzlich erhalten wir nun auch eine Rubrik MEANS, unter welcher wir die geschätzten latenten State-Mittelwerte ablesen können. Es ist erkennbar, dass sich die latenten Depressions-Mittelwerte über die Zeit kaum verändert haben. Das mittlere Depressivitätsniveau der Kinder ist nahezu konstant geblieben ist ($M_1 = 0,31, SD_1 = 0,30; M_2 = 0,28, SD_2 = 0,30; M_3 = 0,29, SD_3 = 0,29$). Ein formaler Test auf Gleichheit der Mittelwerte in der Population wird in Box 4.6 gezeigt.

MODEL RESULTS

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
STATE1	BY				
D11		1.000	0.000	999.000	999.000
D21		0.923	0.021	44.841	0.000
STATE2	BY				
D12		1.000	0.000	999.000	999.000
D22		0.923	0.021	44.841	0.000
STATE3	BY				
D13		1.000	0.000	999.000	999.000
D23		0.923	0.021	44.841	0.000
IS2	BY				
D21		1.000	0.000	999.000	999.000
D22		1.000	0.000	999.000	999.000
D23		1.000	0.000	999.000	999.000
IS2	WITH				
STATE1		0.000	0.000	999.000	999.000
STATE2		0.000	0.000	999.000	999.000
STATE3		0.000	0.000	999.000	999.000
STATE2	WITH				
STATE1		0.064	0.005	13.393	0.000
STATE3	WITH				
STATE1		0.053	0.004	11.940	0.000
STATE2		0.064	0.005	13.683	0.000
Means					
STATE1		0.311	0.013	24.748	0.000
STATE2		0.282	0.013	22.452	0.000
STATE3		0.286	0.012	23.349	0.000
Intercepts					
D11		0.000	0.000	999.000	999.000
D21		-0.004	0.007	-0.545	0.586
D12		0.000	0.000	999.000	999.000
D22		-0.004	0.007	-0.545	0.586
D13		0.000	0.000	999.000	999.000
D23		-0.004	0.007	-0.545	0.586
Variances					
STATE1		0.088	0.006	15.384	0.000
STATE2		0.088	0.006	15.451	0.000
STATE3		0.085	0.005	15.657	0.000
IS2		0.004	0.001	5.671	0.000

Residual Variances				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
STATE1 BY				
D11	0.017	0.002	7.828	0.000
D21	0.010	0.002	5.907	0.000
D12	0.015	0.002	7.807	0.000
D22	0.014	0.002	8.601	0.000
D13	0.011	0.002	6.063	0.000
D23	0.011	0.002	7.001	0.000
STANDARDIZED MODEL RESULTS (STDYX Standardization)				
IS2 WITH				
STATE1	0.000	0.000	999.000	999.000
STATE2	0.000	0.000	999.000	999.000
STATE3	0.000	0.000	999.000	999.000
STATE2 WITH				
STATE1	0.726	0.023	31.683	0.000
STATE3 WITH				
STATE1	0.609	0.029	21.199	0.000
STATE2	0.739	0.022	33.700	0.000
Means				
STATE1	1.050	0.054	19.335	0.000
STATE2	0.951	0.052	18.144	0.000
STATE3	0.981	0.052	18.705	0.000
Intercepts				
D11	0.000	0.000	999.000	999.000
D21	-0.014	0.025	-0.546	0.585
D12	0.000	0.000	999.000	999.000
D22	-0.013	0.024	-0.546	0.585
D13	0.000	0.000	999.000	999.000
D23	-0.014	0.025	-0.546	0.585

Variances				
STATE1	1.000	0.000	999.000	999.000
STATE2	1.000	0.000	999.000	999.000
STATE3	1.000	0.000	999.000	999.000
IS2	1.000	0.000	999.000	999.000
Residual Variances				
D11	0.160	0.020	7.829	0.000
D21	0.113	0.019	5.861	0.000
D12	0.144	0.019	7.656	0.000
D22	0.150	0.018	8.377	0.000
D13	0.117	0.019	6.016	0.000
D23	0.127	0.018	6.919	0.000
R-SQUARE				
Observed				
Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
D11	0.840	0.020	40.984	0.000
D21	0.887	0.019	46.007	0.000
D12	0.856	0.019	45.406	0.000
D22	0.850	0.018	47.396	0.000
D13	0.883	0.019	45.423	0.000
D23	0.873	0.018	47.442	0.000

Box 4.6: Test auf latente Mittelwertsunterschiede über die Zeit

Inhaltlich interessant ist an dieser Stelle die Frage, ob sich der Modellfit signifikant verschlechtert, wenn man auch die Mittelwerte der drei latenten State-Faktoren über die Zeit hinweg gleichsetzt. Wenn diese zusätzliche Restriktion nicht zu einer Verschlechterung des Modelfits führen würde, könnte man schließen, dass sich die mittlere Depressivität der Kinder über die Zeit nicht bedeutsam verändert hat. Das Gleichsetzen der State-Mittelwerte geschieht über folgende Ergänzung im Mplus model-Statement (der vollständige Input zu diesem Modell befindet sich auf der Begleit-CD):

```
[state1-state3] (3);
```

Das Modell mit gleichgesetzten latenten Mittelwerten passt im Vergleich zum Modell mit frei geschätzten latenten Mittelwerten signifikant schlechter ($\Delta\chi^2 = 8,36$, $\Delta df = 2$, $p = 0,02$). Daraus würde man schließen, dass tatsächlich eine signifikante Abnahme in den mittleren Depressionswerten zu verzeichnen ist. Der AIC-Index favorisiert ebenfalls das Modell mit ungleichen Mittelwerten ($AIC = -1265,83$) gegenüber dem Modell gleicher Mittelwerte ($AIC = -1261,48$), wobei die Unterschiede in den AIC-Werten allerdings nicht sehr groß sind.

Wenn man den absoluten Fit des restriktiveren Modells mit gleichen Mittelwerten (siehe Output auf der Begleit-CD) betrachtet, so stellt man fest, dass dieser immer noch als sehr gut zu bewerten ist ($\chi^2 = 14,31$, $df = 11$, $p = 0,22$; $CFI = 0,999$; $RMSEA = 0,02$; $SRMR = 0,020$). Dies deutet darauf hin, dass die gefundenen Unterschiede in den Mittelwerten eher ge-

ring sind. Dafür spricht auch, dass im weniger restriktierten Modell, in welchem die Mittelwerte frei geschätzt wurden, die Mittelwertsunterschiede über die Zeit als eher gering einzustufen sind (s.o.). In der Praxis müsste der Benutzer aufgrund von inhaltlichen Gesichtspunkten sowie der Betrachtung der praktischen Bedeutsamkeit der Größe der Mittelwertsunterschiede entscheiden, ob das Modell mit ungleichen oder das Modell mit gleichen Mittelwerten über die Zeit zu bevorzugen ist.

Schließlich testen wir noch ein Modell strikter faktorieller Invarianz. In diesem Modell werden zusätzlich zu den Ladungen und Intercepts auch noch die Fehlervarianzen $Var(\varepsilon_{ik})$ über die Zeit hinweg gleichgesetzt (wir gehen hierbei vom Modell mit ungleichen Mittelwerten über die Zeit aus). Die Zahl (3) wird für die Gleichsetzung der Fehlervarianzen des ersten Indikators verwendet. Die Zahl (4) dient zur Gleichsetzung der Fehlervarianzen des zweiten Indikators. Das vollständige model-Statement zur Spezifikation des Modells strikter faktorieller Invarianz ist in Abbildung 4.10 dargestellt.

```
model: state1 by d11
        d21 (1); ! Gleichsetzung der State-Ladungen des 1. Indikators
state2 by d12
        d22 (1);
state3 by d13
        d23 (1);

is2 by d21
        d22@1 ! Fixierung aller Ladungen auf
        d23@1; ! dem indikatorsspezifischen Faktor auf 1

is2 with state1-state3@0;

[d11@0 d12@0 d13@0]; ! Fixierung der Intercepts des ersten Indikators auf 0

[state1-state3]; ! Schätzung der Mittelwerte der latenten State-Faktoren

[d21 d22 d23] (2); ! Gleichsetzung der Intercepts des 2. Indikators

d11 d12 d13 (3); ! Gleichsetzung der Fehlervarianzen des 1. Indikators
d21 d22 d23 (4); ! Gleichsetzung der Fehlervarianzen des 2. Indikators
```

Abbildung 4.10: Mplus-model-Befehl zur Spezifikation des LS-Modells mit einem indikatorsspezifischen Faktor (is2) für den zweiten Indikator sowie invarianten Ladungen, Intercepts und Residualvarianzen über die Zeit (Modell strikter faktorieller Invarianz).

Das Modell strikter faktorieller Invarianz bei frei geschätzten latenten Mittelwerten zeigt ebenfalls einen sehr guten absoluten Fit ($\chi^2 = 16,36$, $df = 13$, $p = 0,23$; $CFI = 0,999$; $RMSEA = 0,02$; $SRMR = 0,008$; der vollständige Output zu diesem Modell ist auf der Begleit-CD enthalten). Im Vergleich zum Modell starker faktorieller Invarianz mit unrestringierten latenten Mittelwerten ergibt sich allerdings ein χ^2 -Differenzwert von $\Delta\chi^2 = 10,417$ bei $\Delta df = 4$. Dieser Wert ist auf dem 5%-Niveau signifikant ($p = 0,034$). Mithin wäre nach diesem Kriterium das weniger restriktive Modell starker Invarianz (mit gleichen Ladungen und Intercepts aber ungleichen Fehlervarianzen) zu bevorzugen. Der AIC-Index ist ebenfalls geringfügig günstiger für das Modell starker Invarianz ($AIC = -1265,84$ versus $-1263,42$).

Allerdings ist der absolute Fit des Modells strikter Invarianz wie oben gesehen als sehr gut zu bewerten, und der χ^2 -Differenzentest sowie der AIC-Index weisen nur auf geringe Unterschiede zwischen den Modellen hin. Somit wäre es auch vertretbar (und vermutlich sinnvoller), das sparsamere Modell strikter Invarianz dem Modell starker Invarianz vorzuziehen.

An dieser Stelle beenden wir die Serie von Messinvarianztests für das LS-Modell. Es sei darauf hingewiesen, dass nach dem gleichen Prinzip zusätzliche Invarianztests für die State-Faktorvarianzen und -kovarianzen durchgeführt werden könnten. Insgesamt haben die umfangreichen Analysen des LS-Modells eine Reihe von interessanten inhaltlichen Erkenntnissen geliefert, die wir an dieser Stelle noch einmal kurz zusammenfassen wollen:

- Die verwendeten Skalen zur Erfassung der Depressivität weisen eine hohe Messgenauigkeit (Reliabilität) und geringe Indikatorsspezifität zu allen Messzeitpunkten auf.
- Die Parameter der Messmodelle (und damit die Messeigenschaften der Skalen) haben sich über die Zeit nicht bedeutsam verändert, es liegt starke bzw. strikte Messinvarianz vor.
- Messfehlerbereinigte („wahre“) interindividuelle Unterschiede hinsichtlich der Depressivität sind bei den hier untersuchten Kindern relativ stabil über die Zeit.
- Die Korrelationsstruktur der latenten State-Faktoren weist auf einen autoregressiven Effekt hin.
- Mittelwertsunterschiede über die Zeit sind nur gering, wobei eine leichte Tendenz einer Abnahme der mittleren Depressivität von T1 zu T2 erkennbar ist.

Das LS-Modell ist wie bereits erwähnt ein nützliches Basismodell, wenn es um die Auswertung von Längsschnittdaten mit Hilfe von Strukturgleichungsmodellen geht und zu jedem Messzeitpunkt (möglichst identische) multiple Indikatoren vorliegen. Im folgenden Abschnitt wird die Erweiterung des LS-Modells zu einem LST-Modell anhand des gleichen Datenbeispiels besprochen.

4.2 Latent-State-Trait-Analyse

Eine Einschränkung des LS-Modells besteht darin, dass das Ausmaß an Messgelegenheits-Spezifität bzw. Stabilität interindividueller Unterschiede zwar indirekt über die Höhe der Interkorrelationen der latenten State-Variablen abgeschätzt werden kann, eine direkte Zerlegung der Varianz der beobachteten Variablen in stabile, messgelegenheitsspezifische und Messfehlereinflüsse jedoch in diesem Modell nicht möglich ist (vgl. Steyer, 1988). Die Erweiterung des LS-Modells zu einem Latent-State-Trait-(LST-)Modell kann verwendet werden, wenn man an einer derartigen Zerlegung interessiert ist.

Im LST-Modell werden die beobachteten Variablen Y_{ik} (i = Indikator, k = Messgelegenheit) wie im LS-Modell zerlegt in ein Intercept α_{ik} , einen messgelegenheitsspezifischen latenten State-Faktor (*State k*) und eine messgelegenheitsspezifische Messfehlervariable ε_{ik} :

$$Y_{ik} = \alpha_{ik} + \lambda_{ik} \cdot (\text{State } k) + \varepsilon_{ik},$$

wobei λ_{ik} wiederum die Faktorladung von Y_{ik} auf dem State-Faktor bezeichnet.

Im Unterschied zum LS-Modell werden im LST-Modell die latenten State-Faktoren ihrerseits weiter zerlegt in einen latenten Trait-Faktor, welcher den Einfluss zeitlich stabiler Persönlichkeitsdispositionen charakterisiert und eine messgelegenheitsspezifische latente Residualvariable ζ_k , welche situationsspezifische Abweichungen vom durch den stabilen Trait vorhergesagten latenten Zustandswert (State-Score) einer Person zur Messgelegenheit k kennzeichnet (vgl. Steyer et al., 1992; 1999):

$$\text{State } k = \gamma_{0k} + \gamma_k \cdot \text{Trait} + \zeta_k,$$

wobei γ_{0k} einen latenten Intercept-Term und γ_k die Faktorladung 2. Ordnung auf dem Traitfaktor bezeichnet.

Wir hatten gesehen, dass die Kovarianzen der State-Faktoren im LS-Modell ein Maß für die Stabilität interindividueller Unterschiede über die Zeit darstellen (vgl. Abschnitt 4.1.2). Im LST-Modell wird nun die Kovarianzstruktur der State-Faktoren durch eine eindimensionale Struktur erklärt, d.h. die Zusammenhänge zwischen den State-Faktoren werden auf einen einzelnen Trait-Faktor zurückgeführt. Für $k > 3$ ist das LST-Modell daher restriktiver als das LS-Modell. Der Trait-Faktor bildet die zeitlich stabilen Varianzanteile der State-Faktoren ab.

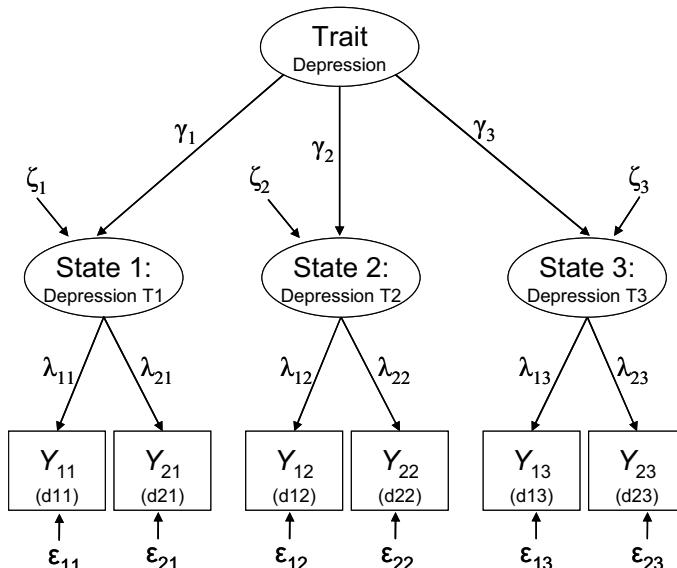


Abbildung 4.11: Latent-State-Trait-Modell für zwei beobachtete Variablen Y_{ik} (i = Indikator, k = Messgelegenheit), welche zu drei Messgelegenheiten erhoben wurden. ε_{ik} : Messfehlervariable; ζ_k : latente State-Residualvariablen; λ_{ik} , γ_k : Faktorladungsparameter. In Klammern sind die Mplus-Variablennamen dargestellt.

Abbildung 4.11 zeigt ein LST-Modell für zwei beobachtete Variablen Y_{1k} und Y_{2k} , welche zu drei Messgelegenheiten ($k = 1, 2, 3$) erhoben wurden. Es zeigt sich, dass das LST-Modell ein Faktorenmodell 2. Ordnung impliziert (vgl. Abschnitt 3.4.2)

Als Beispiel spezifizieren wir ein LST-Modell für die Depressionsdaten (Datensatz **depression.dat**), die wir bereits zur Illustration der LS-Analysen (Abschnitt 4.1.2) verwendet hatten. Ausgangspunkt ist das LS-Modell ohne Invarianzrestriktionen mit einem indikator-spezifischen Faktor für den zweiten Indikator, welches einen guten Fit aufwies (vgl. Abschnitt 4.1.3). Die Erweiterung dieses Modells zu einem LST-Modell wird in Abbildung 4.12 gezeigt. Ein ähnliches LST-Modell wird ausführlich von Eid et al. (1999) diskutiert.

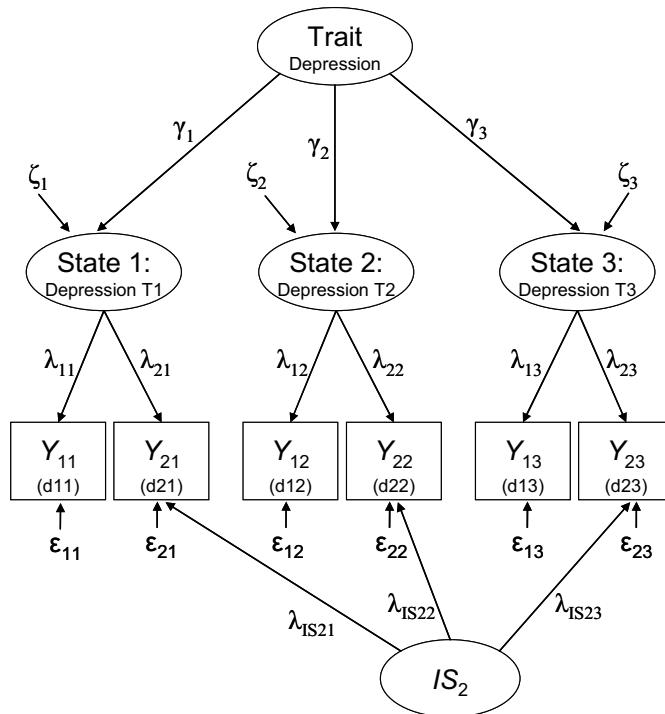


Abbildung 4.12: Latent-State-Trait-Modell für zwei beobachtete Variablen Y_{ik} (i = Indikator, k = Messgelegenheit), welche zu drei Messgelegenheiten erhoben wurden. ϵ_{ik} : Messfehlervariable; ζ_k : latente State-Residualvariablen; λ_{ik} , γ_k : Faktorladungsparameter. Das Modell enthält zusätzlich einen indikator-spezifischen Faktor (IS_2) für den zweiten wiederholt gemessenen Indikator (Y_{2k}) zur Modellierung indikator-spezifischer Effekte über die Zeit. Die Parameter λ_{ISik} bezeichnen die Faktorladungen auf dem indikator-spezifischen Faktor IS_2 . In Klammern sind die Mplus-Variablennamen dargestellt.

Das entsprechende Mplus-model-Statement ist in Abbildung 4.13 dargestellt. Zu beachten ist, dass die Identifikation der latenten Mittelwertsstruktur beim LST-Modell etwas komplexer ist, als dies beim LS-Modell der Fall war, jedenfalls dann, wenn man an der Schätzung

des latenten Trait-Faktormittelwertes interessiert ist (zur Analyse der latenten Mittelwertsstruktur in LST-Modellen siehe Box 4.7).

Will man den Mittelwert des latenten Trait-Faktors sowie die Intercepts γ_{0k} der latenten State-Faktoren identifizieren, so bietet es sich an, wie beim LS-Modell zunächst die Intercepts des ersten Indikators (dessen State-Ladungen von Mplus zu jedem Messzeitpunkt automatisch auf 1 fixiert wird) zu jedem Messzeitpunkt auf Null zu setzen, d.h. $\alpha_{11} = \alpha_{12} = \alpha_{13} = 0$.

Im nächsten Schritt wird der Intercept-Term (γ_{01}) des ersten State-Faktors (*State 1*) ebenfalls auf Null fixiert, d.h. $\gamma_{01} = 0$. Der Faktor *State 1* wird aus Gründen der Einfachheit gewählt, weil er der erste Indikator des Trait-Faktors ist und somit seine Ladung (γ_1) auf dem Trait-Faktor von Mplus automatisch auf 1 fixiert wird (d.h., $\gamma_1 = 1$). Die Konsequenz ist, dass der Trait-Faktormittelwert dann direkt über den Mittelwert der ersten beobachteten Variable (hier: Y_{11} bzw. d_{11}) identifiziert ist und geschätzt werden kann. Im letzten Schritt werden die Intercepts γ_{02} und γ_{03} der beiden übrigen latenten State-Faktoren frei geschätzt. Der Mittelwert des indikatorspezifischen Faktors bleibt wiederum gemäß der Mplus-Voreinstellung auf Null fixiert, da es sich um einen Residuenfaktor handelt.

```
model: state1 by d11 d21; ! Spezifikation der Messmodelle
       state2 by d12 d22; ! für die Latent-State-Variablen
       state3 by d13 d23;

       trait by state1-state3; ! Spezifikation des Traitfaktors
                           ! als Faktor 2. Ordnung

       is2 by d21 d22 d23; ! Indikatorsspezifischer Faktor
                           ! für den 2. Indikator

       is2 with state1-state3@0 trait@0;

! Identifikation der Mittelwertsstruktur der latenten Variablen
[d11@0 d12@0 d13@0]; ! Intercepts des ersten Indikators auf 0 fixieren
[state1@0]; ! Intercept des ersten States auf 0 fixieren
[trait]; ! Mittelwert des Trait-Faktors frei schätzen
[state2 state3]; ! Intercepts der übrigen States frei schätzen
```

Abbildung 4.13: Mplus-model-Befehl zur Spezifikation des LST-Modells mit einem indikator-spezifischen Faktor (is2) für den zweiten Indikator aus Abbildung 4.12

Box 4.7: Mittelwertsstrukturen im LST-Modell

Ob latente Mittelwerte im LST-Modell von Interesse sind, hängt von der jeweiligen Fragestellung ab. In vielen LST-Analysen konzentriert man sich ausschließlich auf die Kovarianzstruktur der Variablen und lässt die Mittelwerte außen vor. Manchmal ist man jedoch auch an der Betrachtung der latenten Mittelwertsstruktur interessiert, beispielsweise, wenn man Mittelwertsveränderungen über die Zeit erwartet oder untersuchen möchte.

Im Text wurde eine Möglichkeit beschrieben, im LST-Modell die latente Mittelwertsstruktur zu identifizieren und den Mittelwert des Trait-Faktors zu schätzen. Es gibt neben der im Text vorgestellten Variante (Identifikation des Trait-Mittelwertes über den Mittelwert des ersten State-Faktors) auch noch andere Möglichkeiten zur Identifikation der latenten Mittelwertsstruktur im LST-Modell, die zu statistisch äquivalenten Modellen führen. Ist man an der Schätzung des latenten Trait-Mittelwertes nicht interessiert, so kann man als eine Alternative den Trait-Mittelwert gemäß Mplus-Voreinstellung auf Null fixiert lassen und dafür die Intercepts aller drei State-Faktoren (γ_{01} , γ_{02} und γ_{03}) schätzen. Die komplette Modellspezifikation vereinfacht sich dann wie folgt:

```
state1 by d11 d21;
state2 by d12 d22;
state3 by d13 d23;
trait by state1-state3;
is2 by d21 d22 d23;
is2 with state1-state3@0 trait@0;
[d11@0 d12@0 d13@0];
[state1 state2 state3];
```

Diese Parametrisierung hat den Vorteil, dass die Intercepts der drei State-Faktoren dann genau den State-Faktor-Mittelwerten entsprechen, was bei der im Text beschriebenen Variante nicht der Fall ist. Ein sinnvoller Vergleich der latenten State-Mittelwerte setzt allerdings zumindest starke faktorielle Invarianz voraus (siehe Abschnitt 4.1.4).

Eine weitere Möglichkeit besteht darin, die Mittelwertsstruktur der latenten Variablen komplett zu ignorieren (alle latenten Mittelwerte auf Null fixiert zu lassen) und nur die Intercepts α_{ik} für die beobachteten Variablen schätzen zu lassen. Diese (ebenfalls äquivalente) Variante ließe sich durch folgende model-Spezifikation erreichen:

```
state1 by d11 d21;
state2 by d12 d22;
state3 by d13 d23;
trait by state1-state3;
is2 by d21 d22 d23;
is2 with state1-state3@0 trait@0;
```

Weitere alternative Formen der Betrachtung latenter Mittelwerte sind möglich. Die Entscheidung für die eine oder andere Art der Spezifikation sollte von den inhaltlichen Fragestellungen abhängig gemacht werden.

Aus den Fit-Statistiken (s.u.) können wir ersehen, dass das LST-Modell in diesem Fall dieselbe Anpassungsgüte aufweist, wie das LS-Modell mit unrestringierten Parametern aus

Abschnitt 4.1.3. Dies ist damit zu erklären, dass in unserem Beispiel nur drei Messzeitpunkte analysiert werden. Bei nur drei Messzeitpunkten ist die Kovarianzstruktur der latenten Variablen saturiert, da der Traitfaktor in diesem Fall nur drei Indikatoren (die drei State-Faktoren 1. Ordnung) hat und die Ladungen (bis auf die erste) unrestringiert sind.

Auch die Mittelwertsstruktur ist saturiert, da keine parameterreduzierenden Restriktionen auf die Mittelwerte oder Intercepts gelegt wurden. Mithin liefert das Strukturmodell in diesem Fall keine zusätzlichen Freiheitsgrade, so dass die modellimplizierte Kovarianz- und Mittelwertsstruktur (und somit auch der Fit) identisch mit der Struktur im LS-Modell mit unrestringierten Parametern ist. Dies wäre anders, wenn wir mehr als drei Messzeitpunkte betrachten würden. Dann würde das LST-Strukturmodell testbare Restriktionen für die latente Kovarianzstruktur implizieren.

Chi-Square Test of Model Fit

Value	1.237
Degrees of Freedom	3
P-Value	0.7441

CFI / TLI

CFI	1.000
TLI	1.003

Information Criteria

Number of Free Parameters	24
Akaike (AIC)	-1258.543
Bayesian (BIC)	-1152.229
Sample-Size Adjusted BIC (n* = (n + 2) / 24)	-1228.425

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.000
90 Percent C.I.	0.000 0.047
Probability RMSEA <= .05	0.959

SRMR (Standardized Root Mean Square Residual)

Value	0.003
-------	-------

Die nachfolgend wiedergegebenen Parameterschätzwerte für das LST-Modell zeigen uns unter anderem, dass die State-Faktoren hoch und statistisch signifikant auf dem Trait-Faktor laden (zu erkennen im Output unter TRAIT BY...). Die z-Werte für die beiden frei geschätzten unstandardisierten Ladungen ($\hat{\gamma}_2 = 1,242$; $\hat{\gamma}_3 = 1,017$) betragen $z = 16,62$ und $z = 16,43$ (standardisierte Ladungen: $\hat{\gamma}_1^{\text{standardisiert}} = 0,774$; $\hat{\gamma}_2^{\text{standardisiert}} = 0,942$; $\hat{\gamma}_3^{\text{standardisiert}} = 0,782$).

Aus der Stärke dieser Ladungen können wir schließen, dass ein bedeutsamer Anteil der wahren Unterschiede (Varianz der State-Faktoren) in der Depressivität zu jedem Messzeitpunkt auf zeitlich stabile Einflüsse (Trait-Einflüsse) zurückgeführt werden kann (dies zeigte sich im datenäquivalenten LS-Modell bereits an den hohen Korrelationen der State-Faktoren untereinander). Den exakten Anteil der durch den Trait erklärbaren State-Faktor-Varianz können wir im Output unter R-SQUARE - Latent variable (unter der standardisierten Lösung ausgegeben) ersehen. Die R^2 -Werte für die latenten Variablen (hier: die State-Faktoren) entsprechen in diesem Fall den quadrierten standardisierten Ladungen der State-Faktoren auf dem Trait-Faktor und betragen 0,60, 0,89 und 0,61. Somit können zwischen 60% und 89% der State-Varianz auf Depressions-Traitunterschiede (stabile Persönlichkeitsdispositionen) zurückgeführt werden. Die restlichen 11-40% der State-Varianz gehen auf messgelegenheitsspezifische Einflüsse zurück. Die Inhomogenität der Faktorladungen 2. Ordnung spiegelt die im LS-Modell bereits bemerkte leichte Tendenz einer autoregressiven Struktur der State-Korrelationen wider.

MODEL RESULTS

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
STATE1	BY				
	D11	1.000	0.000	999.000	999.000
	D21	0.950	0.034	28.155	0.000
STATE2	BY				
	D12	1.000	0.000	999.000	999.000
	D22	0.918	0.030	30.365	0.000
STATE3	BY				
	D13	1.000	0.000	999.000	999.000
	D23	0.913	0.031	29.496	0.000
IS2	BY				
	D21	1.000	0.000	999.000	999.000
	D22	1.040	0.272	3.828	0.000
	D23	1.735	0.670	2.588	0.010
TRAIT	BY				
	STATE1	1.000	0.000	999.000	999.000
	STATE2	1.242	0.075	16.622	0.000
	STATE3	1.017	0.062	16.425	0.000
IS2	WITH				
	STATE1	0.000	0.000	999.000	999.000
	STATE2	0.000	0.000	999.000	999.000
	STATE3	0.000	0.000	999.000	999.000
	TRAIT	0.000	0.000	999.000	999.000

Means				
TRAIT	0.314	0.013	24.301	0.000
Intercepts				
D11	0.000	0.000	999.000	999.000
D21	-0.017	0.013	-1.358	0.175
D12	0.000	0.000	999.000	999.000
D22	0.002	0.011	0.166	0.868
D13	0.000	0.000	999.000	999.000
D23	-0.001	0.011	-0.090	0.929
STATE1	0.000	0.000	999.000	999.000
STATE2	-0.110	0.027	-4.099	0.000
STATE3	-0.033	0.023	-1.437	0.151
Variances				
TRAIT	0.051	0.006	9.213	0.000
IS2	0.003	0.001	2.158	0.031
Residual Variances				
D11	0.018	0.002	7.431	0.000
D21	0.010	0.002	4.763	0.000
D12	0.015	0.002	6.960	0.000
D22	0.015	0.002	7.981	0.000
D13	0.010	0.002	4.711	0.000
D23	0.008	0.003	2.348	0.019
STATE1	0.034	0.003	10.180	0.000
STATE2	0.010	0.003	2.963	0.003
STATE3	0.033	0.003	10.082	0.000
STANDARDIZED MODEL RESULTS (STDYX Standardization)				
			Two-Tailed	
		Estimate	S.E.	Est./S.E.
				P-Value
STATE1	BY			
	D11	0.907	0.014	65.166
	D21	0.926	0.014	68.362
STATE2	BY			
	D12	0.926	0.012	79.024
	D22	0.897	0.013	70.943
STATE3	BY			
	D13	0.944	0.013	75.501
	D23	0.902	0.014	66.504
IS2	BY			
	D21	0.176	0.041	4.315
	D22	0.180	0.042	4.273
	D23	0.308	0.064	4.847
TRAIT	BY			
	STATE1	0.774	0.024	32.629
	STATE2	0.942	0.020	46.859
	STATE3	0.782	0.023	33.820

IS2 WITH				
STATE1	0.000	0.000	999.000	999.000
STATE2	0.000	0.000	999.000	999.000
STATE3	0.000	0.000	999.000	999.000
TRAIT	0.000	0.000	999.000	999.000
Means				
TRAIT	1.390	0.095	14.683	0.000
Intercepts				
D11	0.000	0.000	999.000	999.000
D21	-0.057	0.042	-1.367	0.172
D12	0.000	0.000	999.000	999.000
D22	0.006	0.036	0.166	0.868
D13	0.000	0.000	999.000	999.000
D23	-0.003	0.037	-0.090	0.929
STATE1	0.000	0.000	999.000	999.000
STATE2	-0.370	0.088	-4.195	0.000
STATE3	-0.113	0.078	-1.463	0.144
Variances				
TRAIT	1.000	0.000	999.000	999.000
IS2	1.000	0.000	999.000	999.000
Residual Variances				
D11	0.178	0.025	7.037	0.000
D21	0.112	0.024	4.613	0.000
D12	0.143	0.022	6.601	0.000
D22	0.163	0.022	7.466	0.000
D13	0.108	0.024	4.584	0.000
D23	0.092	0.040	2.320	0.020
STATE1	0.401	0.037	10.920	0.000
STATE2	0.113	0.038	2.974	0.003
STATE3	0.388	0.036	10.709	0.000
R-SQUARE				
Observed Variable				
Estimate S.E. Est./S.E. Two-Tailed P-Value				
D11	0.822	0.025	32.583	0.000
D21	0.888	0.024	36.742	0.000
D12	0.857	0.022	39.512	0.000
D22	0.837	0.022	38.450	0.000
D13	0.892	0.024	37.751	0.000
D23	0.908	0.040	22.806	0.000
Latent Variable				
Estimate S.E. Est./S.E. Two-Tailed P-Value				
STATE1	0.599	0.037	16.314	0.000
STATE2	0.887	0.038	23.430	0.000
STATE3	0.612	0.036	16.910	0.000

Auch die latente Mittelwertsstruktur können wir dem Output entnehmen. In diesem Modell wird der unstandardisierte latente Trait-Mittelwert auf 0,314 geschätzt. Die geschätzten Intercepts von *State 1* und *State 2* betragen in unstandardisierter Form $\gamma_{02} = -0,11$ bzw. $\gamma_{03} = -0,033$.

Auch für das LST-Modell ist es sinnvoll, die Frage der Messinvarianz über die Zeit empirisch zu überprüfen, um ein inhaltlich gehaltvolles und möglichst sparsames Modell zu erhalten. An dieser Stelle soll jedoch auf die Darstellung von Messinvarianztests für das LST-Modell verzichtet werden, da das Prinzip bereits in Abschnitt 4.1.4 für das LS-Modell besprochen wurde und das dort beschriebene Vorgehen auf das LST-Modell übertragen werden kann.

LST-Modelle wurden insbesondere zur Modellierung und Analyse systematischer messgelegenheitsspezifischer Schwankungen von Messwerten um einen konstanten Traitwert entwickelt. Wie wir gesehen haben, können diese Modelle dazu verwendet werden, um Varianzkomponenten zu schätzen (Trennung zeitlich stabiler von messgelegenheitsspezifischer Varianz unter Berücksichtigung von Messfehlereinflüssen). In den folgenden Abschnitten werden wir Längsschnittmodelle besprechen, die sich weniger mit der Analyse messgelegenheitsspezifischer Variabilität um ein zeitlich stabiles Merkmal befassen, sondern zur Analyse von *Merkmalsveränderungen* (Trait-Veränderungen) über die Zeit eingesetzt werden. Trait-Veränderungsmodelle kommen insbesondere dann zum Einsatz, wenn davon ausgegangen werden muss, dass sich ein Merkmal tatsächlich verändert hat und es sich nicht um bloße Schwankungen um einen an sich stabilen Wert handelt. Als „klassische“ Modelle der Veränderungsmessung im Framework der Pfadanalyse bzw. der Strukturgleichungsmodelle besprechen wir zunächst die sogenannten autoregressiven Modelle.

4.3 Autoregressive Modelle

Autoregressive Modelle (z.B. Jöreskog, 1979a; 1979b) galten lange Zeit als erste Wahl bei der Auswertung von Längsschnittdaten. Mit dem Aufkommen latenter Wachstumskurvenmodelle (siehe Kapitel 4.5) sind autoregressive Modelle jedoch etwas aus der Mode gekommen (siehe dazu auch Hertzog & Nesselroade, 1987), obwohl sie nach wie vor zur Beantwortung einer Vielzahl wichtiger Forschungsfragestellungen in Längsschnittstudien sinnvoll eingesetzt werden können.

Autoregressive Modelle basieren auf der Grundidee, dass der beste Prädiktor für gegenwärtiges Erleben und Verhalten zeitlich vorgeordnetes Erleben und Verhalten ist. Aus diesem Grund werden wiederholt gemessene Variablen in autoregressiven Modellen über sogenannte *autoregressive Pfade* miteinander verknüpft (siehe Pfadkoeffizienten β_1 - β_4 in Abb. 4.14). Über die Stärke der autoregressiven Effekte sowie die Höhe der Residualvarianz (Varianz der Residualvariablen ε_{jk} in Abb. 4.14) kann die Stabilität interindividueller Unterschiede über die Zeit abgeschätzt werden. Eine starke Veränderung interindividueller Unterschiede über die Zeit liegt vor, wenn die Residualvarianz im Verhältnis zur allein durch autoregressive Effekte (Stabilität) erklärbaren Varianz hoch ist.

Das Ziel einer Analyse autoregressiver Modelle besteht jedoch zumeist nicht nur darin, das Ausmaß der Stabilität bzw. Veränderung interindividueller Unterschiede über die Zeit

zu bestimmen. Vielmehr wird in der Regel zusätzlich versucht, Veränderung (d.h. die durch den autoregressiven Effekt nicht erklärbare Varianz zu einem gegebenen Messzeitpunkt k) durch andere, im Modell zusätzlich berücksichtigte Variablen zu erklären. Dies geschieht häufig über sogenannte *kreuzverzögerte Effekte* (*cross-lagged effects*; siehe Pfadkoeffizienten β_5 - β_8 in Abb. 4.14). Kreuzverzögerte Effekte sind Effekte anderer zeitlich vorgeordneter Variablen, die zusätzlich zu den autoregressiven Effekten in das Modell aufgenommen werden. Ziel ist es, mit kreuzverzögerten Effekten einen Teil der Varianz zu erklären, der nicht durch die Autoregressionen vorhergesagt werden kann. Es wird also versucht, den „instabilen“ Teil interindividueller Unterschiede durch andere Variablen zu erklären.

4.3.1 Manifeste autoregressive Modelle

Ursprünglich wurden autoregressive Modelle meist als manifeste pfadanalytische Modelle auf der Ebene einzelner wiederholt gemessener manifester Indikatoren spezifiziert („single indicator models“, wie in Abb. 4.14). Auch heutzutage findet man solche Modelle auf Ebene beobachteter Variablen noch häufig in der Literatur. Hierbei handelt es sich um einfache Pfadanalysen ohne latente Variablen (vgl. Abschnitt 3.5.2).

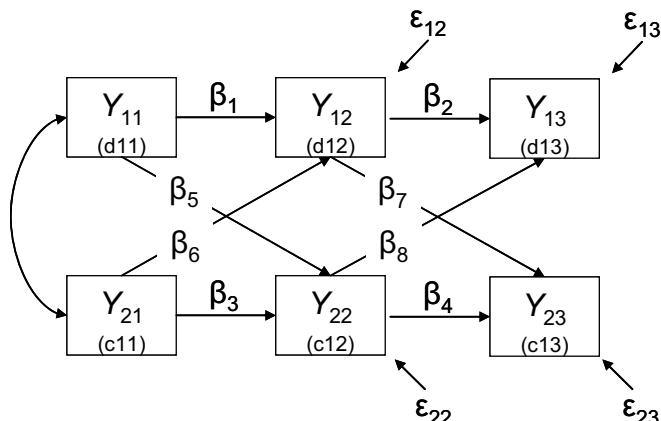
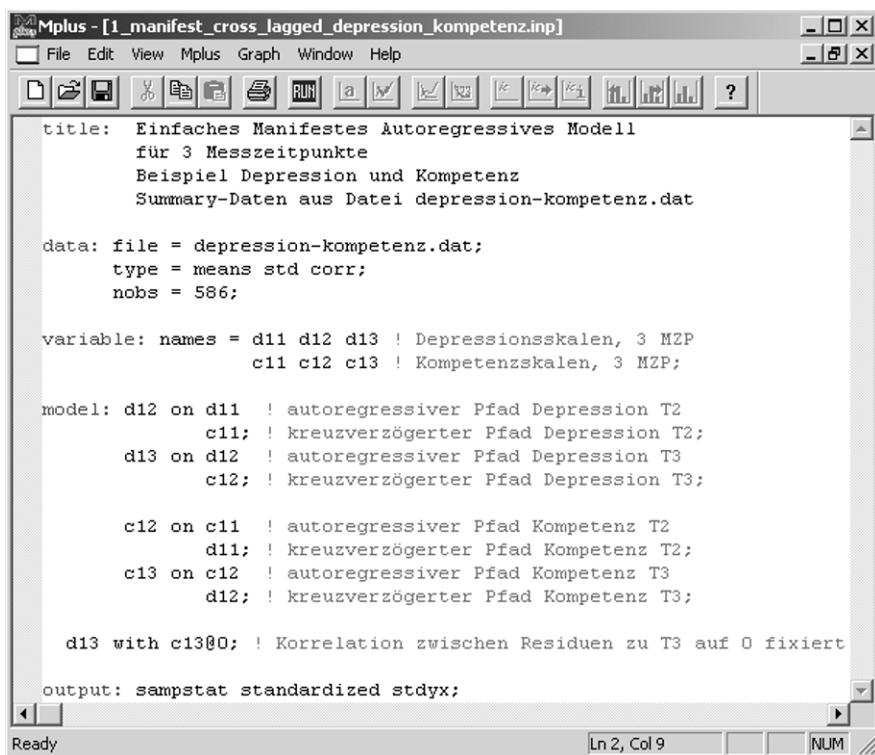


Abbildung 4.14: Manifestes autoregressives Modell für zwei manifester Indikatoren Y_{jk} (j = Konstrukt, k = Messzeitpunkt), welche zwei Konstrukte ($j = 1, 2$; Depression und Kompetenz) zu drei Messzeitpunkten ($k = 1, 2, 3$) repräsentieren. β_1 - β_4 : Pfadkoeffizienten, die autoregressive Effekte repräsentieren. β_5 - β_8 : Pfadkoeffizienten, die kreuzverzögerte Effekte repräsentieren. ϵ_{jk} : Residualvariablen. In Klammern sind die Mplus-Variablennamen dargestellt.

Im Folgenden wird zunächst die Analyse manifester autoregressiver Modelle in Mplus besprochen. In Abschnitt 4.3.2 behandeln wir latente autoregressive Modelle.

Box 4.8: Fallbeispiel manifeste autoregressive Modelle

Als Beispiel analysieren wir einen Längsschnittdatensatz mit den Konstrukten *Depressivität* und *Kompetenz*, welche jeweils zu drei Messgelegenheiten bei $N = 586$ Kindern mittels Fragebögen erhoben wurden (Abstand zwischen den Messzeitpunkten jeweils 6 Monate). Die Daten **depression-kompetenz.dat** (siehe Begleit-CD), enthält Summary-Daten (Mittelwerte, Standardabweichungen und Korrelationen) in sechs Variablen. Die ersten drei Variablen gehören zu einer dreimalig erhobenen Depressionsskala, die letzten drei repräsentieren eine zu den drei Messzeitpunkten ebenfalls erfasste Kompetenzskala. Ziel ist es, ein manifestes autoregressives Modell mit kreuzverzögerten Effekten auf Ebene dieser sechs beobachteten Variablen zu schätzen. In diesem Modell soll die Stabilität der beiden Konstrukte Depressivität und Kompetenz untersucht werden. Außerdem wollen wir kreuzverzögerte Effekte testen.



```
Mplus - [1_manifest_cross_lagged_depression_kompetenz.inp]
File Edit View Mplus Graph Window Help
title: Einfaches Manifestes Autoregressives Modell
       für 3 Messzeitpunkte
       Beispiel Depression und Kompetenz
       Summary-Daten aus Datei depression-kompetenz.dat

data: file = depression-kompetenz.dat;
       type = means std corr;
       nobs = 586;

variable: names = d11 d12 d13 ! Depressionsskalen, 3 MZP
           c11 c12 c13 ! Kompetenzskalen, 3 MZP;

model: d12 on d11  ! autoregressiver Pfad Depression T2
       c11; ! kreuzverzögerner Pfad Depression T2;
       d13 on d12  ! autoregressiver Pfad Depression T3
       c12; ! kreuzverzögerner Pfad Depression T3;

       c12 on c11  ! autoregressiver Pfad Kompetenz T2
       d11; ! kreuzverzögerner Pfad Kompetenz T2;
       c13 on c12  ! autoregressiver Pfad Kompetenz T3
       d12; ! kreuzverzögerner Pfad Kompetenz T3;

       d13 with c13@0; ! Korrelation zwischen Residuen zu T3 auf 0 fixiert

output: sampstat standardized stdyx;
Ready
```

Abbildung 4.15: Mplus-Input-Datei zur Spezifikation des einfachen manifesten autoregressiven Modells für drei Messzeitpunkte aus Abbildung 4.14.

Wir spezifizieren zunächst ein Modell analog zu Abbildung 4.14. Der entsprechende Mplus-Input wird in Abbildung 4.15 gezeigt.

Die im Modell zu schätzenden autoregressiven und kreuzverzögerten Effekte werden durch Partialregressionskoeffizienten repräsentiert (man spricht auch von Pfadkoeffizienten, vgl. Kapitel 3.5 zur Pfadanalyse). Aus diesem Grund kommen bei der Spezifikation in Mplus on-Befehle zum Einsatz. Im Prinzip stellt das Modell eine manifeste Pfadanalyse (vgl. Abschnitt 3.5.2) dar, wobei bei autoregressiven Pfadmodellen u.a. Pfade zwischen wiederholt gemessenen Variablen Gegenstand der Analyse sind.

Ein wichtiger Aspekt ist, dass das Modell in Abbildung 4.14 annimmt, dass die Residualvariablen ε_{jk} untereinander unkorreliert sind. Dies muss man Mplus für die Residuen zum dritten Messzeitpunkt explizit mitteilen, da das Programm diese Residualkorrelation ansonsten per Voreinstellung zulässt und als zusätzlichen Parameter schätzen würde. Aus diesem Grund wurde im Input die Befehlszeile

```
d13 with c13@0;
```

hinzugefügt. Alle übrigen Residuenkorrelationen werden von Mplus per Voreinstellung auf Null fixiert (wie wir in Kürze sehen werden, passt das Modell ohne Residuenkorrelationen in diesem Fall jedoch sehr schlecht auf die Daten, so dass diese Restriktionen in weiteren Analyseschritten gelockert werden).

Die von Mplus berechneten Modellfitindices für das einfache autoregressive Modell (der vollständige Output befindet sich auf der Begleit-CD) weisen auf eine sehr schlechte Datenanpassung hin ($\chi^2 = 201,90, df = 6, p < 0,001; CFI = 0,86; RMSEA = 0,24, SRMR = 0,09$). Die schlechte Datenanpassung kann in diesem Beispiel dadurch erklärt werden, dass im Modell keine Korrelationen zwischen Residualvariablen zum selben Messzeitpunkt zugelassen wurden. Die Schätzung dieser Korrelationen ist im vorliegenden Beispiel nötig, da es messgelegenheitsspezifische Einflüsse gibt, die sich auf beide Konstrukte auswirken. Individuen, die sich zu einem bestimmten Messzeitpunkt depressiver einschätzen, neigen dazu, sich zum selben Zeitpunkt auch als weniger kompetent einzustufen. Dieser situationsbezogene Effekt wird in einem autoregressiven Modell ohne korrelierte Residuen zum selben Messzeitpunkt nicht hinreichend abgebildet.

Zusätzlich ist im vorliegenden Beispiel die Einführung regressiver Pfade zweiter Ordnung erforderlich, um eine zufriedenstellende Datenanpassung zu erreichen. Das in dieser Form erweiterte Modell ist in Abbildung 4.16 als Pfaddiagramm dargestellt. Die korrelierten Residuen sind über Doppelpfeile zwischen ε_{12} und ε_{22} sowie zwischen ε_{13} und ε_{23} dargestellt. Die zusätzlichen autoregressiven Pfade zweiter Ordnung sind mit β_9 und β_{10} bezeichnet. Das entsprechend modifizierte Mplus-model-Statement kann Abbildung 4.17 entnommen werden.

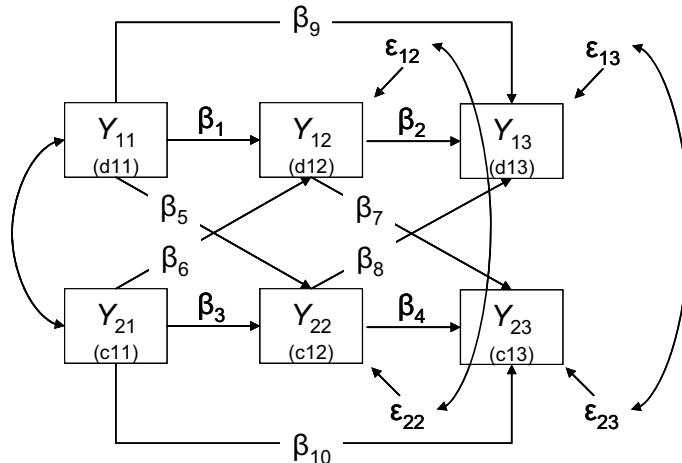


Abbildung 4.16: Manifestes autoregressives Modell für zwei Indikatoren Y_{jk} (j = Konstrukt, k = Messzeitpunkt), welche zwei Konstrukte ($j = 1, 2$) zu drei Messzeitpunkten ($k = 1, 2, 3$) repräsentieren. β_1 - β_4 sowie β_9 - β_{10} : Pfadkoeffizienten, die autoregressive Effekte repräsentieren. β_5 - β_8 : Pfadkoeffizienten, die kreuzverzögerte Effekte repräsentieren. ε_{jk} : Residualvariablen. Im Unterschied zum in Abb. 4.14 dargestellten Modell enthält das Modell zusätzlich korrelierte Residualvariablen zum selben Messzeitpunkt sowie autoregressive Effekte zweiter Ordnung (repräsentiert durch β_9 und β_{10}). In Klammern sind die Mplus-Variablennamen dargestellt.

```
model: d12 on d11      ! autoregressiver Pfad Depression T2
       c11;           ! kreuzverzögter Pfad Depression T2;
d13 on d12 d11      ! autoregressive Pfade Depression T3
       c12;           ! kreuzverzögter Pfad Depression T3;

c12 on c11          ! autoregressiver Pfad Kompetenz T2
       d11;           ! kreuzverzögter Pfad Kompetenz T2;
c13 on c12 c11      ! autoregressive Pfade Kompetenz T3
       d12;           ! kreuzverzögter Pfad Kompetenz T3;

d12 with c12; ! Korrelation zwischen Residuen zu T2 zugelassen
d13 with c13; ! Korrelation zwischen Residuen zu T3 zugelassen
```

Abbildung 4.17: Mplus-model-Befehl zur Spezifikation des erweiterten manifesten autoregressiven Modells mit autoregressiven Effekten zweiter Ordnung und korrelierten Residualvariablen zum selben Messzeitpunkt aus Abbildung 4.16.

Das erweiterte autoregressive Modell passt sehr gut auf die Daten ($\chi^2 = 2,74$, $df = 2$, $p = 0,25$; $CFI = 0,999$; $RMSEA = 0,03$, $SRMR = 0,01$). Es passt zudem signifikant besser als das restriktivere Modell aus Abbildung 4.14, wie ein χ^2 -Differenzentest belegt ($\Delta\chi^2 = 199,16$, $\Delta df = 4$, $p < 0,001$). Der AIC-Koeffizient fällt mit einem Wert von 2493,08 ebenfalls geringer für das erweiterte Modell aus (AIC-Wert des Modells aus Abb. 4.14: 2690,01). Nachfolgend sind die Parameterschätzwerte für das erweiterte Modell wiedergegeben:

MODEL RESULTS

		Estimate	S.E.	Two-Tailed Est./S.E.	P-Value
D12	ON				
D11		0.526	0.037	14.225	0.000
C11		-0.085	0.020	-4.184	0.000
D13	ON				
D12		0.510	0.040	12.782	0.000
D11		0.156	0.035	4.411	0.000
C12		-0.056	0.020	-2.749	0.006
C12	ON				
C11		0.422	0.037	11.352	0.000
D11		-0.392	0.068	-5.786	0.000
C13	ON				
C12		0.386	0.044	8.797	0.000
C11		0.192	0.037	5.181	0.000
D12		-0.436	0.072	-6.066	0.000
D12	WITH				
C12		-0.040	0.005	-8.067	0.000
D13	WITH				
C13		-0.035	0.005	-7.868	0.000
Intercepts					
D12		0.288	0.050	5.789	0.000
D13		0.210	0.050	4.198	0.000
C12		1.335	0.091	14.640	0.000
C13		0.995	0.107	9.322	0.000
Residual Variances					
D12		0.061	0.004	17.117	0.000
D13		0.051	0.003	17.117	0.000
C12		0.206	0.012	17.117	0.000
C13		0.207	0.012	17.117	0.000

STANDARDIZED MODEL RESULTS (STDYX Standardization)

		Estimate	S.E.	Two-Tailed Est./S.E.	P-Value
D12	ON				
D11		0.531	0.033	15.947	0.000
C11		-0.156	0.037	-4.209	0.000

D13	ON			
	D12	0.517	0.037	13.820
	D11	0.160	0.036	4.426
	C12	-0.100	0.036	-2.754
				0.006
C12	ON			
	C11	0.440	0.036	12.214
	D11	-0.224	0.038	-5.872
				0.000
C13	ON			
	C12	0.361	0.040	9.128
	C11	0.187	0.036	5.214
	D12	-0.231	0.038	-6.150
				0.000
D12	WITH			
	C12	-0.353	0.036	-9.778
				0.000
D13	WITH			
	C13	-0.344	0.036	-9.434
				0.000
Intercepts				
	D12	0.909	0.159	5.718
	D13	0.674	0.163	4.146
	C12	2.386	0.197	12.142
	C13	1.663	0.200	8.315
Residual Variances				
	D12	0.610	0.031	19.374
	D13	0.526	0.030	17.590
	C12	0.657	0.032	20.657
	C13	0.579	0.031	18.679
				0.000
R-SQUARE				
Observed Variable		Estimate	S.E.	Two-Tailed P-Value
			Est./S.E.	
D12		0.390	0.031	12.403
D13		0.474	0.030	15.875
C12		0.343	0.032	10.800
C13		0.421	0.031	13.602

Es zeigt sich, dass alle autoregressiven Effekte (sowohl erster als auch zweiter Ordnung) statistisch signifikant sind (siehe geschätzte Pfadkoeffizienten und Signifikanztests unter D12 ON D11, D13 ON D12 D11, C12 ON C11 sowie C13 ON C12 C11). Somit liegt ein statistisch bedeutsames Ausmaß an Stabilität interindividueller Unterschiede hinsichtlich Depressivität und Kompetenz vor.

Allerdings sind interindividuelle Unterschiede hinsichtlich Depression und Kompetenz über die Zeit nicht perfekt stabil. Es gibt interindividuelle Unterschiede in intraindividuellen

Veränderungen, die sowohl für Depressivität als auch für Kompetenz zum Teil durch zeitlich vorgeordnete individuelle Unterschiede des jeweils anderen Konstrukttes erklärt werden können. Dies kann man anhand der signifikanten kreuzverzögerten Effekte erkennen.

Die kreuzverzögerten Effekte lassen sich dahingehend interpretieren, dass weniger kompetente Individuen tendenziell höhere Depressivitätsscores zu späteren Messgelegenheiten zeigen, auch unter Berücksichtigung der Autoregression (Stabilität) sowie der (Residual-) Korrelation der Konstrukte zum selben Messzeitpunkt. Umgekehrt zeigen stärker depressive Individuen zu späteren Messzeitpunkten tendenziell geringere Kompetenzwerte, auch wenn die Stabilität sowie die Korrelationen zur selben Messgelegenheit berücksichtigt werden.

Hierbei ist interessant, dass die kreuzverzögerten Effekte von Depression auf Kompetenz (geschätzte standardisierte Pfadkoeffizienten $\hat{\beta}^{\text{standardisiert}} = -0,224$ bzw. $-0,231$) vom Betrag her größer ausfallen, als die kreuzverzögerten Effekte von Kompetenz auf Depression (standardisierte Pfadkoeffizienten $\hat{\beta}^{\text{standardisiert}} = -0,156$ bzw. $-0,100$). Insgesamt weisen die relativ kleinen standardisierten Pfadkoeffizienten darauf hin, dass die kreuzverzögerten Effekte (im Vergleich zu den autoregressiven Pfaden) bei beiden Konstrukten nur wenig zusätzliche Varianz erklären können.

Die Korrelationen der Residualvariablen zu T2 und T3 ($r = -0,35$ / $-0,34$, $ps < 0,001$) sind statistisch bedeutsam und von der Größe her substantiell. Die mittelhohen Korrelationen der Residualvariablen zur selben Messgelegenheit indizieren, dass es zusätzliche gemeinsame Ursachen von Depressivität und Kompetenz zum selben Messzeitpunkt gibt (gemeinsame messgelegenheitsspezifische Varianz), welche weder durch die Autoregression noch durch die kreuzverzögerten Effekte erklärbar sind. Verantwortlich für das Zustandekommen dieser Korrelationen könnten zum Beispiel situationsspezifische Einflüsse sein, welche sich zu einem gegebenen Messzeitpunkt auf beide Konstrukte auswirken.

Die R^2 -Werte geben die im Modell insgesamt erklärte Varianz jeder endogenen Variablen an. Es zeigt sich, dass insgesamt (d.h., durch Autoregression und kreuzverzögerte Effekte zusammen) zwischen 34,3% und 47,4% der Depressions- bzw. Kompetenzvarianz zu T2 bzw. T3 aufgeklärt werden können. Der restliche, an 100% fehlende Teil der Varianz geht auf andere Einflüsse zurück, die im Modell nicht explizit berücksichtigt wurden (z.B. situationspezifische Effekte). Hierbei ist zu beachten, dass auch zufällige Messfehler nicht berücksichtigt wurden, da die Analyse auf der Ebene beobachteter (und nicht latenter) Variablen stattfand. Im nächsten Abschnitt besprechen wir die Spezifikation von *latenten autoregressiven Modellen*, in welchen Messfehler explizit Berücksichtigung finden.

4.3.2 Latente autoregressive Modelle

In den im vorigen Abschnitt behandelten manifesten autoregressiven Modellen wurden ausschließlich beobachtete Variablen verwendet. Somit haben manifeste autoregressive Modelle wie andere manifeste Pfadmodelle auf manifester Ebene den Nachteil, dass Messfehlereinflüsse in diesen Modellen nicht explizit modelliert und somit auch nicht bei der Berechnung der Pfadkoeffizienten berücksichtigt werden. Dies kann zu einer Verzerrung der Ergebnisse und dadurch beispielsweise zu einer Fehleinschätzung des Ausmaßes an

Stabilität bzw. Veränderung sowie einer verzerrten Schätzung kreuzverzögter Effekte führen. Diese Problematik ist identisch zur Messfehlerproblematik in manifesten Regressions- und Pfadmodellen für Querschnittsdaten (vgl. Abschnitt 3.2 und 3.5.2).

Autoregressive Modelle mit latenten Variablen (im Folgenden *latente autoregressive Modelle* genannt) berücksichtigen Messfehler bei der Schätzung der Pfadkoeffizienten. Sie erlauben außerdem die Testung von Messmodellen, die Analyse indikatorsspezifischer Effekte (vgl. Abschnitt 4.1.3) sowie die Bestimmung des Grades an Messinvarianz über die Zeit (vgl. Abschnitt 4.1.4). Diese Modelle werden u.a. von Jöreskog (1979a, 1979b) sowie Herzog & Nesselroade (1987) im Detail beschrieben.

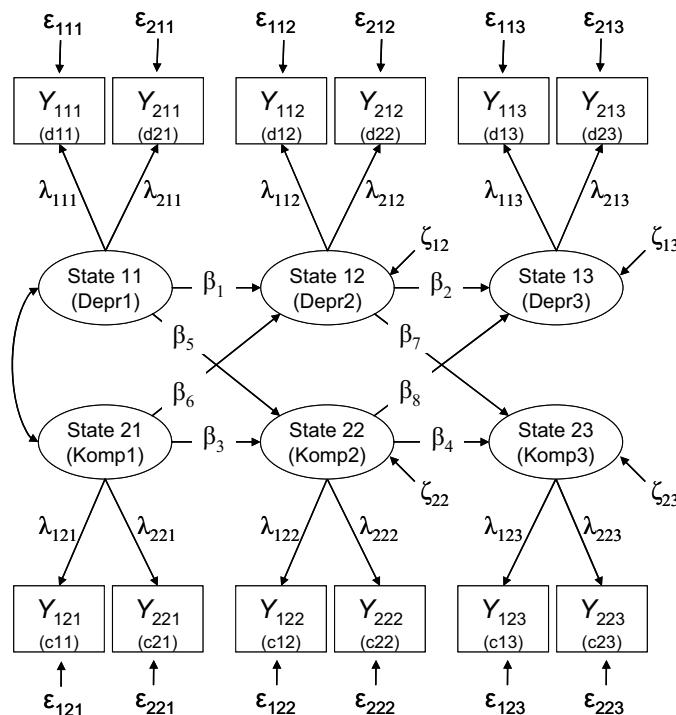


Abbildung 4.18: Latentes autoregressives Modell für zwölf Indikatoren Y_{ijk} (i = Indikator, j = Konstrukt, k = Messzeitpunkt), welche zwei Konstrukte ($j = 1, 2$; Depression und Kompetenz) zu drei Messzeitpunkten ($k = 1, 2, 3$) repräsentieren. $\beta_1\text{--}\beta_4$: Latente Pfadkoeffizienten, die autoregressive Effekte repräsentieren. $\beta_5\text{--}\beta_8$: Latente Pfadkoeffizienten, die kreuzverzögerte Effekte repräsentieren. ϵ_{ijk} : Messfehlervariablen. ζ_{jk} : Latente Residualvariablen. In Klammern sind die Mplus-Variablennamen dargestellt.

Latente autoregressive Modelle basieren auf dem in Abschnitt 4.1.2 besprochenen LS-Modell. Dabei wird auf die Kovarianzstruktur der latenten State-Variablen, welche im LS-Modell in der Regel unrestringiert ist, eine autoregressive Struktur angepasst. Die latenten State-Variablen werden wie die beobachteten Variablen im manifesten autoregressiven Modell über autoregressive und ggf. über kreuzverzögerte Pfade miteinander verknüpft.

Abbildung 4.18 zeigt ein latentes autoregressives Modell mit kreuzverzögerten Effekten für zwei Konstrukte, welche jeweils zu drei Messzeitpunkten mit je zwei Indikatoren Y_{ijk} (i = Indikator, j = Konstrukt, k = Messgelegenheit) gemessen wurden. Man beachte, dass das Strukturmodell (also der Teil, der die Zusammenhänge zwischen den latenten State-Variablen untereinander betrifft) äquivalent zum manifesten Modell in Abbildung 4.14 ist.

Ein guter Ausgangspunkt für die Analyse eines latenten autoregressiven Modells ist somit das (weniger restriktive) LS-Modell (siehe Kapitel 4.1). Dieses Modell kann u.a. dazu dienen, zunächst Fragen der Indikatorsspezifität und Messinvarianz über die Zeit zu untersuchen (siehe Abschnitte 4.1.3 bzw. 4.1.4). Das entsprechende LS-Modell kann anschließend zu einem autoregressiven Modell erweitert werden.

Box 4.9: Fallbeispiel latente autoregressive Modelle

Im Folgenden wird gezeigt, wie ein latentes autoregressives Modell für das Datenbeispiel „Depression und Kompetenz“ spezifiziert werden kann. Im latenten autoregressiven Modell wird für jedes Konstrukt zu jedem Messzeitpunkt ein Messmodell mit multiplen Indikatoren spezifiziert. Die entsprechenden Summary-Daten von zwei Depressions- und zwei Kompetenzskalen, welche jeweils zu drei Messgelegenheiten erhoben wurden ($N = 569$) befinden sich in der Datei **depression-kompetenz-latent.dat** auf der Begleit-CD.

Abbildung 4.15 zeigt den Mplus-Input zur Schätzung des latenten autoregressiven Modells aus Abbildung 4.18. Der Einfachheit halber wird zunächst eine einfache Modellvariante ohne Messinvarianzrestriktionen und ohne indikatorsspezifische Faktoren demonstriert (also das Modell in der Form, wie es in Abb. 4.18 dargestellt ist). Weiter unten wird noch eine komplexere Modellvariante mit indikatorsspezifischem Faktor sowie zusätzlichen Effekten besprochen.

The screenshot shows the Mplus software window with the title bar 'Mplus - [4_latent_cross_lagged_depression_kompetenz_ohne_invarianz.inp]'. The menu bar includes File, Edit, View, Mplus, Graph, Window, and Help. Below the menu is a toolbar with various icons. The main area contains the Mplus input command language. The input specifies a model with three latent variables (depression and competence at three time points) and their cross-lagged paths. It also includes structural equations and correlation constraints.

```

title: Einfaches latentes autoregressives Modell
für 3 Messzeitpunkte (MZP)
2 Indikatoren pro Konstrukt und MZP
Beispiel Depression und Kompetenz
Summary-Daten aus Datei depression-kompetenz-latent.dat
Keine Messinvarianzrestriktionen
Kein indikatorsspezifischer Faktor
Keine Autoregressionen 2. Ordnung
Keine korrelierten Residuen zum selben MZP

data: file = depression-kompetenz-latent.dat;
type = means std corr;
nobs = 569;

variable: names = d11 d21 ! Depressionsskalen T1
           d12 d22 ! Depressionsskalen T2
           d13 d23 ! Depressionsskalen T3
           c11 c21 ! Kompetenzskalen T1
           c12 c22 ! Kompetenzskalen T2
           c13 c23 ! Kompetenzskalen T3

model: ! Messmodelle Depression und Kompetenz
       depr1 by d11
              d21; ! Depression State 1
       depr2 by d12
              d22; ! Depression State 2
       depr3 by d13
              d23; ! Depression State 3

       komp1 by c11
              c21; ! Kompetenz State 1
       komp2 by c12
              c22; ! Kompetenz State 2
       komp3 by c13
              c23; ! Kompetenz State 3

! Strukturmodell: Autoregressives Modell mit kreuzverzögerten Effekten
       depr2 on depr1 komp1; ! Autoregressiver und kreuzverzögter Effekt Depression T2
       depr3 on depr2 komp2; ! Autoregressiver und kreuzverzögter Effekt Depression T3
       komp2 on komp1 depr1; ! Autoregressiver und kreuzverzögter Effekt Kompetenz T2
       komp3 on komp2 depr2; ! Autoregressiver und kreuzverzögter Effekt Kompetenz T3

       depr3 with komp3@0; ! Korrelation der latenten Residualvariablen zu T3 auf 0 fixieren

output: sampstat standardized stdyx;

```

Abbildung 4.19: Mplus-Input-Datei zur Spezifikation des einfachen latenten autoregressiven Modells aus Abbildung 4.18.

Aus dem Input geht hervor, dass die grundlegende Spezifikation der Messmodelle identisch ist zur Spezifikation im LS-Modell, mit dem Unterschied, dass wir hier zwei Konstrukte simultan betrachten. Im Strukturmodell werden die autoregressiven und kreuzverzögerten Pfade auf latenter Ebene wiederum mittels on-Befehlen spezifiziert. Der Befehl

`depr3 with komp3@0;`

setzt die Korrelation zwischen den latenten Residualvariablen zum dritten Messzeitpunkt auf Null (diese würde ansonsten wie im manifesten autoregressiven Modell per Mplus-Voreinstellung frei geschätzt werden).

Die Analyse der Modellgüte ergibt, dass das Modell sehr schlecht auf die Daten passt ($\chi^2 = 408,42$, $df = 45$, $p < 0,001$; $CFI = 0,93$; $RMSEA = 0,12$; $SRMR = 0,05$). Dies wird vor allem anhand des sehr hohen χ^2 -Tests und des ebenfalls hohen $RMSEA$ -Koeffizienten deutlich (lediglich der $SRMR$ -Wert ist akzeptabel). Aus diesem Grund wird hier auf eine Darstellung der Parameterschätzergebnisse zu diesem Modell verzichtet (der Output befindet sich auf der Begleit-CD).

Der schlechte Modellfit kann zum einen darauf zurückgeführt werden, dass indikator-spezifische Effekte über die Zeit (vgl. Abschnitt 4.1.3) in dieser Modellvariante nicht berücksichtigt wurden. Zudem hatte unsere Analyse des manifesten autoregressiven Modells für Depression und Kompetenz bereits ergeben, dass zusätzlich zu den autoregressiven Effekten erster Ordnung autoregressive Effekte zweiter Ordnung sowie Residuenkorrelationen zum selben Messzeitpunkt geschätzt werden müssen, um einen adäquaten Datenfit zu erzielen. Das Fehlen dieser zusätzlichen Effekte macht sich auch im latenten autoregressiven Modell durch den schlechten Fit bemerkbar.

Das latente autoregressive Modell lässt sich in analoger Weise erweitern. Die erweiterte Version mit autoregressiven Pfaden zweiter Ordnung (Pfadkoeffizienten β_9 und β_{10}), korrelierten latenten Residualvariablen ζ_{jk} zum selben Messzeitpunkt sowie indikatorspezifischen Faktoren für den jeweils zweiten Indikator ist in Abbildung 4.20 dargestellt.

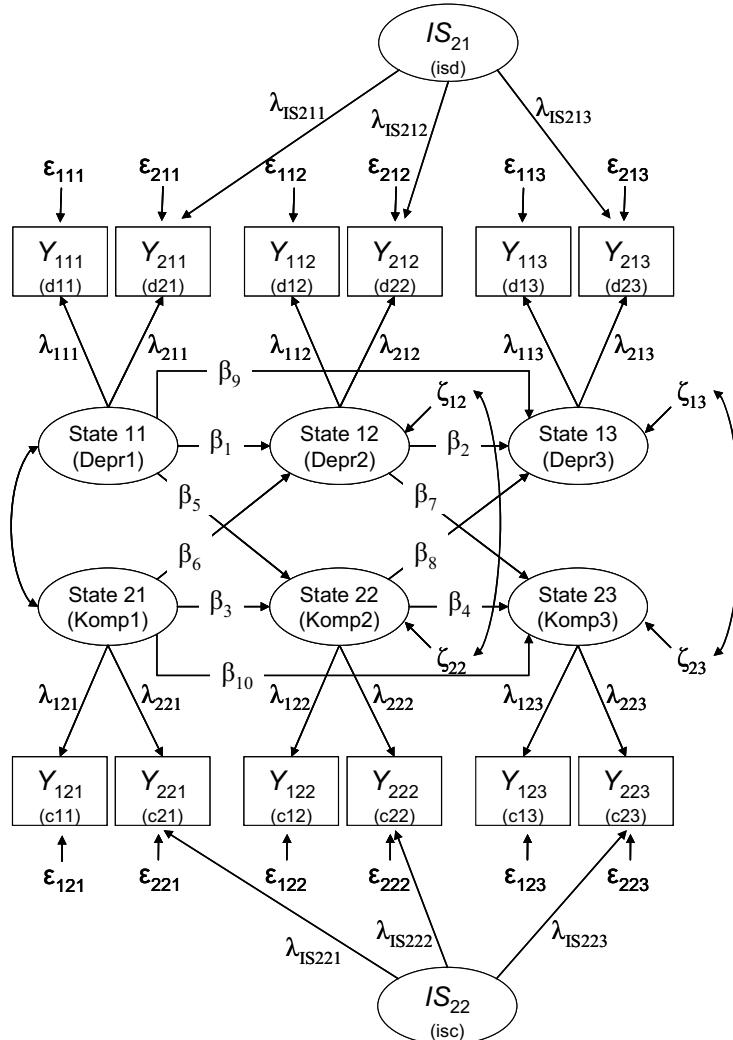


Abbildung 4.20: Latentes autoregressives Modell für zwölf Indikatoren Y_{ijk} (i = Indikator, j = Konstrukt, k = Messzeitpunkt), welche zwei Konstrukte ($j = 1, 2$; Depression, Kompetenz) zu drei Messzeitpunkten ($k = 1, 2, 3$) repräsentieren. λ_{ijk} : State-Faktorladungen. λ_{isijk} : Ladungen auf dem indikatorspezifischen Faktor IS_{ij} . β_1 - β_4 : Latente Pfadkoeffizienten, die autoregressive Effekte 1. Ordnung repräsentieren. β_5 - β_{10} : Latente Pfadkoeffizienten, die autoregressive Effekte 2. Ordnung repräsentieren. β_5 - β_8 : Latente Pfadkoeffizienten, die kreuzverzögerte Effekte repräsentieren. ϵ_{ijk} : Messfehlervariablen. ζ_{jk} : Latente Residualvariablen. Im Unterschied zum Modell in Abb. 4.18 enthält das Modell zusätzlich autoregressive Effekte zweiter Ordnung (repräsentiert durch β_9 und β_{10}), korrelierte latente Residualvariablen zum selben Messzeitpunkt und indikatorspezifische Faktoren für den 2. Indikator. Aus Gründen der Übersichtlichkeit sind nicht alle zulässigen Korrelationen eingezeichnet. In Klammern sind die Mplus-Variablennamen dargestellt.

Abbildung 4.21 zeigt die Mplus-Spezifikation des erweiterten Modells aus Abbildung 4.20. Aus Platzgründen wird nur das model-Statement gezeigt:

```
model:
! Messmodelle Depression und Kompetenz
depr1 by d11
    d21; ! Depression State 1
depr2 by d12
    d22; ! Depression State 2
depr3 by d13
    d23; ! Depression State 3

isd by d21 d22 d23; ! Indikatorsspez. Faktor Depression
isd with depr1-depr3@0; ! mit Depressions-State-Faktoren und -residuen unkorreliert
isd with komp1 komp2 komp3; ! darf mit Kompetenz-State-Faktoren
    ! und -residuen korrelieren

komp1 by c11
    c21; ! Kompetenz State 1
komp2 by c12
    c22; ! Kompetenz State 2
komp3 by c13
    c23; ! Kompetenz State 3

isc by c21 c22 c23; ! Indikatorsspez. Faktor Kompetenz
isc with komp1-komp3@0; ! mit Kompetenz-State-Faktoren und -residuen unkorreliert
isc with depr1 depr2 depr3; ! darf mit Depressions-State-Faktoren
    ! und -residuen korrelieren

! Strukturmodell: Autoregressives Modell mit kreuzverzögerten Effekten
! Mit autoregressiven Effekten 2. Ordnung
depr2 on depr1 komp1; ! Autoregressiver und kreuzverzögter Effekt Depression T2
depr3 on depr2 depr1 komp2; ! Autoregressive und kreuzverzögerte Effekte Depression T3
komp2 on komp1 depr1; ! Autoregressiver und kreuzverzögter Effekt Kompetenz T2
komp3 on komp2 komp1 depr2; ! Autoregressive und kreuzverzögerte Effekte Kompetenz T3

depr2 with komp2; ! Korrelation der latenten Residualvariablen zu T2 schätzen
depr3 with komp3; ! Korrelation der latenten Residualvariablen zu T3 schätzen
```

Abbildung 4.21: Mplus-model-Befehl zur Spezifikation des latenten autoregressiven Modells mit autoregressiven Effekten zweiter Ordnung, korrelierten latenten Residualvariablen zum selben Messzeitpunkt und indikatorsspezifischen Faktoren für den zweiten Indikator aus Abbildung 4.20.

Das erweiterte Modell weist zwar immer noch einen signifikanten χ^2 -Wert auf, zeigt im Vergleich zum einfacheren Modell aus jedoch eine wesentlich bessere Datenanpassung ($\chi^2 = 46,72, df = 28, p = 0,02; CFI = 0,996; RMSEA = 0,03; SRMR = 0,02$). Ein Vergleich der AIC-Werte ergibt ebenfalls, dass das komplexere Modell aus ($AIC = 2601,53$) dem Modell aus Abbildung 4.18 ($AIC = 2929,24$) vorzuziehen ist. Nachfolgend sind die Parameterschätzwerte für das komplexere Modell dargestellt.

MODEL RESULTS

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
DEPR1	BY				
	D11	1.000	0.000	999.000	999.000
	D21	0.957	0.032	29.972	0.000
DEPR2	BY				
	D12	1.000	0.000	999.000	999.000
	D22	0.909	0.030	30.500	0.000
DEPR3	BY				
	D13	1.000	0.000	999.000	999.000
	D23	0.924	0.028	32.900	0.000
ISD	BY				
	D21	1.000	0.000	999.000	999.000
	D22	1.093	0.277	3.941	0.000
	D23	1.385	0.425	3.256	0.001
KOMP1	BY				
	C11	1.000	0.000	999.000	999.000
	C21	1.031	0.059	17.501	0.000
KOMP2	BY				
	C12	1.000	0.000	999.000	999.000
	C22	0.913	0.054	16.840	0.000
KOMP3	BY				
	C13	1.000	0.000	999.000	999.000
	C23	0.857	0.049	17.346	0.000
ISC	BY				
	C21	1.000	0.000	999.000	999.000
	C22	1.167	0.165	7.084	0.000
	C23	1.151	0.172	6.685	0.000
DEPR2	ON				
	DEPR1	0.720	0.055	13.141	0.000
	KOMP1	-0.030	0.034	-0.878	0.380
DEPR3	ON				
	DEPR2	0.590	0.063	9.406	0.000
	DEPR1	0.153	0.051	3.033	0.002
	KOMP2	-0.049	0.033	-1.471	0.141
KOMP2	ON				
	KOMP1	0.693	0.071	9.715	0.000
	DEPR1	-0.208	0.100	-2.080	0.037

KOMP3	ON			
	KOMP2	0.597	0.101	5.930
	KOMP1	0.195	0.073	2.679
	DEPR2	-0.243	0.104	-2.338
				0.000
				0.007
				0.019
ISD	WITH			
	DEPR1	0.000	0.000	999.000
	DEPR2	0.000	0.000	999.000
	DEPR3	0.000	0.000	999.000
	KOMP1	0.000	0.002	0.151
	KOMP2	-0.002	0.002	-1.089
	KOMP3	0.000	0.002	-0.242
				0.880
				0.276
				0.809
ISC	WITH			
	KOMP1	0.000	0.000	999.000
	KOMP2	0.000	0.000	999.000
	KOMP3	0.000	0.000	999.000
	DEPR1	-0.011	0.004	-2.823
	DEPR2	-0.007	0.003	-2.601
	DEPR3	-0.008	0.003	-3.122
	ISD	-0.002	0.001	-1.622
				0.005
				0.009
				0.002
				0.105
DEPR2	WITH			
	KOMP2	-0.035	0.004	-8.310
				0.000
DEPR3	WITH			
	KOMP3	-0.036	0.004	-8.584
				0.000
KOMP1	WITH			
	DEPR1	-0.093	0.008	-11.127
				0.000
Intercepts				
D11		0.306	0.013	22.757
D21		0.274	0.013	21.672
D12		0.274	0.013	20.497
D22		0.253	0.013	20.079
D13		0.283	0.013	21.352
D23		0.256	0.013	20.152
C11		2.074	0.024	85.370
C21		2.121	0.026	82.201
C12		2.091	0.024	88.301
C22		2.115	0.025	83.180
C13		2.082	0.025	82.382
C23		2.184	0.025	85.819
				0.000
				0.000
				0.000
Variances				
DEPR1		0.085	0.006	13.491
ISD		0.003	0.001	2.508
KOMP1		0.225	0.021	10.628
ISC		0.054	0.012	4.578
				0.000
				0.012
				0.000

Residual Variances				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
D11	0.018	0.002	8.220	0.000
D21	0.010	0.002	5.108	0.000
D12	0.014	0.002	6.928	0.000
D22	0.014	0.002	7.801	0.000
D13	0.010	0.002	5.251	0.000
D23	0.008	0.002	3.802	0.000
C11	0.111	0.012	9.473	0.000
C21	0.086	0.011	7.648	0.000
C12	0.084	0.011	7.546	0.000
C22	0.094	0.012	8.178	0.000
C13	0.087	0.013	6.982	0.000
C23	0.085	0.011	7.693	0.000
DEPR2	0.039	0.003	11.783	0.000
DEPR3	0.038	0.003	12.569	0.000
KOMP2	0.096	0.013	7.650	0.000
KOMP3	0.102	0.013	7.901	0.000
STANDARDIZED MODEL RESULTS (STDYX Standardization)				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
DEPR1 BY				
D11	0.909	0.013	72.457	0.000
D21	0.925	0.012	75.929	0.000
DEPR2 BY				
D12	0.929	0.011	82.978	0.000
D22	0.896	0.013	71.363	0.000
DEPR3 BY				
D13	0.948	0.011	90.229	0.000
D23	0.914	0.012	76.546	0.000
ISD BY				
D21	0.193	0.038	5.013	0.000
D22	0.211	0.041	5.215	0.000
D23	0.266	0.046	5.721	0.000
KOMP1 BY				
C11	0.818	0.022	36.435	0.000
C21	0.794	0.023	34.255	0.000
KOMP2 BY				
C12	0.858	0.021	40.148	0.000
C22	0.729	0.026	28.125	0.000
KOMP3 BY				
C13	0.871	0.021	42.281	0.000
C23	0.741	0.026	28.907	0.000

ISC	BY			
C21	0.377	0.040	9.400	0.000
C22	0.446	0.043	10.403	0.000
C23	0.439	0.043	10.224	0.000
DEPR2	ON			
DEPR1	0.708	0.046	15.362	0.000
KOMP1	-0.048	0.055	-0.880	0.379
DEPR3	ON			
DEPR2	0.583	0.057	10.175	0.000
DEPR1	0.149	0.049	3.048	0.002
KOMP2	-0.078	0.053	-1.475	0.140
KOMP2	ON			
KOMP1	0.678	0.057	11.966	0.000
DEPR1	-0.125	0.060	-2.080	0.038
KOMP3	ON			
KOMP2	0.551	0.084	6.591	0.000
KOMP1	0.176	0.065	2.699	0.007
DEPR2	-0.137	0.059	-2.330	0.020
ISD	WITH			
DEPR1	0.000	0.000	999.000	999.000
DEPR2	0.000	0.000	999.000	999.000
DEPR3	0.000	0.000	999.000	999.000
KOMP1	0.010	0.064	0.150	0.880
KOMP2	-0.091	0.082	-1.111	0.267
KOMP3	-0.020	0.083	-0.241	0.809
ISC	WITH			
KOMP1	0.000	0.000	999.000	999.000
KOMP2	0.000	0.000	999.000	999.000
KOMP3	0.000	0.000	999.000	999.000
DEPR1	-0.169	0.051	-3.295	0.001
DEPR2	-0.157	0.059	-2.691	0.007
DEPR3	-0.185	0.056	-3.307	0.001
ISD	-0.152	0.087	-1.741	0.082
DEPR2	WITH			
KOMP2	-0.570	0.051	-11.108	0.000
DEPR3	WITH			
KOMP3	-0.570	0.049	-11.625	0.000
KOMP1	WITH			
DEPR1	-0.675	0.032	-21.093	0.000

Intercepts				
D11	0.954	0.051	18.870	0.000
D21	0.909	0.050	18.254	0.000
D12	0.859	0.049	17.521	0.000
D22	0.842	0.049	17.264	0.000
D13	0.895	0.050	18.049	0.000
D23	0.845	0.049	17.323	0.000
C11	3.579	0.114	31.374	0.000
C21	3.446	0.110	31.414	0.000
C12	3.702	0.117	31.514	0.000
C22	3.487	0.111	31.443	0.000
C13	3.454	0.111	31.226	0.000
C23	3.598	0.114	31.647	0.000
Variances				
DEPR1	1.000	0.000	999.000	999.000
ISD	1.000	0.000	999.000	999.000
KOMP1	1.000	0.000	999.000	999.000
ISC	1.000	0.000	999.000	999.000
Residual Variances				
D11	0.174	0.023	7.652	0.000
D21	0.107	0.022	4.907	0.000
D12	0.136	0.021	6.538	0.000
D22	0.152	0.021	7.249	0.000
D13	0.101	0.020	5.060	0.000
D23	0.091	0.025	3.703	0.000
C11	0.331	0.037	9.004	0.000
C21	0.227	0.032	7.146	0.000
C12	0.265	0.037	7.219	0.000
C22	0.256	0.033	7.652	0.000
C13	0.241	0.036	6.702	0.000
C23	0.230	0.032	7.209	0.000
DEPR2	0.450	0.034	13.290	0.000
DEPR3	0.425	0.031	13.604	0.000
KOMP2	0.410	0.043	9.547	0.000
KOMP3	0.369	0.038	9.642	0.000
R-SQUARE				
Observed Variable		Estimate	S.E.	Two-Tailed P-Value
D11		0.826	0.023	36.229
D21		0.893	0.022	41.054
D12		0.864	0.021	41.489
D22		0.848	0.021	40.347
D13		0.899	0.020	45.115
D23		0.909	0.025	36.837
C11		0.669	0.037	18.218
C21		0.773	0.032	24.327

C12	0.735	0.037	20.074	0.000
C22	0.744	0.033	22.263	0.000
C13	0.759	0.036	21.141	0.000
C23	0.770	0.032	24.147	0.000
Latent Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
DEPR2	0.550	0.034	16.267	0.000
DEPR3	0.575	0.031	18.403	0.000
KOMP2	0.590	0.043	13.765	0.000
KOMP3	0.631	0.038	16.494	0.000

Die Parameterschätzwerte des Strukturmodells im latenten autoregressiven Modell weisen im Vergleich zum manifesten autoregressiven Modell einige bemerkenswerte Unterschiede auf. Global fällt auf, dass die geschätzten R^2 -Werte für das Strukturmodell (R-SQUARE - Latent Variable) im latenten autoregressiven Modell deutlich höher sind ($0,55 \leq R^2 \leq 0,63$) als im manifesten Modell ($0,39 \leq R^2 \leq 0,47$). Somit kann durch das latente Modell mehr Varianz zum zweiten und dritten Messzeitpunkt vorhergesagt werden.

Allerdings fallen die kreuzverzögerten Effekte von Kompetenz auf Depression, welche im manifesten autoregressiven Modell statistisch von Null verschieden waren, im latenten autoregressiven Modell nicht signifikant aus. Im Gegensatz dazu sind die kreuzverzögerten Effekte von Depression auf Kompetenz auch im latenten autoregressiven Modell statistisch bedeutsam. Die höheren R^2 -Werte des Strukturmodells im latenten Modell können vor allem dadurch erklärt werden, dass die autoregressiven Effekte erster Ordnung im latenten Modell deutlich größer sind [standardisierte autoregressiven Pfadkoeffizienten erster Ordnung: $0,55 \leq \hat{\beta}^{\text{standardisiert}} \leq 0,71$, siehe ON-Befehle unter STANDARDIZED MODEL RESULTS (STDYX Standardization)] als im manifesten Modell ($0,36 \leq \hat{\beta}^{\text{standardisiert}} \leq 0,53$). Dies deutet auf eine Unterschätzung der zeitlichen Stabilität der Merkmale im manifesten autoregressiven Modell hin, was möglicherweise durch Nichtberücksichtigung von Messfehlerinflüssen erklärt werden kann.

Die Befunde verdeutlichen, dass die Berücksichtigung von Messfehlereinflüssen durch die Verwendung latenter Variablen gerade in Längsschnittanalysen zur präziseren Analyse von Zusammenhangsstrukturen vorteilhaft ist. Gerade die Kompetenz-Skalen weisen im vorliegenden Beispiel eher durchschnittliche Reliabilitäten auf ($0,67 \leq R^2 \leq 0,77$; siehe Werte unter R-SQUARE - Observed Variable, Variablen C11 bis C23). Auch die Depressions-Indikatoren sind nicht perfekt reliabel ($0,83 \leq R^2 \leq 0,91$; siehe Werte unter R-SQUARE - Observed Variable, Variablen D11 bis D23). Dies kann eine Erklärung für die Unterschiede zwischen den Ergebnissen im manifesten und im latenten autoregressiven Modell sein.

Auch für das latente autoregressive Modell sollte in weitergehenden Analyseschritten der Grad der Invarianz von Parametern des Messmodells über die Zeit ermittelt werden. Die Möglichkeit, Annahmen bezüglich der Messinvarianz über die Zeit zu testen ist ein weiterer wesentlicher Vorteil des latenten gegenüber dem manifesten autoregressiven Mo-

dell. Im manifesten Modell kann die Annahme der Messinvarianz nicht – bzw. nur sehr eingeschränkt – statistisch überprüft werden.

Wir verzichten hier auf eine ausführliche Darstellung von Messinvarianztests für das latente autoregressive Modell, da das Prinzip ausführlich für das LS-Modell besprochen wurde (siehe Abschnitt 4.1.4) und die für das LS-Modell gezeigten Schritte der Messinvarianztestung auf das latente autoregressive Modell übertragen werden können. Als Hilfe befindet sich auf der Begleit-CD ein Beispiel-Input zur Spezifikation eines latenten autoregressiven Modells mit starker Messinvarianz.

4.4 Latent-Change-Modelle

In Latent-Change-(LC-)Modellen (McArdle, 1988; McArdle & Hamagami, 2001; Steyer et al., 1997, 2000) wird – im Unterschied zu autoregressiven Modellen – Veränderung über die Zeit nicht indirekt über Residuen von Autoregressionen modelliert, sondern direkt in Form sogenannter *latenter Differenzvariablen*. Latente Differenzvariablen repräsentieren interindividuelle Unterschiede in intraindividuellen Veränderungen auf latenter, d.h. messfehlerfreier Ebene. Aus diesem Grund wurden die Modelle von Steyer et al. (1997) auch als *True-Change-Modelle* bezeichnet. Da latente Differenzvariablen in diesem Ansatz zentral sind, findet man häufig auch den Begriff *Latent-Difference-Modelle*.

Latente Differenzvariablen lassen sich relativ einfach Weise durch eine simple Reformulierung von LS-Modellen (siehe Kapitel 4.1) in Strukturgleichungsmodelle integrieren. Abbildung 4.22 illustriert die Umformulierung („Reparametrisierung“) eines LS-Modells (Abb. 4.22A) zu einem LC-Modell (Abb. 4.22B) anhand eines einfachen Beispiels (ein Konstrukt gemessen mit zwei Indikatoren zu zwei Messzeitpunkten). Wir können uns dabei vorstellen, dass sich die latente State-Variable zum zweiten Messzeitpunkt (*State 2*) zusammensetzt aus Ausgangswert (*State 1*) plus Veränderungswert (Differenz *State 2–State 1*; d.h. Zuwachs bzw. Abnahme):

$$State\ 2 = 1 \cdot State\ 1 + 1 \cdot (State\ 2 - State\ 1)$$

Diese Zerlegung ist in keiner Weise restriktiv: Wir addieren lediglich eine Variable (den latenten Faktor *State 1*) zu einer anderen Variablen (dem latenten Faktor *State 2*) hinzu und ziehen diese Variable gleichzeitig wieder ab. Die Zerlegung in Ausgangsvariable und Veränderungsvariable kann im LS-Modell durch die Aufnahme der latenten Differenzvariable (*State 2 – State 1*) implementiert werden (siehe Abb. 4.22B). Dadurch erhält man die latente Differenz zwischen den beiden Messzeitpunkten als zusätzlichen Modellbestandteil (zusätzliche latente Variable) und kann so direkt interindividuelle Unterschiede in intraindividuellen Veränderungen auf messfehlerfreier Ebene untersuchen.

Box 4.10: Fallbeispiel Latent-Change-Modelle

Als inhaltliches Beispiel können wir uns vorstellen, dass wir in unserem LS-Modell für Depression (vgl. Kapitel 4.1.2) den latenten State-Faktor für Depression zu Messzeitpunkt 2 in Ausgangswert (State-Faktor Depression T1) plus latente Differenz (State-Faktor T2 minus State Faktor T1) zerlegen:

$$\text{Depression T2} = \text{Depression T1} + (\text{Depression T2} - \text{Depression T1}).$$

Mit dieser Zerlegung können wir beispielsweise das Ausmaß interindividueller Unterschiede in intraindividuellen Veränderungen hinsichtlich Depression untersuchen (über die Varianz des Differenzenfaktors) und/oder interindividuelle Unterschiede in intraindividuellen Veränderungen zu anderen Variablen in Beziehung setzen (durch Aufnahme von Kovariaten).

Eine latente Differenzvariable kann wie jede andere latente Variable in einem Strukturgleichungsmodell behandelt werden. Das bedeutet, die latente Differenzvariable kann mit anderen Variablen korreliert werden. Sie kann darüber hinaus sowohl als exogene als auch als endogene Variable in erweiterten Modellen fungieren. Beispielsweise ist man im Rahmen von Längsschnittstudien häufig an der Erklärung von interindividuellen Unterschieden hinsichtlich Veränderung über die Zeit interessiert (warum verändern sich einige Menschen in ihrer Depressivität stärker als andere?). Die latente Differenzvariable kann in einem solchen Fall als endogene Variable genutzt werden, die durch andere Variablen vorhergesagt („erklärt“) wird. Auch kann der Mittelwert der latenten Differenzvariable geschätzt werden, um Mittelwertsunterschiede zwischen zwei Messzeitpunkten zu untersuchen (beispielsweise auf Signifikanz zu testen).

Man beachte, dass in der obigen Gleichung die Variable *State 2* auf den ersten State (Variable *State 1*) und den Veränderungswert (Variable *State 2 – State 1*) zurückgeführt wird. Man kann dies als eine „perfekte Regression“ auffassen, da der zweite State durch Ausgangswert und Veränderungswert vollständig determiniert ist (die Gleichung enthält kein Residuum). Aus diesem Grund ist in der pfadanalytischen Darstellung in Abbildung 4.22B das Residuum für den zweiten State auf Null gesetzt. Dieser Aspekt ist für die Modellspezifikation in Mplus von Bedeutung, da hier bei der entsprechenden Spezifikation die Residualvarianz für den zweiten State explizit auf Null gesetzt werden muss.

Wichtig ist, dass bezüglich des Messmodells wenigstens starke faktorielle Invarianz (gleiche Ladungen und gleiche Intercepts der Indikatoren) über die Zeit vorliegen muss (vgl. Abschnitt 4.1.4), da starke faktorielle Invarianz eine wichtige Voraussetzung für die Interpretierbarkeit der latenten Differenzvariablen ist. Diese Form der Messinvarianz stellt sicher, dass man in Hinblick auf die Differenzen der latenten State-Variablen nicht „Äpfel von Birnen“ subtrahiert.

In Abbildung 4.22 ist die Annahme invarianten Ladungen dadurch erkennbar, dass die Ladungsparameter λ_i nur den Index i für die Variable, jedoch keinen Index k für den Messzeitpunkt tragen (die Intercepts sind nicht dargestellt). Die Implementierung der Annahme starker Messinvarianz muss bei der Spezifikation des Modells in Mplus beachtet werden. Es

müssen hierbei Gleichheitsrestriktionen (analog zu den in Abschnitt 4.1.4 diskutierten LS-Modellen mit invarianten Parametern) eingeführt werden.

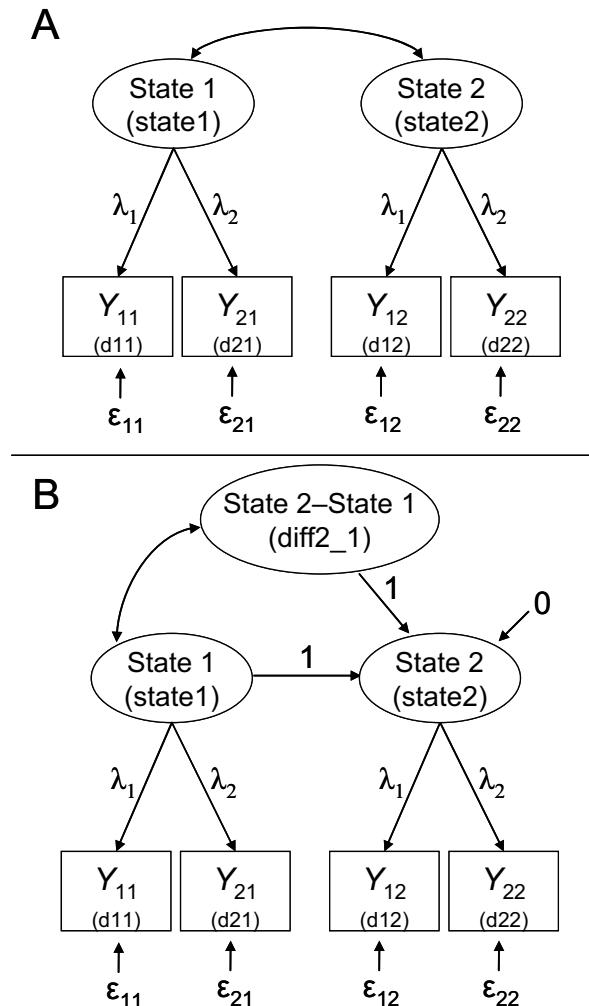


Abbildung 4.22: Latent State- und Latent-Change-Modell für Depression mit zwei beobachteten Variablen Y_{ik} (i = Indikator, k = Messgelegenheit), welche zu zwei Messgelegenheiten erhoben wurden. ε_{ik} : Messfehlervariable; λ_i : zeitlich invariante Faktorladungsparameter. A: Latent-State-Modell. B: Latent-Change-Modell. In Klammern sind die in Mplus verwendeten Variablennamen dargestellt.

Steyer et al. (1997, 2000) unterscheiden in Hinblick auf Latent-Change-Modelle für mehr als zwei Messzeitpunkte zwischen einer *Baseline-Change-Version* des Modells und einer *Neighbor-Change-Version*. Im Baseline-Change-Modell wird Veränderung immer in Bezug auf den ersten Messzeitpunkt untersucht. Das bedeutet, der erste Messzeitpunkt dient hier als Refe-

renz („Baseline“), und es wird die Veränderung vom ersten zum zweiten, vom ersten zum dritten, vom ersten zum vierten Messzeitpunkt usw. analysiert.

Im Neighbor-Change-Modell ist Veränderung zwischen unmittelbar aufeinanderfolgenden („benachbarten“) Messzeitpunkten Gegenstand der Analyse. Es wird somit Veränderung vom ersten zum zweiten, vom zweiten zum dritten, vom dritten zum vierten Messzeitpunkt usw. mittels latenter Differenzvariablen untersucht. Abbildung 4.23A zeigt ein Baseline-Change- und Abbildung 4.23B ein Neighbor-Change-Modell für je drei Messzeitpunkte.

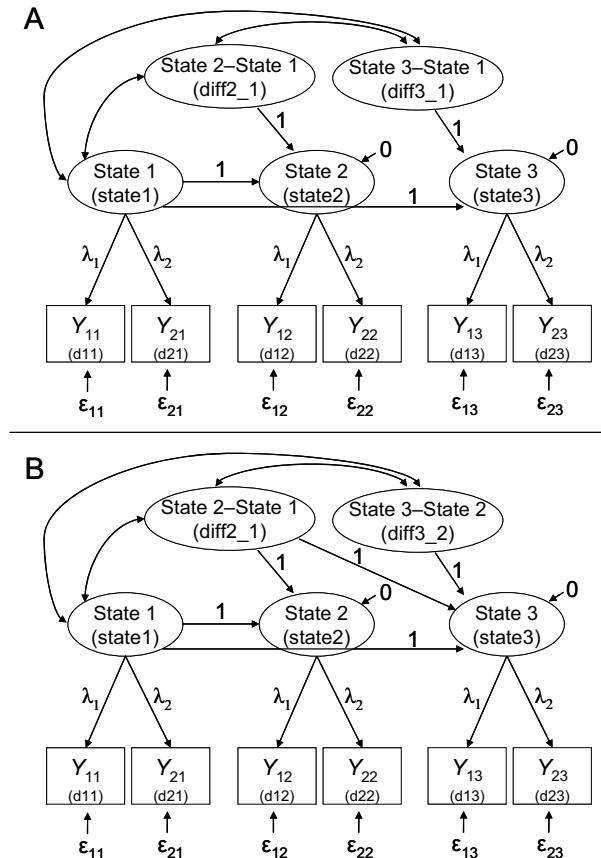


Abbildung 4.23: Latent-Change-Modelle für zwei beobachtete Variablen Y_{ik} (i = Indikator, k = Messgelegenheit), welche zu drei Messgelegenheiten erhoben wurden. ε_{ik} : Messfehlervariable; λ : zeitlich invariante Faktorladungsparameter. A: Baseline-Change-Modell. B: Neighbor-Change-Modell. Die Modelle sind statistisch äquivalent zueinander und zu einem Latent-State-Modell mit invarianten Parametern. Der (häufig benötigte) indikatorsspezifische Faktor für den zweiten Indikator ist aus Platzgründen nicht dargestellt. In Klammern sind die in Mplus verwendeten Variablennamen angegeben.

Der Unterschied zwischen dem Baseline-Change-Modell und dem Neighbor-Change-Modell besteht darin, dass der State-Faktor zum dritten Messzeitpunkt (State 3) im Baseline-Change-Modell in Ausgangswert (*State 1*) + Veränderungswert (*State 3 – State 1*) zerlegt wird:

$$\text{State 3} = 1 \cdot \text{State 1} + 1 \cdot (\text{State 3} - \text{State 1}),$$

während *State 3* sich im Neighbor-Change-Modell aus Ausgangswert + Veränderungswert (*State 2 – State 1*) + Veränderungswert (*State 3 – State 2*) zusammensetzt (was auf statistisch äquivalente Modelle hinausläuft):

$$\text{State 3} = 1 \cdot \text{State 1} + 1 \cdot (\text{State 2} - \text{State 1}) + 1 \cdot (\text{State 3} - \text{State 2}).$$

Box 4.11: Datenbeispiel Latent-Change-Modelle in Mplus

Als Basis für die exemplarische Spezifikation von LC-Modellen in Mplus greifen wir auf das LS-Modell für die Depressionsdaten aus Abschnitt 4.1.4 (Datensatz **depression.dat**, $N = 620$) zurück. Ausgangspunkt unserer Spezifikation ist die LS-Modellvariante für drei Messzeitpunkte mit einem indikatorspezifischen Faktor für den zweiten Indikator und starker faktorieller Invarianz (Faktorladungen und Intercepts über die Zeit hinweg gleichgesetzt). Da in diesem Fall die Annahme der Messinvarianz über die Zeit bereits getestet wurde (siehe Abschnitt 4.1.4), können wir direkt in die Analyse einsteigen. Normalerweise müsste an dieser Stelle zunächst eine Überprüfung dieser Annahme analog zum Vorgehen in Abschnitt 4.1.4 erfolgen. Das vorliegende LS-Modell ist statistisch äquivalent zu einem Baseline-Change- und einem Neighbor-Change-Modell mit invarianten Ladungen und Intercepts der Indikatoren, d.h. alle drei Modelle implizieren dieselbe Mittelwerts- und Kovarianzstruktur. Dies liegt daran, dass es sich bei den LC-Modellen lediglich um Reparametrisierungen des LS-Modells handelt. Somit ist auch der Fit der drei Modelle identisch (s.u.).

Abbildung 4.24 zeigt den Mplus-model-Befehl zur Spezifikation des Baseline-Change-Modells (die Spezifikation des Neighbor-Change-Modells wird weiter unten dargestellt). Der restliche Teil des Inputs unterscheidet sich nicht vom Input zur Spezifikation des LS-Modells (vgl. Abb. 4.9).

```

model: state1 by d11
        d21 (1); ! Gleichsetzung der State-Ladungen des 2. Indikators
state2 by d12
        d22 (1);
state3 by d13
        d23 (1);

diff2_1 by d11@0; ! Einführung der Variablennamen für die
diff3_1 by d11@0; ! latenten Differenzvariablen

state2 on state1@1 diff2_1@1; ! (perfekte) Regression des zweiten States
state2@0;                      ! auf Ausgangswert T1 und Veränderungswert T2-T1

state3 on state1@1 diff3_1@1; ! (perfekte) Regression des dritten States
state3@0;                      ! auf Ausgangswert T1 und Veränderungswert T3-T1

state2 with state3@0 diff3_1@0;
state3 with diff2_1@0;

! Indikatorsspezifischer (Residual-)Faktor für den 2. Indikator
is2 by d21
        d22@1 ! Fixierung aller Ladungen auf
        d23@1; ! dem indikatorsspezifischen Faktor auf 1

is2 with state1-state3@0 diff2_1@0 diff3_1@0; ! Indikatorsspez. Faktor ist mit den
                                                ! übrigen Faktoren unkorreliert

[d11@0 d12@0 d13@0]; ! Fixierung der Intercepts des ersten Indikators auf 0

[state1 diff2_1 diff3_1]; ! Schätzung des State-Mittelwertes zu T1 sowie der
                           ! Mittelwerte der latenten Differenzvariablen

[d21 d22 d23] (2); ! Gleichsetzung der Intercepts des 2. Indikators

```

Abbildung 4.24: Mplus-Modellspezifikation zur Schätzung des Baseline-Change-Modells aus Abbildung 4.23A. Es wurde zusätzlich ein indikatorsspezifischer Faktor für den zweiten Indikator eingeführt, welcher in Abbildung 4.23 aus Platzgründen nicht dargestellt ist.

Die Messmodelle zur Spezifikation der drei latenten Depressions-State-Variablen werden analog zum LS-Modell mit invarianten Ladungen formuliert. Die latenten Differenzvariablen sind gemäß der Parametrisierung in Abbildung 4.23 nicht direkt mit beobachteten Indikatoren verbunden (Steyer et al., 1997, 2000), diskutieren eine alternative Parametrisierung von LC-Modellen, in welcher die latenten Differenzvariablen direkt mit beobachteten Variablen verknüpft sind. Aus didaktischen Gründen wurde diese Parametrisierung hier nicht gewählt. Auf der Begleit-CD sind jedoch Mplus-Inputs für beide Arten von Parametrisierungen enthalten).

Zur Spezifikation der latenten Differenzvariablen `diff2_1` und `diff3_1` wurde daher ein Trick angewendet. Um die Differenzvariablen auch ohne „eigene“ manifeste Indikatoren einführen zu können, wurden „Pseudomessmodelle“ für die latenten Differenzvariablen spezifiziert. Diese Pseudomessmodelle dienen dazu, die Namen der Differenzvariablen in Mplus einzuführen. Dies geschieht über die Befehle:

```

diff2_1 by d11@0;
diff3_1 by d11@0;

```

Sie bedeuten, dass die Variable d11 auf beiden Differenzfaktoren eine Ladung von Null hat. Der by-Befehl wird somit aus rein technischen Gründen benötigt, um die Namen der latenten Differenzvariablen (diff2_1 und diff3_1) einzuführen.

Die on-Befehle in Verbindung mit dem Nullsetzen der Residualvarianz für die zweite und dritte latente State-Variablen dienen dazu, die „perfekten“ Regressionen der State-Variablen zu T2 und T3 auf die T1-State-Variable und die latenten Differenzvariablen zu spezifizieren:

```
state2 on state1@1 diff2_1@1;
state2@0;

state3 on state1@1 diff3_1@1;
state3@0;
```

Man beachte, dass die latenten Pfadkoeffizienten von state1 und diff2_1 auf state2 sowie von state1 und diff3_1 auf state2 jeweils mittels @ auf 1 fixiert wurden, so wie es der Modellspezifikation in Abbildung 4.23 entspricht. Mit Hilfe der folgenden Befehle werden die Residualvarianzen der latenten State-Faktoren zum zweiten und dritten Messzeitpunkt auf Null fixiert (Mplus würde ansonsten per Voreinstellung versuchen, diese zu schätzen):

```
state2@0;
state3@0;
```

Die Residualvarianzen müssen im Modell unbedingt auf Null gesetzt werden, da die State-Faktoren state2 und state3 im Modell durch Ausgangswert (state1) und Veränderungswert (diff2_1 bzw. diff3_1) perfekt determiniert sind. Durch die on-Befehle in Verbindung mit den auf Null fixierten Residualvarianzen erhalten die Variablen diff2_1 und diff3_1 ihre Bedeutung als latente Differenzvariablen.

Da die State-Faktoren state2 und state3 im Latent-Change-Modell perfekt durch andere Variablen determiniert sind, können sie nicht mit anderen Faktoren korreliert sein. Deshalb werden im nächsten Schritt die Korrelationen von state2 und state3 mit den übrigen State- und Differenzfaktoren auf Null fixiert:

```
state2 with state3@0 diff3_1@0;
state3 with diff2_1@0;
```

Die weiteren Befehle dienen wie im LS-Modell dazu, einen indikatorspezifischen Faktor (is2) für den zweiten Indikator zur Modellierung indikatorspezifischer Effekte über die Zeit einzuführen (vgl. Abschnitt 4.1.3). Alle Ladungen auf diesem Faktor werden auf 1 fixiert, um sie invariant über die Zeit zu setzen. Über den with-Befehl wird festgelegt, dass dieser Faktor mit allen State- und State-Differenzfaktoren unkorreliert ist.

```

is2 by d21
  d22@1
  d23@1;

is2 with state1-state3@0 diff2_1@0 diff3_1@0;

```

Die folgenden Befehle beziehen sich auf die Mittelwertsstruktur. Zunächst werden die Intercepts des Referenzindikators auf Null fixiert, um die latenten Mittelwerte identifizieren zu können (vgl. Abschnitt 4.1.4):

```
[d11@0 d12@0 d13@0];
```

Anschließend wird die Schätzung der Mittelwerte von `state1` sowie der latenten Differenzvariablen `diff2_1` und `diff3_1` angefordert (ansonsten würde Mplus diese latenten Mittelwerte per Voreinstellung auf Null fixieren):

```
[state1 diff2_1 diff3_1];
```

Schließlich werden noch die Intercepts des zweiten Indikators über die Zeit hinweg gleichgesetzt, um der Annahme starker Messinvarianz für alle Indikatoren gerecht zu werden:

```
[d21 d22 d23] (2);
```

Es zeigt sich wie erwartet, dass der Fit des Baseline-Change-Modells mit invarianten Ladungen und Intercepts identisch zum Fit des LS-Modells mit starker Messinvarianz ist: $\chi^2 = 5,95$, $df = 9$, $p = 0,75$; $CFI = 1,00$; $RMSEA = 0,00$; $SRMR = 0,006$ (vgl. Abschnitt 4.1.4; der vollständige Output kann der Begleit-CD entnommen werden).

In der Tabelle MODEL RESULTS können wir erkennen, dass sich an den Parametern des Messmodells (den Ladungen, Intercepts und Fehlervarianzen) im Vergleich zum LS-Modell mit starker Invarianz nichts verändert hat. Der Output liefert uns nun zusätzlich die Kovarianzen (unter MODEL RESULTS) bzw. die Korrelationen (unter STANDARDIZED MODEL RESULTS) der latenten Differenzvariablen mit dem Ausgangswert sowie die Korrelation zwischen den beiden Differenzvariablen.

MODEL RESULTS (Baseline-Change-Modell)

		Estimate	S.E.	Two-Tailed	
				Est./S.E.	P-Value
STATE1	BY				
D11		1.000	0.000	999.000	999.000
D21		0.923	0.021	44.840	0.000
STATE2	BY				
D12		1.000	0.000	999.000	999.000
D22		0.923	0.021	44.840	0.000

STATE3	BY			
D13		1.000	0.000	999.000
D23		0.923	0.021	44.840
DIFF2_1	BY			
D11		0.000	0.000	999.000
DIFF3_1	BY			
D11		0.000	0.000	999.000
IS2	BY			
D21		1.000	0.000	999.000
D22		1.000	0.000	999.000
D23		1.000	0.000	999.000
STATE2	ON			
STATE1		1.000	0.000	999.000
DIFF2_1		1.000	0.000	999.000
STATE3	ON			
STATE1		1.000	0.000	999.000
DIFF3_1		1.000	0.000	999.000
STATE2	WITH			
STATE3		0.000	0.000	999.000
DIFF3_1		0.000	0.000	999.000
STATE3	WITH			
DIFF2_1		0.000	0.000	999.000
IS2	WITH			
STATE1		0.000	0.000	999.000
STATE2		0.000	0.000	999.000
STATE3		0.000	0.000	999.000
DIFF2_1		0.000	0.000	999.000
DIFF3_1		0.000	0.000	999.000
DIFF2_1	WITH			
STATE1		-0.024	0.003	-6.914
DIFF3_1	WITH			
STATE1		-0.035	0.004	-8.686
DIFF2_1		0.035	0.003	10.208
Means				
STATE1		0.311	0.013	24.748
DIFF2_1		-0.029	0.010	-2.834
DIFF3_1		-0.025	0.011	-2.186
Intercepts				
D11		0.000	0.000	999.000
D21		-0.004	0.007	-0.545
				0.586

D12	0.000	0.000	999.000	999.000
D22	-0.004	0.007	-0.545	0.586
D13	0.000	0.000	999.000	999.000
D23	-0.004	0.007	-0.545	0.586
Variances				
STATE1	0.088	0.006	15.384	0.000
DIFF2_1	0.048	0.004	12.628	0.000
DIFF3_1	0.068	0.005	13.823	0.000
IS2	0.004	0.001	5.671	0.000
Residual Variances				
D11	0.017	0.002	7.828	0.000
D21	0.010	0.002	5.907	0.000
D12	0.015	0.002	7.807	0.000
D22	0.014	0.002	8.601	0.000
D13	0.011	0.002	6.063	0.000
D23	0.011	0.002	7.001	0.000
STATE2	0.000	0.000	999.000	999.000
STATE3	0.000	0.000	999.000	999.000
STANDARDIZED MODEL RESULTS (STDYX Standardization)				
		Estimate	S.E.	Two-Tailed P-Value
STATE1	BY			
D11		0.916	0.011	81.969 0.000
D21		0.916	0.012	79.229 0.000
STATE2	BY			
D12		0.925	0.010	90.812 0.000
D22		0.896	0.011	80.519 0.000
STATE3	BY			
D13		0.940	0.010	90.846 0.000
D23		0.908	0.011	80.793 0.000
DIFF2_1	BY			
D11		0.000	0.000	999.000 999.000
DIFF3_1	BY			
D11		0.000	0.000	999.000 999.000
IS2	BY			
D21		0.220	0.020	10.962 0.000
D22		0.215	0.020	10.952 0.000
D23		0.222	0.020	10.972 0.000
STATE2	ON			
STATE1		0.998	0.033	30.487 0.000
DIFF2_1		0.740	0.033	22.316 0.000

STATE3	ON			
STATE1		1.016	0.037	27.745
DIFF3_1		0.892	0.037	24.226
				0.000
STATE2	WITH			
STATE3		999.000	999.000	999.000
DIFF3_1		999.000	999.000	999.000
				999.000
STATE3	WITH			
DIFF2_1		999.000	999.000	999.000
				999.000
IS2	WITH			
STATE1		0.000	0.000	999.000
STATE2		999.000	999.000	999.000
STATE3		999.000	999.000	999.000
DIFF2_1		0.000	0.000	999.000
DIFF3_1		0.000	0.000	999.000
				999.000
DIFF2_1	WITH			
STATE1		-0.368	0.041	-8.895
				0.000
DIFF3_1	WITH			
STATE1		-0.457	0.036	-12.669
DIFF2_1		0.619	0.032	19.227
				0.000
Means				
STATE1		1.050	0.054	19.335
DIFF2_1		-0.130	0.046	-2.820
DIFF3_1		-0.096	0.044	-2.181
				0.005
				0.029
Intercepts				
D11		0.000	0.000	999.000
D21		-0.014	0.025	-0.546
D12		0.000	0.000	999.000
D22		-0.013	0.024	-0.546
D13		0.000	0.000	999.000
D23		-0.014	0.025	-0.546
				0.585
Variances				
STATE1		1.000	0.000	999.000
DIFF2_1		1.000	0.000	999.000
DIFF3_1		1.000	0.000	999.000
IS2		1.000	0.000	999.000
				999.000
Residual Variances				
D11		0.160	0.020	7.829
D21		0.113	0.019	5.861
D12		0.144	0.019	7.656
D22		0.150	0.018	8.377
D13		0.117	0.019	6.016
D23		0.127	0.018	6.919
				0.000

STATE2	0.000	999.000	999.000	999.000
STATE3	0.000	999.000	999.000	999.000
R-SQUARE				
Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
D11	0.840	0.020	40.984	0.000
D21	0.887	0.019	46.007	0.000
D12	0.856	0.019	45.406	0.000
D22	0.850	0.018	47.396	0.000
D13	0.883	0.019	45.423	0.000
D23	0.873	0.018	47.442	0.000
Latent Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
STATE2	1.000	999.000	999.000	999.000
STATE3	1.000	999.000	999.000	999.000

Außerdem erhalten wir die latenten Mittelwerte von state1, diff2_1 sowie diff3_1. Wenn wir diese Werte mit den geschätzten Mittelwerten für die latenten Variablen state1, state2 und state3 im LS-Modell mit starker Invarianz ($M_1 = 0,311$, $M_2 = 0,282$, $M_3 = 0,286$) vergleichen, so sehen wir, dass der Mittelwert von diff2_1 ($M_{2-1} = -0,029$) genau der Differenz der Mittelwerte von state2 minus state1 und der Mittelwert von diff3_1 ($M_{3-1} = -0,025$) genau der Mittelwertsdifferenz von state3 minus state1 entspricht. Die negativen Vorzeichen der Differenzmittelwerte zeigen uns, dass Depressivität im Mittel von T1 zu T2 und von T1 zu T3 abgenommen hat.

Anhand der geschätzten *p*-Werte für die Mittelwerte der Differenzvariablen können wir erkennen, dass diese Mittelwertsunterschiede signifikant von Null verschieden sind (Alpha = 0,05, zweiseitiger Test). Die *z*-Werte liegen bei -2,834 (*p* = 0,005) und -2,186 (*p* = 0,029). Somit liegt eine statistisch bedeutsame Abnahme der mittleren Depressivität über die Zeit vor.

Die geschätzten Varianzen für state1 (0,088) sowie für diff2_1 (0,048) und diff3_1 (0,068) zeigen uns, dass es interindividuelle Unterschiede im Depressivitätsausgangswert sowie in den latenten Veränderungswerten gibt. Man beachte, dass die *R*²-Werte für die beiden latenten State-Variablen state2 und state3 gleich 1 sind, da die beiden Variablen im Modell perfekt determiniert sind.

Abbildung 4.25 zeigt das model-Statement zur Spezifikation des Neighbor-Change-Modells für dasselbe Datenbeispiel. Die Spezifikation der latenten Differenzvariable diff2_1 ist unverändert. Diese Variable repräsentiert nach wie vor latente Veränderung vom ersten zum zweiten Messzeitpunkt. Die zweite latente Differenzvariable heißt nun diff3_2, um zu verdeutlichen, dass in der Neighbor-Change-Version latente Veränderung zwischen dem zweiten und dem dritten Messzeitpunkt betrachtet wird. Gemäß der Spezifikation in Abbildung 4.23B stehen auf der rechten Seite des on-Statements für die Variable state3 nun drei Komponenten: state1, diff2_1 und diff3_2. Alle Pfade des on-

Befehls werden wiederum mit @ auf 1 fixiert. Die übrigen Befehle sind analog zur Spezifikation des Baseline-Change-Modells.

```

model: state1 by d11
        d21 (1); ! Gleichsetzung der State-Ladungen des 2. Indikators
state2 by d12
        d22 (1);
state3 by d13
        d23 (1);

diff2_1 by d11@0; ! Einführung der Variablennamen für die
diff3_2 by d11@0; ! latenten Differenzvariablen

state2 on state1@1 diff2_1@1; ! (perfekte) Regression des zweiten States
state2@0; ! auf Ausgangswert T1 und Veränderungswert T2-T1

state3 on state1@1 diff2_1@1 diff3_2@1; ! (perfekte) Regression des dritten States
state3@0; ! auf Ausgangswert T1, Veränderungswert T2-T1 und Veränderungswert T3-T2

state2 with state3@0 diff3_2@0;
state3 with diff2_1@0;

! Indikatorsspezifischer (Residual-)Faktor für den 2. Indikator
is2 by d21
        d22@1 ! Fixierung aller Ladungen auf
        d23@1; ! dem indikatorsspezifischen Faktor auf 1

is2 with state1-state3@0 diff2_1@0 diff3_2@0; ! Indikatorsspez. Faktor ist mit den
                                                ! übrigen Faktoren unkorreliert

[d11@0 d12@0 d13@0]; ! Fixierung der Intercepts des ersten Indikators auf 0

[state1 diff2_1 diff3_2]; ! Schätzung des State-Mittelwertes zu T1 sowie der
                           ! Mittelwerte der latenten Differenzvariablen

[d21 d22 d23] (2); ! Gleichsetzung der Intercepts des 2. Indikators

```

Abbildung 4.25: Mplus-Modellspezifikation zur Schätzung des Neighbor-Change-Modells aus Abbildung 4.23B. Es wird zusätzlich ein indikatorsspezifischer Faktor für den zweiten Indikator spezifiziert, welcher in Abbildung 4.23 aus Platzgründen nicht dargestellt ist.

Der Modellfit des Neighbor-Change-Modells entspricht wiederum dem Fit des LS-Modells mit starker Messinvarianz sowie dem Fit der Baseline-Change-Version (siehe Output auf der Begleit-CD). Die Parameterschätzungen für das Messmodell (Ladungen, Intercepts, Fehlervarianzen) sind ebenfalls identisch zu den Schätzwerten im LS-Modell sowie im Baseline-Change-Modell. Im Strukturmodell erhalten wir nun die Kovarianzen bzw. Korrelationen zwischen der Ausgangsvariable `state1` und den Differenzvariablen `diff2_1` und `diff3_2` sowie die Kovarianzen bzw. Korrelationen zwischen `diff2_1` und `diff3_2`. Außerdem gibt Mplus die Mittelwerte von `state1`, `diff2_1` sowie `diff3_2` aus. Die Mittelwerte von `state1` und `diff2_1` sind identisch zu den Werten in der Baseline-Change-Version ($M_1 = 0,311$, $M_{2-1} = -0,029$). Der Mittelwert der Differenzvariable `diff3_2` wird auf 0,004 geschätzt. Dies entspricht genau der Differenz zwischen den Mittelwerten von `state3` minus `state2` im LS-Modell ($0,286 - 0,282$). Dieser Mittelwertsunterschied ist nicht bedeutsam von Null verschieden ($z = 0,375$, $p = 0,707$). Daraus können wir schließen, dass sich die mittlere latente Depressivität zwischen dem zweiten und dem dritten Messzeitpunkt nicht bedeutsam verändert hat.

Außerdem erhalten wir wiederum Schätzwerte für die Varianzen von state1 (0,088), diff2_1 (0,048) sowie diff3_2 (0,045). Diese zeigen uns, dass sich die Kinder in ihren Ausgangswerten sowie ihren Veränderungswerten unterscheiden. Man könnte in weiteren Analyseschritten zusätzliche Variablen (Kovariaten, z.B. Veränderungen hinsichtlich Kompetenzen, vgl. autoregressives Modell in Abschnitt 4.3.2) in das Modell aufnehmen, die diese interindividuellen Unterschiede im Ausgangswert und in den Veränderungswerten möglicherweise aufklären können (siehe z.B. Steyer et al., 2000).

MODEL RESULTS (Neighbor-Change-Modell)

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
STATE1	BY				
	D11	1.000	0.000	999.000	999.000
	D21	0.923	0.021	44.841	0.000
STATE2	BY				
	D12	1.000	0.000	999.000	999.000
	D22	0.923	0.021	44.841	0.000
STATE3	BY				
	D13	1.000	0.000	999.000	999.000
	D23	0.923	0.021	44.841	0.000
DIFF2_1	BY				
	D11	0.000	0.000	999.000	999.000
DIFF3_2	BY				
	D11	0.000	0.000	999.000	999.000
IS2	BY				
	D21	1.000	0.000	999.000	999.000
	D22	1.000	0.000	999.000	999.000
	D23	1.000	0.000	999.000	999.000
STATE2	ON				
STATE1		1.000	0.000	999.000	999.000
DIFF2_1		1.000	0.000	999.000	999.000
STATE3	ON				
STATE1		1.000	0.000	999.000	999.000
DIFF2_1		1.000	0.000	999.000	999.000
DIFF3_2		1.000	0.000	999.000	999.000
STATE2	WITH				
STATE3		0.000	0.000	999.000	999.000
DIFF3_2		0.000	0.000	999.000	999.000

STATE3 WITH				
DIFF2_1	0.000	0.000	999.000	999.000
IS2 WITH				
STATE1	0.000	0.000	999.000	999.000
STATE2	0.000	0.000	999.000	999.000
STATE3	0.000	0.000	999.000	999.000
DIFF2_1	0.000	0.000	999.000	999.000
DIFF3_2	0.000	0.000	999.000	999.000
DIFF2_1 WITH				
STATE1	-0.024	0.003	-6.914	0.000
DIFF3_2 WITH				
STATE1	-0.011	0.003	-3.665	0.000
DIFF2_1	-0.013	0.003	-4.865	0.000
Means				
STATE1	0.311	0.013	24.748	0.000
DIFF2_1	-0.029	0.010	-2.834	0.005
DIFF3_2	0.004	0.010	0.375	0.707
Intercepts				
D11	0.000	0.000	999.000	999.000
D21	-0.004	0.007	-0.545	0.586
D12	0.000	0.000	999.000	999.000
D22	-0.004	0.007	-0.545	0.586
D13	0.000	0.000	999.000	999.000
D23	-0.004	0.007	-0.545	0.586
Variances				
STATE1	0.088	0.006	15.384	0.000
DIFF2_1	0.048	0.004	12.628	0.000
DIFF3_2	0.045	0.004	12.559	0.000
IS2	0.004	0.001	5.671	0.000
Residual Variances				
D11	0.017	0.002	7.828	0.000
D21	0.010	0.002	5.907	0.000
D12	0.015	0.002	7.807	0.000
D22	0.014	0.002	8.601	0.000
D13	0.011	0.002	6.063	0.000
D23	0.011	0.002	7.001	0.000
STATE2	0.000	0.000	999.000	999.000
STATE3	0.000	0.000	999.000	999.000
STANDARDIZED MODEL RESULTS (STDYX Standardization)				
		Two-Tailed		
Estimate		S.E.	Est./S.E.	P-Value

STATE1	BY				
D11		0.916	0.011	81.969	0.000
D21		0.916	0.012	79.230	0.000
STATE2	BY				
D12		0.925	0.010	90.813	0.000
D22		0.896	0.011	80.519	0.000
STATE3	BY				
D13		0.940	0.010	90.846	0.000
D23		0.908	0.011	80.793	0.000
DIFF2_1	BY				
D11		0.000	0.000	999.000	999.000
DIFF3_2	BY				
D11		0.000	0.000	999.000	999.000
IS2	BY				
D21		0.220	0.020	10.962	0.000
D22		0.215	0.020	10.952	0.000
D23		0.222	0.020	10.972	0.000
STATE2	ON				
STATE1		0.998	0.033	30.487	0.000
DIFF2_1		0.740	0.033	22.316	0.000
STATE3	ON				
STATE1		1.016	0.037	27.745	0.000
DIFF2_1		0.753	0.036	20.660	0.000
DIFF3_2		0.729	0.033	22.010	0.000
STATE2	WITH				
STATE3		999.000	999.000	999.000	999.000
DIFF3_2		999.000	999.000	999.000	999.000
STATE3	WITH				
DIFF2_1		999.000	999.000	999.000	999.000
IS2	WITH				
STATE1		0.000	0.000	999.000	999.000
STATE2		999.000	999.000	999.000	999.000
STATE3		999.000	999.000	999.000	999.000
DIFF2_1		0.000	0.000	999.000	999.000
DIFF3_2		0.000	0.000	999.000	999.000
DIFF2_1	WITH				
STATE1		-0.368	0.041	-8.895	0.000
DIFF3_2	WITH				
STATE1		-0.179	0.047	-3.815	0.000
DIFF2_1		-0.277	0.048	-5.730	0.000

Means				
STATE1	1.050	0.054	19.335	0.000
DIFF2_1	-0.130	0.046	-2.820	0.005
DIFF3_2	0.017	0.046	0.375	0.707
Intercepts				
D11	0.000	0.000	999.000	999.000
D21	-0.014	0.025	-0.546	0.585
D12	0.000	0.000	999.000	999.000
D22	-0.013	0.024	-0.546	0.585
D13	0.000	0.000	999.000	999.000
D23	-0.014	0.025	-0.546	0.585
Variances				
STATE1	1.000	0.000	999.000	999.000
DIFF2_1	1.000	0.000	999.000	999.000
DIFF3_2	1.000	0.000	999.000	999.000
IS2	1.000	0.000	999.000	999.000
Residual Variances				
D11	0.160	0.020	7.829	0.000
D21	0.113	0.019	5.860	0.000
D12	0.144	0.019	7.656	0.000
D22	0.150	0.018	8.377	0.000
D13	0.117	0.019	6.016	0.000
D23	0.127	0.018	6.919	0.000
STATE2	0.000	999.000	999.000	999.000
STATE3	0.000	999.000	999.000	999.000
R-SQUARE				
Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
D11	0.840	0.020	40.984	0.000
D21	0.887	0.019	46.007	0.000
D12	0.856	0.019	45.406	0.000
D22	0.850	0.018	47.396	0.000
D13	0.883	0.019	45.423	0.000
D23	0.873	0.018	47.442	0.000
Latent Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
STATE2	1.000	999.000	999.000	999.000
STATE3	1.000	999.000	999.000	999.000

Der besondere Reiz von LC-Modellen besteht darin, dass in diesen messfehlerbereinigte Veränderungsvariablen analysiert und zu anderen Variablen in Beziehung gesetzt werden können. So kann beispielsweise untersucht werden, welche Variablen geeignet sind zu er-

klären, warum die Depressivität bei manchen Kindern über die Zeit stärker abnimmt als bei anderen. Außerdem können Veränderungsvariablen auch als unabhängige oder als Mediatorvariablen eingesetzt werden.

Technisch können derartige Analysen einfach durchgeführt werden, indem weitere Variablen und zusätzliche on-Befehle aufgenommen werden, in welchen die Differenzvariablen als abhängige oder unabhängige Variablen spezifiziert werden. Zum Beispiel könnte man untersuchen, ob es Geschlechtsunterschiede in der latenten Veränderung zwischen T1 und T2 gibt, indem man die Variable `diff2_1` auf Geschlecht zurückführt (sofern Geschlecht als Variable mit im Datensatz enthalten wäre; in unserem Beispiel ist dies nicht der Fall). Die Spezifikation könnte dann z.B. folgendermaßen aussehen, wobei `geschl` eine dichotome (Gruppen-)Variable (z.B. dummykodiert) ist, die das Geschlecht indiziert (es könnten ebenso gut auch kontinuierliche Variablen als Prädiktoren verwendet werden):

```
diff2_1 on geschl;
```

Ein Beispiel für eine Latent-Change-Analyse mit multiplen Konstrukten findet man bei Reuter, Ziegelmann, Wiedemann, Geiser, Lippke, Schüz & Schwarzer (2010).

4.5 Latente Wachstumskurvenmodelle

Latente Wachstumskurvenmodelle (*latent growth curve models*, LGCM; Bollen & Curran, 2006; Duncan et al., 2006; Meredith & Tisak, 1984, 1990) werden in den Sozialwissenschaften zunehmend häufiger zur Analyse von Längsschnittdaten eingesetzt. Im Gegensatz zu den im vorigen Abschnitt besprochenen LC-Modellen steht bei LGCM die Untersuchung der *Form* der Veränderung über die Zeit im Vordergrund. Die zentrale Frage bei LGCM ist, mit welcher Funktion sich die individuellen Wachstumskurven von Personen beschreiben lassen (z.B. lineares oder u-förmiges Wachstum). Wichtige Forschungsfragestellungen, die mit LGCM beantwortet werden können, sind unter anderem:

- Wie groß ist der latente Mittelwert von Personen hinsichtlich eines Konstruktions zum ersten Messzeitpunkt (mittlerer Ausgangswert)?
- Gibt es interindividuelle Unterschiede im Ausgangswert („Intercept“) und wenn ja, wie groß sind diese?
- Findet eine Veränderung des Merkmals über die Zeit statt und wenn ja, welche Form und Stärke weist diese Veränderung auf? Liegt im Mittel ein Anstieg oder eine Abnahme über die Zeit vor? Ist die Veränderung z.B. linear oder kurvenlinear?
- Wie groß ist der mittlere Anstieg („Slope“) bzw. Abfall des Merkmals über die Zeit?
- Gibt es interindividuelle Unterschiede hinsichtlich der Stärke des Anstiegs oder Abfalls und wenn ja, wie groß sind diese?
- Hängen Ausgangswert und Veränderungswert/Slope miteinander zusammen? Wenn ja, welche Richtung und Stärke weist dieser Zusammenhang auf?
- Gibt es andere Variablen („Kovariaten“), die Unterschiede im Ausgangs- und/oder im Veränderungswert/Slope erklären können?

Somit sind die Fragestellungen, die mit LGCM untersucht werden können, sehr ähnlich und teilweise sogar identisch zu denen, die man mit LC-Modellen (vgl. Kapitel 4.4) analysieren kann. Beispielsweise ist man auch im Rahmen einer LC-Analyse an Mittelwertsveränderungen über die Zeit, am Ausmaß interindividueller Unterschiede in Ausgangs- und Veränderungswerten sowie an Zusammenhängen zwischen Ausgangs- und Veränderungswerten untereinander und mit anderen Variablen interessiert.

Der wesentliche Unterschied zwischen LC-Modellen und LGCM besteht darin, dass in LGCM normalerweise explizite Annahmen über die *Form der Veränderung* über die Zeit (linear, quadratisch usw.) gemacht werden, was in LC-Modellen nicht der Fall ist. In diesem Sinne sind LC-Modelle weniger restriktiv, da die Veränderung über die Zeit in diesen Modellen keiner bestimmten Form folgen muss. Wir besprechen zunächst die Theorie und praktische Anwendung von latenten Wachstumskurvenmodellen 1. Ordnung. In Abschnitt 4.5.2 werden wir LGCM 2. Ordnung kennen lernen.

4.5.1 Latente Wachstumskurvenmodelle 1. Ordnung

LGCM 1. Ordnung basieren pro Konstrukt auf einem einzelnen wiederholt gemessenen Indikator (z.B. dem mehrfach erhobenen Summenwert einer Ängstlichkeitsskala), welcher direkt als Indikator für einen oder mehrere latente Wachstumsfaktoren dient. LGCM 2. Ordnung können (und sollten!) dagegen dann eingesetzt werden, wenn für jeden Messzeitpunkt mehrere Indikatoren pro Konstrukt vorliegen, wobei die Indikatoren zu jedem Messzeitpunkt dieselben sein sollten. LGCM 2. Ordnung weisen einige wichtige Vorteile gegenüber LGCM 1. Ordnung auf, die wir in Abschnitt 4.5.2 besprechen werden.

4.5.1.1 Analyse linearer Wachstumskurvenmodelle 1. Ordnung

Der Einfachheit halber betrachten wir zunächst ein *lineares* LGCM 1. Ordnung. In Abschnitt 4.5.1.3 wird die Erweiterung zu einem quadratischen LGCM gezeigt. In Abbildung 4.26 sehen wir ein lineares LGCM 1. Ordnung für ein Konstrukt, welches zu vier Messzeitpunkten jeweils mit derselben beobachteten Variable Y_k erhoben wurde, wobei der Index k den Messzeitpunkt angibt.

Box 4.12: Fallbeispiel Wachstumskurvenmodelle 1. Ordnung

Um die Modellierung von Längsschnittdaten mit Hilfe von LGCM 1. Ordnung zu veranschaulichen, stellen wir uns vor, wir hätten die Ängstlichkeit von Kindern zu vier Messzeitpunkten T1-T4 mit demselben Ängstlichkeitsfragebogen gemessen. Weiterhin nehmen wir an, dass die zeitlichen Abstände zwischen den Messungen jeweils gleich groß waren (z.B. eine Messung alle 3 Monate). Bezogen auf Abbildung 4.26 würden die Variablen Y_1-Y_4 die wiederholt erfassten Ängstlichkeitsscores der Kinder repräsentieren.

Als Beispiel für die Analyse latenter Wachstumskurvenmodelle 1. Ordnung in Mplus dient uns der Datensatz **aengstlichkeit.dat**. Er enthält die Individualdaten von $N = 485$ Kindern auf vier manifesten Variablen (a1-a4). Die vier Variablen repräsentieren Ängstlichkeitsselbstberichte (Fragebogensummenscores) der Kinder, welche zu vier Messzeitpunkten erhoben wurden. Der Datensatz enthält keine fehlenden Werte.

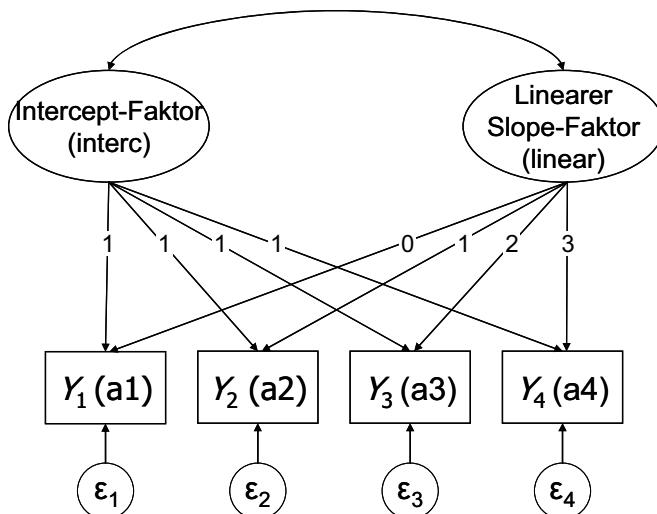


Abbildung 4.26: Latentes Wachstumskurvenmodell 1. Ordnung für eine beobachtete Variable Y_k ($k = \text{Messgelegenheit}$), welche zu vier Messgelegenheiten ($k = 1, 2, 3, 4$) erhoben wurde. ε_k : Messfehlervariable. Das Modell nimmt gleiche zeitliche Abstände der Messungen und lineares Wachstum an. In Klammern sind die in Mplus verwendeten Variablennamen angegeben.

Im linearen LGCM in Abbildung 4.26 haben zwei latente Faktoren einen direkten Einfluss auf die gemessenen Variablen Y_1-Y_4 . Dies sind die sogenannten latenten Wachstums-(Growth-) Faktoren „Intercept-Faktor“ und „Linearer Slope-Faktor“. Zusätzlich nimmt das LGCM noch an, dass alle Messungen von zufälligen Messfehlern beeinflusst werden. Diese werden durch die Messfehlervariablen $\varepsilon_1-\varepsilon_4$ repräsentiert, welche ebenfalls einen direkten Effekt auf die beobachteten Variablen Y_1-Y_4 haben.

In Abbildung 4.26 ist zu erkennen, dass für alle Faktorladungen in diesem Modell *a priori* bestimmte Werte angenommen werden (d.h. die Ladungen werden nicht als freie Parameter geschätzt, sondern sind *a priori* fixiert). Das liegt daran, dass die Form des Veränderungsprozesses in LGCM zumeist *vor* der Analyse gemäß theoretischer Annahmen festgelegt und dann empirisch überprüft wird. Wie nachfolgend noch genauer erklärt wird, wurden die Ladungen in diesem Fall so gewählt, dass der Slopefaktor *lineares* Wachstum repräsentiert.

Wir können erkennen, dass alle beobachteten Ängstlichkeitsvariablen im Modell durch den latenten Interceptfaktor beeinflusst werden. Dieser Faktor repräsentiert interindividuelle Unterschiede in den Ausgangswerten (d.h. hinsichtlich der latenten Ängstlichkeit zum ersten Messzeitpunkt = T1). Dies wird dadurch erreicht, dass die Ladungen aller beobachteten Variablen auf dem Interceptfaktor auf 1 und alle beobachteten Intercepts auf 0 fixiert sind.

Der Interceptfaktor hat somit eine klare Bedeutung: Es handelt sich um die *True-Score-Variable* (τ_1) von Y_1 , also um die Variable, die die „wahren“ (messfehlerfreien) Ängstlichkeitswerte der Kinder zum ersten Messzeitpunkt repräsentiert. Dies folgt daraus, dass die Variable Y_1 ein Intercept von 0, eine Ladung von 1 auf dem Interceptfaktor und eine Ladung von 0 auf dem Slope-Faktor aufweist. Bezogen auf unser Beispiel repräsentiert der latente Intercept-Faktor also die messfehlerbereinigten Ängstlichkeitswerte der Kinder zu T1.

Der Faktor, der in Abbildung 4.26 mit „Linear Slope-Faktor“ bezeichnet ist, repräsentiert im Gegensatz zum Interceptfaktor wahre interindividuelle Unterschiede in Hinblick auf *lineare Veränderung* über die Zeit. Auch dieser Faktor hat eine klare Bedeutung: Er repräsentiert die *Differenz der True-Score-Variablen* der Indikatoren Y_1 und Y_2 , kann also als latente True-Score-Differenzvariable ($\tau_2 - \tau_1$) interpretiert werden.

In diesem Beispiel steht der Slopefaktor wie bereits erwähnt für *lineares* Wachstum. Dies ergibt sich daraus, dass die Variablen Y_1-Y_4 Faktorladungen von 0, 1, 2 und 3 auf dem Slopefaktor aufweisen. Da wir in unserem Beispiel gleiche zeitliche Abstände zwischen den einzelnen Messungen annehmen, impliziert dieses Ladungsmuster die Annahme der Linearität der Veränderung.

Bezogen auf das Ängstlichkeitsbeispiel zeigt ein Kind mit einem hohen positiven Wert auf dem Slopefaktor einen größeren linearen Anstieg seiner Ängstlichkeit über die Zeit als ein Kind mit einem geringeren positiven Slopefaktorwert. Kinder mit negativen Slopefaktorwerten zeigen abnehmende Ängstlichkeitswerte über die Zeit. Der Doppelpfeilbogen, welcher zwischen dem Intercept- und Slopefaktor eingezeichnet ist, symbolisiert die Kovarianz bzw. Korrelation zwischen den Wachstumsfaktoren (Zusammenhang zwischen Ausgangswert und linearer Veränderung), die im Modell ebenfalls geschätzt werden kann.

Wie können wir nun mit Hilfe des linearen LGCM die oben diskutierten Fragestellungen untersuchen? Der Mittelwert des Interceptfaktors liefert uns eine Antwort auf die Frage nach dem mittleren latenten Ängstlichkeitsniveau in der Stichprobe zum ersten Messzeitpunkt, während die Varianz dieses Faktors ein Maß für die Stärke interindividueller Unterschiede im latenten Ängstlichkeitsausgangswert darstellt. Ein hoher Mittelwert des Interceptfaktors deutet auf ein hohes latentes Ängstlichkeitsniveau zu T1 hin. Eine große Varianz dieses Faktors bedeutet, dass es große Unterschiede zwischen Kindern hinsichtlich der latenten Ängstlichkeitsausgangswerte gibt.

Im Gegensatz dazu geben Mittelwert und Varianz des Slopefaktors Aufschluss über den mittleren linearen Anstieg oder Abfall der latenten Ängstlichkeit der Kinder über die Zeit sowie die Frage nach dem Ausmaß interindividueller Unterschiede im linearen Anstieg bzw. Abfall über die Zeit. Ein *positiver* Mittelwert des Slopefaktors würde darauf hinweisen, dass die Ängstlichkeit der Kinder im Durchschnitt (d.h. über alle individuellen Wachstumskurven hinweg gemittelt) über die Zeit *linear zunimmt*. Im Gegensatz dazu würde ein *negativer* Mittelwert des Slopefaktors anzeigen, dass die Ängstlichkeit im Mittel über die Zeit *linear abnimmt*.

Die Frage nach dem Zusammenhang zwischen latenter Ausgangswert und Slope lässt sich über die Kovarianz bzw. Korrelation des Intercept- und Slopefaktors beantworten. Eine *positive* Korrelation zwischen Intercept- und Slope-Faktor bedeutet, dass Kinder mit *höheren* latenten Ängstlichkeitswerten zu T1 tendenziell einen *stärkeren Anstieg* ihrer Ängstlichkeit zeigen, während bei Kindern mit geringeren Ausgangswerten eher ein *Rückgang* der Ängstlichkeit zu erwarten ist. Eine negative Korrelation würde dagegen bedeuten, dass *höhere* Ausgangswerte im Schnitt mit einem *Rückgang* verbunden sind, während *geringere* T1-Werte tendenziell mit einem *Anstieg* assoziiert wären.

The screenshot shows the Mplus software interface with the title bar 'Mplus - [2_First_Order_Linear_Growth.inp]'. The menu bar includes File, Edit, View, Mplus, Graph, Window, and Help. Below the menu is a toolbar with various icons. The main window displays the Mplus input syntax:

```

title: Lineares latentes Wachstumskurvenmodell 1. Ordnung
       Individualdaten aus Datei "aengstlichkeit.dat"
       4 beobachtete Variablen:
       Ängstlichkeits-Selbstbericht gemessen mit 1 Indikator
       zu 4 Messgelegenheiten
       keine fehlenden Werte

data: file = aengstlichkeit.dat;

variable: names = a1-a4;

model: interc linear | a1@0 a2@1 a3@2 a4@3; ! spezielle Syntax für LGCM

output: sampstat standardized stdyx;

plot: type = plot3; ! Liefert graph. Darstellung der Wachstumskurven
      series = a1 (linear) a2 (linear) a3 (linear) a4 (linear);

```

The status bar at the bottom indicates 'Ready' and 'Ln 10, Col 25'.

Abbildung 4.27: Mplus-Input-Datei zur Spezifikation des linearen LGCM 1. Ordnung für Ängstlichkeit gemessen mit einem Indikator zu 4 Messzeitpunkten (vgl. Abbildung 4.26) unter Verwendung der speziellen Syntax für LGCM.

Die Frage nach dem Einfluss anderer Variablen (z.B. Alter, Geschlecht, Persönlichkeitsvariablen, sozioökonomischer Status etc.) auf den Veränderungsprozess lässt sich durch die Aufnahme von Kovariaten in das Modell untersuchen. Die Regression der Growthfaktoren auf mögliche Kovariaten ist – ähnlich wie die Analyse von Kovariaten in LC-Modellen (sie-

he Abschnitt 4.4) – problemlos möglich. Darüber hinaus können die Growthfaktoren in Modellerweiterungen auch als unabhängige Variablen fungieren. Zum Beispiel könnte es interessant sein, zu klären, ob eine lineare Veränderung in der Ängstlichkeit Veränderungen in anderen psychiatrisch relevanten Symptomen (z.B. Depressivität) vorhersagen kann. Abbildung 4.27 zeigt den Mplus-Input-File zur Spezifikation des linearen LGCM 1. Ordnung für die vier Ängstlichkeitsvariablen a1, a2, a3 und a4.

Aus dem model-Befehl ist ersichtlich, dass in Mplus eine spezielle Syntax für Wachstumskurvenmodelle zur Verfügung steht, welche die Spezifikation dieser Modelle vereinfacht:

```
model: interc linear | a1@0 a2@1 a3@2 a4@3;
```

Die latenten Variablen `interc` und `linear` repräsentieren den Intercept- und den linearen Slope-Faktor. Der senkrechte Strich (das „|“-Zeichen) hinter `interc` und `linear` dient in Mplus als Symbol für *Random Slopes* (vgl. Kapitel 5). Random Slope bedeutet im Zusammenhang mit LGCM, dass der Anstieg bzw. Abfall der latenten Wachstumskurve für jede Person unterschiedlich ausfallen kann (d.h. es kann Variabilität zwischen Personen in den Slope-Faktor-Werten geben).

Ebenso verhält es sich in Hinblick auf den latenten Ausgangswert (den Intercept-Faktor-Wert), welcher ebenfalls für jede Person unterschiedlich sein kann. Daher spricht man auch von *Random Intercept*. Die Begriffe *Random Intercept* und *Random Slope* sowie das „|“-Zeichen werden uns in Abschnitt 5.5 bei der Besprechung von Mehrebenenregressionsmodellen mit *Random Coefficients* (Random Intercepts und Random Slopes) erneut begegnen.

Hinter dem „|“-Zeichen werden die wiederholt gemessenen Variablen aufgelistet, für welche latentes Wachstum modelliert werden soll. In diesem Fall handelt es sich um die vier Ängstlichkeitsscores a1-a4. Die @-Zeichen beziehen sich auf die zu fixierenden Ladungen der vier Indikatoren auf dem Slope-Faktor. Gemäß der Annahme linearen Wachstums wird die Slope-Faktorladung des ersten Indikators (a1) auf 0 fixiert, die Slope-Ladung von a2 auf 1, die Slope-Ladung von a3 auf 2 und die Slope-Ladung von a4 auf 3 (vgl. Abb. 4.26).

Mplus fixiert bei dieser Spezifikation automatisch alle Ladungen auf dem Intercept-Faktor (`interc`) auf 1 und setzt alle Intercept-Parameter der beobachteten Variablen auf 0. Zudem schätzt Mplus automatisch die Mittelwerte der latenten Wachstumsfaktoren `interc` und `linear`, ohne dass man dies gesondert anfordern müsste (bei Verwendung der konventionellen Mplus-Syntax [s.u.] müssen latente Mittelwerte angefordert und die Intercepts der manifesten Variablen explizit auf 0 fixiert werden).

Box 4.13: Spezifikation von Wachstumskurvenmodellen mittels konventioneller Mplus-Syntax-Befehle für CFAs

Man beachte, dass LGCM in Mplus auch auf konventionelle Weise, also mit Hilfe der normalen CFA-Spezifikation unter Verwendung von by-Befehlen spezifiziert werden können (da es sich ja bei LGCM auch um konfirmatorische Faktorenmodelle handelt). Dies ist allerdings etwas komplizierter, da man bei der konventionellen Spezifikation selbst daran denken muss, die Intercepts aller Indikatoren auf Null zu fixieren und die Schätzung der latenten Mittelwerte gesondert anzufordern. Diese Schritte vollzieht Mplus bei Verwendung der obigen Spezial-Syntax (siehe Abb. 4.27) automatisch. Nachfolgend zeigen wir die Befehle der „konventionellen“ Spezifikation (die zu denselben Ergebnissen führen). Der entsprechende vollständige Input befindet sich auf der Begleit-CD.

```
model: interc by a1-a4@1; ! Intercept-Faktor
       linear by a1@0 a2@1 a3@2 a4@3; ! Slope-Faktor

       [a1-a4@0]; ! Intercepts der Indikatoren auf Null fixieren

       [interc linear]; ! Mittelwerte des Intercept- und Slope-Faktors anfordern
```

Die plot-Option ist bei der Analyse von LGCM in Mplus sehr nützlich, da sie u.a. die beobachteten und die modellimplizierten individuellen und mittleren Wachstumskurven liefert. Dies ist zur Analyse der individuellen Verläufe sowie zur Veranschaulichung der geschätzten Modellparameter des LGCM sehr hilfreich. Man beachte, dass die plot-Option nur bei Verwendung von Individual-Daten (nicht jedoch bei Summary-Daten) zur Veranschaulichung der individuellen Wachstumskurven verwendet werden kann (der Unterschied zwischen Individual- und Summary-Daten wurde in Kapitel 2 erklärt).

Zu diesem Zweck wird der Plot-Typ 3 angefordert (TYPE = PLOT3) und mit Hilfe der SERIES-Option spezifiziert, dass die Werte auf den Variablen a1-a4 als eine Funktion der Zeit dargestellt werden sollen („linear“ bezieht sich hierbei auf den Namen für den linearen Slope-Faktor):

```
plot: type = plot3;
      series = a1 (linear) a2 (linear) a3 (linear) a4 (linear);
```

Man erhält so verschiedene Arten von Grafiken, die man nach erfolgreicher Durchführung einer LCGM-Analyse im Mplus-Output unter der Menüoption **Graph → View Graphs** anzeigen lassen kann (s. u.).

Wir wenden uns vor der Betrachtung der Grafiken zunächst den numerischen Ergebnissen im Mplus-Output zu. Anhand der Fit-Indices können wir erkennen, dass das lineare LGCM für die Ängstlichkeitsdaten einen eher mäßigen Datenfit aufweist. Dies wird insbesondere durch den χ^2 -Test und den RMSEA-Koeffizienten angezeigt: $\chi^2 = 27,288$, $df = 5$, $p < 0,001$; $CFI = 0,98$; $RMSEA = 0,096$; $SRMR = 0,061$.

Zum Vergleich kann noch ein „Intercept-Only“-Modell geschätzt werden, welches nur den Intercept-Faktor, jedoch keinen Slope-Faktor enthält. Dieses Modell impliziert gleichbleibende latente Ängstlichkeitswerte über die Zeit, d.h. es handelt sich um ein „No-

Change-Modell“, in welchem die messfehlerfreien Ängstlichkeitswerte über die Zeit weder ansteigen noch abfallen. Anhand eines Vergleichs des Intercept-Only-Modells mit dem linearen LGCM kann man untersuchen, ob das lineare LGCM zumindest besser passt, als ein Modell, das überhaupt keine Veränderung über die Zeit annimmt. Abbildung 4.28 zeigt den Mplus-Input zur Spezifikation des Intercept-Only-Modells.

```

Mplus - [1_First_Order_Intercept_Only.inp]
File Edit View Mplus Graph Window Help
D E F G H I J K L M N P R U V W X Y Z ? 
title: Intercept-Only-Modell 1. Ordnung
      Individualdaten aus Datei "aengstlichkeit.dat"
      4 beobachtete Variablen:
      Ängstlichkeits-Selbstbericht gemessen mit 1 Indikator
      zu 4 Messgelegenheiten
      keine fehlenden Werte

data: file = aengstlichkeit.dat;

variable: names = a1-a4;

model: interc linear | a1@0 a2@1 a3@2 a4@3; ! spezielle Syntax für LGCM
       linear@0; ! Varianz des Slope-Faktors auf Null setzen
       [linear@0]; ! Mittelwert des Slope-Faktors auf Null setzen
       interc with linear@0; ! Kovarianz auf Null setzen

output: sampstat standardized stdyx;

plot: type = plot3; ! Liefert graph. Darstellung der Wachstumskurven
      series = a1 (linear) a2 (linear) a3 (linear) a4 (linear);

```

Abbildung 4.28: Mplus-Input-Datei zur Spezifikation des Intercept-Only-Modells 1. Ordnung für 4 Messzeitpunkte unter Verwendung der speziellen Syntax für LGCM.

Die grundlegende Modellspezifikation des linearen LGCM 1. Ordnung kann zur Spezifikation des Intercept-Only-Modells beibehalten werden. Drei zusätzlich aufgenommene Befehle sorgen dafür, dass Varianz, Mittelwert und Kovarianz des linearen Slope-Faktors auf Null fixiert werden, so dass de facto nur noch der Intercept-Faktor im Modell verbleibt.

Die Plot-Option wird unverändert beibehalten. Man erhält so zusätzlich zu den beobachteten Wachstumskurven auch die vom Intercept-Only-Modell implizierten „Wachstums-“ Kurven zum Vergleich (s. u.).

Die Modellgütestatistiken für das Intercept-Only-Modell zeigen, dass dieses wesentlich schlechter auf die Daten passt, als das lineare LGCM ($\chi^2 = 176,355$, $df = 8$, $p < 0,001$; $CFI = 0,857$; $RMSEA = 0,208$; $SRMR = 0,141$; der vollständige Output befindet sich auf der Begleit-CD). Auch ein Vergleich der AIC-Werte führt zu dem Schluss, dass die Annahme linearer Veränderung über die Zeit der Annahme keiner Veränderung vorzuziehen ist ($AIC_{lineares\ LGCM} = 1153$; $AIC_{Intercept-Only-Modell} = 1296$).

Nachfolgend besprechen wir deshalb trotz des mäßigen Fits die wichtigsten Ergebnisse für das lineare LCGM. Unter MODEL RESULTS sehen wir, dass Mittelwert und Varianz (*Var*) beider Wachstumsfaktoren statistisch signifikant von Null verschieden sind. Somit ist der mittlere latente Ängstlichkeitsausgangswert (der Mittelwert der True-Score-Variable zu T1) bedeutsam von Null verschieden ($M_{\text{Intercept}} = 0,698$, $z = 34,866$, $p < 0,001$). Außerdem gibt es signifikante Variabilität im latenten Ausgangswert ($Var_{\text{Intercept}} = 0,151$, $z = 11,996$, $p < 0,001$).

In Bezug auf den Slope-Faktor ist erkennbar, dass im Mittel eine (lineare) Abnahme der Ängstlichkeit über die Zeit vorliegt, da der Mittelwert des Slope-Faktors negativ ist ($M_{\text{Linear}} = -0,062$, $z = -10,405$, $p < 0,001$). Die mittlere modellimplizierte (lineare) Wachstumskurve fällt also ab. Zudem gibt es interindividuelle Unterschiede in der Stärke des linearen Anstiegs/Abfalls der einzelnen Wachstumskurven ($Var_{\text{Linear}} = 0,007$, $z = 4,667$, $p < 0,001$). Das heißt, die Kinder unterscheiden sich nicht nur in ihrer anfänglichen Ängstlichkeit, sondern auch im Verlauf (Stärke des Anstiegs/Abfalls) ihrer Ängstlichkeitswerte über die Zeit.

MODEL RESULTS

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
INTERC					
A1		1.000	0.000	999.000	999.000
A2		1.000	0.000	999.000	999.000
A3		1.000	0.000	999.000	999.000
A4		1.000	0.000	999.000	999.000
LINEAR					
A1		0.000	0.000	999.000	999.000
A2		1.000	0.000	999.000	999.000
A3		2.000	0.000	999.000	999.000
A4		3.000	0.000	999.000	999.000
LINEAR WITH INTERC		-0.011	0.003	-3.485	0.000
Means					
INTERC		0.698	0.020	34.866	0.000
LINEAR		-0.062	0.006	-10.405	0.000
Intercepts					
A1		0.000	0.000	999.000	999.000
A2		0.000	0.000	999.000	999.000
A3		0.000	0.000	999.000	999.000
A4		0.000	0.000	999.000	999.000
Variances					
INTERC		0.151	0.013	11.996	0.000
LINEAR		0.007	0.001	4.667	0.000
Residual Variances					
A1		0.067	0.008	8.675	0.000
A2		0.048	0.004	10.965	0.000
A3		0.048	0.004	11.180	0.000
A4		0.040	0.006	6.545	0.000

Anhand der standardisierten Modellschätzungen [STANDARDIZED MODEL RESULTS (STDYX Standardization)] sehen wir u.a., dass Intercept- und Slope-Faktor in diesem Beispiel zu $r = -0,349$ korreliert sind ($z = -5,001$, $p < 0,001$) und dass zwischen 69,4 und 78,5% der Varianz der beobachteten Ängstlichkeitsvariablen durch die latenten Wachstumsfaktoren erklärt werden kann (siehe R-SQUARE OBSERVED VARIABLE).

STANDARDIZED MODEL RESULTS (STDYX Standardization)				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
INTERC				
A1	0.833	0.019	43.608	0.000
A2	0.906	0.023	40.204	0.000
A3	0.911	0.030	29.912	0.000
A4	0.903	0.040	22.597	0.000
LINEAR				
A1	0.000	0.000	999.000	999.000
A2	0.193	0.021	9.160	0.000
A3	0.389	0.042	9.259	0.000
A4	0.578	0.065	8.861	0.000
LINEAR WITH				
INTERC	-0.349	0.070	-5.001	0.000
Means				
INTERC	1.796	0.091	19.789	0.000
LINEAR	-0.753	0.108	-6.965	0.000
Intercepts				
A1	0.000	0.000	999.000	999.000
A2	0.000	0.000	999.000	999.000
A3	0.000	0.000	999.000	999.000
A4	0.000	0.000	999.000	999.000
Variances				
INTERC	1.000	0.000	999.000	999.000
LINEAR	1.000	0.000	999.000	999.000
Residual Variances				
A1	0.306	0.032	9.622	0.000
A2	0.264	0.024	11.037	0.000
A3	0.265	0.023	11.747	0.000
A4	0.215	0.031	6.812	0.000
R-SQUARE				
Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
A1	0.694	0.032	21.804	0.000
A2	0.736	0.024	30.789	0.000
A3	0.735	0.023	32.541	0.000
A4	0.785	0.031	24.941	0.000

Man könnte das Modell erweitern, indem man zusätzlich Prädiktorvariablen (Kovariaten) aufnimmt, die möglicherweise erklären können, warum sich Kinder in den Ängstlichkeitsausgangswerten und/oder in der Stärke der linearen Veränderung der Ängstlichkeit über die Zeit unterscheiden. Dies könnten zum Beispiel Variablen wie „elterliche Unterstützung“ oder andere Veränderungsvariablen sein (wie z.B. „Veränderung in personalen Kompetenzen“ etc.). Die Aufnahme von Prädiktorvariablen des Wachstumsprozesses geschieht einfach dadurch, dass die Wachstumsfaktoren in on-Befehlen als abhängige Variablen spezifiziert werden, die auf die Prädiktorvariablen zurück geführt werden, z.B.:

```
interc linear on x y z;
```

wobei die Variablen x, y und z hier für mögliche Prädiktorvariablen der Ausgangs- und der linearen Veränderungswerte stehen.

Andererseits könnte man die Wachstumsfaktoren interc und linear auch als unabhängige Variablen in Bezug auf andere (Veränderungs-)Variablen einsetzen. Sie würden dann nicht links, sondern rechts von einem on-Befehl stehen.

4.5.1.2 Grafische Darstellung von Wachstumskurven

Als nächstes wollen wir auf die Verwendung der plot-Option im Zusammenhang mit LCGM eingehen. Hat man im Input plot3 angefordert, so kann man verschiedene Diagramme zur Visualisierung individueller und mittlerer Wachstumskurven im Output unter **Graph → View graphs** abrufen (siehe Abb. 4.29). Alternativ kann dazu auch auf das Menü-

Symbol  geklickt werden.

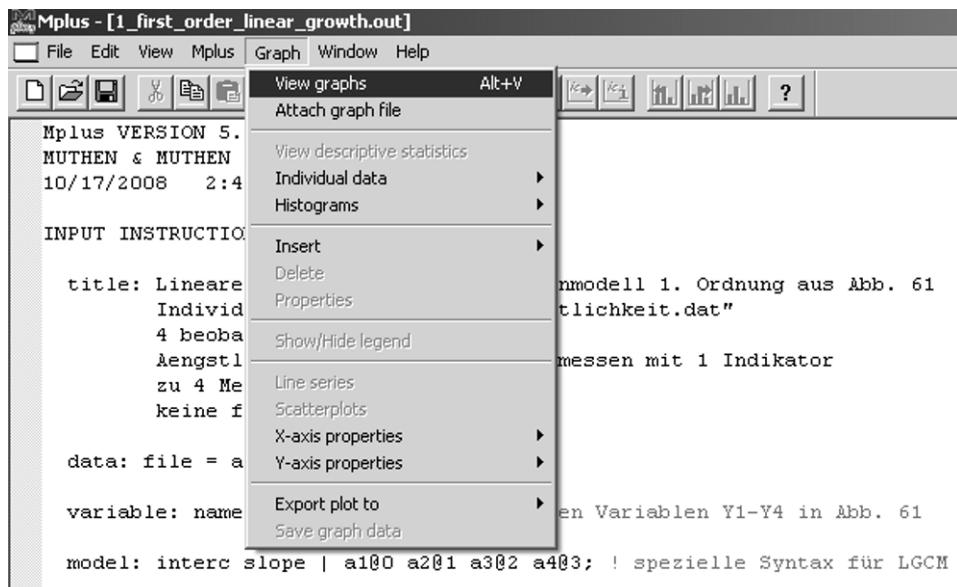


Abbildung 4.29: Öffnen der angeforderten Grafiken im Mplus-Output über **Graph → View Graphs**.

Abbildung 4.30 zeigt verschiedene Optionen: Die Option „Sample means“ liefert die mittlere, auf der Basis der beobachteten Daten berechnete Wachstumskurve (hier: beobachtete Ängstlichkeitsmittelwerte). „Estimated means“ zeigt die mittlere modellimplizierte Wachstumskurve. Mit der Option „Sample and estimated means“ erhält man sowohl die beobachtete als auch die modellimplizierte mittlere Wachstumskurve in einem einzigen Diagramm.

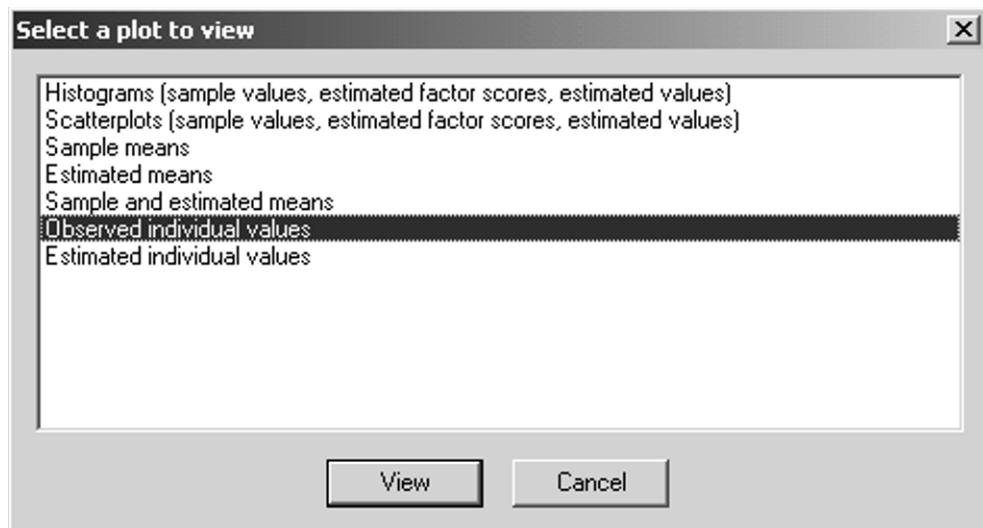


Abbildung 4.30: Auswahl einer bestimmten Grafik. Hier: Beobachtete individuelle Wachstumskurven.

Die Anzeige von „Observed individual values“ ist eine nützliche Option, um die Wachstumskurven für jede einzelne Person inspizieren zu können. Wählt man diese Option, so erscheint das in Abbildung 4.31 gezeigte Dialogfenster „Properties for viewing individual curves“, in welchem man u.a. die Reihenfolge, Anzahl und Art der in einem Diagramm darzustellenden individuellen Wachstumskurven festlegen kann. Die Voreinstellung ist, dass die Wachstumskurven gemäß der Reihenfolge im Datenfile („Consecutive order“) angezeigt werden und dass – zur Wahrung der Übersichtlichkeit – pro Diagramm nur zehn Kurven dargestellt werden („Number of curves: 10“). Außerdem werden standardmäßig die beobachteten Werte gezeigt („Type of curves for observed data → Individual data“). Es wird keine besondere Anpassung vorgenommen, diese könnte man jedoch unter „Type of curves for observed data → Individually-fitted curves“ auswählen.

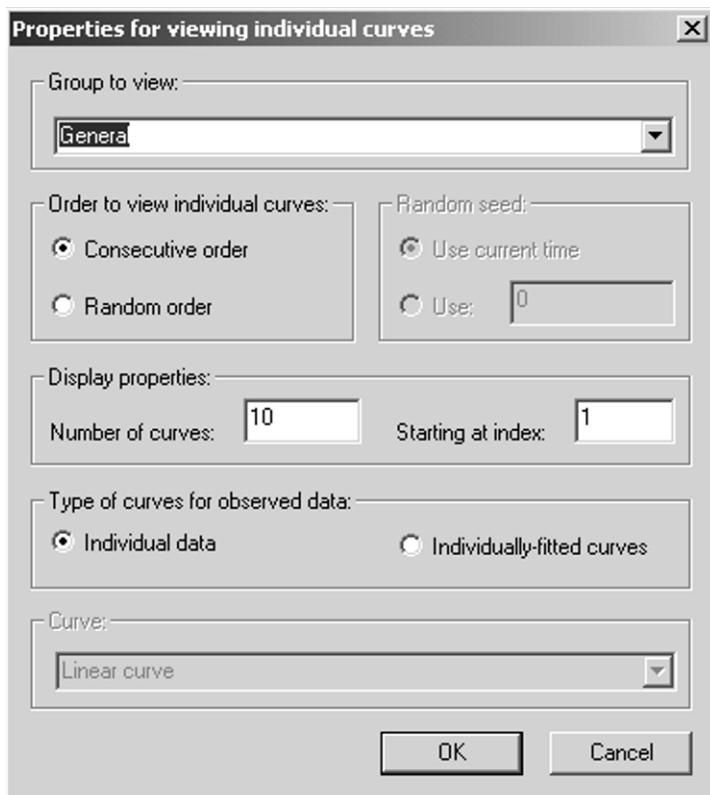


Abbildung 4.31: Auswahl des Typs, der Anzahl und der Reihenfolge der darzustellenden individuellen Wachstumskurven. Hier: Beobachtete Daten (**Individual data**), Darstellung von zehn Kurven pro Diagramm, (**Number of curves: 10**) und Reihenfolge wie im Datensatz, beginnend mit der ersten Person im Datensatz (**Consecutive order; Starting at index: 1**).

Nach dieser Voreinstellung erhalten wir für das Ängstlichkeitsbeispiel die in Abbildung 4.32 dargestellten beobachteten Wachstumskurven der ersten zehn Kinder im Datensatz **aengstlichkeit.dat**. Durch Auswahl eines bestimmten Datenpunktes mit dem Mauszeiger können wir uns die Fallnummer sowie die konkreten Werte des betreffenden Kindes anzeigen lassen. Dies ist praktisch, beispielsweise wenn es um die Identifikation von Fällen mit Extremwerten oder extremen Verläufen geht.

Wir können uns die Verläufe für die übrigen Kinder anschauen, indem wir im Mplus-Menü auf das Symbol klicken. So können wir schrittweise durch den gesamten Datensatz „wandern“. Mit Hilfe des Menü-Symbols können wir jederzeit wieder zum Dialogfenster „Properties for viewing individual curves“ wechseln, um die Einstellungen dort zu verändern.

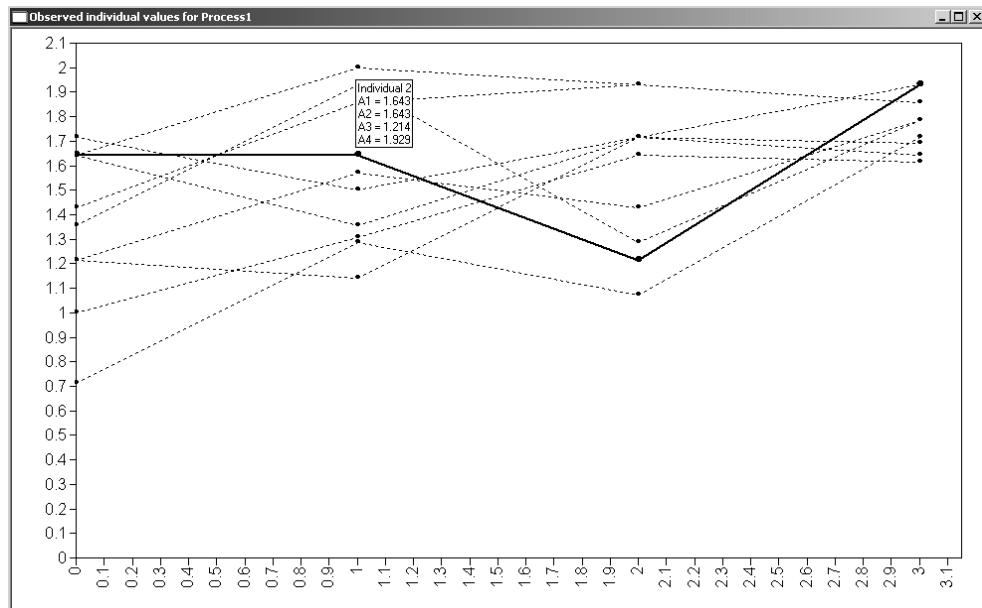


Abbildung 4.32: Individuelle beobachtete Wachstumskurven der ersten zehn Kinder im Datensatz **aengstlichkeit.dat**. Die Hervorhebung der Wachstumskurve des zweiten Kindes im Datensatz sowie die Anzeige der Fallnummer und der beobachteten Werte dieses Kindes werden durch Auswahl eines Datenpunktes mit dem Mauszeiger erreicht.



Durch klicken auf das Symbol kann man sich eine andere Art von Plot anzeigen lassen, beispielsweise die modellimplizierten individuellen oder die mittleren Wachstumskurven. In Abbildung 4.33 wurden die mittleren beobachteten und modellimplizierten Wachstumskurven ausgewählt, welche in Abbildung 4.34 für das lineare LGCM und in Abbildung 4.35 für das Intercept-Only-Modell dargestellt sind.

Es besteht die Möglichkeit, die Mplus-Grafiken als Bilddateien (im *.DIB-, *.EMF- oder *.JPEG-Format) zu exportieren. Dies geschieht über die Menüoption **Graph → Export plot to** (siehe Abb. 4.36).

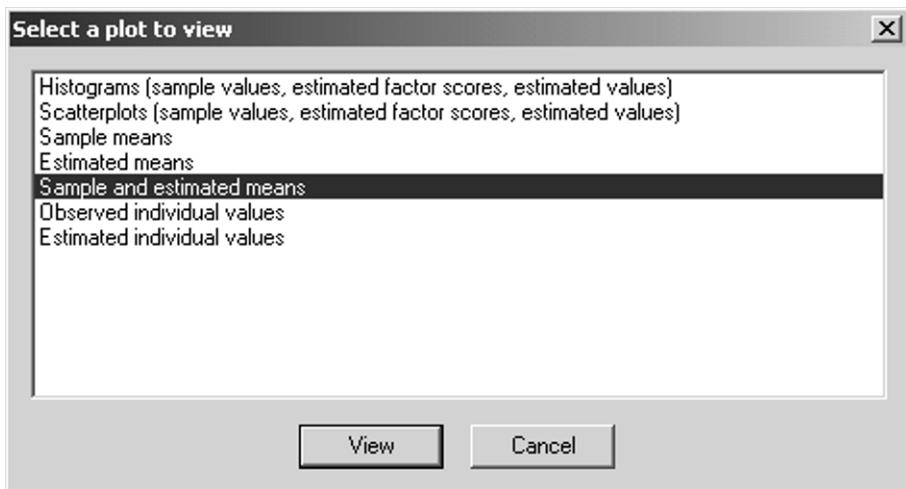


Abbildung 4.33: Auswahl einer bestimmten Grafik. Hier: Beobachtete (Sample) und modellimplizierte (estimated) mittlere Wachstumskurven.

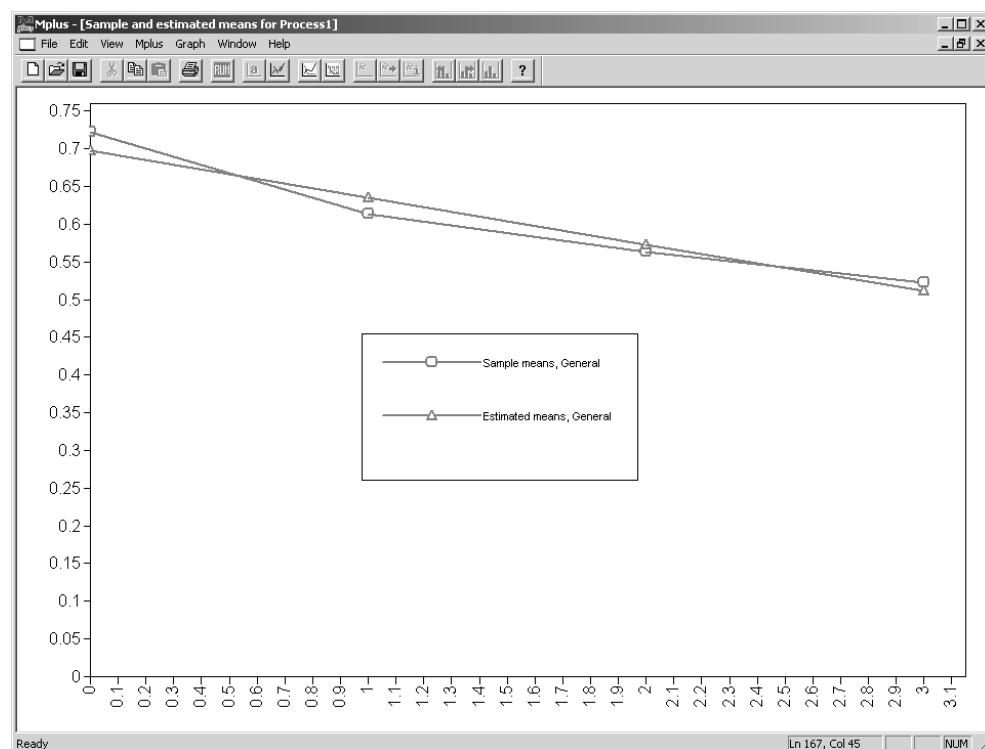


Abbildung 4.34: Mittlere beobachtete (sample) und mittlere modellimplizierte (estimated) Wachstumskurve des linearen LGCM 1. Ordnung für die Ängstlichkeitsdaten.

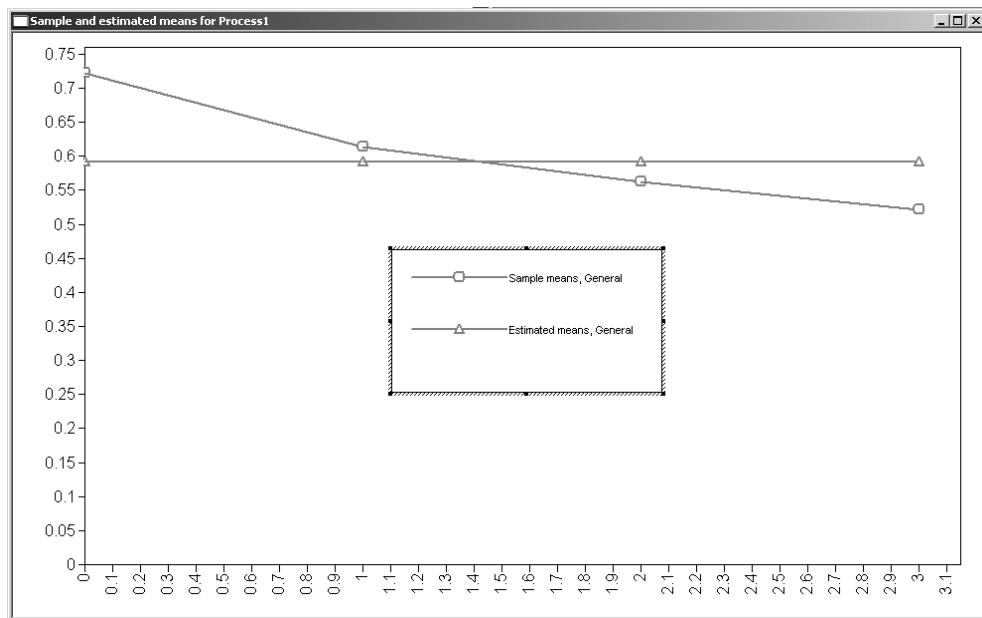


Abbildung 4.35: Mittlere beobachtete und mittlere modellimplizierte Wachstumskurve des Intercept-Only-Modells 1. Ordnung für die Ängstlichkeitsdaten.

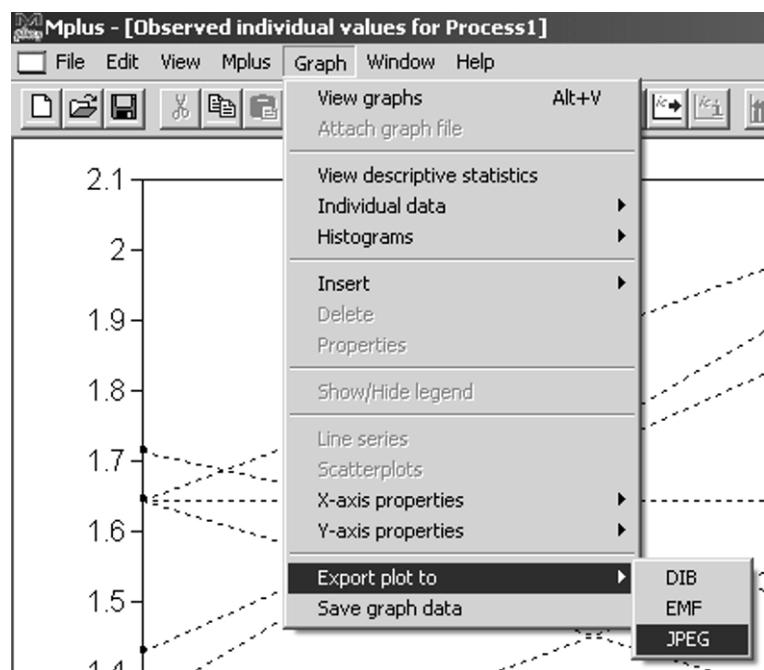


Abbildung 4.36: Export eines Mplus-Diagramms. Hier: Abspeicherung als JPEG (Bild-)Datei.

4.5.1.3 Analyse quadratischer Wachstumskurvenmodelle 1. Ordnung

Das bislang besprochene LGC-Modell enthält nur einen *linearen* Slopefaktor und nimmt somit an, dass die Form der Veränderung für alle Personen linear ist. Um andere (nichtlineare) Formen von Veränderung zu modellieren, können zusätzliche Slopefaktoren in das Modell aufgenommen werden. Möchte man z.B. die Hypothese testen, dass sich Ängstlichkeit über die Zeit nicht linear, sondern kurvenlinear verändert, so kann man einen weiteren Slopefaktor in das Modell hinzufügen, welcher quadratisches Wachstum abbildet. Eine Erweiterung des Modells mit einem zusätzlichen quadratischen Slopefaktor („Quadratischer Slope-Faktor“) wird in Abbildung 4.37 gezeigt.

Der quadratische Wachstumsfaktor wird wiederum durch eine entsprechende Festlegung der Faktorladungen definiert. Verbessert sich die Modellanpassung durch Aufnahme des quadratischen Slopefaktors im Vergleich zum sparsameren linearen LGCM, und ist die Varianz des quadratischen Slopefaktors signifikant von Null verschieden, so sollte das quadratische LGCM bevorzugt werden, sofern dies auch inhaltlich gerechtfertigt erscheint. Zusätzlich zu quadratischen Trends können bei entsprechenden Hypothesen auch noch andere Formen von Wachstum spezifiziert werden (siehe dazu Bollen & Curran, 2006 sowie Duncan et al., 2006).

LGCM sind sehr flexibel und können für zahlreiche Spezialfälle entsprechend erweitert werden (z.B. Modelle für nicht gleichabständige Messungen sowie verschiedene Wachstumskomponenten). Eine ausführliche Behandlung verschiedener LGCM findet man in den Lehrbüchern von Bollen und Curran (2006) sowie Duncan et al. (2006).

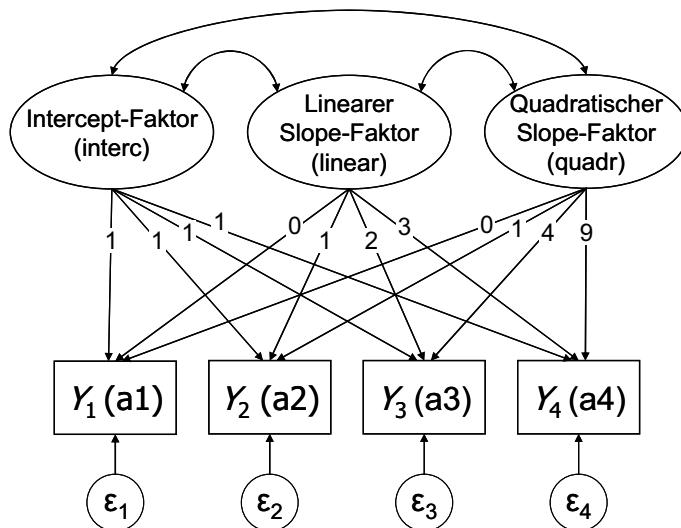


Abbildung 4.37: Latentes Wachstumskurvenmodell 1. Ordnung für eine beobachtete Variable Y_k ($k = \text{Messgelegenheit}$), welche zu vier Messgelegenheiten ($k = 1, 2, 3, 4$) erhoben wurde. ε_k : Messfehlervariable. Das Modell nimmt gleiche zeitliche Abstände der Messungen an und beinhaltet zusätzlich zum linearen auch einen quadratischen Slope-Faktor.

Abbildung 4.38 zeigt die Mplus-Spezifikation des quadratischen LGCM 1. Ordnung für das Ängstlichkeitsbeispiel.

The screenshot shows the Mplus software window with the title bar 'Mplus - [3_First_Order_Quadratic_Growth.inp]'. The menu bar includes File, Edit, View, Mplus, Graph, Window, and Help. Below the menu is a toolbar with various icons. The main area contains the Mplus input syntax:

```

title: Quadratisches latentes Wachstumskurvenmodell 1. Ordnung
Individualdaten aus Datei "aengstlichkeit.dat"
4 beobachtete Variablen:
Aengstlichkeit-Selbstbericht gemessen mit 1 Indikator
zu 4 Messgelegenheiten
keine fehlenden Werte

data: file = aengstlichkeit.dat;

variable: names = a1-a4;

model: interc linear quadr | a1@0 a2@1 a3@2 a4@3; ! spezielle Syntax für LGCM

output: sampstat standardized stdyx;

plot: type = plot3; ! Liefert graph. Darstellung der Wachstumskurven
      series = a1 (linear) a2 (linear) a3 (linear) a4 (linear);

```

The status bar at the bottom left says 'Ready' and at the bottom right shows 'Ln 10, Col 25' and 'NUM'.

Abbildung 4.38: Mplus-Input-Datei zur Spezifikation des quadratischen LGCM 1. Ordnung für 4 Messzeitpunkte aus Abbildung 4.37 unter Verwendung der speziellen Syntax für LGCM.

Es zeigt sich, dass der einzige Unterschied zur Spezifikation des linearen LGCM darin besteht, dass ein zusätzlicher Wachstumsfaktor (hier `quadr` genannt) in der Zeile

`model: interc linear quadr | a1@0 a2@1 a3@2 a4@3;`

spezifiziert wird. Mplus behandelt den dritten Faktor automatisch als quadratischen Wachstumsfaktor.

Bei der Schätzung der Modellparameter des quadratischen LGCM für die Ängstlichkeitsdaten tritt allerdings ein Schätzproblem auf, auf das Mplus im Output mit der nachfolgenden Warnmeldung hinweist:

```

THE MODEL ESTIMATION TERMINATED NORMALLY

WARNING: THE LATENT VARIABLE COVARIANCE MATRIX (PSI) IS NOT POSITIVE
DEFINITE. THIS COULD INDICATE A NEGATIVE VARIANCE/RESIDUAL VARIANCE
FOR A LATENT VARIABLE, A CORRELATION GREATER OR EQUAL TO ONE BETWEEN
TWO LATENT VARIABLES, OR A LINEAR DEPENDENCY AMONG MORE THAN TWO
LATENT VARIABLES. CHECK THE TECH4 OUTPUT FOR MORE INFORMATION. PROB-
LEM INVOLVING VARIABLE QUADR.

```

Diese Warnmeldung sagt aus, dass die Kovarianzmatrix der latenten Variablen (in diesem Fall der Wachstumsfaktoren) *nicht positiv definit* ist. Dies kann z.B. bedeuten, dass die Vari-

anz einer latenten Variable im Modell auf einen negativen Wert oder die Korrelation zwischen zwei latenten Variablen auf einen Wert größer |1| geschätzt wurde (vgl. die Ergebnisse für das CFA-Modell 2. Ordnung in Abschnitt 3.4).

Unter MODEL RESULTS können wir erkennen, welches konkrete Problem hier aufgetreten ist. Unter STANDARDIZED MODEL RESULTS (STDYX Standardization) erkennen wir, dass die Korrelation zwischen dem linearen und dem quadratischen Wachstumsfaktor auf -1,205 (und somit auf einen unzulässigen „out-of-range“ Wert; vgl. Box 3.9 in Kapitel 3) geschätzt wurde (der Output der standardisierten Parameterschätzungen wird unten gekürzt dargestellt).

Außerdem wird bei Betrachtung der unstandardisierten Parameterschätzungen deutlich, dass der zusätzlich ins Modell aufgenommene quadratische Wachstumsfaktor nur eine sehr geringe und statistisch nicht signifikante Varianz aufweist ($Var_{QUADR} = 0,001$, $z = 0,673$, $p = 0,501$). Die Ursache für die Schätzprobleme liegt somit offenbar in einer Überfaktorisierung durch die Aufnahme des quadratischen Faktors. Eine visuelle Inspektion der modellimplizierten Wachstumskurven (nicht dargestellt) deutet in diesem Beispiel darauf hin, dass – wenn überhaupt – nur ein sehr geringer kurvenlinearer Trend vorliegt, was die geringe Varianz dieses Faktors erklärt. Man würde somit im vorliegenden Fall aufgrund dieser Ergebnisse das sparsamere lineare LGCM dem quadratischen LGCM vorziehen.

MODEL RESULTS

				Two-Tailed	
		Estimate	S.E.	Est. / S.E.	P-Value
INTERC					
A1		1.000	0.000	999.000	999.000
A2		1.000	0.000	999.000	999.000
A3		1.000	0.000	999.000	999.000
A4		1.000	0.000	999.000	999.000
LINEAR					
A1		0.000	0.000	999.000	999.000
A2		1.000	0.000	999.000	999.000
A3		2.000	0.000	999.000	999.000
A4		3.000	0.000	999.000	999.000
QUADR					
A1		0.000	0.000	999.000	999.000
A2		1.000	0.000	999.000	999.000
A3		4.000	0.000	999.000	999.000
A4		9.000	0.000	999.000	999.000
LINEAR	WITH				
INTERC		-0.026	0.021	-1.201	0.230
QUADR	WITH				
INTERC		0.002	0.005	0.374	0.709
LINEAR		-0.008	0.005	-1.565	0.117

Means				
INTERC	0.720	0.020	35.617	0.000
LINEAR	-0.114	0.018	-6.452	0.000
QUADR	0.016	0.005	3.113	0.002
Intercepts				
A1	0.000	0.000	999.000	999.000
A2	0.000	0.000	999.000	999.000
A3	0.000	0.000	999.000	999.000
A4	0.000	0.000	999.000	999.000
Variances				
INTERC	0.163	0.022	7.441	0.000
LINEAR	0.048	0.023	2.110	0.035
QUADR	0.001	0.001	0.673	0.501
Residual Variances				
A1	0.036	0.019	1.870	0.061
A2	0.053	0.007	7.720	0.000
A3	0.035	0.006	5.498	0.000
A4	0.072	0.019	3.890	0.000
STANDARDIZED MODEL RESULTS (STDYX Standardization)				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
LINEAR WITH INTERC	-0.290	0.166	-1.746	0.081
QUADR WITH INTERC	0.154	0.370	0.415	0.678
QUADR WITH LINEAR	-1.205	0.471	-2.558	0.011
Means				
INTERC	1.784	0.128	13.928	0.000
LINEAR	-0.518	0.146	-3.552	0.000
QUADR	0.514	0.415	1.237	0.216

4.5.2 Latente Wachstumskurvenmodelle 2. Ordnung

Eine wichtige Voraussetzung für die sinnvolle Anwendung von LGCM ist die uns bereits mehrfach begegnete Annahme der Messinvarianz über die Zeit (vgl. Abschnitt 4.1.4). Messinvarianz setzt zum einen voraus, dass zu jedem Messzeitpunkt dasselbe Messinstrument (z.B. Test oder Fragebogen) verwendet wurde und zum anderen, dass sich die Messeigenschaften des Messinstrumentes nicht verändert haben (Meredith, 1993; Meredith & Horn, 2001; Millsap & Meredith, 2007). Bei Verletzung eines Mindestgrades an Messinvarianz können die Parameter des LGCM (ebenso wie die eines LC-Modells) nicht sinnvoll interpretiert werden, da in diesem Fall unklar wäre, ob etwaige Veränderungen auf eine tatsächliche

Entwicklung der zugrundeliegenden latenten Personenmerkmale oder auf die Verwendung unterschiedlicher Messverfahren bzw. die Veränderung der Messeigenschaften eines Messinstruments zurückzuführen sind.

In den im vorangegangenen Abschnitt diskutierten LGCM 1. Ordnung wird die Annahme der Messinvarianz über die Zeit implizit getroffen, ohne dass diese in den Modellen testbar wäre. Eine wesentliche Stärke von LGCM 2. Ordnung (*Second order growth curve models, curve-of-factors models*; Hancock, Kuo & Lawrence, 2001; McArdle, 1988; Sayer & Cumsille, 2001) liegt darin, dass diese es erlauben, die wichtige Frage der Messinvarianz über die Zeit empirisch zu testen.

Ein LGCM 2. Ordnung für ein Konstrukt (z.B. Ängstlichkeit), welches zu vier Messgelegenheiten mit je zwei Indikatoren (z.B. zwei unterschiedliche Skalen zur Messung von Ängstlichkeit) erfasst wurde, ist in Abbildung 4.39 dargestellt. Abbildung 4.39 verdeutlicht, dass LGCM 2. Ordnung nicht wie LGCM 1. Ordnung nur *einen* Indikator pro Messgelegenheit verwenden, sondern jeweils *multiple* Indikatoren.

In LCGM 2. Ordnung sind die beobachteten Indikatoren jedoch nicht direkt mit den latenten Wachstumsfaktoren verknüpft, sondern es sind latente State-Variablen (vgl. Abschnitt 4.1) „zwischen geschaltet“. Die Wachstumskurven werden auf Ebene der latenten State-Variablen (mittels Wachstumsfaktoren 2. Ordnung) und nicht auf Ebene der beobachteten Variablen angepasst.

Die Verwendung mehrerer Indikatoren pro Messgelegenheit ermöglicht es, die Messinvarianzannahme empirisch zu testen, da zu jedem Messzeitpunkt – wie in einem LS-Modell – ein eigenes Messmodell spezifiziert wird. Für die Indikatoren der State-Faktoren kann die Invarianz der Intercepts, Faktorladungen und Residualvarianzen getestet werden (zur Spezifikation verschiedener Grade an Messinvarianz vgl. Abschnitt 4.1.4). Der zur sinnvollen Interpretation von Wachstumsfaktoren 2. Ordnung benötigte Mindestgrad an Messinvarianz ist starke faktorielle Invarianz (gleiche Ladungen und Intercepts über die Zeit; vgl. Box 4.3).

Der in LGCM interessierende Wachstumsprozess wird in LGCM 2. Ordnung wie bereits erwähnt auf der Ebene der latenten State-Faktoren modelliert. Dies hat den zusätzlichen Vorteil, dass so, ähnlich wie in einem LST-Modell (vgl. Abschnitt 4.2), messgelegenheitsspezifische Einflüsse von Messfehlereinflüssen sowie vom eigentlich interessierenden Wachstumsprozess getrennt werden können. Dies ist in LGCM 1. Ordnung nicht möglich.

Messgelegenheitsspezifische Einflüsse werden in LGCM 2. Ordnung durch die latenten Residualvariablen ($\zeta_1-\zeta_4$ in Abb. 4.39) repräsentiert, während sie in LGCM 1. Ordnung nicht von Messfehlereinflüssen trennbar (und somit Bestandteil der Fehlervariablen ϵ_k) sind. Dies führt in LGCM 1. Ordnung häufig zu einer Unterschätzung der Reliabilitäten der Indikatoren. Im Modell in Abbildung 4.39 wurde wiederum ein indikatorspezifischer Faktor (IS_2) zusätzlich aufgenommen, welcher die Abweichung des zweiten Indikators (Y_{2k}) vom ersten (Y_{1k}) abbildet (vgl. Abschnitt 4.1.3).

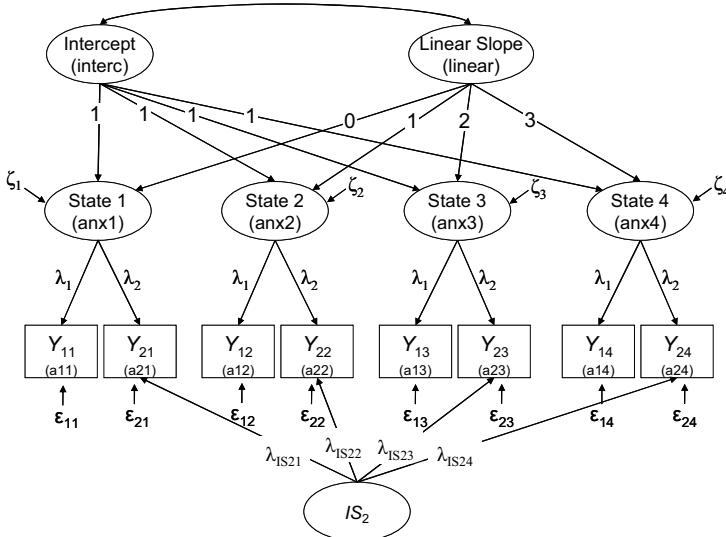


Abbildung 4.39: Latentes Wachstumskurvenmodell 2. Ordnung für zwei beobachtete Variablen Y_{ik} ($i = \text{Indikator}$, $k = \text{Messgelegenheit}$), welche zu vier Messgelegenheiten erhoben wurden. ϵ_{ik} : Messfehlervariable; λ_i : zeitlich invariante State-Faktorladungen; Λ_{IS_2} : Ladungen auf dem indikator spezifischen Faktor IS_2 ; ζ_{ik} : messgelegenheitsspezifische latente Residualvariablen. Der Veränderungsprozess wird auf Ebene der latenten State-Variablen mittels Intercept- und Slope-Faktoren 2. Ordnung modelliert. Das Modell nimmt gleiche zeitliche Abstände der Messungen und lineares Wachstum an.

Box 4.14: Fallbeispiel Wachstumskurvenmodelle 2. Ordnung

Wir illustrieren die Analyse von LGCM 2. Ordnung in Mplus anhand des Ängstlichkeitsbeispiels, welches wir bereits im Zusammenhang mit der Spezifikation von LGCM 1. Ordnung kennengelernt haben (vgl. Abschnitt 4.5.1). Allerdings verwenden wir nun für jeden der vier Messzeitpunkte nicht nur einen Indikator für Ängstlichkeit, sondern zwei Indikatoren. Zwei Indikatoren sind wenigstens erforderlich, um ein Wachstumsmodell 2. Ordnung spezifizieren zu können. Die entsprechenden Daten befinden sich in der Datei **aengstlichkeit2.dat**, welche die Werte von $N = 484$ Kindern auf acht manifesten Variablen (a11-a24) enthält.

Abbildung 4.40 zeigt den Mplus-Input zur Spezifikation eines linearen LGCM 2. Ordnung für die Ängstlichkeitsdaten. Die Mplus-Variablennamen für die Ängstlichkeit-Indikatoren wurden gemäß der Darstellung in Abbildung 4.39 gewählt. Der Buchstabe „a“ steht für Ängstlichkeit, die erste Zahl steht für den Indikator (i), die zweite Zahl indiziert den Messzeitpunkt (k).

In der Praxis bietet es sich an, vor einer Analyse des Veränderungsprozesses mittels LGCM 2. Ordnung zunächst den vorliegenden Grad an Messinvarianz mit Hilfe des weniger restriktiven LS-Modells zu testen, in welchem noch keine Annahmen bezüglich der Art

des Veränderungsprozesses gemacht werden (vgl. Diskussion in Kapitel 4.1). Da die Testung von Messinvarianz im Rahmen des LS-Modells bereits in Abschnitt 4.1.4 ausführlich diskutiert wurde, wird an dieser Stelle auf diesen zusätzlichen Analyseschritt verzichtet, und es wird von Vornherein ein State-Modell für Ängstlichkeit (State-Faktoren $anx1-anx4$) zu Grunde gelegt, das von starker faktorieller Invarianz (gleiche Ladungen und Intercepts der Indikatoren a11-a24) ausgeht.

The screenshot shows the Mplus software interface with the title bar 'Mplus - [1_Second_Order_Linear_Growth.inp]'. The menu bar includes File, Edit, View, Mplus, Graph, Window, and Help. Below the menu is a toolbar with various icons. The main window displays the Mplus input syntax. The code specifies a second-order linear growth model (LGCM 2. Order) with four latent variables (anx1, anx2, anx3, anx4) and their loadings on two state factors (a11, a12, a13, a14, a21, a22, a23, a24). It also includes intercept specifications and plotting options. The status bar at the bottom indicates 'Ready' and 'Ln 39, Col 1'.

```

title: Lineares latentes Wachstumskurvenmodell 2. Ordnung
Individualdaten aus Datei "ängstlichkeit2.dat"
8 beobachtete Variablen:
Ängstlichkeits-Selbstbericht gemessen mit 2 Indikatoren
zu 4 Messgelegenheiten
keine fehlenden Werte

data: file = aengstlichkeit2.dat;

variable: names = a11 a21 a12 a22 a13 a23 a14 a24;

model: ! State-Faktoren 1. Ordnung
anx1 by a11
      a21 (1); ! Ladungen über die Zeit gleichgesetzt

anx2 by a12
      a22 (1);

anx3 by a13
      a23 (1);

anx4 by a14
      a24 (1);

! Intercepts der Referenzindikatoren werden auf Null fixiert
[a11@0 a12@0 a13@0 a14@0];

! Intercepts der restlichen Indikatoren werden über die Zeit gleichgesetzt
[a21 a22 a23 a24] (2);

! Spezifikation der Wachstumsfaktoren als Faktoren 2. Ordnung
interc linear | anx1@0 anx2@1 anx3@2 anx4@3;

! Spezifikation eines indikatorsspez. Faktors für den 2. Indikator
is2 by a21 a22@1 a23@1 a24@1;

! Indspez. Faktor ist mit allen übrigen Faktoren unkorreliert
is2 with anx1-anx4@0 interc@0 linear@0;

! Mittelwert des Intercept-Faktors schätzen lassen
[interc];

plot: type = plot3; ! Graph. Darstellung der mittleren latenten Wachstumskurve
      series = anx1 (linear) anx2 (linear) anx3 (linear) anx4 (linear);

output: sampstat standardized stdyx;

```

Abbildung 4.40: Mplus-Input-Datei zur Spezifikation des linearen LGCM 2. Ordnung für 4 Messzeitpunkte aus Abbildung 4.39 unter Verwendung der speziellen Syntax für LGCM.

Der erste Teil der Modellspezifikation ist identisch zur Spezifikation eines LS-Modells mit starker faktorieller Invarianz über die Zeit (vgl. Abschnitt 4.1.4). Hierbei ist zu beachten, dass zunächst die State-Faktor-Ladungen der Indikatoren über die Zeit gleichgesetzt werden:

```

anx1 by a11
    a21 (1);

anx2 by a12
    a22 (1);

anx3 by a13
    a23 (1);

anx4 by a14
    a24 (1);

```

Weiterhin werden die Intercepts der Referenzindikatoren a11, a12, a13 und a14 auf Null fixiert, um die latente Mittelwertsstruktur zu identifizieren:

```
[a11@0 a12@0 a13@0 a14@0];
```

Die Intercepts der restlichen Indikatoren werden über die Zeit hinweg gleichgesetzt, was in Verbindung mit den gleichgesetzten Ladungen die Annahme starker Messinvarianz impliziert:

```
[a21 a22 a23 a24] (2);
```

Die Wachstumsfaktoren `interc` (Interceptfaktor) und `linear` (linearer Slopefaktor) werden analog zum linearen LGCM 1. Ordnung mit Hilfe der speziellen Mplus-Syntax für Wachstumskurvenmodelle spezifiziert. Der Unterschied zur Spezifikation im LGCM 1. Ordnung besteht darin, dass nun nicht die beobachteten Variablen direkt als Indikatoren für die Wachstumsfaktoren dienen, sondern die latenten Ängstlichkeits-State-Faktoren 1. Ordnung:

```
interc linear | anx1@0 anx2@1 anx3@2 anx4@3;
```

Um indikatorspezifische Effekte über die Zeit zu modellieren, wird wiederum ein indikatorspezifischer Faktor für den zweiten Indikator mit in die Modellspezifikation eingeschlossen (vgl. Abschnitt 4.1.3):

```
is2 by a21 a22@1 a23@1 a24@1;
```

Die Ladungen werden alle auf 1 fixiert, um auch sie invariant über die Zeit zu halten. Der indikatorspezifische Faktor ist ein Residualfaktor und daher mit allen übrigen Faktoren unkorreliert:

```
is2 with anx1-anx4@0 interc@0 linear@0;
```

Schließlich muss man Mplus noch mitteilen, dass der Mittelwert des latenten Intercept-Faktors 2. Ordnung geschätzt werden soll, da dieser (aus unerfindlichen Gründen) ansonsten vom Programm auf Null fixiert (und nicht geschätzt) wird:

```
[interc];
```

Der Mittelwert des Slope-Faktors 2. Ordnung wird dagegen automatisch geschätzt, ohne dass man dies gesondert anfordern muss. Der Mplus-Output zeigt uns, dass das LGCM 2. Ordnung einen passablen Datenfit aufweist. Zwar ist der χ^2 -Test signifikant ($\chi^2 = 55,372, df = 24, p = 0,0003$), allerdings weisen *CFI* (0,993), *RMSEA* (0,052) und *SRMR* (0,050) auf eine akzeptable Anpassung des Modells an die Daten hin.

Die Schätzergebnisse des Wachstumskurventeils des Modells (wir betrachten zunächst die unstandardisierten Parameterschätzungen) ergeben ein ähnliches Bild wie die Analysen auf der Basis des LGCM 1. Ordnung (vgl. Abschnitt 4.5.1). Der latente Ängstlichkeitsausgangswert beträgt 0,693 und ist signifikant von Null verschieden ($z = 35,827, p < 0,001$). Ebenso liegt signifikante Varianz im Ausgangswert vor ($Var_{INTERC} = 0,141, z = 11,993, p < 0,001$). Der Mittelwert des linearen Slope-Faktors ist wie im LGCM 1. Ordnung negativ und signifikant von Null verschieden ($M_{LINEAR} = -0,059, z = -10,593, p < 0,001$). Dies weist darauf hin, dass im Mittel eine Abnahme der Ängstlichkeit bei den Kindern zu verzeichnen ist. Zudem zeigen sich auch im LGCM 2. Ordnung wieder signifikante Unterschiede in der Stärke der linearen Veränderung über die Zeit ($Var_{LINEAR} = 0,007, z = 5,408, p < 0,001$). Das Ängstlichkeitsniveau sinkt also nicht bei allen Kindern in gleichem Maße.

MODEL RESULTS

				Two-Tailed
		Estimate	S.E.	P-Value
INTERC				
ANX1		1.000	0.000	999.000
ANX2		1.000	0.000	999.000
ANX3		1.000	0.000	999.000
ANX4		1.000	0.000	999.000
LINEAR				
ANX1		0.000	0.000	999.000
ANX2		1.000	0.000	999.000
ANX3		2.000	0.000	999.000
ANX4		3.000	0.000	999.000
ANX1	BY			
A11		1.000	0.000	999.000
A21		0.987	0.018	55.720
ANX2	BY			
A12		1.000	0.000	999.000
A22		0.987	0.018	55.720

ANX3	BY			
A13		1.000	0.000	999.000
A23		0.987	0.018	55.720
ANX4	BY			
A14		1.000	0.000	999.000
A24		0.987	0.018	55.720
IS2	BY			
A21		1.000	0.000	999.000
A22		1.000	0.000	999.000
A23		1.000	0.000	999.000
A24		1.000	0.000	999.000
IS2	WITH			
ANX1		0.000	0.000	999.000
ANX2		0.000	0.000	999.000
ANX3		0.000	0.000	999.000
ANX4		0.000	0.000	999.000
INTERC		0.000	0.000	999.000
LINEAR		0.000	0.000	999.000
LINEAR	WITH			
INTERC		-0.008	0.003	-2.875
				0.004
Means				
INTERC		0.693	0.019	35.827
LINEAR		-0.059	0.006	-10.593
Intercepts				
A11		0.000	0.000	999.000
A21		0.028	0.012	2.263
A12		0.000	0.000	999.000
A22		0.028	0.012	2.263
A13		0.000	0.000	999.000
A23		0.028	0.012	2.263
A14		0.000	0.000	999.000
A24		0.028	0.012	2.263
ANX1		0.000	0.000	999.000
ANX2		0.000	0.000	999.000
ANX3		0.000	0.000	999.000
ANX4		0.000	0.000	999.000
Variances				
IS2		0.009	0.001	7.336
INTERC		0.141	0.012	11.993
LINEAR		0.007	0.001	5.408

Residual Variances				
A11	0.021	0.004	5.561	0.000
A21	0.025	0.004	6.720	0.000
A12	0.019	0.003	6.491	0.000
A22	0.023	0.003	7.844	0.000
A13	0.019	0.003	7.209	0.000
A23	0.017	0.003	6.718	0.000
A14	0.020	0.003	7.352	0.000
A24	0.016	0.003	6.003	0.000
ANX1	0.048	0.007	7.063	0.000
ANX2	0.031	0.004	7.922	0.000
ANX3	0.030	0.004	8.199	0.000
ANX4	0.017	0.005	3.339	0.001

Die standardisierten Parameterschätzungen unter STANDARDIZED MODEL RESULTS (STDYX Standardization) zeigen uns, dass die Korrelation zwischen dem Intercept- und dem linearen Slope-Faktor auf -0,259 geschätzt wurde. Mithin fällt sie vom Betrag her etwas geringer aus als die Korrelation im LGCM 1. Ordnung.

Insgesamt können durch das Veränderungsmodell 2. Ordnung (latenter Ausgangswert = Faktor INTERC sowie linearer Wachstumsprozess = Faktor LINEAR) zwischen 74,6 und 90,3% der Varianz der messfehlerbereinigten Ängstlichkeitszustandswerte (State-Faktoren anx1-anx4) aufgeklärt werden (siehe Rubrik R-SQUARE Latent Variable). Dies spricht für eine gute Anpassung des linearen Modells an die vorliegenden Daten. Der an 100% fehlende Anteil ist auf messgelegenheitsspezifische Einflüsse auf die latenten Ängstlichkeitsscores zurückzuführen.

In Hinblick auf die Parameter des Messmodells ist festzustellen, dass beide Indikatoren zu jedem Messzeitpunkt sehr hohe standardisierte Ladungen (Werte zwischen 0,911 und 0,947) auf den Ängstlichkeits-State-Faktoren anx1-anx4 aufweisen. Dies zeigt, dass die Indikatoren eher homogen und zudem sehr reliabel sind. Dies wird durch eine Inspektion der standardisierten Ladungen des zweiten Indikators auf dem indikatorsspezifischen Faktor is2 bestätigt. Diese sind zwar alle statistisch signifikant von Null verschieden, jedoch vom Betrag her als eher gering einzustufen (zwischen 0,208 und 0,224). Somit liegt der reliable aber spezifische Varianzanteil der Indikatoren (quadrierte standardisierte Ladungen auf dem indikatorsspezifischen Faktor) lediglich zwischen 0,043 und 0,050 (der indikatorsspezifische Faktor erklärt also nur zwischen 4,3 und 5,0% der Varianz der Indikatoren).

Die Reliabilitäten der Indikatoren (R-SQUARE Observed Variable) werden auf Werte zwischen 0,880 und 0,918 geschätzt, also deutlich höher als im LGCM 1. Ordnung (dort lagen sie nur zwischen 0,694 und 0,785). Dies liegt daran, dass das LGCM 2. Ordnung Messfehlervarianz von reliabler messgelegenheitsspezifischer Varianz trennt und somit die Reliabilitäten der Indikatoren in diesem Modell korrekter geschätzt werden können. Die standardisierten Residualvarianzen der Faktoren anx1-anx3 zeigen uns, dass 9,7-25,4% der reliablen Varianz durch messgelegenheitsspezifische Einflüsse determiniert wird.

STANDARDIZED MODEL RESULTS (STDYX Standardization)				
		Estimate	S.E.	Two-Tailed Est./S.E. P-Value
INTERC				
ANX1		0.864	0.019	46.518 0.000
ANX2		0.932	0.021	43.447 0.000
ANX3		0.922	0.030	31.217 0.000
ANX4		0.909	0.039	23.222 0.000
LINEAR				
ANX1		0.000	0.000	999.000 999.000
ANX2		0.205	0.020	10.387 0.000
ANX3		0.405	0.038	10.669 0.000
ANX4		0.600	0.059	10.193 0.000
ANX1	BY			
A11		0.949	0.010	99.540 0.000
A21		0.918	0.010	90.264 0.000
ANX2	BY			
A12		0.947	0.009	110.162 0.000
A22		0.911	0.009	97.051 0.000
ANX3	BY			
A13		0.947	0.008	125.147 0.000
A23		0.927	0.009	108.099 0.000
ANX4	BY			
A14		0.946	0.008	123.294 0.000
A24		0.932	0.009	106.026 0.000
IS2	BY			
A21		0.208	0.015	13.758 0.000
A22		0.223	0.016	13.928 0.000
A23		0.224	0.016	13.857 0.000
A24		0.223	0.016	13.532 0.000
IS2	WITH			
ANX1		0.000	0.000	999.000 999.000
ANX2		0.000	0.000	999.000 999.000
ANX3		0.000	0.000	999.000 999.000
ANX4		0.000	0.000	999.000 999.000
INTERC		0.000	0.000	999.000 999.000
LINEAR		0.000	0.000	999.000 999.000
LINEAR	WITH			
INTERC		-0.259	0.071	-3.659 0.000
Means				
INTERC		1.844	0.092	20.014 0.000
LINEAR		-0.720	0.095	-7.582 0.000

Intercepts				
A11	0.000	0.000	999.000	999.000
A21	0.059	0.027	2.230	0.026
A12	0.000	0.000	999.000	999.000
A22	0.064	0.029	2.226	0.026
A13	0.000	0.000	999.000	999.000
A23	0.064	0.029	2.221	0.026
A14	0.000	0.000	999.000	999.000
A24	0.064	0.029	2.219	0.027
ANX1	0.000	0.000	999.000	999.000
ANX2	0.000	0.000	999.000	999.000
ANX3	0.000	0.000	999.000	999.000
ANX4	0.000	0.000	999.000	999.000
Variances				
IS2	1.000	0.000	999.000	999.000
INTERC	1.000	0.000	999.000	999.000
LINEAR	1.000	0.000	999.000	999.000
Residual Variances				
A11	0.099	0.018	5.460	0.000
A21	0.115	0.017	6.746	0.000
A12	0.104	0.016	6.363	0.000
A22	0.120	0.015	7.793	0.000
A13	0.103	0.014	7.176	0.000
A23	0.090	0.014	6.465	0.000
A14	0.106	0.015	7.310	0.000
A24	0.082	0.014	5.696	0.000
ANX1	0.254	0.032	7.910	0.000
ANX2	0.188	0.023	8.188	0.000
ANX3	0.180	0.021	8.648	0.000
ANX4	0.097	0.028	3.446	0.001
R-SQUARE				
Observed Variable				
Estimate S.E. Est./S.E. Two-Tailed P-Value				
A11	0.901	0.018	49.770	0.000
A21	0.885	0.017	52.066	0.000
A12	0.896	0.016	55.081	0.000
A22	0.880	0.015	57.118	0.000
A13	0.897	0.014	62.574	0.000
A23	0.910	0.014	65.336	0.000
A14	0.894	0.015	61.647	0.000
A24	0.918	0.014	63.977	0.000
Latent Variable				
Estimate S.E. Est./S.E. Two-Tailed P-Value				
ANX1	0.746	0.032	23.259	0.000
ANX2	0.812	0.023	35.357	0.000
ANX3	0.820	0.021	39.311	0.000
ANX4	0.903	0.028	32.101	0.000

Mit Hilfe der verwendeten `plot`-Option wurde von Mplus u.a. eine Grafik zur Darstellung der mittleren modellimplizierten Wachstumskurve erzeugt. Dieses Diagramm kann man sich wiederum über **Graph → View graphs/Estimated means** ausgeben lassen (siehe Abb. 4.41). Diagramme mit den individuellen Wachstumskurven (basierend auf Faktorwerten) sind in Mplus für LGCM 2. Ordnung leider derzeit nicht erhältlich.

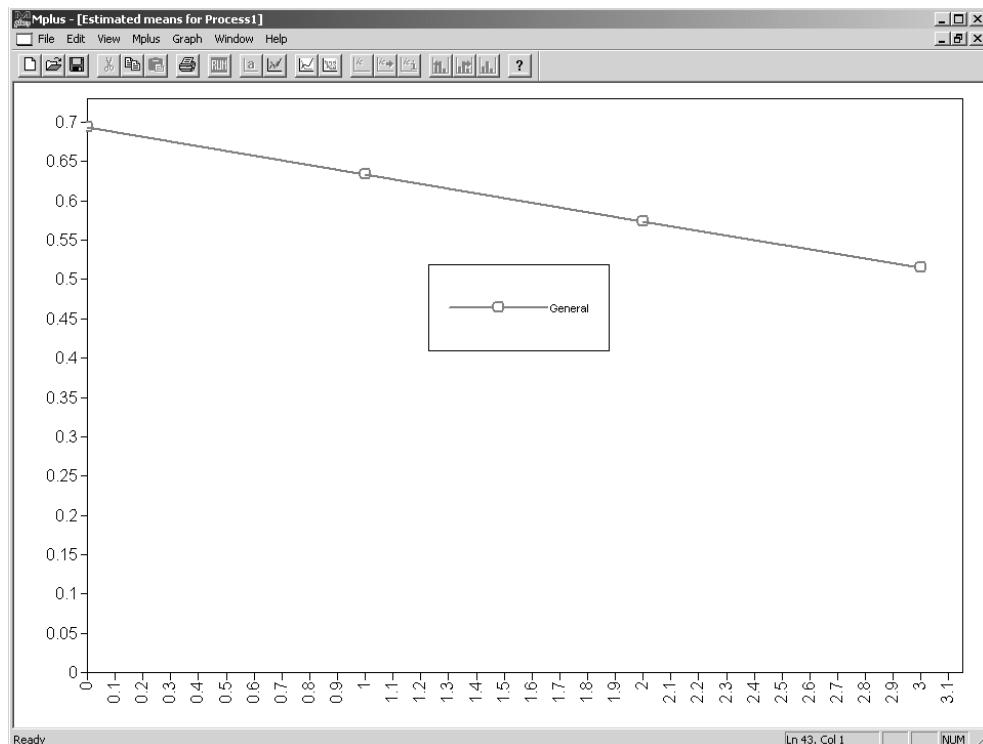


Abbildung 4.41: Mittlere modellimplizierte Wachstumskurve des linearen LGCM 2. Ordnung für die Ängstlichkeitsdaten, erzeugt über die Mplus-`plot`-Option.

Ebenso wie LGCM 1. Ordnung lassen sich auch LGCM 2. Ordnung in vielfältiger Weise erweitern. So kann z.B. bei Bedarf zusätzlich zu einem linearen Trend auch ein quadratischer Wachstumsfaktor mit in das Modell aufgenommen werden. Die Spezifikation erfolgt analog zum LGCM 1. Ordnung (im vorliegenden Beispiel führt die Aufnahme eines zusätzlichen quadratischen Faktors wie im LGCM 1. Ordnung zu einer Überfaktorisierung).

Außerdem können wie im LGCM 1. Ordnung Kovariaten des Veränderungsprozesses mit in das Modell aufgenommen werden. In dieser Hinsicht bieten LGCM 2. Ordnung ebenfalls mehr Möglichkeiten als LGCM 1. Ordnung, da in LGCM 2. Ordnung beispielsweise auch Kovariaten zur Erklärung messgelegenheitspezifischer Einflüsse mit aufgenommen werden könnten.

5 Mehrebenenregressionsmodelle

5.1 Einführung in die Mehrebenenanalyse

Mehrebenenmodelle (Synonyme: *Multilevelmodelle*, *Hierarchische lineare Modelle*, *Random Coefficient Models*, *Mixed Models*) dienen dazu, geschachtelte Daten (Daten mit einer hierarchischen Struktur oder Clusterstruktur) auszuwerten. Beispielsweise liegt in Untersuchungen mit Schülern häufig eine hierarchische Struktur vor, und zwar dann, wenn in der gezogenen Stichprobe die Schüler in Schulklassen „geschachtelt“ (man sagt auch *geclustert*) sind. Man unterscheidet in diesem Zusammenhang zwischen *Level-1-* und *Level-2-Einheiten*. Im Schülerbeispiel stellen die Schüler die Level-1-Einheiten (Einheiten der unteren Ebene, der sog. *Micro-Ebene*) und die Schulklassen die übergeordneten Level-2-Einheiten (Einheiten der *Macro-Ebene*) dar. Je nach Untersuchungsdesign können weitere übergeordnete Ebenen (z.B. Schulen, Stadtteile, Gemeinden, Kreise, Bundesländer, usw.) hinzukommen. In unserer Betrachtung werden wir uns jedoch auf zwei Ebenen beschränken.

Auch bei längsschnittlich erhobenen Daten liegt eine geschachtelte Datenstruktur vor (Datenpunkte geschachtelt in Individuen). Somit können Mehrebenenanalysen auch zur Auswertung von Längsschnittdaten verwendet werden (siehe z.B. Bryk & Raudenbush, 1987; Singer & Willett, 2003).

Die Beachtung der hierarchischen Struktur von Daten ist besonders aus zwei Gründen wichtig. Zum einen verletzt die Clusterung in Gruppen die Annahme der Unabhängigkeit der Beobachtungseinheiten voneinander. Zum anderen sind mögliche Einflussvariablen (Prädiktoren) häufig auf verschiedenen Ebenen angesiedelt, was in der Analyse adäquat berücksichtigt werden muss.

Die Gültigkeit der Annahme der Unabhängigkeit der Beobachtungseinheiten setzen viele statistische Verfahren voraus, so z.B. die konventionelle Ordinary-Least-Squares-(OLS-) Regressionsanalyse (Cohen et al., 2003). In Zufallsstichproben (ohne Clusterstruktur) ist die Unabhängigkeit der Beobachtungen gegeben. In Clusterstichproben liegt jedoch häufig eine Abhängigkeit von Beobachtungen innerhalb eines Clusters vor. Bezogen auf das Schüler-Beispiel können wir uns vorstellen, dass Schüler innerhalb ein- und derselben Schulkasse einander in Bezug auf bestimmte Merkmale (z.B. Schulleistung) ähnlicher sind als Schülern, die zu einer anderen Klasse gehören, da Schüler derselben Klasse beispielsweise bestimmten gemeinsamen Einflüssen ausgesetzt sind (gemeinsamer Unterricht, soziale Interaktionen innerhalb der Klasse etc.).

Die Verwendung konventioneller statistischer Verfahren und die damit verbundene Nichtbeachtung solcher Abhängigkeiten können zu verzerrten Ergebnissen führen. Insbesondere besteht die Gefahr der Unterschätzung der Standardfehler von Modellparametern (z.B. Regressionskoeffizienten), da durch die Abhängigkeiten in den Daten die effektive Stichprobengröße überschätzt wird (siehe z.B. Cohen et al., 2003; Snijders & Bosker, 1999).

Eine Folge der Unterschätzung der Größe der Standardfehler sind verzerrte statistische Inferenz (Erhöhung der Alpha-Fehlerrate) und inkorrekt geschätzte Konfidenzintervalle. Die *p*-Werte (z.B. für die Signifikanztests von Regressionsgewichten) werden zu gering geschätzt, was dazu führt, dass Nullhypotesen häufiger als statistisch erwartet zu Unrecht verworfen und Parameter häufiger zu Unrecht als „signifikant“ angesehen werden. Eine weitere Folge zu gering geschätzter Standardfehler sind verzerrte Konfidenzintervallschätzungen, die eine höhere Präzision der Parameterschätzung „vorgaukeln“ als in Wahrheit gegeben ist.

Der zweite Grund, aus welchem die Mehrebenenstruktur der Daten berücksichtigt werden sollte, liegt darin, dass die Modellierung von Variablen auf mehreren Ebenen häufig von inhaltlichem Interesse ist. Beispielsweise sind in vielen Fällen sowohl Variablen auf dem Individuallevel (z.B. Schülermerkmale) als auch Variablen auf dem Clusterlevel (z.B. Merkmale auf Klassenebene) zur Vorhersage von Variablen auf der Individualebene („Outcomes“) von Interesse. Auf Schülerebene könnten Persönlichkeitsmerkmale von Schülern wie Intelligenz, Fleiß und Motivation zur Erklärung von Schulleistung herangezogen werden. Auf Klassenebene könnten z.B. Klassengröße, Klassenklima und Lehrermerkmale (z.B. Ausbildung, Unterrichtsart) interessante Prädiktorvariablen darstellen. Auch können Interaktionen zwischen Variablen auf verschiedenen Ebenen (sogenannte „Cross-Level-Interactions“, siehe z.B. Luke, 2004 sowie Abschnitt 5.5.2) von Interesse sein.

Box 5.1: Zusammenfassung der wesentlichen Vorteile von Mehrebenenanalysen

Grundlegende Vorteile der Mehrebenenanalyse lassen sich folgendermaßen zusammenfassen:

- Adäquate Berücksichtigung der durch die hierarchische Datenstruktur (Clusterung) entstehenden Abhängigkeiten in den Daten
- Vermeidung von
 - Standardfehlerbias (zu gering geschätzten Standardfehlern)
 - inflationiertem Alpha-Fehler-Risiko (zu liberalen Signifikanztests)
 - zu eng geschätzten Konfidenzintervallen
- Betrachtung von Einflussvariablen auf Mikro- und Makroebene sowie Möglichkeit zur Analyse von Cross-Level-Interaktionen
- Hohe Flexibilität und weniger strenge Voraussetzungen als bei anderen statistischen Verfahren (z.B. Varianzanalyse mit Messwiederholung).

Detaillierte Einführungen und Weiterführendes zur Mehrebenenanalyse findet man in den einschlägigen Lehrbüchern von Hox (2002), Kreft und de Leeuw (1998), Luke (2004), Raudenbush und Bryk (2002), Singer und Willett (2003) sowie Snijders und Bosker (1999). Ein hervorragendes Einführungskapitel zur Mehrebenenregressionsanalyse liefern außerdem Cohen et al. (2003, Kapitel 14). Die Verwendung der Mehrebenenanalyse zur Auswertung von Längsschnittdaten wird u.a. von Bryk und Raudenbush (1987) sowie Singer und Willett (2003) behandelt.

5.2 Spezifikation von Mehrebenenmodellen in Mplus

Die Spezifikation von Mehrebenenmodellen geschieht in Mplus über die Option analysis: type = twolevel. Um type = twolevel benutzen zu können, benötigt man das Multilevel- oder das Combination-Add-On zusätzlich zum Mplus-Base-Programm (siehe www.statworld.com). Mplus kann bei Mehrebenenanalysen (derzeit Version 5) maximal zwei Hierarchieebenen berücksichtigen. Eine Ausnahme stellen Längsschnittdaten dar, da zur Analyse von Längsschnittdaten ein multivariater Ansatz verwendet werden kann. Multivariat bedeutet hierbei, dass bei der Auswertung von Längsschnittdaten mittels Wachstumskurvenmodellen zu jedem Messzeitpunkt (wenigstens) eine abhängige Variable modelliert wird. Die in Längsschnittdaten vorhandene Mehrebenenstruktur (Messzeitpunkte geschachtelt in Individuen) wird in Mplus über die Auswertung mit latenten Wachstumskurvenmodellen (Single-Level-Analyse; siehe Abschnitt 4.5) berücksichtigt. Weitere Hierarchieebenen (z.B. Individuen geschachtelt in Organisationen) können bei Längsschnittdaten dann wiederum über type = twolevel berücksichtigt werden, so dass man bei der Auswertung von Längsschnittdaten in Mplus bis zu drei Hierarchieebenen modellieren kann. Ab Mplus 6 sind generell 3 Ebenen möglich.

Box 5.2: Fallbeispiel Mehrebenenregressionsanalyse

Wir wollen uns die Durchführung von Mehrebenenregressionsanalysen anhand eines einfachen querschnittlichen Datenbeispiels veranschaulichen. Hierbei geht es um die Vorhersage der Mathematikleistung von Schülern. Im Beispiel wurden 503 Schülerinnen und Schüler (Level-1-Einheiten) aus 34 Schulklassen (Level-2-Einheiten) hinsichtlich ihrer individuellen Mathematikleistung (Variable **mathe**) und ihres Abschneidens in sechs Subtests (V1, V3, Q1, Q3, N1 und N3) des Kognitiven-Fähigkeits-Tests (KFT, Heller et al., 1976) untersucht. Die Leistungen in den sechs KFT-Subtests wurden zu einem KFT-Gesamtsummenwert zusammengefasst (Variable **kft**). Die Variablen **mathe** und **kft** stellen Level-1-Variablen dar, da sie auf Ebene der Schüler erfasst wurden und für jede Schülerin und jeden Schüler unterschiedliche Werte annehmen können.

Zusätzlich wurde der Schultyp (Sekundarschule versus Gymnasium) als Level-2-Variable mit erhoben (Variable **schultyp**) als Dummyvariable mit den Kategorien 0 = Sekundarschule, 1 = Gymnasium). Schultyp ist eine Level-2-Variable, da der Schultyp für Schüler innerhalb derselben Schulkasse nicht variiert (sondern nur zwischen Schulklassen).

Abbildung 5.1 zeigt einen Ausschnitt aus dem entsprechenden SPSS-Datensatz **Mathe_KFT.sav**. Zu beachten ist hierbei, dass für eine Mehrebenenanalyse in Mplus ein Individualdatensatz benötigt wird, der eine sogenannte *Clustervariable* enthält, welche die Level-2-Einheiten (hier: die Schulklassen) definiert. In unserem Beispiel ist dies die Variable **klasse**, welche die Clusterzugehörigkeit (Klasse Nr. 1-34) kennzeichnet.

*Mathe_KFT.sav [DatenSet1] - SPSS Daten-Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Fenster Hilfe

28 : schultyp 0 Sichtbar: 4

	klasse	mathe	kft	schultyp	var	var
1	1	4	37	Sekundarschule		
2	1	4	51	Sekundarschule		
3	1	4	53	Sekundarschule		
4	1	4	65	Sekundarschule		
5	1	5	33	Sekundarschule		
6	1	6	33	Sekundarschule		
7	1	6	33	Sekundarschule		
8	1	6	37	Sekundarschule		
9	1	8	66	Sekundarschule		
10	1	9	67	Sekundarschule		
11	1	10	75	Sekundarschule		
12	1	12	46	Sekundarschule		
13	1	18	70	Sekundarschule		
14	1	21	64	Sekundarschule		
15	1	24	91	Sekundarschule		
16	2	1	52	Sekundarschule		
17	2	2	32	Sekundarschule		
18	2	4	44	Sekundarschule		
19	2	5	53	Sekundarschule		
20	2	6	38	Sekundarschule		
21	2	6	41	Sekundarschule		
22	2	6	47	Sekundarschule		
23	2	7	55	Sekundarschule		
24	2	10	77	Sekundarschule		
25	2	12	54	Sekundarschule		
26	2	15	90	Sekundarschule		

Datenansicht Variablenansicht SPSS Prozessor ist bereit

Abbildung 5.1: Ausschnitt aus dem Beispieldatensatz Mathe_KFT.sav. Die Variable **klasse** stellt die in Mplus benötigte Clustervariable da, die die Level-2-Einheiten (Schulklassen) indiziert. Die Variable **mathe** ist die kontinuierliche abhängige (Level-1-)Variable, **kft** ist ein kontinuierlicher Level-1-Prädiktor und **schultyp** ein dichotomer Level-2-Prädiktor.

Box 5.3: Verwendete Notation für Multilevel-Regressionsmodelle

Die folgende, in der Literatur gängige Notation für Mehrebenenmodelle wird in diesem Kapitel verwendet. Man beachte, dass konkrete, aus Stichprobendaten geschätzte Werte zusätzlich mit einem „Dach“ (^) versehen werden, um auszudrücken, dass es sich um *geschätzte* Populationsparameter handelt.

Y_{ij} : Wert einer Person i aus Cluster j auf der abhängigen Variablen auf Level 1

β_{0j} : Random Intercept (Achsenabschnitt) in der Regression auf Level-1

β_{1j} : Random Slope (Steigungskoeffizient) in der Regression auf Level-1

X_{ij} : Wert einer Person i aus Cluster j auf einer Level-1-Prädiktorvariable

r_{ij} : Residualwert auf Level-1

$\sigma_{r_{ij}}^2$: Residualvarianz auf Level 1 (im Intercept-Only-Modell auch mit σ_W^2 für „within-clusters“ bezeichnet)

γ_{00} : Grand Mean bzw. Intercept für den Random Intercept auf Level 2

γ_{01} : Steigungskoeffizient für den Random Intercept auf Level 2

γ_{10} : Intercept für den Random Slope auf Level 2

γ_{11} : Steigungskoeffizient für den Random Slope auf Level 2

W_j : Wert eines Clusters j auf einer Level-2-Prädiktorvariable

u_{0j} : Residualwert für den Random Intercept auf Level 2

u_{1j} : Residualwert für den Random Slope auf Level 2

$\sigma_{u_{0j}}^2$: Residualvarianz im Random Intercept auf Level 2 (im Intercept-Only-Modell auch mit σ_B^2 für „between-clusters“ bezeichnet)

$\sigma_{u_{1j}}^2$: Residualvarianz im Random Slope auf Level 2

τ_{01} : Kovarianz zwischen Random Intercept und Random Slope auf Level 2

5.3 Option `twolevel basic`

Zu Beginn einer Mehrebenenanalyse in Mplus führt man am besten zunächst eine `twolevel basic`-Analyse durch, um das korrekte Einlesen der Daten zu überprüfen (vgl. Kapitel 2) und sich einen Überblick über die Art der Clusterstruktur und den Grad der Abhängigkeit der Beobachtungen zu verschaffen. Als Maß für den Grad der Abhängigkeit der

Beobachtungen wird meistens die *Intraklassenkorrelation* ρ_{IC} herangezogen, die bei einer Mehrebenenanalyse in Mplus praktischerweise automatisch ausgegeben wird. In anderen Mehrebenenanalyseprogrammen muss die Intraklassenkorrelation vom Benutzer selbst aus Varianzparameterschätzungen eines sogenannten Nullmodells (vgl. Abschnitt 5.4.1) von Hand ausgerechnet werden.

Die Intraklassenkorrelation ist definiert als das Verhältnis der Varianz zwischen den Clustern (Level-2-Varianz σ_B^2 ; „B“ steht für „between clusters“) zur Gesamtvarianz ($\sigma_W^2 + \sigma_B^2$), wobei σ_W^2 die (Level-1-)Varianz innerhalb der Cluster („W“ für „within clusters“) charakterisiert:

$$\rho_{IC} = \frac{\sigma_B^2}{(\sigma_W^2 + \sigma_B^2)}.$$

Bereits sehr geringe Intraklassenkorrelationen von 0,05 oder 0,01 können zu einer bedeutsamen Verzerrung der Ergebnisse von Signifikanztests in konventionellen OLS-Regressionsanalysen führen, in welchen die Abhängigkeiten nicht berücksichtigt werden (Cohen et al., 2003, S. 538). In der twolevel basic-Analyse gibt Mplus sowohl ρ_{IC} als auch die geschätzten Varianzen und Kovarianzen getrennt für Level 1 (WITHIN) und Level 2 (BETWEEN) aus.

The screenshot shows the Mplus software interface with the title bar 'Mplus - [0_Twolevel_basic.inp]'. The menu bar includes File, Edit, View, Mplus, Graph, Window, and Help. Below the menu is a toolbar with various icons. The main window displays the Mplus input command language. The code is as follows:

```

title: Berechnung der Intraklassenkorrelationen
      für Mathe- und KFT-Leistung unter Benutzung
      der Mplus-Option "type = twolevel basic"
      Clustervariable: Schulkasse

data: file = Mathe_KFT.dat;

variable: names = klasse ! Clustervariable (Schulkasse)
          kft mathe ! KFT-Score, KFT-Score zentriert, Mathescore
          schultyp; ! 0 = Sekundarschule, 1 = Gymnasium

usevar = kft mathe;

cluster = klasse; ! Definition der hierarchischen Struktur

analysis: type = twolevel basic;

```

The status bar at the bottom left says 'Ready' and at the bottom right says 'Ln 14, Col 69'.

Abbildung 5.2: Mplus-Input-Datei zur Spezifikation der twolevel basic-Analyse für die Level-1-Variablen **kft** und **mathe**.

Die Individualdaten aus Abbildung 5.1 wurden zur Verwendung in Mplus in die ASCII-Datei **Mathe_KFT.dat** exportiert (zum Aufbereiten und Abspeichern von Individualdaten siehe Abschnitt 2.1). Der Mplus-Input zur Spezifikation einer `twolevel basic`-Analyse für die Variablen **mathe** und **kft** kann Abbildung 5.2 entnommen werden.

Die Clustervariable **klasse** wird hierbei unter `variable` in einem separaten Befehl angegeben, um Mplus mitzuteilen, wie die Level-2-Einheiten (hier die Schulklassen) definiert sind: `cluster = klasse;`

Diese Clustervariable **klasse** muss nicht extra unter `usevar` aufgelistet werden, da es sich um eine Variable mit einer speziellen Funktion („variable with special function“) handelt.

Nachfolgend ist der wichtigste Teil des Mplus-Outputs für die `twolevel basic`-Analyse dargestellt. Unter **SUMMARY OF ANALYSIS** können wir u.a. erkennen, dass 503 Schüler (Level-1-Einheiten) in die Analyse einfließen und dass 2 Variablen (**mathe** und **kft**) Gegenstand der Analyse sind. Mplus teilt uns zur Kontrolle außerdem mit, dass es sich bei der Variable **klasse** um die verwendete Clustervariable handelt. So können wir nochmals überprüfen, ob unsere Level-2-Einheiten korrekt definiert sind (die Level-2-Variable **schultyp** wird erst weiter unten mit in die Analyse einbezogen).

SUMMARY OF ANALYSIS		
Number of groups		1
Number of observations		503
Number of dependent variables		2
Number of independent variables		0
Number of continuous latent variables		0
Observed dependent variables		
Continuous		
MATHE		KFT
Variables with special functions		
Cluster variable		KLASSE

Unter der Rubrik **SUMMARY OF DATA** erhalten wir genauere Informationen über die Clusterstruktur der Daten. Es liegen 34 Cluster (Schulklassen = Level-2-Einheiten) vor. Mplus teilt uns anschließend mit, wie die Clustergrößen verteilt sind. Wir können erkennen, dass es eine Klasse gibt, die fünf Schüler enthält (die Schulklasse mit der Nummer 22), eine Klasse, der acht Schüler angehören (Schulklasse Nr. 18) und so weiter. Der Modalwert liegt bei 13 Schülern in einem Cluster (Schulklassennummern 14, 2, 5, 23 und 13) und die Schulklasse, die am meisten Schüler liefert, ist die Klasse mit der Nummer 7 (24 Schüler stammen aus dieser Klasse). Unter diesem Plot wird die mittlere Clustergröße (Average cluster size) ausgegeben (hier: im Mittel 14,794 Schüler pro Klasse).

SUMMARY OF DATA

Number of clusters	34
Size (s) Cluster ID with Size s	
5	22
8	18
9	19
10	16 21 12 25
11	20 3
12	8
13	14 2 5 23 13
14	4
15	17 24 1
16	6 28 30
17	11 27
18	29 26 35 36
19	33
20	31
21	34
22	10
23	32
24	7
Average cluster size	14.794

Mplus gibt dann die geschätzten Intraklassenkorrelationen $\hat{\rho}_{IC}$ für die beiden Variablen **mathe** und **kft** aus. Die Werte von $\hat{\rho}_{IC} = 0,44$ und $\hat{\rho}_{IC} = 0,378$ deuten auf eine starke Abhängigkeit der Beobachtungen vom Cluster hin. Somit ist ein beträchtlicher Anteil der Varianz der Mathematik- und KFT-Leistung auf Unterschiede zwischen den Klassen zurückzuführen. Dies ist auch ein Hinweis darauf, dass es in diesem Beispiel Sinn macht, Prädiktoren auf Klassenebene (Level-2-Prädiktoren) zu betrachten (dies wäre nicht der Fall, wenn die Varianz auf Clusterebene gleich Null wäre).

Estimated Intraclass Correlations for the Y Variables			
Variable	Intraclass Correlation	Variable	Intraclass Correlation
MATHE	0.440	KFT	0.378

Anschließend liefert Mplus Deskriptivstatistiken (Mittelwerte, Kovarianzen und Korrelationen) getrennt für Level 1 (**WITHIN**) und Level 2 (**BETWEEN**). Mittelwerte (die Gesamtmittelwerte oder „Grand Means“, γ_{00}) werden nur auf Level 2 geschätzt. Die Level-1-(=Cluster-)Mittelwerte ergeben sich als Funktionen von Level-2-Parametern (den Grand Means sowie der clusterspezifischen Abweichungen von den Grand Means). Die geschätzten Grand Means betragen $\hat{\gamma}_{00} = 11,930$ für die Mathematikleistung und $\hat{\gamma}_{00} = 66,716$ für den KFT-

Score. Die geschätzten Varianzen der Variablen auf Individualebene betragen $\hat{\sigma}_W^2 = 26,484$ für den Mathescore und $\hat{\sigma}_W^2 = 235,735$ für den KFT-Score. Auf Klassenebene liegen die Varianzen für die Variable **mathe** bei $\hat{\sigma}_B^2 = 20,806$ und für die Variable **kft** bei $\hat{\sigma}_B^2 = 142,985$.

RESULTS FOR BASIC ANALYSIS

NOTE: The sample statistics for within and between refer to the maximum-likelihood estimated within and between covariance matrices, respectively.

ESTIMATED SAMPLE STATISTICS FOR WITHIN

Means		
	MATHE	KFT
1	0.000	0.000
 Covariances		
	MATHE	KFT
MATHE	26.484	
KFT	52.327	235.735
 Correlations		
	MATHE	KFT
MATHE	1.000	
KFT	0.662	1.000

ESTIMATED SAMPLE STATISTICS FOR BETWEEN

Means		
	MATHE	KFT
1	11.930	66.716
 Covariances		
	MATHE	KFT
MATHE	20.806	
KFT	49.983	142.985
 Correlations		
	MATHE	KFT
MATHE	1.000	
KFT	0.916	1.000

Die Varianzschätzungen können zur Berechnung der Intraklassenkorrelation verwendet werden. Wir können so die von Mplus weiter oben ausgegebenen Schätzungen für die Intraklassenkorrelation ($\hat{\rho}_{IC}$) überprüfen:

$$\text{Für Mathe: } \hat{\rho}_{IC} = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_W^2 + \hat{\sigma}_B^2} = \frac{20,806}{26,484 + 20,806} = 0,44.$$

$$\text{Für KFT: } \hat{\rho}_{IC} = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_W^2 + \hat{\sigma}_B^2} = \frac{142,985}{235,735 + 142,985} = 0,378.$$

Ferner sehen wir, dass die Korrelation zwischen dem Mathe- und dem KFT-Score auf Individualebene $r = 0,662$ beträgt, während der Zusammenhang auf Ebene der Schulklassen mit $r = 0,916$ noch deutlich höher liegt.

Zusammenfassend stellen wir fest, dass die Verwendung einer Mehrebenenanalyse zur Auswertung der vorliegenden Daten unbedingt geboten ist, da die Intraklassenkorrelationen auf starke Abhängigkeiten in den Daten hinweisen. Würden wir eine konventionelle Regressionsanalyse ohne Berücksichtigung der Clusterstruktur durchführen, so müssten wir mit einer Unterschätzung der Standardfehler unserer Regressionskoeffizienten und damit verbunden mit einer Alpha-Fehler-Inflationierung bei Signifikanztests sowie zu eng geschätzten Konfidenzintervallen rechnen. Nachfolgend werden wir verschiedene Modellklassen der Mehrebenenanalyse in Anlehnung an Luke (2004, Table 2.1) betrachten. Wir beginnen mit sogenannten *Random-Intercept-Modellen* und dabei mit dem einfachsten Random-Intercept-Modell, dem sogenannten *Nullmodell* oder *Intercept-Only-Modell* (vgl. Luke, 2004).

5.4 Random-Intercept-Modelle

5.4.1 Nullmodell (Intercept-Only-Modell)

Das Null- oder Intercept-Only-Modell wird häufig im ersten Schritt einer Mehrebenenanalyse geschätzt, u.a. da es zur Bestimmung der Level-1- und Level-2-Varianzen sowie der Intraklassenkorrelation der abhängigen Variable verwendet werden kann. Dies ist in Mplus eigentlich nicht erforderlich, da die Level-1- und Level-2-Varianzen sowie die Intraklassenkorrelation unter `type = twolevel` direkt ausgegeben werden. Aus didaktischen Gründen wird dieses Modell hier dennoch als erstes besprochen, da es das einfachste Mehrebenenmodell darstellt und das Verständnis dieses grundlegenden Modells auch für das Verständnis komplexerer Mehrebenenmodelle hilfreich ist.

Im Intercept-Only-Modell wird für die abhängige Variable (Outcome) keine Prädiktorvariable eingeführt, weder auf Level 1, noch auf Level 2. Auf Level 1 wird lediglich eine Zerlegung der individuellen Messwerte Y_{ij} ($i = \text{Individuum}, j = \text{Cluster}$) in Clustermittelwert β_{0j} und Abweichung des individuellen Messwertes vom Clustermittelwert (r_{ij}) vorgenommen:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + r_{ij}.$$

Bezogen auf unser Datenbeispiel zerlegen wir also den Mathescore Y_{ij} eines Schülers i (z.B. Franz) aus Schulklasse j in den Mittelwert der betreffenden Schulklasse (Clustermittelwert) plus die Abweichung von Franz' Mathescore vom Clustermittelwert. Die Varianz $\sigma_{r_{ij}}^2$ der Variable r_{ij} kennzeichnet die Varianz auf Level 1 (dem Within-Level):

$$\sigma_{r_{ij}}^2 = \sigma_W^2,$$

d.h. das Ausmaß der Unterschiedlichkeit der Schüler innerhalb der Klasse. Große Varianzen $\sigma_{r_{ij}}^2$ weisen darauf hin, dass es zwischen Schülern derselben Schulklasse große Unterschiede in der Mathematikleistung gibt. Aus unserer twolevel basic-Analyse (vgl. Abschnitt 5.3) wissen wir bereits, dass $\hat{\sigma}_{r_{ij}}^2 = \hat{\sigma}_W^2 = 26,484$ für den Mathetestscore ist. Die gleiche Schätzung erhalten wir auch auf Basis des Intercept-Only-Modells (s.u.).

Auf Level 2 werden die Clustermittelwerte β_{0j} zerlegt in den Gesamtmittelwert über alle Schulklassen (Grand Mean γ_{00}) plus die Abweichung des Clustermittelwertes vom Gesamtmittelwert (u_{0j}):

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j}.$$

Die Clustermittelwerte sind im Intercept-Only-Modell also Variablen auf Level 2. Sie haben auf Level 2 einen Mittelwert (γ_{00}) und eine Varianz $\sigma_{u_{0j}}^2$. Dies erklärt den Namen „Random-Intercept-Modell“: Während im normalen OLS-Regressionsmodell der Achsenabschnitt β_0 eine Konstante (d.h. ein *fixed effect*) ist, kann der Intercept β_{0j} im Random-Intercept-Modell über die Cluster hinweg unterschiedliche Werte annehmen (*random effect*).

In unserem Beispiel können also die Mathetestmittelwerte von Schulklasse zu Schulklasse variieren. Auf Level 2 wird zum Beispiel der Mathematiktestmittelwert in Franz' Klasse weiter zerlegt in den Mathe-Gesamtmittelwert (Grand Mean) über alle Klassen ($\hat{\gamma}_{00} = 11,930$; siehe twolevel basic-Analyse in Abschnitt 5.3) plus die Abweichung des Mathe-Mittelwertes in Franz' Klasse vom Mathe-Gesamtmittelwert.

Die Varianz $\sigma_{u_{0j}}^2$ des Residuums u_{0j} kennzeichnet die Varianz in den Clustermittelwerten auf Level 2 (dem Between-Level):

$$\sigma_{u_{0j}}^2 = \sigma_B^2,$$

d.h. das Ausmaß an Varianz zwischen den Schulklassen. Große Varianzen $\sigma_{u_{0j}}^2$ zeigen, dass es große Unterschiede in der mittleren Mathematikleistung zwischen den Schulklassen gibt.

Aus unserer twolevel basic-Analyse wissen wir, dass $\hat{\sigma}_{u_{0j}}^2 = \hat{\sigma}_B^2 = 20,806$ für den Mathetestscore ist. Eine ähnliche Schätzung erhalten wir auch im Intercept-Only-Modell (s.u.).

Da in diesem Modell die Variabilität in den Mathematikmittelwerten über die Klassen hinweg nicht durch andere Variablen erklärt wird (es gibt in der Level-2-Gleichung keine Prädiktorvariablen), bezeichnet man das Intercept-Only-Modell häufig auch als *Unconditional Cell Means Model*, *Unconstrained Model*, *Empty Model* oder *Nullmodell*.

Zusammenfassend erhalten wir im Intercept-Only-Modell von Mplus Schätzungen für die folgenden drei Modellparameter:

- Auf Level 1 (WITHIN):
 - die Varianz der individuellen Messwerte um den Clustermittelwert, $\sigma_{ij}^2 = \sigma_W^2$.
- Auf Level 2 (BETWEEN):
 - den Gesamtmittelwert (Grand Mean) γ_{00}
 - die Varianz der Clustermittelwerte um den Gesamtmittelwert, $\sigma_{u_{0j}}^2 = \sigma_B^2$.

Abbildung 5.3 zeigt den Mplus-Input zur Schätzung des Intercept-Only-Modells für die Variable **mathe**. Es muss dabei lediglich die Variable **mathe** unter **usevariables** aufgeführt werden, ein **model**-Statement ist zur Schätzung der Parameter des Intercept-Only-Modells nicht erforderlich. Es muss lediglich unter **analysis: type = twolevel** gewählt werden.

```
Mplus - [1_Intercept_Only_Modell.inp]
File Edit View Mplus Graph Window Help
□ ☐ ×
□ ☐ ×
title: Intercept-Only-Modell ("Unconditional Cell Means Model", "Empty Model") ▲
      für den Mathematikscore;

data: file = Mathe_KFT.dat;

variable: names = klasse ! Clustervariable (Schulklasse)
          mathe kft ! Mathe- und KFT-Score
          schultyp; ! 0 = Sekundarschule, 1 = Gymnasium

usevar = mathe;

cluster = klasse; ! Definition der hierarchischen Struktur

analysis: type = twolevel; ! Fordert Mehrebenenanalyse an

output: sampstat;

Ready Ln 14, Col 58 NUM
```

Abbildung 5.3: Mplus-Input-Datei zur Spezifikation des Intercept-Only-Modells für die Variable **mathe**.

Die von Mplus zunächst ausgegebene Warnmeldung

```
*** WARNING in MODEL command
  All variables are uncorrelated with all other variables in the
model.
  Check that this is what is intended.
  1 WARNING(S) FOUND IN THE INPUT INSTRUCTIONS
```

ist in diesem Fall für uns irrelevant und kann ignoriert werden. Sie bezieht sich darauf, dass im Modell keine Beziehungen zwischen Variablen spezifiziert wurden. Dies ist jedoch im Nullmodell auch gar nicht intendiert, so dass die entsprechende Warnung All variables are uncorrelated with all other variables in the model hier ausnahmsweise für uns nicht von Bedeutung ist.

Der Mplus-Output liefert unter WITHIN LEVEL die Schätzung der Level-1-Varianz ($\hat{\sigma}_{r_{ij}}^2 = 26,480$). Unter BETWEEN LEVEL erhalten wir Schätzungen für den Grand Mean ($\hat{\gamma}_{00} = 11,933$) und die Level-2-Varianz $\hat{\sigma}_{b_{0j}}^2 = 20,853$. Die geschätzten Werte weichen nur geringfügig von den in der twolevel basic-Analyse geschätzten Werten (siehe Abschnitt 5.3) ab. Es werden im Intercept-Only-Modell zusätzlich die Standardfehler der Parameterschätzungen (S.E.) sowie z- und p-Werte ausgegeben.

Box 5.4: Schätzung der Standardfehler unter type = twolevel in Mplus

Die Standardfehler der Parameterschätzungen werden in Mplus bei Wahl des Analysetyps type = twolevel per Voreinstellung über ein robustes Maximum-Likelihood-Schätzverfahren (estimator = mlr) geschätzt. Konventionelle Maximum-Likelihood-Schätzung kann über die Spezifikation

analysis: estimator = ml;

angefordert werden.

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
Variances				
MATHE	26.480	2.450	10.809	0.000
Between Level				
Means				
MATHE	11.933	0.808	14.776	0.000
Variances				
MATHE	20.853	6.017	3.466	0.001

Das Intercept-Only-Modell liefert außer den zur Schätzung der Intraklassenkorrelation benötigten Level-1- und Level-2-Varianzen (welche wir auch über die *twolevel basic-Analyse* erhalten können) keine weiteren Informationen, da es sich um ein „leeres“ Modell ohne Prädiktorvariablen handelt. Im Folgenden werden zunächst einfache Erweiterungen des Intercept-Only-Modells besprochen, in welchen Prädiktorvariablen entweder nur auf Level 1 (*Random-Effects-ANCOVA-Modelle*) oder nur auf Level 2 (sogenannte *Means-as-Outcomes-Modelle*) eingeführt werden.

5.4.2 One-Way-Random-Effects-ANCOVA

Wir besprechen zunächst den Fall, dass ein Forscher ein Level-1-Outcome (Y_{ij} ; z.B. die individuelle Mathematikleistung) durch einen Level-1-Prädiktor (X_{ij} ; individuelle KFT-Leistung) vorhersagen und dabei für Unterschiede in den KFT-Mittelwerten zwischen den Schulklassen kontrollieren will, ohne Unterschiede im Slope der Level-1-Regression über Cluster hinweg anzunehmen. Das resultierende *twolevel-Modell* weist eine Level-1-Regression mit einem Random Intercept (β_{0j}) und einem konstanten Slope ($\beta_{1j} = \gamma_{10}$) auf und ist äquivalent zu einer einfaktoriellen Kovarianzanalyse (ANCOVA) mit Zufallsfaktoren (sogenannte One-Way-Random-Effects-ANCOVA, vgl. Luke, 2004):

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j} \cdot X_{ij} + r_{ij} .$$

$$\begin{aligned}\text{Level 2: } \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} .\end{aligned}$$

Die Level-2-Gleichungen zeigen, dass in diesem Modell der Intercept-Term β_{0j} aus der Level-1-Regression der Mathematikleistung auf die KFT-Leistung über die Schulklassen (Level-2-Einheiten) hinweg variieren kann. Im Gegensatz dazu wird für den Steigungskoeffizienten (Slope β_{1j}) angenommen, dass dieser für jede Schulkasse gleich ist. Dies ist daran zu erkennen, dass in der Level-2-Gleichung für den Slope β_{1j} *kein* Abweichungswert u_{1j} enthalten ist, während für β_{0j} ein solcher Term vorhanden ist (u_{0j}).

Somit ist das Modell eher restriktiv in Bezug auf den Slope der Regression von **mathe** auf **kft**, da es gut möglich ist, dass dieser für verschiedene Schulklassen ebenfalls unterschiedlich ausfällt. In Abschnitt 5.5 besprechen wir sogenannte *Random-Intercept-and-Slope-Modelle*, die Varianz sowohl in den Intercepts also auch in den Slopes der Level-1-Regression zulassen. Zusammenfassend erhalten wir im Random-Intercept-Modell ohne Random Slope von Mplus Schätzungen für die folgenden Modellparameter:

- Auf Level 1 (WITHIN):
 - den konstanten Steigungskoeffizienten (fixed slope) $\beta_{1j} = \gamma_{10}$
 - die Residualvarianz $\sigma_{r_{ij}}^2$.

- Auf Level 2 (BETWEEN):
 - den mittleren Intercept über alle Cluster, γ_{00}
 - die Varianz in den Intercepts über die Cluster, $\sigma_{u_{0j}}^2$.

Bevor wir die Analyse des Modells in Mplus besprechen, wollen wir uns mit dem in der Mehrebenenanalyse sehr wichtigen Konzept der Zentrierung von Prädiktorvariablen befassen. Wir hatten uns bereits bei der Besprechung der OLS-Regressionsanalyse in Kapitel 3.2 kurz mit der Zentrierung befasst (vgl. Box 3.4). Der Einfachheit halber kommen wir zur Erläuterung der Zentrierungsproblematik zunächst auf die in Kapitel 3.2 besprochene simple OLS-Regressionsgleichung ohne Berücksichtigung der Clusterstruktur zurück. Allgemein kann die bivariate OLS-Regressionsgleichung für eine Kriteriums- und eine Prädiktorvariable für die Population wie folgt notiert werden (vgl. Kapitel 3.2):

$$Y = \beta_0 + \beta_1 \cdot X + \epsilon,$$

wobei β_0 den Achsenabschnitt (Intercept) und β_1 den Steigungskoeffizienten (Slope) für die Regression der Kriteriumsvariable Y auf die Prädiktorvariable X bezeichnet und ϵ eine Residualvariable darstellt. Die geschätzte OLS-Regressionsgleichung für unsere Kriteriumsvariable **mathe** (für den Gesamtdatensatz der $N = 503$ Kinder ohne Berücksichtigung der Clusterstruktur berechnet) lautet:

$$\text{mathe} = -6,486 + 0,279 \times \text{kft} + \text{Residuum}.$$

Der negative geschätzte Intercept-Term ($\hat{\beta}_0 = -6,486$) der Regressionsgleichung bedeutet inhaltlich, dass für Schüler mit einem KFT-Score von Null ein Mathetestwert von -6,486 vorhergesagt wird. Negative Scores liegen für den Mathematiktest jedoch außerhalb des möglichen Wertebereichs und auch ein KFT-Rohwert von Null ist im vorliegenden Datensatz für keinen Schüler aufgetreten. Somit ist eine inhaltlich sinnvolle Interpretation des Intercepts in dieser Form nicht möglich. Aus diesem Grund empfiehlt es sich, die unabhängige Variable (**kft**) vor der Analyse zu *zentrieren*, um den Nullpunkt dieser Variablen bedeutsam zu machen. Zentrieren bedeutet, dass von jedem einzelnen Messwert X_{ij} eine bedeutungsvolle Konstante (meist eine Art von Mittelwert) abgezogen wird. In der Analyse werden dann die Abweichungswerte (die zentrierten Scores) anstelle der Rohwerte verwendet. Die häufigste Form der Zentrierung besteht darin, den Gesamtmittelwert $\bar{X}_{..}$ von allen Rohwerten zu subtrahieren (sog. *Zentrierung am Grand Mean* oder *grand-mean-centering*):

$$X_{ij}^Z = X_{ij} - \bar{X}_{..},$$

wobei X_{ij}^Z für die zentrierte Variable steht.

Hat eine Person beispielsweise einen Score von Null auf der Grand-Mean-zentrierten KFT-Variable, so bedeutet dies, dass die KFT-Leistung der Person genau dem Durchschnitt in dieser Stichprobe entspricht. Negative Werte auf der Variable X_{ij}^Z bedeuten unterdurchschnittliche, positive Werte auf X_{ij}^Z überdurchschnittliche Scores.

In unserem OLS-Regressionsmodell führt die Zentrierung wie in Kapitel 3.2 beschrieben dazu, dass wir den geschätzten Intercept-Term $\hat{\beta}_0$ sinnvoller interpretieren können (am Slope β_1 ändert sich durch die Zentrierung im einfachen Regressionsmodell nichts). Der Intercept in einer Regression mit einer zentrierten Prädiktorvariablen gibt den erwarteten Wert der abhängigen Variablen für eine Person mit einem *durchschnittlichen* Wert auf der unabhängigen Variablen an. Die OLS-Regressionsgleichung für das Mathe-KFT-Beispiel verändert sich wie folgt, wenn statt der unzentrierten KFT-Variable die Grand-Mean-zentrierte KFT-Variable verwendet wird:

$$\text{mathe} = 12,533 + 0,279 \times \text{kft_zentriert} + \text{Residuum} .$$

Wir sehen, dass sich nur der Intercept, nicht aber der Slope der Regressionsgleichung ändert, wenn anstelle der unzentrierten die zentrierte Prädiktorvariable verwendet wird. Der Intercept von 12,533 besagt, dass für einen Schüler mit einem durchschnittlichen KFT-Wert ein Mathetestscore von 12,533 vorhergesagt wird, was eine inhaltlich sinnvolle Aussage ist.

In Mplus kann die Zentrierung von Variablen praktischerweise über einen einzigen zusätzlichen Befehl (*centering*) vorgenommen werden. Die Variablen müssen somit nicht vor der Analyse in Mplus zentriert werden. Es besteht dabei in Mplus die Wahl zwischen einer Zentrierung am Gesamtmittelwert (*Grand Mean centering*) und einer Zentrierung am Clustermittelwert (*Group Mean centering*). Grand Mean centering ist die häufiger verwendete Variante und wird in Mplus durch folgende Spezifikation erreicht:

```
variable: centering = grandmean(kft) ;
```

Group Mean centering kann in Mplus über folgenden Befehl angefordert werden:

```
variable: centering = groupmean(kft) ;
```

Ausführliche Diskussionen der Vor- und Nachteile verschiedener Zentrierungsarten in Multilevelmodellen findet man in den Arbeiten von Enders und Tofghi (2007), Hofmann und Gavin (1998) sowie Kreft, de Leeuw und Aiken (1995).

Abbildung 5.4 zeigt den Mplus-Input zur Spezifikation des Random-Intercept-Modells ohne Random Slope. Die Prädiktorvariable **kft** muss zunächst mit folgendem Befehl als Level-1-Variable definiert werden, wenn diese (wie in unserem Modell) nur auf Level 1 verwendet werden soll:

```
within = kft;
```

Außerdem wird **kft** vor der Schätzung der Modellparameter am Gesamtmittelwert zentriert (s.o.). Im **model**-Befehl wird spezifiziert, dass auf Level 1 (%within%) die Regression von **mathe** auf **kft** geschätzt werden soll. Da auf Level 2 nur ein Random Intercept vorhanden

und das Level-2-Modell ansonsten unstrukturiert ist, wird hier kein gesonderter Befehl für das Between-Level benötigt. Mplus schätzt automatisch die Parameter γ_{00} und $\sigma_{u_{0j}}^2$ auf Level 2. Der Befehl **stdyx** in der output-Zeile fordert zusätzlich die vollstandardisierte Lösung an.

```

Mplus - [2a_Random_Intercept_One_Way_Random_Effects_ANCOVA.inp]
File Edit View Mplus Graph Window Help
□ □ X
□ □ X
title: Random-Intercept-Modell für den Mathematikscore;
  1 Prädiktor (KFT-Score) auf Level 1
  Kein Random Slope, kein Prädiktor auf Level 2
  ("One Way Random Effects ANCOVA")

data: file = Mathe_KFT.dat;

variable: names = klasse ! Clustervariable (Schulklasse)
           mathe kft ! Mathe- und KFT-Score
           schultyp; ! 0 = Sekundarschule, 1 = Gymnasium

usevar = mathe kft;

cluster = klasse; ! Definition der hierarchischen Struktur

within = kft; ! definiert kft als Level-1-Variable

centering = grandmean(kft); ! zentriert kft am Gesamtmittelwert

analysis: type = twolevel;

model: %within%
      mathe on kft;

output: sampstat stdyx;

```

Ready Ln 25, Col 6 NUM

Abbildung 5.4: Mplus-Input-Datei zur Spezifikation des One-Way-Random-Effects-ANCOVA-Modells für die Variable **mathe** mit dem Level-1-Prädiktor **kft**.

Die wichtigsten Ergebnisse sind nachfolgend wiedergegeben. Unter MODEL RESULTS liefert uns Mplus zunächst den geschätzten Slope $\hat{\beta}_{1j} = \hat{\gamma}_{10} = 0,234$ und die geschätzte Residualvarianz $\hat{\sigma}_{r_{ij}}^2 = 14,883$ als Parameter des Within-Level. Auf dem Between-Level wurden der mittlere Intercept $\hat{\gamma}_{00} = 12,310$ und die Varianz $\hat{\sigma}_{u_{0j}}^2 = 5,403$ geschätzt.

MODEL RESULTS

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level					
MATHE	ON				
KFT					
		0.234	0.012	20.042	0.000
Residual Variances					
MATHE		14.883	1.052	14.141	0.000
Between Level					
Means					
MATHE		12.310	0.437	28.142	0.000
Variances					
MATHE		5.403	1.612	3.352	0.001

Es zeigt sich, dass die Level-1-Residualvarianz mit $\hat{\sigma}_{ij}^2 = 14,883$ in diesem Modell deutlich geringer ausfällt als im Intercept-Only-Modell, in welchem wir keinen Level-1-Prädiktor berücksichtigt hatten (im Intercept-Only-Modell wurde $\hat{\sigma}_{ij}^2$ auf 26,48 geschätzt, vgl. Abschnitt 5.4.1). Daraus können wir schließen, dass ein beträchtlicher Teil der Varianz in der Mathematikleistung auf Individual- (Schüler-)ebene durch Unterschiede in den individuellen KFT-Leistungen erklärt werden kann. Genaue Informationen zur Höhe der Varianzaufklärung erhalten wir unter der standardisierten Lösung (STDYX). Wir sehen, dass der standardisierte Level-1-Regressionskoeffizient für die Regression von **mathe** auf **kft** 0,773 beträgt. Der R^2 -Wert für diese Level-1-Regression ist mit 0,598 recht groß und zeigt, dass ca. 60% der Unterschiede zwischen Schülern in der Mathematikleistung auf Individualebene durch Unterschiede in den KFT-Scores erklärt werden können.

STANDARDIZED MODEL RESULTS (STDYX Standardization)

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level					
MATHE	ON				
KFT					
		0.773	0.020	38.617	0.000
Residual Variances					
MATHE		0.402	0.031	12.987	0.000

Between Level				
Means				
MATHE	5.296	0.763	6.938	0.000
Variances				
MATHE	1.000	0.000	999.000	999.000
R-SQUARE				
Within Level				
Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
MATHE	0.598	0.031	19.309	0.000

5.4.3 Means-as-Outcomes-Modell

Im vorherigen Abschnitt wurde ein Random-Intercept-Modell mit einem Level-1-Prädiktor ohne Level-2-Prädiktoren besprochen. Wir wollen nun ein Modell betrachten, in welchem auf Level 1 kein Prädiktor vorhanden ist, auf Level 2 jedoch eine oder mehrere Prädiktorvariablen eingeführt werden. Ein solches Modell ist dann nützlich, wenn man ausschließlich daran interessiert ist, Mittelwertsunterschiede zwischen Clustern in Bezug auf die abhängige Variable (z.B. Unterschiede in den Mathematikmittelwerten zwischen Schulklassen) durch Makrovariablen (z.B. Schultyp) zu erklären. Die Level-1- und Level-2-Gleichungen für ein Modell ohne Level-1- aber mit einem Level-2-Prädiktor W_j lauten:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + r_{ij} .$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01} \cdot W_j + u_{0j} .$$

Auf Level 1 wird wie im Intercept-Only-Modell (siehe Abschnitt 5.4.1) lediglich eine Zerlegung der Messwerte in Clustermittelwert β_{0j} plus individuelle Abweichung vom Clustermittelwert (r_{ij}) vorgenommen. Auf Level 2 werden Unterschiede in den Clustermittelwerten durch den Level-2-Prädiktor W_j erklärt. Die Parameter γ_{00} und γ_{01} sind Regressionskoeffizienten (Intercept und Slope, fixed effects) für die Regression der Clustermittelwerte auf den Level-2-Prädiktor. Der Term u_{0j} steht für den Anteil eines Clustermittelwertes β_{0j} , der nicht durch W_j vorhergesagt werden kann. Da in diesem Modell der Fokus auf der Erklärung von Mittelwertsunterschieden zwischen Clustern durch Level-2-Variablen liegt, wird das Modell als *Means-as-Outcomes-Modell* bezeichnet (Luke, 2004).

Übertragen auf unser Datenbeispiel können wir die Frage untersuchen, ob Unterschiede in der Mathematikleistung auf Klassenebene durch die Zugehörigkeit der Klassen zu

unterschiedlichen Schultypen (Sekundarschule versus Gymnasium) erklärt werden können. Die Variable Schultyp dient als Level-2-Prädiktorvariable W_j , welche zur Vorhersage von Mittelwertsunterschieden hinsichtlich der Mathematikleistung zwischen den Klassen herangezogen wird.

Mplus liefert uns für das Means-as-Outcomes-Modell Schätzungen für die folgenden Modellparameter:

- Auf Level 1 (WITHIN):
 - die Residualvarianz $\sigma_{r_{ij}}^2$.
- Auf Level 2 (BETWEEN):
 - den Intercept γ_{00} für die Regression von β_{0j} auf W_j
 - den Steigungskoeffizienten γ_{01} für die Regression von β_{0j} auf W_j
 - die Residualvarianz $\sigma_{u_{0j}}^2$.

In unserem Beispiel haben γ_{00} und γ_{01} eine spezielle Bedeutung, da unser Level-2-Prädiktor Schultyp eine dichotome Variable mit den Kategorien 0 = Sekundarschule und 1 = Gymnasium ist. Aus diesem Grund gibt γ_{00} den erwarteten Mittelwert der Mathematikleistung in der Gruppe Sekundarschule an. Der Steigungskoeffizient γ_{01} repräsentiert den erwarteten Mittelwertsunterschied in der Mathematikleistung (Gymnasium minus Sekundarschule). Es sei darauf hingewiesen, dass die Interpretation der Koeffizienten bei Wahl eines anderen Kodierschemas für die Variable Schultyp (z.B. Effekt- oder Kontrastkodierung) anders ausfallen würde (siehe dazu Cohen et al. 2003, Kapitel 8).

Abbildung 5.5 zeigt den Mplus-Input zur Spezifikation des Means-as-Outcomes-Modells für die Mathematikleistung. Wichtig ist, dass man Mplus zunächst mitteilen muss, dass es sich bei **schultyp** um eine Level-2-(Between-)Variable handelt. Dies geschieht über folgenden zusätzlichen Befehl unter **variable**:

```
between = schultyp;
```

Zu beachten ist, dass *between*-Variablen nur *zwischen*, jedoch nicht *innerhalb* der Cluster unterschiedliche Werte annehmen dürfen (andernfalls gibt Mplus Fehlermeldungen aus). Im **model**-Statement wird festgelegt, dass die Mathematikleistung (Variable **mathe**) auf Level 2 (also auf Klassenebene) durch die Variable **schultyp** vorhergesagt werden soll. Da auf Level 1 nur ein Random Intercept vorhanden ist und das Level-1-Modell ansonsten unstrukturiert ist, wird hier kein Statement für das *within*-Level benötigt. Mplus schätzt für das *within*-Level automatisch die Residualvarianz $\sigma_{r_{ij}}^2$. Der Befehl **stdyx** in der *output*-Zeile fordert zusätzlich die vollstandardisierte Lösung an.

Abbildung 5.5: Mplus-Input-Datei zur Spezifikation des Means-as-Outcomes-Modells für die Variable **mathe** mit dem Level-2-Prädiktor **schultyp**.

Im Output erhalten wir folgende Schätzergebnisse:

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
Variances				
MATHE	26.467	2.450	10.802	0.000
Between Level				
MATHE	ON			
SCHULTYP		10.641	1.188	8.957
Intercepts				
MATHE		10.003	0.441	22.671
Residual Variances				
MATHE		3.980	0.887	4.488

Für das Within-Level schätzt Mplus nur die Level-1-Residualvarianz ($\hat{\sigma}_{r_{ij}}^2 = 26,467$). Für das Between-Level wird der Slope-Koeffizient auf $\hat{\gamma}_{01} = 10,641$ geschätzt. Das bedeutet, dass der erwartete Mittelwertsunterschied in der Mathematikleistung zwischen Gymnasium und Sekundarschule bei 10,641 Punkten liegt. Dieser Wert ist statistisch signifikant ($z = 8,957, p < 0,001$). Somit ist der Mittelwert der Mathematikleistung im Gymnasium signifikant größer als in der Sekundarschule.

Der Intercept-Term der Level-2-Regression wird von Mplus auf $\hat{\gamma}_{00} = 10,003$ geschätzt. Das bedeutet, dass der auf Basis des Modells geschätzte Mittelwert im Mathematiktest in der Sekundarschule 10,003 Punkte beträgt. Der geschätzte Mittelwert für den Schultyp Gymnasium kann über die Summe der Level-2-Intercept- und Slope-Koeffizienten bestimmt werden und beträgt $\hat{\gamma}_{00} + \hat{\gamma}_{01} = 10,003 + 10,641 = 20,644$.

Die Level-2-Residualvarianz wird in diesem Modell auf $\hat{\sigma}_{u_{0j}}^2 = 3,980$ geschätzt. Somit ist die Level-2-Residualvarianz nach Aufnahme des Prädiktors Schultyp im Vergleich zu der im Intercept-Only-Modell geschätzten Level-2-Varianz von $\hat{\sigma}_{u_{0j}}^2 = 20,853$ (vgl. Abschnitt 5.4.1) deutlich reduziert. Dies zeigt, dass der Level-2-Prädiktor Schultyp einen beträchtlichen Teil der Unterschiede in den Mathematikleistungen zwischen Schulklassen erklären kann.

Das Fehlerbalkendiagramm in Abbildung 5.6 veranschaulicht diesen Sachverhalt. Die Kreise bzw. Kästchen zeigen die (ungewichteten) Mittelwerte in den 34 Schulklassen (Clustermittelwerte). Die Fehlerbalken indizieren jeweils eine Standardabweichung ober- bzw. unterhalb des entsprechenden Clustermittelwertes. Kreise beziehen sich auf die Clustermittelwerte der Sekundarschulklassen, während die Kästchen die Mittelwerte der Gymnasialklassen zeigen. Die durchgängige Linie illustriert den im Means-as-Outcomes-Modell geschätzten Intercept-Koeffizienten $\hat{\gamma}_{00}$ (also den geschätzten Sekundarschulmittelwert). Die gestrichelte Linie zeigt den geschätzten Gymnasialmittelwert (die Summe der Koeffizienten $\hat{\gamma}_{00} + \hat{\gamma}_{01}$). Das Diagramm veranschaulicht den geschätzten großen Mittelwertsunterschied von 10,641 Punkten in der Mathematikleistung zwischen Sekundarschule und Gymnasium. Dieser Effekt führt in unserem Means-as-Outcomes-Modell zu dem signifikanten $\hat{\gamma}_{01}$ -Koeffizienten und dazu, dass sich die Level-2-Varianz durch Aufnahme des Prädiktors Schultyp stark reduziert.

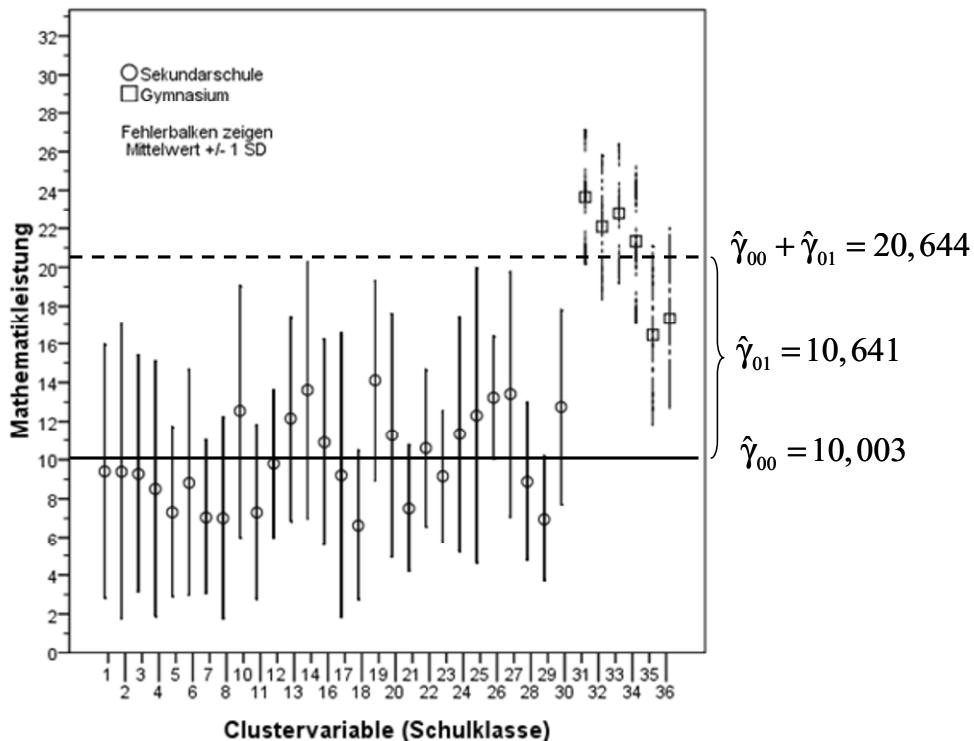


Abbildung 5.6: Fehlerbalkendiagramm zur Veranschaulichung des Zusammenhangs zwischen Schultyp und Mathematikleistung sowie der im Means-as-Outcomes-Modell geschätzten Level-2-Regressionskoeffizienten.

Das genaue Ausmaß der Varianzaufklärung können wir der standardisierten Lösung (STDYX, s.u.) entnehmen. Wir können zunächst erkennen, dass der standardisierte Level-2-Regressionskoeffizient mit 0,897 sehr groß ist. Der R^2 -Wert wird von Mplus auf 0,805 geschätzt. Somit können ca. 80,5% der Unterschiede in den Mathematiktestmittelwerten zwischen den Schulklassen durch die Zugehörigkeit zu einem der beiden Schultypen (Gymnasium vs. Sekundarschule) erklärt werden.

STANDARDIZED MODEL RESULTS (STDYX Standardization)				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
Variances				
MATHE	1.000	0.000	999.000	999.000
Between Level				
MATHE	ON SCHULTYP	0.897	0.033	27.184 0.000
Intercepts				
MATHE		2.213	0.246	9.013 0.000
Residual Variances				
MATHE		0.195	0.059	3.287 0.001
R-SQUARE				
Between Level				
Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
MATHE	0.805	0.059	13.592	0.000

5.5 Random-Intercept-and-Slope-Modelle

Bislang haben wir Mehrebenenregressionsmodelle besprochen, in welchen nur für den Intercept-Term einer Level-1-Regression Varianz auf Level 2 zugelassen wurde. Häufig ist es jedoch auch sinnvoll, Level-1-Steigungskoeffizienten (Slopes) über Cluster hinweg variieren zu lassen, d.h. Modelle mit Random Slopes zu spezifizieren.

Abbildung 5.7 zeigt auf deskriptiver Ebene, dass einfache lineare Regressionen der Matheleistung auf den KFT-Score in verschiedenen Schulklassen unterschiedlich ausfallen und zwar sowohl in Hinblick auf den Intercept der Regressionen (die Geraden schneiden die Y-Achse an unterschiedlichen Stellen) als auch in Hinblick auf den Slope (die Geraden haben unterschiedliche Steigungen). Somit erscheint es sinnvoll, auch für den Slope dieser Regression Varianz auf Level 2 zuzulassen. Modelle, in welchen sowohl Intercept als auch Slopes über Level-2-Einheiten hinweg variieren können, werden häufig als *Random-Coefficient-Regressionsmodelle* bezeichnet (Cohen et al., 2003; Luke, 2004). Modelle, in welchen Unterschiede in den Intercepts und Slopes zwischen Clustern durch Level-2-Variablen erklärt werden, nennt Luke (2004) *Intercepts-and-Slopes-as-Outcomes-Modelle*.

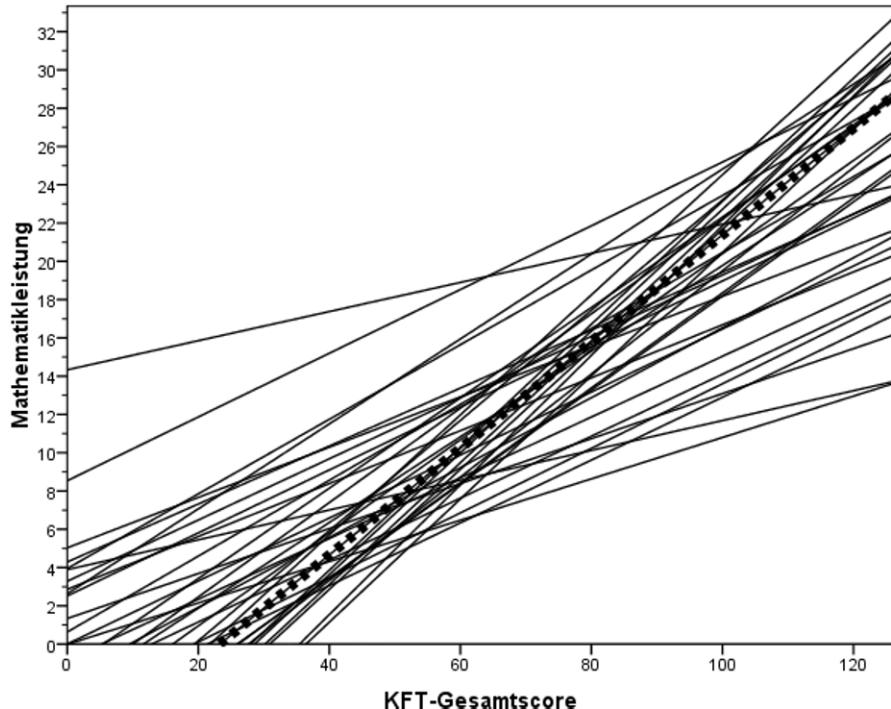


Abbildung 5.7: OLS-Regression der Mathematikleistung auf KFT-Score getrennt nach Schulklassen. Jede Linie steht für die lineare OLS-Regression in einer der 34 Schulklassen. Die dicke gestrichelte Linie repräsentiert die (unzentrierte) OLS-Regressionsgerade über alle Schulklassen hinweg.

5.5.1 Random-Coefficient-Regressionsanalyse

Wenn wir sowohl in Bezug auf den Intercept als auch in Bezug auf den Slope einer Level-1-Regression Unterschiede zwischen Clustern zulassen wollen (zunächst ohne dabei Prädiktorenvariablen auf Level 2 zu berücksichtigen), erweitert sich das One-Way-Random-Effects-ANCOVA-Modell (siehe Abschnitt 5.4.2) wie folgt:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j} \cdot X_{ij} + r_{ij} .$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} .$$

Ein solches Random-Coefficient-Regressionsmodell ohne Level-2-Prädiktoren ist beispielsweise dann nützlich, wenn man ausschließlich an Level-1-Prädiktoren interessiert ist, jedoch Daten mit Clusterstruktur untersucht. Würde man in einem solchen Fall eine konventionelle OLS-Regressionsanalyse ohne Berücksichtigung der mit der Clusterung verbundenen Ab-

hängigkeiten in den Daten durchführen, so müsste man mit einer Unterschätzung der Standardfehler der Regressionskoeffizienten und in der Folge mit verzerrten Signifikanztests und Konfidenzintervallen rechnen. Im Random-Coefficient-Regressionsmodell werden die Abhängigkeiten hingegen berücksichtigt und die Standardfehler, p -Werte und Konfidenzintervalle korrekt geschätzt.

Man beachte, dass der einzige Unterschied zum One-Way-Random-Effects-ANCOVA-Modell darin besteht, dass nun auch ein Residualterm (u_{1j}) in der Level-2-Gleichung für den Level-1-Slope-Koeffizienten aufgenommen wurde. Dies bedeutet, dass nun auch Varianz ($\sigma_{u_{1j}}^2$) in den Slopes über Cluster hinweg zugelassen wird. Für unser Beispiel bedeutet dies, dass wir zulassen, dass die Regressionsgeraden für die Regression des Mathematiktestscores auf den KFT-Score in jeder Schulkasse sowohl einen anderen Intercept als auch einen anderen Anstieg haben können (vgl. Abb. 5.7). Mplus liefert für das Random-Coefficient-Regressionsmodell Schätzungen der folgenden Parameter:

- Auf Level 1 (WITHIN):
 - die Residualvarianz $\sigma_{\eta_j}^2$.
- Auf Level 2 (BETWEEN):
 - den mittleren Intercept über alle Cluster, γ_{00}
 - die Varianz in den Intercepts über Cluster hinweg, $\sigma_{u_{0j}}^2$
 - den mittleren Steigungskoeffizienten (Slope), γ_{10}
 - die Varianz in den Slopes über Cluster hinweg, $\sigma_{u_{1j}}^2$
 - die Kovarianz zwischen Intercept und Slope, τ_{01} (muss explizit angefordert werden).

Abbildung 5.8 zeigt die Spezifikation eines Random-Coefficient-Regressionsmodells für unser Datenbeispiel.

The screenshot shows the Mplus software window with the title bar 'Mplus - [3_Random_Coefficient_Regression.inp]'. The menu bar includes File, Edit, View, Mplus, Graph, Window, and Help. Below the menu is a toolbar with various icons. The main area contains the Mplus input command language. The code is as follows:

```

title: Random-Coefficient-Regressionsanalyse
       mit 1 Level-1-Prädiktor (KFT-Score)
       Modell mit Random Intercept und Random Slope
       ohne Level-2-Prädiktoren

data: file = Mathe_KFT.dat;

variable: names = klasse ! Clustervariable (Schulklasse)
           mathe kft ! Mathe- und KFT-Scores
           schultyp; ! 0 = Sekundarschule, 1 = Gymnasium

usevar = mathe kft;

cluster = klasse; ! Definition der hierarchischen Struktur

within = kft; ! definiert kft als Level-1-Variable

centering = grandmean(kft); ! zentriert den KFT-Score am Gesamtmittelwert

analysis: type = twolevel random; ! Option "random" wird benötigt,
           ! wenn Random Slopes geschätzt werden sollen

model: %within%
       Beta1j | mathe on kft; ! Spezifikation der Level-1-Regression mit Random Slope

       %between%
       mathe with Beta1j; ! Kovarianz zwischen Random Intercept und Random Slope

output: sampstat;

```

The status bar at the bottom left says 'Ready' and at the bottom right says 'Ln 14, Col 69'.

Abbildung 5.8: Mplus-Input-Datei zur Spezifikation des Random-Coefficient-Regressionsmodells mit Random Intercept und Random Slope für die Level-1-Regression der Variable **mathe** auf den Prädiktor **kft**.

Die Prädiktorvariable **kft** wird wiederum mit `within =` als Level-1-Variable definiert und in zentrierter Form (hier: Zentrierung am Grand Mean) aufgenommen, um eine bessere Interpretierbarkeit der Ergebnisse zu ermöglichen (vgl. Abschnitt 5.4.2):

```

within = kft;
centering = grandmean(kft);

```

Auf Level 1 wird die Mathematikleistung durch den KFT-Score vorhergesagt. Intercept und Slope der Level-1-Regression sind Random Coefficients, die auf Level 2 variieren können. Um in Mplus Random Slopes analysieren zu können, muss im `analysis`-Befehl `type = twolevel random` gewählt werden. Anschließend wird der entsprechende Random Slope im `model`-Befehl definiert und von Mplus als latente Variable behandelt, deren Varianz dann automatisch auf Level 2 geschätzt wird. Dies geschieht folgendermaßen:

```

%within%
Beta1j | mathe on kft;

```

wobei **Beta1j** der frei gewählte Name für den Random Slope ist. Diese Befehlszeile bedeutet, dass der Slope **Beta1j** der Level-1-Regression des Mathematikscores auf den KFT-Score eine latente Variable ist, die über die Level-2-Einheiten hinweg variieren kann. Auf Level 2 fordern wir die Schätzung der Kovarianz τ_{01} zwischen dem Random Intercept und dem Random Slope an, indem wir folgenden Befehl aufnehmen:

```
%between%
mathe with Beta1j;
```

Man beachte, dass für `type = twolevel random` in Mplus keine standardisierten Parameterschätzungen erhältlich sind. Im Mplus-Output erhalten wir die folgenden Parameterschätzungen:

MODEL RESULTS

	Estimate	S.E.	Two-Tailed Est. / S.E.	P-Value
Within Level				
Residual Variances				
MATHE	14.862	1.091	13.627	0.000
Between Level				
MATHE WITH BETA1J	0.050	0.032	1.572	0.116
Means				
MATHE	12.195	0.456	26.771	0.000
BETA1J	0.233	0.012	19.401	0.000
Variances				
MATHE	4.801	1.356	3.542	0.000
BETA1J	0.001	0.001	0.672	0.502

Die Level-1-Residualvarianz wird auf $\hat{\sigma}_{r_{ij}}^2 = 14,862$ geschätzt. Dies ist ein deutlich geringerer Wert als im Intercept-Only-Modell (20,853), in welchem der Prädiktor **kft** nicht enthalten war. Für das Between-Level erhalten wir zunächst Schätzungen für den mittleren Intercept ($\hat{\gamma}_{00} = 12,195$) und den mittleren Slope ($\hat{\gamma}_{10} = 0,233$) über alle Cluster. Die über alle Schulklassen hinweg gemittelte Regressionsgleichung mit dem zentrierten KFT-Prädiktor lautet somit gemäß Mplus-Random-Coefficient-Regressionsanalyse:

$$\text{mathe} = 12,195 + 0,233 \times \text{kft_zentriert} + \text{Residuum} .$$

Zum Vergleich betrachten wir noch einmal die gemäß OLS-Regression ohne Berücksichtigung der Clusterstruktur geschätzte Gleichung:

$$\text{mathe} = 12,533 + 0,279 \times \text{kft_zentriert} + \text{Residuum}.$$

Wir sehen, dass die Ergebnisse recht ähnlich sind. Dennoch ist das Random-Coefficient-Regressionsmodell hier die zu bevorzugende Analysemethode, da es die Clusterstruktur der Daten berücksichtigt, die Standardfehler korrekt schätzt sowie Unterschiede in den Intercepts und Slopes über Schulklassen hinweg zulässt.

Diese Unterschiede können wir über die von Mplus unter Between Level ausgegebenen Varianzschätzungen analysieren. Wir erhalten von Mplus Schätzungen für die Varianz des Intercepts ($\hat{\sigma}_{u_{0j}}^2 = 4,801$) und des Slopes ($\hat{\sigma}_{u_{1j}}^2 = 0,001$) auf Level 2. Es zeigt sich, dass die Varianz in den Slopes eher gering ausfällt. Dennoch sollte man in einem solchen Fall nicht auf den Versuch verzichten, mögliche Unterschiede in den Slopes zwischen Schulklassen durch Level-2-Variablen zu erklären, sofern es dazu inhaltliche Hypothesen gibt. Dies geschieht im *Intercepts-and-Slopes-as-Outcomes-Modell*, welches wir im folgenden Abschnitt besprechen. Die von Mplus geschätzte Kovarianz zwischen Intercept und Slope beträgt $\hat{\tau}_{01} = 0,050$ und ist statistisch nicht signifikant ($z = 1,572$, $p = 0,116$).

5.5.2 *Intercepts-and-Slopes-as-Outcomes-Modell*

Häufig ist man bei der Analyse von Mehrebenendaten sowohl an Einflussvariablen auf Individualebene („Micro“-Variablen, Level-1-Prädiktoren) als auch an Einflussvariablen auf Clusterebene („Macro“-Variablen, Level-2-Prädiktoren) interessiert. Bislang haben wir nur Modelle besprochen, in welchen es entweder gar keine Prädiktorvariablen gab (Intercept-Only-Modell, Abschnitt 5.4.1) oder aber Prädiktorvariablen nur auf Level 1 (One-Way-Random-Effects-ANCOVA, Abschnitt 5.4.2; Random-Coefficient-Regression, Abschnitt 5.5.1) bzw. nur auf Level 2 (Means-as-Outcomes-Modell, Abschnitt 5.4.3). Wir wollen nun ein allgemeineres Modell kennenlernen, in welchem sowohl Random Intercepts als auch Random Slopes spezifiziert werden und in welchem sowohl auf Level 1 als auch auf Level 2 Prädiktorvariablen vorhanden sind. Dieses Modell bezeichnen wir gemäß Luke (2004) als *Intercepts-and-Slopes-as-Outcomes-Modell*. Zum Beispiel können wir mit diesem Modell für unsere abhängige Variable Mathematikleistung sowohl Prädiktorvariablen auf Individualebene (X_{ij} , z.B. individuelle KFT-Leistung) als auch auf Clusterebene (W_j ; z.B. Schultyp) untersuchen:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j} \cdot X_{ij} + r_{ij}.$$

$$\begin{aligned} \text{Level 2: } \beta_{0j} &= \gamma_{00} + \gamma_{01} \cdot W_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} \cdot W_j + u_{1j}. \end{aligned}$$

In unserem Datenbeispiel betrachten wir wiederum auf Level 1 die Regression der individuellen Mathestestscores auf die KFT-Scores. Der Intercept β_{0j} und der Slope β_{1j} dieser Regression können wie zuvor im Random-Coefficient-Regressionsmodell (siehe Ab-

schnitt 5.5.1) über die Schulklassen hinweg variieren. Der Residualterm r_{ij} steht für Unterschiede in der Mathematikleistung zwischen Schülern, die durch die im Modell vorhandenen Prädiktoren nicht erklärt werden können.

Auf Level 2 werden mögliche Unterschiede im Intercept β_{0j} und im Slope β_{1j} zwischen Schulklassen auf den Level-2-Prädiktor Schultyp (W_j) zurückgeführt. Es sei darauf hingewiesen, dass man auch Modelle spezifizieren kann, in welchen nur für einen der Parameter (z.B. nur für den Intercept-Term) Prädiktoren aufgenommen werden. Der Parameter γ_{00} kennzeichnet den erwarteten Intercept, wenn der Level-2-Prädiktor den Wert Null annimmt. In unserem Beispiel steht γ_{00} also für den Intercept in der mit Null kodierten Gruppe (Sekundarschule).

Der Parameter γ_{01} gibt den erwarteten Anstieg im Intercept an, wenn der Level-2-Prädiktor um eine Einheit erhöht wird. In unserem Fall gibt γ_{01} somit die erwartete Differenz in den Intercepts zwischen der Gruppe Gymnasium (**schultyp** = 1) und der Gruppe Sekundarschule (**schultyp** = 0) an.

Im Falle $\gamma_{01} = 0$ gibt es keinen Unterschied in den Intercepts zwischen den Schultypen. Wenn $\gamma_{01} > 0$, so ist der Intercept in der Gruppe Gymnasium größer als in der Gruppe Sekundarschule. Wenn $\gamma_{01} < 0$, so ist der Intercept in der Gruppe Sekundarschule größer als in der Gruppe Gymnasium.

Die Interpretation der Parameter γ_{10} und γ_{11} erfolgt nach demselben Prinzip, nur dass es hierbei um die Steigung der Regressionsgeraden geht. Der Parameter γ_{10} gibt den erwarteten Slope in der mit Null kodierten Gruppe (Sekundarschule) an. Der Parameter γ_{11} steht für die Differenz in den Slopes (Gymnasium – Sekundarschule). Im Falle $\gamma_{11} = 0$ gibt es keinen Unterschied in den Slopes zwischen den Schultypen. Wenn $\gamma_{11} > 0$, so ist der Slope in der Gruppe Gymnasium größer als in der Gruppe Sekundarschule. Wenn $\gamma_{11} < 0$, so ist der Slope in der Gruppe Sekundarschule größer als in der Gruppe Gymnasium.

Der Parameter γ_{11} ist häufig von besonderem Interesse, da dieser Koeffizient für einen möglichen Interaktionseffekt zwischen einem Level-1-Prädiktor (hier: **kft**) und einem Level-2-Prädiktor (hier: **schultyp**) steht. Dies bezeichnet man als *Cross-Level-Interaktion*. Eine Cross-Level-Interaktion bedeutet, dass sich ein Level-1-Prädiktor auf verschiedenen Stufen eines Level-2-Prädiktors unterschiedlich auf die Kriteriumsvariable auswirkt. In unserem Beispiel hieße das, dass die Regression der Matheleistung auf die KFT-Leistung für verschiedene Schultypen unterschiedliche Steigungen aufweist. Ein signifikanter γ_{11} -Koeffizient indiziert eine signifikante Cross-Level-Interaktion.

Die Residualterme u_{0j} und u_{1j} stehen für Unterschiede im Intercept und im Slope zwischen Schulklassen, welche nicht durch den Level-2-Prädiktor Schultyp erklärt werden können. Zusammengefasst erhalten wir in Mplus für das Intercepts-and-Slopes-as-Outcomes-Modell Schätzungen für die folgenden Modellparameter:

- Auf Level 1 (WITHIN):
 - die Residualvarianz $\sigma_{r_{ij}}^2$.
- Auf Level 2 (BETWEEN):
 - den Intercept für die Regression des Random Intercepts β_{0j} auf W_i , γ_{00}
 - den Slope für die Regression des Random Intercepts β_{0j} auf W_i , γ_{01}
 - den Intercept für die Regression des Random Slopes β_{1j} auf W_i , γ_{10}
 - den Slope für die Regression des Random Slopes β_{1j} auf W_i , γ_{11}
 - die Residualvarianz in den Intercepts über Cluster hinweg, $\sigma_{u_{0j}}^2$
 - die Residualvarianz in den Slopes über Cluster hinweg, $\sigma_{u_{1j}}^2$
 - die Residualkovarianz zwischen Intercept und Slope, τ_{01} (muss explizit angefordert werden).

Der Mplus-Inputfile zur Schätzung der Parameter des Intercepts-and-Slopes-as-Outcomes-Modells ist in Abbildung 5.9 dargestellt.

```
Mplus - [4_Intercepts_and_Slopes_as_Outcomes_Modell.inp]
File Edit View Mplus Graph Window Help
title: Intercepts-and-Slopes-as-Outcomes-Modell
       mit 1 Level-1-Prädiktor (KFT-Score)
       und 1 Level-2-Prädiktor (Schultyp)
       Modell mit Random Intercept und Random Slope

data: file = Mathe_KFT.dat;

variable: names = klasse ! Clustervariable (Schulklasse)
           mathe kft ! Mathe- und KFT-Scores
           schultyp; ! 0 = Sekundarschule, 1 = Gymnasium

usevar = mathe kft schultyp;

cluster = klasse; ! Definition der hierarchischen Struktur

within = kft; ! definiert kft als Level-1-Variable

centering = grandmean(kft); ! zentriert den KFT-Score am Gesamtmittelwert

between = schultyp; ! definiert schultyp als Level-2-Variable

analysis: type = twolevel random; ! Option "random" wird benötigt,
           ! wenn Random Slopes geschätzt werden sollen

model: %within%
       Beta1j | mathe on kft; ! Spezifikation der Level-1-Regression mit Random Slope

       %between%
       mathe on schultyp; ! Vorhersage des Intercepts durch den Level-2-Prädiktor
       Beta1j on schultyp; ! Vorhersage des Slopes durch den Level-2-Prädiktor
       mathe with Beta1j; ! Residualkovarianz zwischen Random Intercept & Random Slope

output: sampstat;
```

Abbildung 5.9: Mplus-Input-Datei zur Spezifikation des Intercepts-and-Slopes-as-Outcomes-Modells für die Regression der Variable **mathe** auf **kft** (Level 1) und **schultyp** (Level 2).

Man beachte, dass die Variable **kft** wiederum mittels `within = kft;` als Level-1-Variable definiert und über den `centering=Befehl` am Grand Mean zentriert wurde, um eine sinnvolle Interpretation der Schätzergebnisse zu gewährleisten. Die Variable **schultyp** wurde mit Hilfe des `between = -Befehls` im `variable=Statement` als Level-2-Variable definiert, um diese Variable auf Level 2 (`between`) benutzen zu können. Im Modellstatement wird auf Level 1 (`%within%`) wiederum ein Random Slope (`Beta1j`) für die Regression von **mathe** auf **kft** definiert. Auf Level 2 (`%between%`) wird zunächst die Regression des Random Intercepts auf den Level-2-Prädiktor **schultyp** angefordert (`mathe on schultyp;`). Anschließend wird spezifiziert, dass auch der Random Slope **Beta1j** auf **schultyp** zurückgeführt werden soll (`Beta1j on schultyp;`). Der Befehl `mathe with Beta1j;` fordert die Residualkovarianz ($\hat{\tau}_{01}$) zwischen Intercept und Slope an. Im Mplus-Output erhalten wir die folgenden Parameterschätzungen:

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
Residual Variances				
MATHE	14.608	1.070	13.658	0.000
Between Level				
BETA1J ON				
SCHULTYP	-0.058	0.018	-3.139	0.002
MATHE ON				
SCHULTYP	5.394	0.969	5.568	0.000
MATHE WITH				
BETA1J	0.035	0.022	1.567	0.117
Intercepts				
MATHE	11.536	0.396	29.149	0.000
BETA1J	0.226	0.013	17.258	0.000
Residual Variances				
MATHE	3.022	0.724	4.171	0.000
BETA1J	0.001	0.001	0.509	0.611

Die Level-1-Residualvarianz wird von Mplus auf $\hat{\sigma}_{r_{ij}}^2 = 14,608$ geschätzt. Auf Level 2 gibt Mplus zunächst den geschätzten Regressionskoeffizienten $\hat{\gamma}_{11}$ für die Regression des Random Slopes β_{1j} auf **schultyp** aus. Der geschätzte Koeffizient $\hat{\gamma}_{11}$ gibt die Differenz in den geschätzten mittleren Slopes (Gymnasium minus Sekundarschule) an. Der Wert beträgt $\hat{\gamma}_{11} = -0,058$ und ist statistisch signifikant von Null verschieden ($z = -3,139, p = 0,002$). Dies bedeutet, dass der Anstieg der Regressionsgeraden in der Regression der Mathematikleistung auf den KFT-Score im Gymnasium signifikant geringer ausfällt als in der Sekundarschule.

Dieser Cross-Level-Interaktions-Effekt wird in Abbildung 5.10 anhand eines Streudiagramms veranschaulicht, in welchem die (OLS-)Regressionsgeraden getrennt nach Schultyp dargestellt sind. Abbildung 5.10 zeigt, dass die Regressionslinie im Gymnasium (gestrichelt) einen flacheren Anstieg (geringeren Slope) aufweist als in der Sekundarschule. Der signifikante $\hat{\gamma}_{11}$ -Koeffizient zeigt uns, dass es sich hierbei um einen signifikanten Unterschied handelt.

Als nächstes wird von Mplus der geschätzte Regressionskoeffizient $\hat{\gamma}_{01}$ für die Regression des Random Intercept β_{0j} auf **schultyp** berichtet. Der Koeffizient $\hat{\gamma}_{01}$ gibt die Differenz in den geschätzten mittleren Intercepts (Gymnasium minus Sekundarschule) an. Der Wert beträgt $\hat{\gamma}_{01} = 5,394$ und ist ebenfalls statistisch signifikant ($z = 5,568, p < 0,001$). Somit

kann der Prädiktor **schultyp** nicht nur Unterschiede im Slope, sondern auch Unterschiede im Intercept erklären. Da der Prädiktor **kft** am Gesamtmittelwert zentriert wurde, bedeutet der geschätzte Wert von 5,394, dass für Gymnasiasten mit einem mittleren KFT-Score 5,394 Punkte mehr im Mathetest erwartet werden als für Sekundarschüler mit mittlerem KFT-Wert. Den Unterschied in den Intercepts können wir uns ebenfalls anhand von Abbildung 5.10 veranschaulichen: Die Regressionslinien der beiden Schultypen schneiden den Nullpunkt der zentrierten KFT-Variable an unterschiedlichen Stellen (die gestrichelte Linie des Gymnasiums schneidet den Nullpunkt oberhalb der durchgehenden Linie der Sekundarschule).

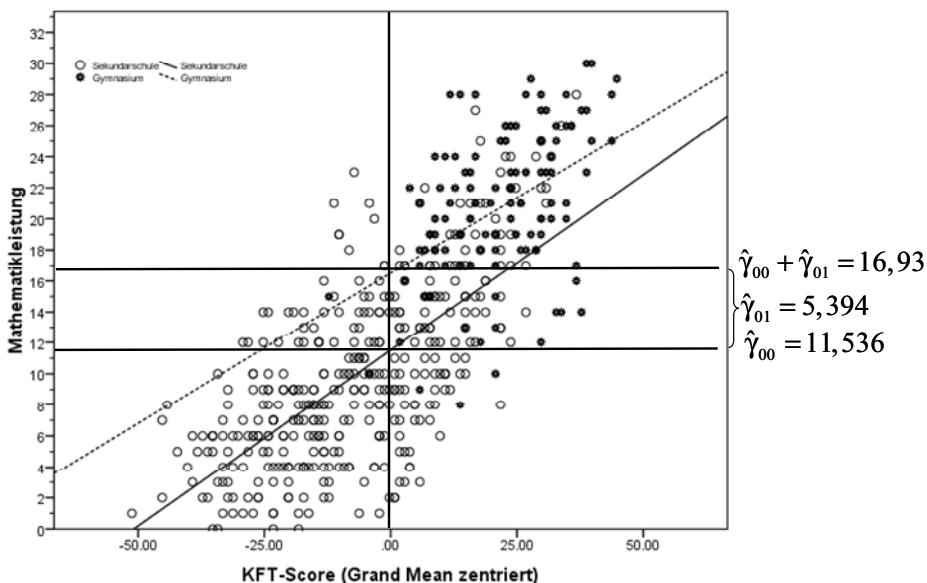


Abbildung 5.10: OLS-Regression der Mathematikleistung auf den am Grand Mean zentrierten KFT-Score getrennt nach Schultyp.

Unter „Intercepts“ wird zunächst der geschätzte Intercept $\hat{\gamma}_{00}$ für die Regression des Random Intercepts β_{0j} auf **schultyp** mitgeteilt ($\hat{\gamma}_{00} = 11,536$). An diesem Wert können wir erkennen, dass der geschätzte Intercept für die Regression der Matheleistung auf den KFT-Score in der Sekundarschule 11,536 beträgt (vgl. Abb. 5.10). Somit wird für Sekundarschüler mit durchschnittlichem KFT-Wert ein Mathetestwert von 11,536 erwartet. Aus der Summe der Koeffizienten $\hat{\gamma}_{00}$ und $\hat{\gamma}_{01}$ können wir den erwarteten Intercept in der Gruppe Gymnasium ausrechnen: $\hat{\gamma}_{00} + \hat{\gamma}_{01} = 11,536 + 5,394 = 16,93$. Mithin erwarten wir für Gymnasiasten mit durchschnittlichem KFT-Score einen Mathetestwert von 16,93.

Außerdem wird der geschätzte Intercept $\hat{\gamma}_{10}$ für die Regression des Random Slopes β_{1j} auf **schultyp** ausgegeben ($\hat{\gamma}_{10} = 0,226$). Dieser Wert gibt in unserem Beispiel die ge-

schätzte Steigung der Regressionsgerade für die Regression der Matheleistung auf den KFT-Score in der Gruppe Sekundarschule an ($\hat{\gamma}_{10}$ entspricht in etwa der Steigung der durchgehenden Regressionsgerade in Abb. 5.10). Somit können wir nun auch den erwarteten Slope in der Gruppe Gymnasium ausrechnen, und zwar aus der Summe $\hat{\gamma}_{10} + \hat{\gamma}_{11} = 0,226 + (-0,058) = 0,168$.

Auf Level 2 werden von Mplus außerdem noch die geschätzten Residualvarianzen für den Intercept ($\hat{\sigma}_{u_{0j}}^2 = 3,022$) und den Slope ($\hat{\sigma}_{u_{1j}}^2 = 0,001$) sowie die geschätzte Residualkoeffizienten zwischen Intercept und Slope ($\hat{\tau}_{01} = 0,035$) ausgegeben.

Abschließend sei noch darauf hingewiesen, dass man in Mplus auch Multilevelregressionsmodelle mit mehr als nur einem Level-1- und Level-2-Prädiktor analysieren kann. In diesen Modellen können einige oder alle Intercepts und Slopes als Random Coefficients spezifiziert werden. Für Prädiktoren, deren Slopes nicht als Random Coefficients spezifiziert werden sollen, wird in einem separaten on-Befehl einfach das | Zeichen sowie der Name für den Random Slope vor dem | weggelassen. Als Beispiel betrachten wir den folgenden Befehl:

```
%within%
Beta1j | y on x1;
y on x2;
```

Hierbei würde in der Level-1-Regression der Variable y auf x1 und x2 nur der Slope von x1 als Random Coefficient spezifiziert. Der Slope von x2 bleibt hier eine Konstante, die nicht zwischen Clustern variiert.

Eine Stärke von Mplus liegt darin, dass auch multivariate Mehrebenenmodelle analysiert werden können. Das heißt, dass bei Vorliegen einer Mehrebenenstruktur z.B. auch Pfadanalysen, konfirmatorische Faktorenanalysen, Strukturgleichungsmodelle (siehe Kapitel 3), Wachstumskurvenmodelle (siehe Kapitel 4.5) und Latent-Class-Modelle (siehe Kapitel 6) mit Hilfe der Option type = twolevel als Mehrebenenmodelle spezifiziert werden können. Die Analyse von Mehrebenen-Strukturgleichungsmodellen in Mplus wird im Fortsetzungsbuch von Geiser et al. (in Vorb.) besprochen.

6 Latent-Class-Analyse

6.1 Einführung in die Latent-Class-Analyse (LCA)

Die *Latent-Class-Analyse* (LCA) oder *Latent-Structure-Analysis* (Goodman, 1974; Lazarsfeld & Henry, 1968) ist ein statistisches Verfahren, das zur Klassifizierung von Personen in homogene Subgruppen („latente Klassen“) eingesetzt werden kann. Ausgangspunkt für die Klassifizierung sind die beobachteten Antwortmuster von Personen über eine Reihe von kategorialen (nominalen oder ordinalen) Test- oder Fragebogenitems (z.B. Intelligenztestaufgabe gelöst/nicht gelöst; Symptom vorhanden ja/nein; Fragebogenitems gemessen mittels 4-stufiger Likert-Skala). In einer LCA werden Zusammenhänge zwischen den Items durch das Vorhandensein von *a priori* unbekannten Subpopulationen (latenten Klassen) erklärt. Anders ausgedrückt: Die für ein Set von Items beobachteten interindividuellen Unterschiede in den Antwortmustern werden durch die Zugehörigkeit zu latenten Klassen mit klassenspezifischen Antwortprofilen erklärt.

Zum Beispiel könnten Personen, die nur bestimmte Items eines Tests zur Messung des räumlichen Vorstellungsvermögens mit hoher Wahrscheinlichkeit lösen, einer Klasse mit einer besonderen Lösungsstrategie zugeordnet werden. Personen, die im Test zufällige Antworten geben, könnten mit Hilfe der LCA zu einer Klasse von „Ratern“ zusammengefasst werden. Im Bereich der klinischen Psychologie könnten Patienten anhand unterschiedlicher Symptommuster bestimmten Störungs- oder Krankheitstypen zugeordnet werden. Ein Beispiel aus dem Bereich der Persönlichkeitsforschung wäre die Klassifizierung von Personen gemäß bestimmter Persönlichkeitstypen. Wesentliche Ziele einer LCA werden in Box 6.1 zusammen gefasst.

Diese Auflistung verdeutlicht Ähnlichkeiten zwischen der LCA und der exploratorischen Faktorenanalyse (EFA). Beide Verfahren können zur Datenreduktion eingesetzt werden. In der LCA werden die Antwortmuster zahlreicher Personen auf wenige latente Klassen „reduziert“. In der EFA werden Zusammenhänge zwischen kontinuierlichen Variablen auf wenige kontinuierliche Faktoren zurückgeführt. Ein wesentlicher Unterschied zwischen beiden Verfahren ist, dass in der EFA kontinuierliche latente Variablen („Faktoren“) und metrische Indikatoren angenommen werden, wohingegen in der LCA eine kategoriale latente Variable (Latent-Class-Variable mit mehreren Kategorien = latenten Klassen) zur Aufklärung der Zusammenhänge zwischen kategorialen Indikatoren eingeführt wird.

Box 6.1: Wesentliche Ziele einer Latent-Class-Analyse (LCA)

- Bestimmung der Anzahl an latenten Klassen, die nötig ist, um Unterschiede in den beobachteten Antwortmustern zufriedenstellend und inhaltlich bedeutsam erklären zu können. Hierbei sollte das Modell möglichst sparsam sein, d.h. mit möglichst wenigen latenten Klassen auskommen.
- Bestimmung der Charakteristika der gefundenen Klassen: Was ist das Spezifische für eine bestimmte Klasse? Worin unterscheiden sich die Klassen? Inwiefern können die Klassen (z.B. im Sinne einer Typologie) inhaltlich interpretiert werden?
- Überprüfung einer Theorie über das Vorliegen typologischer Unterschiede (konfirmatorische LCA)
- Bestimmung der Zuverlässigkeit (Reliabilität) der Klassifikation auf der Basis der gewählten Klassenlösung
- ggf. Bestimmung der individuellen Klassenzugehörigkeit für jedes Individuum für diagnostische Zwecke
- ggf. Analyse von Zusammenhängen der Klassenzugehörigkeit mit externen Variablen (z.B. Alter, Geschlecht, Persönlichkeitsvariablen, Intelligenz usw.)

Die LCA kann nicht nur exploratorisch, sondern auch konfirmatorisch eingesetzt werden. So kann man beispielsweise mit Hilfe der LCA Theorien über typologische Unterschiede zwischen Personen (z.B. unterschiedliche Bindungsstile, Persönlichkeitstypen, Lösungsstrategien etc.) testen, indem man a priori eine bestimmte Anzahl von Klassen vorgibt und ggf. weitere Parameterrestriktionen (z.B. hinsichtlich der Eigenschaften der Klassen oder der Klassengrößen) einführt. Über Modelltests oder Modellvergleiche kann dann ähnlich wie bei linearen Strukturgleichungsmodellen (vgl. Kapitel 3) geprüft werden, ob das zur Theorie korrespondierende Klassenmodell die Daten (in diesem Fall die beobachteten Antwortmuster) adäquat reproduzieren kann.

Der Einfachheit halber werden wir im vorliegenden Kapitel nur die Analyse dichotomer Variablen mittels exploratorischer LCA besprechen. Das Vorgehen in Mplus lässt sich jedoch problemlos auf mehrstufig-kategoriale Variablen übertragen. Für eine dichotome Variable lässt sich das Latent-Class-Modell für ein Item wie folgt notieren (siehe z.B. Rost, 2004):

$$p(X_{vi} = 1) = \sum_{g=1}^G \pi_g \pi_{ig}.$$

In diesem Modell bezeichnet $p(X_{vi} = 1)$ die Wahrscheinlichkeit, dass eine Person v auf einem Item i einen Wert von 1 hat (z.B. ein Item löst oder eine Frage bejaht). Der Parameter π_g steht für die Wahrscheinlichkeit, der latenten Klasse g anzugehören, wobei $\sum_{g=1}^G \pi_g = 1$ und kann als *Klassengrößenparameter* interpretiert werden. Das Modell nimmt an, dass jede Person einer (und nur einer) Klasse angehören muss.

Der Parameter π_{ig} gibt die Wahrscheinlichkeit für den Wert 1 bei einem Item i in Klasse g (klassenbedingte Antwort- bzw. Lösungswahrscheinlichkeit) an. Anders formuliert, π_{ig} gibt uns an, wie wahrscheinlich es ist, dass eine Person das Item i löst, wenn wir wissen, welcher Klasse g die Person angehört. Das Modell sagt also aus, dass die Lösungswahrscheinlichkeit (bzw. Antwortwahrscheinlichkeit) für ein Item davon abhängt, welcher Klasse eine Person angehört und wie hoch die Lösungswahrscheinlichkeit für das betreffende Item in dieser Klasse ist. Box 6.2 gibt einen Überblick über die in einem LCA-Modell geschätzten Parameter.

Box 6.2: Parameter, die in einer LCA für dichotome Variablen geschätzt werden

- π_g : Die Klassengrößenparameter (*class proportions*). Es werden $g-1$ Klassengrößenparameter π_g geschätzt (die Größe der letzten Klasse ist kein unabhängiger Modellparameter, da sie sich aufgrund der Modellannahme

$$\sum_{g=1}^G \pi_g = 1 \text{ als } 1 - \sum_{g=1}^{G-1} \pi_g \text{ ergibt.}$$
 - π_{ig} : Die bedingten Antwortwahrscheinlichkeiten für jedes Item in jeder latenten Klasse (*conditional response probabilities*). Bei dichotomen Indikatoren werden $g \times i$ klassenbedingte Antwortwahrscheinlichkeiten π_{ig} geschätzt (die Gegenwahrscheinlichkeit für jeden π_{ig} -Parameter ergibt sich bei dichotomen Variablen als $1 - \pi_{ig}$). Mplus gibt im Output neben den Wahrscheinlichkeitsparametern π_g und π_{ig} auch die Parameter der äquivalenten *Logit-Parametrisierung* der LCA aus (siehe z.B. Hagenaars, 1993). Hierbei handelt es sich um sogenannte *Thresholds* oder logische *Schwellenparameter*
- Die Umrechnung der Thresholdparameter in Wahrscheinlichkeiten wird im Rahmen des Datenbeispiels unten gezeigt.

Man beachte, dass die *Anzahl der Klassen*, die benötigt wird, um die beobachteten Antwortmuster adäquat abbilden zu können, kein im Modell zu schätzender Parameter ist (Rost, 2004). In der exploratorischen LCA wird die Anzahl der Klassen in der Praxis meist über Modellvergleiche bestimmt. Dabei werden Modelle mit unterschiedlichen Klassenzahlen auf denselben Datensatz angepasst. Mit Hilfe statistischer Indices wird die Datenanpassung der Modelle miteinander verglichen. Das Modell mit der besten Datenanpassung wird ausgewählt, sofern es sich dabei auch um eine von den Parameterschätzungen her adäquate sowie inhaltlich gut interpretierbare Lösung handelt. Verschiedene Kriterien zur Beurteilung der Modellgüte und zur Durchführung von Modellvergleichen werden in Abschnitt 6.3 besprochen.

Ausführlichere Besprechungen und Weiterführendes zur LCA findet man bei Clogg (1995), Collins und Lanza (2010), Hagenaars (1993), Hagenaars und McCutcheon (2002),

Langeheine und Rost (1988, 1993), McCutcheon (1987), Rost (2004, 2006), Rost und Langeheine (1997) sowie auf der Homepage von John Uebersax (<http://ourworld.compuserve.com/homepages/jrsuebersax/>). Die Verwendung der LCA für multiple Gruppenvergleiche (Multigruppen-LCA) besprechen Eid, Langeheine und Diener (2003). Ein Anwendungsbeispiel für eine Multigruppen-LCA kann der Arbeit von Geiser, Lehmann und Eid (2006) entnommen werden, die auch eine Mplus-Beispielsyntax enthält.

6.2 Spezifikation von LCA-Modellen in Mplus

Die Spezifikation von LCA-Modellen wird in Mplus über das Modul zur Analyse von Mischverteilungsmodellen (Option analysis: type = mixture) ermöglicht. Aus diesem Grund ist das Mixture- oder das Combination-Add-On zur Analyse von LCA-Modellen in Mplus erforderlich (die Base-Version von Mplus ist nicht ausreichend).

In Box 6.3 wird das Datenbeispiel zur LCA eingeführt. Abbildung 6.1 zeigt den Mplus-Input für ein LCA-Modell mit 3 latenten Klassen.

Box 6.3: Fallbeispiel LCA

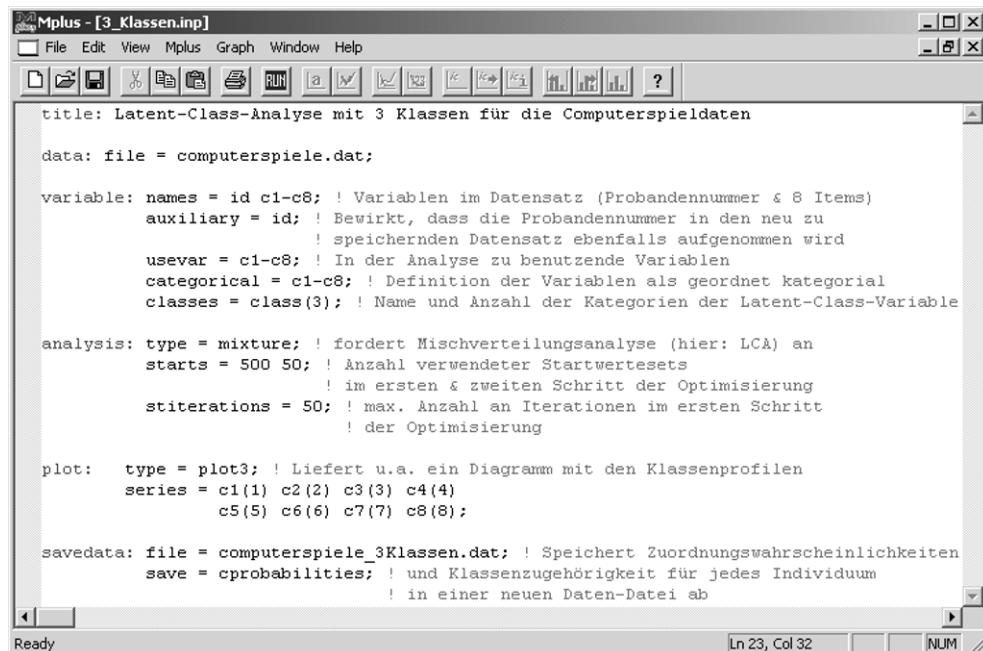
Wir veranschaulichen die Durchführung einer LCA in Mplus anhand des Datensatzes **computerspiele.dat**. Dieser Datensatz enthält die Itemantworten von $N = 861$ Kindern, die den folgenden Fragebogen zu Ihren Computerspielpräferenzen ausgefüllt haben:

	nie	selten	oft	sehr oft
1. Abenteuerspiele z.B. KingsQuest, Myst	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Actionspiele z.B. Quake, Doom	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Sportspiele z.B. Fifa, NHL	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Rollenspiele z.B. Betrayal of Antara, Ultima	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Denk- und Logikspiele z.B. Schach, Block out	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Geschicklichkeitsspiele z.B. PacMan, Tetris	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Simulationsspiele z.B. Siedler, Alpha	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Fahrzeugsimulatoren z.B. RedBaron, Motocross Madness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Mit Hilfe einer LCA soll herausgefunden werden, ob sich auf Basis der acht Computerspielitems verschiedene Typen von Computerspielern identifizieren lassen. Konkret

möchten wir wissen, wie viele latente Klassen wir annehmen müssen, um die Antwortmuster der Kinder zufriedenstellend abbilden zu können und was jede Klasse (jeden Computerspieltyp) auszeichnet (Details zu dieser Studie können in Quaiser-Pohl, Geiser & Lehmann, 2006, nachgelesen werden). Außerdem sind wir daran interessiert zu erfahren, wie zuverlässig wir die Kinder mit Hilfe des LCA-Modells einem bestimmten Computerspieltyp zuordnen können. Dies können wir anhand der Klassenzuordnungswahrscheinlichkeiten untersuchen, die von Mplus ebenfalls geschätzt werden.

Vor der Analyse wurden zur Vereinfachung unseres Beispiels die Kategorien „nie“ und „selten“ sowie die Kategorien „oft“ und „sehr oft“ des Fragebogens jeweils zu einer Kategorie zusammengefasst. Die Kategorie „nie/selten“ wurde mit dem Wert 0 und die Kategorie „oft/sehr oft“ mit dem Wert 1 kodiert. Man könnte die Items auch in 4-stufiger Form einer LCA unterziehen. Dies wäre in der Praxis zur Vermeidung eines Informationsverlustes auch zu empfehlen, sofern genügend Daten vorliegen. Um das Beispiel übersichtlich und anschaulich zu halten, wurde hier jedoch darauf verzichtet, und es wurden die dichotomisierten Varianten der Items analysiert.



The screenshot shows the Mplus software interface with the title bar 'Mplus - [3_Klassen.inp]'. The menu bar includes File, Edit, View, Mplus, Graph, Window, and Help. Below the menu is a toolbar with various icons. The main window displays the Mplus input command language. The code in the input file is as follows:

```

title: Latent-Class-Analyse mit 3 Klassen für die Computerspieldaten

data: file = computerspiele.dat;

variable: names = id c1-c8; ! Variablen im Datensatz (Probandennummer & 8 Items)
           auxiliary = id; ! Bewirkt, dass die Probandennummer in den neu zu
                           ! speichernden Datensatz ebenfalls aufgenommen wird
           usevar = c1-c8; ! In der Analyse zu benutzende Variablen
           categorical = c1-c8; ! Definition der Variablen als geordnet kategorial
           classes = class(3); ! Name und Anzahl der Kategorien der Latent-Class-Variable

analysis: type = mixture; ! fordert Mischverteilungsanalyse (hier: LCA) an
           starts = 500 50; ! Anzahl verwendeter Startwertesets
                           ! im ersten & zweiten Schritt der Optimisierung
           stiterations = 50; ! max. Anzahl an Iterationen im ersten Schritt
                           ! der Optimisierung

plot:   type = plot3; ! Liefert u.a. ein Diagramm mit den Klassenprofilen
       series = c1(1) c2(2) c3(3) c4(4)
               c5(5) c6(6) c7(7) c8(8);

savedata: file = computerspiele_3Klassen.dat; ! Speichert Zuordnungswahrscheinlichkeiten
           save = cprobabilities; ! und Klassenzugehörigkeit für jedes Individuum
                           ! in einer neuen Daten-Datei ab

```

The status bar at the bottom shows 'Ready', 'Ln 23, Col 32', and 'NUM'.

Abbildung 6.1: Mplus-Input-Datei zur Spezifikation einer LCA mit drei Klassen für die Computerspieldaten.

Der verwendete Datensatz heißt **computerspiele.dat**. Er enthält eine Probandenidentifikationsnummer (Variable **id**) sowie die acht zu analysierenden dichotomen Computerspieltags (Variablennamen: **c1-c8**). Mit dem Befehl

```
auxiliary = id;
```

teilen wir Mplus mit, dass es sich bei der Variable **id** um eine „Hilfsvariable“ handelt, die in der statistischen Analyse selbst keine Rolle spielt, die wir aber in den von Mplus neu zu generierenden Datensatz mit den individuellen Klassenzuordnungswahrscheinlichkeiten und der Klassenzuordnung (siehe Beschreibung des `savedata`-Befehls weiter unten) übernehmen möchten. Ohne diesen Befehl würde die Probandennummer nicht in den neuen Datensatz mit abgespeichert werden. Diese wäre ein Nachteil, wenn man die LCA beispielsweise für individuelle diagnostische Zwecke nutzen möchte.

Der Befehl

```
categorical = c1-c8;
```

legt fest, dass es sich bei den acht Computerspielitems um geordnet kategoriale (hier dichotome) Variablen handelt. Den gleichen Befehl würde man auch verwenden, wenn man mehrstufig (nicht dichotome, sondern polytome) geordnet kategoriale (ordinale) Items analysieren wollte (für ungeordnet kategoriale = nominale Items würde man den Befehl `nominal = verwenden`). Mit Hilfe des Befehls

```
classes = class(3);
```

legen wir fest, dass eine Latent-Class-Variable mit dem Namen `class` und drei Kategorien (d.h. drei latenten Klassen) zur Erklärung der Zusammenhänge zwischen den acht Items herangezogen werden soll (dass eine 3-Klassen-Lösung extrahiert werden soll). Über den Befehl

```
analysis: type = mixture;
```

teilen wir Mplus mit, dass wir eine *Mischverteilungsanalyse* (hier LCA) durchführen wollen. Mischverteilung bedeutet, dass wir annehmen, dass sich die Population der Schüler aus unterschiedlichen Subgruppen (latenten Klassen, Computerspieltypen) zusammensetzt.

Die Modellgütestatistiken (siehe Abschnitt 6.3) und Parameterschätzwerte der LCA werden in Mplus mit Hilfe eines iterativen Schätzverfahrens unter Verwendung der Maximum-Likelihood-(ML-)Schätzmethode bestimmt. Ziel ist die Maximierung der Likelihoodfunktion, d.h. die Bestimmung der Lösung mit dem größtmöglichen *Loglikelihood-Wert*. Der Loglikelihood-Wert ist ein Maß für die Wahrscheinlichkeit der Daten unter dem Modell und dient als Grundlage für die Berechnung zahlreicher Modelfittests bzw. -indices (siehe Abschnitt 6.3).

Eine Komplikation, welche bei Latent-Class-Analysen (und Mischverteilungsanalysen allgemein) besonders zu beachten ist, besteht darin, dass es häufig nicht gelingt, mit nur einem einzigen Set von Parameter-Startwerten die Lösung mit dem größten Loglikelihood-Wert zu finden. Die Konsequenz in einem solchen Fall ist, dass der Modellschätzprozess bei einem sogenannten *lokalen Likelihood-Maximum* endet, welches mit einer suboptimalen Lösung assoziiert ist. Eine solche lokale Lösung (*local solution*) kann mit inkorrekten Fitstatistiken und Parameterschätzwerten einhergehen (siehe z.B. Uebersax, 2000) und sollte daher unbedingt vermieden werden.

Die Gefahr, in einer LCA auf ein lokales Likelihood-Maximum zu treffen, ist groß, wenn nur ein einziges oder nur wenige Startwertesets verwendet werden. Lokale Likelihood-Maxima treten in praktischen Anwendungen bereits bei Modellen mit nur 4 oder 5 Klassen häufig auf (Uebersax, 2000). Die Verwendung einer ausreichenden Menge an zufälligen Startwertesets im ersten Schritt der Optimierung stellt sicher, dass letztendlich die optimale Lösung mit dem größten Loglikelihood-Wert vom Programm gefunden wird.

Die wichtigste Maßnahme zur Vermeidung von lokalen Lösungen ist die Verwendung einer großen Anzahl von zufällig generierten Parameterstartwerten (vgl. Box 6.4). Die Voreinstellung bei Mplus ist, dass 10 Sets von zufälligen Startwerten vom Programm generiert und bei der Modellschätzung verwendet werden. Diese relativ geringe Anzahl an Startwertesets ist in der Praxis jedoch häufig nicht ausreichend, insbesondere dann, wenn mehr als zwei oder drei latente Klassen extrahiert werden sollen.

Um auf Nummer sicher zu gehen, sollte bei Modellen mit mehr als zwei Klassen nicht die Mplus-Standardeinstellung (nur 10 Startwertesets) verwendet werden. Besser ist es, wenigstens 500 (bei komplexeren Modellen noch mehr!) Startwertesets für den ersten Schritt der Optimierung zu verwenden. Außerdem ist es wichtig, eine ausreichende Anzahl an Iterationen im ersten Schritt der Optimierung zu wählen (Uebersax, 2000). Der Befehl `starts =` kann dazu benutzt werden, bei der LCA-Modellschätzung in Mplus die Anzahl der zu verwendenden Startwertesets zu erhöhen, um die Gefahr eines lokalen Maximums zu verringern. Außerdem kann man mit dem Befehl `stiterations` die Anzahl der Iterationen im ersten Schritt der Optimierung erhöhen (Mplus-Voreinstellung: 10 Iterationen im ersten Schritt). Die Spezifikation

```
starts = 500 50;  
stiterations = 50;
```

bewirkt, dass Mplus im ersten Schritt der Optimierung der Likelihoodfunktion 500 Startwertesets verwendet und jedes dieser Sets mit bis zu 50 Iterationen (`stiterations = 50 ;`) „bearbeitet“. Die 50 Startwertesets mit dem größten Loglikelihood-Wert werden dann im zweiten Schritt der Optimierung bis zur Erreichung eines Konvergenzkriteriums (in Mplus ist die Voreinstellung 0.000001) weiter verwendet, und die Parameterschätzungen der Lösung mit dem größten Loglikelihood-Wert werden als beste Lösung ausgegeben. Box 6.4 fasst die wichtigsten Maßnahmen zur Vermeidung lokaler Likelihood-Maxima bei einer LCA zusammen.

Box 6.4: Vermeidung von lokalen Likelihood-Maxima in der LCA

Praktische Erfahrungen zeigen, dass lokale Lösungen in der LCA häufig auftreten. Die Beachtung der folgenden Punkte kann helfen, solche „local solutions“ zu vermeiden (siehe auch Uebersax, 2000):

- **Ausreichend viele Startwertesets verwenden.** Bei der Schätzung von LCA-Modellen mit mehr als 2 Klassen sollten ausreichend viele Startwertesets verwendet werden (mindestens 500 im ersten Schritt und davon die 50 besten im zweiten Schritt, bei komplexen Modellen lieber 1000/100 oder noch mehr!) → `starts = 500 50;` bzw. `starts = 1000 100;`
- **Ausreichend viele Iterationen im ersten Optimierungsschritt wählen.** Es ist außerdem empfehlenswert, die Anzahl der Iterationen im ersten Optimierungsschritt hochzusetzen (z.B. `stiterations = 50;`).
- **Replizierbarkeit des besten Loglikelihood-Wertes überprüfen.** Man sollte darauf achten, dass der größte (beste) gefundene Loglikelihood-Wert wenigstens durch zwei unterschiedliche Startwertesets ermittelt werden konnte. Dies kann man im Mplus-Output unter der Rubrik RANDOM STARTS RESULTS RANKED FROM THE BEST TO THE WORST LOGLIKELIHOOD VALUES: Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers leicht überprüfen (s.u.).
- **Modelle mit weniger Klassen bevorzugen.** Wenn möglich, sollte die Extraktion von zu vielen Klassen vermieden werden, da lokale Maxima bei Mehrklassenlösungen häufiger auftreten, als bei Lösungen mit weniger Klassen. Bereits bei 4-5 Klassen treten lokale Maxima wesentlich häufiger auf.
- **Ein kleines Konvergenzkriterium verwenden.** Das Mplus-Kriterium für Konvergenz liegt bei 0,000001. Im Zweifelsfall sollte man ein geringeres Konvergenzkriterium wählen, um ein lokales Maximum zu vermeiden (z.B. `convergence = 0.0000001;`).
- **Skeptisch sein bei Modellen mit vielen Boundary Estimates.** Werden viele klassenbedingte Antwortwahrscheinlichkeiten auf exakt 0 oder 1 geschätzt (sog. *Boundary Estimates*; vgl. Box 6.5), so kann dies ein Zeichen für eine unvalide Lösung und/oder die Extraktion von zu vielen Klassen sein.
- **Modell in einem anderen LCA-Programm replizieren.** Im Zweifelsfall (z.B. bei komplexen Modellen) ist zu empfehlen, das gleiche Klassenmodell zusätzlich in einem oder zwei anderen LCA-Computerprogrammen schätzen zu lassen. Dort sollte sich der in Mplus gefundene Loglikelihood-Wert exakt replizieren lassen.
- **Modell mit anderen Daten replizieren.** Ein Zeichen für die Validität einer Klassenslösung ist auch, wenn sich diese in einem neuen Datensatz replizieren lässt.

Die `plot`-Option ist hilfreich, um sich die Klassenlösung zu veranschaulichen. Dies geschieht in der LCA meist über ein Linendiagramm, in welchem auf der X-Achse die Items und auf der Y-Achse die klassenbedingten Antwortwahrscheinlichkeiten abgetragen werden. So erhält man für jede Klasse ein Profil über die Items, das zur Interpretation der Klassen nützlich ist (s.u.). Ein solches Diagramm kann man in Mplus über die Option

```
plot: type = plot3;
```

erstellen lassen. Mit Hilfe der `series`-Option wird festgelegt, dass die Items auf der X-Achse abgetragen werden:

```
series = c1(1) c2(2) c3(3) c4(4)
        c5(5) c6(6) c7(7) c8(8);
```

Für jede Person kann in einer LCA die Wahrscheinlichkeit bestimmt werden, mit der diese Person jeder der extrahierten latenten Klassen angehört. Auf der Basis der individuellen Zuordnungswahrscheinlichkeiten kann man so für diagnostische Zwecke die Klassenzugehörigkeit einer Person schätzen lassen. Hierbei wird die Person derjenigen Klasse zugeordnet, für welche sie auf Grundlage ihres Antwortmusters die maximale Zuordnungswahrscheinlichkeit hat.

Der `savedata`-Befehl kann verwendet werden, wenn außer den Modellparametern auch die Klassenzuordnungswahrscheinlichkeiten für jedes Individuum sowie die manifeste Klassenzuordnung von Interesse ist. Mplus schätzt bei Verwendung des `savedata`-Befehls einerseits die individuellen Zuordnungswahrscheinlichkeiten für jede Person und jede Klasse und andererseits die individuelle Klassenzugehörigkeit aufgrund der jeweiligen maximalen Zuordnungswahrscheinlichkeit. Diese Werte werden in einer neuen Datendatei abgespeichert, die man für weitere Analysen verwenden kann (s.u.). Zur Speicherung dieser Werte in einer neuen Textdatei mit dem Namen **computerspiele_3Klassen.dat** spezifizieren wir im Input

```
savedata: file = computerspiele_3Klassen.dat;
          save = cprobabilities;
```

Mplus speichert uns aufgrund dieses Befehls eine neue Datei mit den Computerspielitems **c1-c8**, der Personenidentifikationsvariable **id** (da wir diese weiter oben unter `variable` als `auxiliary = defined` haben) sowie den individuellen Klassenzuordnungswahrscheinlichkeiten (Variablen **CPROB1-CPROB3**) und der manifesten Klassenzugehörigkeit (Variable **CLASS**) in dasselbe Verzeichnis, in das wir auch den Input-File abgelegt haben.

Box 6.5: Boundary Estimates

Unter Boundary Estimates versteht man in der LCA Parameter, die auf die Grenzwerte des zulässigen Wertebereichs geschätzt wurden. Bei Wahrscheinlichkeiten liegen die Grenzen bei 0 und 1. Dies entspricht in der Mplus-Logit-Parametrisierung den Werten -15 bzw. +15. Derartige extreme Schätzwerte treten tendenziell umso häufiger auf, je mehr Klassen extrahiert werden. Sie sind häufig ein Zeichen für die Extraktion von zu vielen Klassen. Sie können außerdem auf das Vorliegen eines lokalen Maximums (vgl. Box 6.4) hindeuten. Aus diesem Grunde sollten Klassenlösungen mit Boundary Estimates mit Vorsicht interpretiert werden. In Mplus wird das Auftreten von Boundary Estimates mit folgender Warnung im Output angezeigt (hier für die 5-Klassenlösung des Computerspielbeispiels):

```
IN THE OPTIMIZATION, ONE OR MORE LOGIT THRESHOLDS APPROACHED AND
WERE SET AT THE EXTREME VALUES. EXTREME VALUES ARE -15.000 AND
15.000.
```

THE FOLLOWING THRESHOLDS WERE SET AT THESE VALUES:

- * THRESHOLD 1 OF CLASS INDICATOR C1 FOR CLASS 2 AT ITERATION 35
- * THRESHOLD 1 OF CLASS INDICATOR C4 FOR CLASS 2 AT ITERATION 38
- * THRESHOLD 1 OF CLASS INDICATOR C2 FOR CLASS 2 AT ITERATION 58
- * THRESHOLD 1 OF CLASS INDICATOR C3 FOR CLASS 3 AT ITERATION 66
- * THRESHOLD 1 OF CLASS INDICATOR C8 FOR CLASS 3 AT ITERATION 69

Per Voreinstellung erhält man für eine LCA in Mplus ML-Parameterschätzwerte mit robusten Standardfehlern (Default: `estimator = mlr;`). Konventionelle ML-Schätzung kann über den Befehl `analysis: estimator = ml;` angefordert werden.

Nachfolgend ist der wichtigste Teil des Mplus-Outputs zur 3-Klassenlösung für die Computerspielitems dargestellt. Wir erhalten zunächst die unbedingten Wahrscheinlichkeiten (proportions) für jedes Item und jede Kategorie in der Gesamtstichprobe. Es zeigt sich, dass über alle $N = 861$ Kinder betrachtet bei allen Items die erste Kategorie „nie/selten“ überwiegt. Interessant ist für uns, wie sich die bedingten Antwortwahrscheinlichkeiten in den drei extrahierten Subgruppen (den latenten Klassen) darstellen. Dies erfahren wir weiter unten bzw. über das angeforderte Klassenprofildiagramm.

SUMMARY OF CATEGORICAL DATA PROPORTIONS

C1	
Category 1	0.822
Category 2	0.178
C2	
Category 1	0.727
Category 2	0.273
C3	
Category 1	0.746
Category 2	0.254

```

C4
  Category 1      0.816
  Category 2      0.184
C5
  Category 1      0.739
  Category 2      0.261
C6
  Category 1      0.746
  Category 2      0.254
C7
  Category 1      0.649
  Category 2      0.351
C8
  Category 1      0.674
  Category 2      0.326

```

Als nächstes erscheint die Auflistung der Loglikelihood-Werte für die 50 Startwertesets mit dem größten Loglikelihood-Wert aus dem ersten Schritt der Optimierung (geordnet vom besten zum schlechtesten Loglikelihood-Wert) mit den zugehörigen Seeds (Seed = Ausgangspunkt für die Generierung zufälliger Startwerte) sowie den Nummern des jeweiligen Startwertesets. Es zeigt sich, dass in diesem Modell für alle 50 Startwertesets derselbe Loglikelihood-Wert (-3463.814) gefunden wurde. Dies ist ein gutes Zeichen, da es darauf hindeutet, dass die Lösung eindeutig ist und lokale Maxima offensichtlich bei der Schätzung unseres Modells nicht auftraten. Ein solches Ergebnis ist typisch für Modelle mit wenigen latenten Klassen.

```
RANDOM STARTS RESULTS RANKED FROM THE BEST TO THE WORST LOGLIKELIHOOD VALUES
```

```
Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers:
```

-3463.814	605358	321
-3463.814	57226	208
-3463.814	714455	476
-3463.814	965639	463
-3463.814	798821	423
-3463.814	85734	411
-3463.814	926797	406
-3463.814	973369	202
-3463.814	7959	256
-3463.814	987090	70
-3463.814	137305	379
-3463.814	732596	320
-3463.814	900268	327
-3463.814	319575	499
-3463.814	576596	99
-3463.814	772131	407

-3463.814	100874	108
-3463.814	377504	294
-3463.814	694303	282
-3463.814	741888	138
-3463.814	582296	452
-3463.814	344422	296
-3463.814	340112	126
-3463.814	937225	394
-3463.814	848890	95
-3463.814	695155	150
-3463.814	455617	242
-3463.814	168762	200
-3463.814	846194	93
-3463.814	347515	24
-3463.814	970689	266
-3463.814	21345	199
-3463.814	652266	490
-3463.814	971853	402
-3463.814	81117	305
-3463.814	417035	149
-3463.814	570908	98
-3463.814	655497	376
-3463.814	637345	19
-3463.814	736574	414
-3463.814	923437	398
-3463.814	782821	272
-3463.814	315029	471
-3463.814	850840	232
-3463.814	972873	157
-3463.814	860102	495
-3463.814	793035	187
-3463.814	414828	322
-3463.814	483369	270
-3463.814	529496	343

Bei Modellen mit mehr latenten Klassen findet man in der Mplus-Auflistung der Loglikelihood-Werte häufig auch unter den besten (größten) Werten zumindest leichte Unterschiede. In diesem Fall ist es gut, wenn der größte Loglikelihood-Wert zumindest für zwei Startwertesets gefunden wurde (d.h. der beste Wert in der Optimierung wenigstens einmal repliziert werden konnte). Ist dies nicht der Fall, so sollte man die Anzahl der Startwertesets und/oder die Anzahl der Iterationen im ersten Schritt der Optimierung (über STARTS bzw. STITERATIONS, s.o.) erhöhen. Außerdem ist in diesem Fall dringend zu empfehlen, die Lösung noch in einem weiteren LCA-Programm zu replizieren (vgl. Box 6.4).

Mplus weist unter der Ausgabe der Loglikelihood-Werte für jedes der 50 Startwertesets auf die Gefahr von lokalen Likelihood-Maxima mit folgender Meldung hin:

WARNING: WHEN ESTIMATING A MODEL WITH MORE THAN TWO CLASSES, IT MAY BE NECESSARY TO INCREASE THE NUMBER OF RANDOM STARTS USING THE STARTS OPTION TO AVOID LOCAL MAXIMA.

Im vorliegenden Beispiel kann diese Meldung ignoriert werden, da wir eine ausreichende Anzahl von Startwertesets verwendet haben.

Box 6.6: Exkurs optseed-Option

Die Auflistung der Loglikelihood-Werte in Mplus ist u.a. auch deshalb praktisch, weil man sie nutzen kann, um in nachfolgenden Analyseschritten die Lösung mit dem besten Loglikelihood-Wert zu replizieren, ohne erneut Startwertesets generieren lassen zu müssen (was z.T. sehr viel Rechenzeit in Anspruch nehmen kann). Dies ist z.B. dann nützlich, wenn man nach Auswahl einer Klassenlösung für diese Lösung zusätzlichen Output anfordern möchte, den man zunächst nicht spezifiziert hat. Mit Hilfe der sogenannten optseed-Option kann in einem solchen Fall eine bestimmte Lösung auf Basis des entsprechenden Seeds repliziert werden. Zu diesem Zweck wählt man im Input zum Beispiel:

```
analysis: optseed = 605358;
```

Hierbei entspricht die Zahl 605358 dem aus dem oben gezeigten Output entnommenen Seed für das Startwerteset Nr. 321 (in der Liste der Loglikelihood-Werte ganz oben zu finden). Dieses Startwerteset ist eines derjenigen, das den besten Loglikelihood-Wert von -3463.814 hervorgebracht hat (in diesem Beispiel hätte man jedes der 50 seeds auswählen können, da alle Startwertesets denselben Loglikelihood-Wert produziert haben; dies ist jedoch nicht immer der Fall). Mplus repliziert mit Hilfe des optseed-Befehls exakt die Lösung, die auf Basis des Starwertesets mit dem Seed 605358 ermittelt wurde.

Anschließend sollte im Output ausschließlich die Meldung

THE MODEL ESTIMATION TERMINATED NORMALLY

angezeigt werden. Bei Konvergenz des Modells erhält man dann nochmals den Loglikelihood-Wert für die endgültige Lösung (hier -3463.814). Dieser Wert ist identisch mit dem Wert, der in der obigen Auflistung ganz oben erscheint. Außerdem werden verschiedene Modellgütestatistiken ausgegeben, welche wir in Abschnitt 6.3 im Detail besprechen.

TESTS OF MODEL FIT

Loglikelihood

H0 Value	-3463.814
H0 Scaling Correction Factor for MLR	1.125

Information Criteria

Number of Free Parameters	26
Akaike (AIC)	6979.628
Bayesian (BIC)	7103.338
Sample-Size Adjusted BIC	7020.769
(n* = (n + 2) / 24)	

Chi-Square Test of Model Fit for the Binary and Ordered Categorical (Ordinal) Outcomes

Pearson Chi-Square

Value	386.883
Degrees of Freedom	229
P-Value	0.0000

Likelihood Ratio Chi-Square

Value	340.605
Degrees of Freedom	229
P-Value	0.0000

Unter diesen Statistiken erhalten wir Informationen über die Klassengrößen. Zunächst werden die aufgrund des Modells geschätzten Klassengrößenparameter $\hat{\pi}_g$ für jede der drei Klassen ausgegeben (hier $\hat{\pi}_1 = 0,17362$, $\hat{\pi}_2 = 0,38063$ und $\hat{\pi}_3 = 0,44575$). Somit gehören der größten Klasse 3 ca. 44,6% (entspricht etwa $n = 384$), Klasse 2 ca. 38,1% (etwa $n = 328$) und der kleinsten Klasse 1 ca. 17,4% der Kinder (etwa $n = 150$) an.

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODELLatent
Classes

1	149.48776	0.17362
2	327.71918	0.38063
3	383.79305	0.44575

Darunter sehen wir die Klassengrößen basierend auf den geschätzten *posterior probabilities*. Diese sind hier nahezu identisch mit den aufgrund des Modells geschätzten Klassengrößen:

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASS PATTERNS
BASED ON ESTIMATED POSTERIOR PROBABILITIES

Latent
Classes

1	149.48771	0.17362
2	327.71923	0.38063
3	383.79307	0.44575

Das Entropy-Maß sagt etwas über die Güte der Klassifikation aus. Werte nahe 1 zeigen eine große Sicherheit der Klassifikation an, während Werte nahe Null für eine geringe Sicherheit sprechen. Das Entropy-Maß wird in der Praxis jedoch eher selten als Indikator für den Modellfit verwendet. Man betrachtet als Maß für die Zuverlässigkeit der Klassifikation meist eher die mittleren Klassenzuordnungswahrscheinlichkeiten von Individuen, die einer bestimmten Klasse zugeordnet wurden (s.u.).

CLASSIFICATION QUALITY

Entropy	0.731
---------	-------

Für jede Person kann in einer LCA die Wahrscheinlichkeit berechnet werden, mit der diese Person jeder der extrahierten Klassen angehört. Für diagnostische (oder andere) Zwecke kann eine Person derjenigen Klasse zugeordnet werden, für die sie aufgrund ihres Antwortmusters die höchste Zuordnungswahrscheinlichkeit hat.

Mplus gibt zunächst die auf Grundlage dieser manifesten Klassifikation berechneten Klassengrößen aus (unter CLASSIFICATION OF INDIVIDUALS BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP). Die Klassengrößen auf Grundlage der manifesten Zuordnung weichen in unserem Beispiel (geringfügig) von den geschätzten Klassengrößen $\hat{\pi}_g$ ab. Eine gewisse Abweichung zwischen diesen Werten tritt in den meisten Fällen auf, da bei der manifesten Zuordnung von Personen anhand der maximalen Zuordnungswahrscheinlichkeiten wiederum Schätzfehler auftreten. Aus diesem Grund sollten die weiter oben ausgegebenen $\hat{\pi}_g$ -Parameter als Schätzungen für die latenten Klassengrößen bevorzugt werden.

CLASSIFICATION OF INDIVIDUALS BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

Class Counts and Proportions

Latent
Classes

1	152	0.17654
2	317	0.36818
3	392	0.45528

Als weiteres Maß für die Zuverlässigkeit der Klassifikation der Kinder mittels der 3-Klassenlösung erhalten wir von Mplus die geschätzten mittleren Klassenzuordnungswahrscheinlichkeiten. Diese sind ein wichtiger Indikator für die Güte eines LCA-Modells. Werte nahe 1 auf der Hauptdiagonalen der Klassifikationsmatrix sprechen für eine hohe „Treffsicherheit“ (oder Reliabilität) der Klassifikation. Bei hohen Werten können Individuen anhand ihres Antwortmusters im Mittel sehr sicher einer bestimmten latenten Klasse zugeordnet werden. Bei geringen Werten besteht im Mittel eine hohe Unsicherheit über die Klassenzuordnung.

Die mittleren Klassenzuordnungswahrscheinlichkeiten für dieselbe Klasse (Werte auf der Hauptdiagonale) sollten für jede Klasse möglichst > 0,80 ausfallen (Rost, 2006). In unserem Beispiel erhalten wir für Klasse 2 und Klasse 3 sehr gute mittlere Klassenzuordnungswahrscheinlichkeiten von 0,913 und 0,922, während die Wahrscheinlichkeit für Klasse 1 mit 0,76 nur befriedigend ausfällt. Anscheinend ist die Zuordnung von Kindern zu Klasse 1 etwas weniger eindeutig als die Zuordnung zu Klasse 2 und 3. Kinder in Klasse 1 haben im Schnitt auch eine relativ hohe Wahrscheinlichkeit, Klasse 2 anzugehören (0,146). Somit besteht eine gewisse „Unschärfe“ zwischen der Zuordnung zu Klasse 1 und Klasse 2, wobei festzustellen ist, dass die Klassifikationsgüte auch für Klasse 1 noch zufriedenstellend ist.

Je mehr latente Klassen extrahiert werden, desto geringer fallen in der Regel die mittleren Zuordnungswahrscheinlichkeiten aus, da bei einer Vielzahl extrahierter Klassen mehr Möglichkeiten für eine Fehlklassifikation bestehen und die Klassen häufig schlechter von einander abgrenzbar sind.

Average Latent Class Probabilities for Most Likely Latent Class Membership (Row)
by Latent Class (Column)

	1	2	3
1	0.760	0.146	0.094
2	0.061	0.913	0.025
3	0.037	0.041	0.922

Unter MODEL RESULTS erhalten wir zunächst die Parameterschätzungen (Estimate) und Standardfehler (S.E.) in der Logit-Parametrisierung der LCA (einfacher zu interpretieren sind die Ergebnisse in der Wahrscheinlichkeitsmetrik, die weiter unten folgen). Es erscheinen zunächst unter Thresholds die geschätzten Schwellenparameter $\hat{\tau}_{ig}$ für jedes Item in jeder Klasse. Darunter werden unter Categorical Latent Variables - Means die beiden geschätzten Schwellenparameter $\hat{\tau}_g$ für die Klassengrößen ausgegeben (unter CLASS#1 und CLASS#2). Zur Umrechnung dieser Parameter in Wahrscheinlichkeiten siehe Muthén und Muthén (1998-2007, Kapitel 13).

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Latent Class 1				
Thresholds				
C1\$1	1.398	0.506	2.761	0.006
C2\$1	1.920	0.706	2.717	0.007
C3\$1	1.675	0.423	3.961	0.000
C4\$1	1.489	0.523	2.850	0.004
C5\$1	-0.981	0.406	-2.413	0.016
C6\$1	-1.480	0.654	-2.263	0.024
C7\$1	0.187	0.357	0.524	0.600
C8\$1	0.685	0.334	2.052	0.040
Latent Class 2				
Thresholds				
C1\$1	0.555	0.144	3.852	0.000
C2\$1	-0.591	0.191	-3.101	0.002
C3\$1	-0.097	0.163	-0.593	0.553
C4\$1	0.463	0.143	3.244	0.001
C5\$1	1.154	0.247	4.681	0.000
C6\$1	1.542	0.291	5.290	0.000
C7\$1	-0.628	0.152	-4.134	0.000
C8\$1	-0.576	0.159	-3.633	0.000
Latent Class 3				
Thresholds				
C1\$1	4.582	0.780	5.872	0.000
C2\$1	4.334	0.804	5.388	0.000
C3\$1	2.723	0.322	8.455	0.000
C4\$1	4.583	0.927	4.947	0.000
C5\$1	2.216	0.300	7.391	0.000
C6\$1	2.166	0.366	5.924	0.000
C7\$1	2.873	0.326	8.817	0.000
C8\$1	2.845	0.410	6.937	0.000

Categorical Latent Variables

Means

CLASS#1	- 0.943	0.247	-3.817	0.000
CLASS#2	- 0.158	0.151	-1.048	0.294

Anschaulicher und leichter interpretierbar ist die Ausgabe der LCA-Ergebnisse in Form der geschätzten Wahrscheinlichkeitsparameter $\hat{\pi}_g$ und $\hat{\pi}_{ig}$. Die Schätzungen für die Klassengrößenparameter ($\hat{\pi}_g$) haben wir bereits weiter oben unter FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES BASED ON THE ESTIMATED MODEL erhalten. Die Schätzungen für die bedingten Antwortwahrscheinlichkeiten $\hat{\pi}_{ig}$ (sowie die Gegenwahrscheinlichkeiten $1 - \hat{\pi}_{ig}$) erhalten wir unter der Rubrik RESULTS IN PROBABILITY SCALE. Unter Category 1 finden wir jeweils die Wahrscheinlichkeit für die Kategorie 0 „nie/selten“, d.h. $1 - \hat{\pi}_{ig}$. Unter Category 2 wird die Wahrscheinlichkeit für die größere Kategorie (hier Kategorie 1) „oft/sehr oft“ ($\hat{\pi}_{ig}$) ausgegeben.

RESULTS IN PROBABILITY SCALE

Latent Class 1

C1

Category 1	0.802	0.080	9.968	0.000
Category 2	0.198	0.080	2.463	0.014

C2

Category 1	0.872	0.079	11.068	0.000
Category 2	0.128	0.079	1.623	0.105

C3

Category 1	0.842	0.056	14.991	0.000
Category 2	0.158	0.056	2.808	0.005

C4

Category 1	0.816	0.078	10.399	0.000
Category 2	0.184	0.078	2.345	0.019

C5

Category 1	0.273	0.081	3.383	0.001
Category 2	0.727	0.081	9.020	0.000

C6

Category 1	0.185	0.099	1.878	0.060
Category 2	0.815	0.099	8.247	0.000

C7

Category 1	0.547	0.088	6.179	0.000
Category 2	0.453	0.088	5.125	0.000

C8

Category 1	0.665	0.074	8.941	0.000
Category 2	0.335	0.074	4.508	0.000

Latent Class 2				
C1	Category 1	0.635	0.033	19.021
	Category 2	0.365	0.033	10.915
C2	Category 1	0.356	0.044	8.152
	Category 2	0.644	0.044	14.721
C3	Category 1	0.476	0.041	11.727
	Category 2	0.524	0.041	12.916
C4	Category 1	0.614	0.034	18.144
	Category 2	0.386	0.034	11.423
C5	Category 1	0.760	0.045	16.916
	Category 2	0.240	0.045	5.332
C6	Category 1	0.824	0.042	19.466
	Category 2	0.176	0.042	4.165
C7	Category 1	0.348	0.034	10.094
	Category 2	0.652	0.034	18.917
C8	Category 1	0.360	0.037	9.847
	Category 2	0.640	0.037	17.521
Latent Class 3				
C1	Category 1	0.990	0.008	126.531
	Category 2	0.010	0.008	1.295
C2	Category 1	0.987	0.010	96.062
	Category 2	0.013	0.010	1.259
C3	Category 1	0.938	0.019	50.383
	Category 2	0.062	0.019	3.309
C4	Category 1	0.990	0.009	106.680
	Category 2	0.010	0.009	1.090
C5	Category 1	0.902	0.027	33.919
	Category 2	0.098	0.027	3.699
C6	Category 1	0.897	0.034	26.596
	Category 2	0.103	0.034	3.049
C7	Category 1	0.947	0.017	57.362
	Category 2	0.053	0.017	3.242
C8	Category 1	0.945	0.021	44.386
	Category 2	0.055	0.021	2.580

Anhand der klassenbedingten Antwortwahrscheinlichkeiten können wir erkennen, dass Kinder, welche Klasse 1 zugeordnet wurden, für alle Items geringe Wahrscheinlichkeiten für die Kategorie 2 (oft/sehr oft) haben, außer bei den Items C5 (Denk- und Logikspiele) und C6 (Geschicklichkeitsspiele). Es handelt sich also offenbar um eine Gruppe von Kindern, die besonders häufig Geschicklichkeitss- und Denk- und Logikspiele spielen, während sie andere Arten von Computerspielen eher selten bzw. nie spielen. Wie wir weiter oben anhand der geschätzten Klassengrößenparameter gesehen haben, gehören laut diesem Modell ungefähr 17,4% der Kinder zu dieser latenten Klasse.

Die Kinder in Klasse 2 zeigen ein anderes Muster bedingter Antwortwahrscheinlichkeiten. Generell haben sie bei allen Items eine höhere Wahrscheinlichkeit für die Kategorie 2 (oft/sehr oft). In Klasse 2 zeigen sich darüber hinaus vor allem höhere Antwortwahrscheinlichkeiten für Kategorie 2 bei den Items C2 (Actionspiele), C3 (Sportspiele), C7 (Simulationsspiele) und C8 (Fahrzeugsimulatoren). Es handelt sich somit um eine Gruppe von Kindern, die generell häufiger spielen und dabei insbesondere Action- und Simulationsspiele bevorzugen. Wie wir anhand der Klassengrößenparameter gesehen haben, gehören etwa 38,1% dieser Gruppe an.

Die größte latente Klasse (Klasse 3, 44,6%) zeichnet sich dadurch aus, dass die Antwortwahrscheinlichkeiten für die Kategorie 2 (oft/sehr oft) bei allen Items sehr gering ausfallen. Somit können wir Klasse 3 als eine Klasse von „Nichtspieler“ interpretieren bzw. als eine Gruppe von Kindern, die alle abgefragten Computerspielarten selten oder nie spielen.

Übersichtlich und hilfreich für die Interpretation einer Klassenlösung ist es, sich die geschätzten klassenbedingten Antwortwahrscheinlichkeiten $\hat{\pi}_{ig}$ der einzelnen Klassen grafisch mit Hilfe eines Klassenprofildiagramms zu veranschaulichen. Auf diese Weise kann man die Güte der Klassenlösung sowie die Bedeutung der Klassen rascher erfassen und beurteilen. Wir können das über die `plot`-Option angeforderte Liniendiagramm mit den bedingten Antwortwahrscheinlichkeiten über **Graph → View graphs** im Mplus-Output aufrufen. Unter **Select a plot to view** wählt man dann **Estimated probabilities** aus (siehe Abb. 6.2).

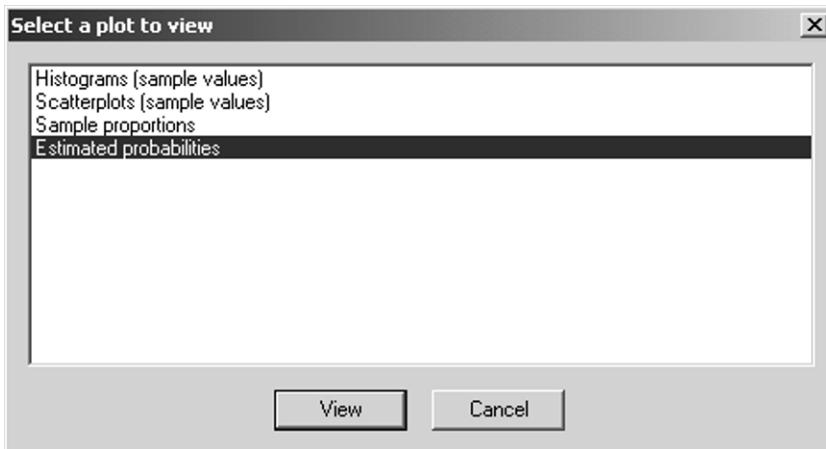


Abbildung 6.2: Auswahl des Klassenprofildiagramms mit den klassenbedingten Antwortwahrscheinlichkeiten über das Mplus-Menü Graph → View graphs.

Anschließend muss man spezifizieren, ob die Wahrscheinlichkeiten für die erste oder die zweite Kategorie angezeigt werden sollen (siehe Abb. 6.3). Wir wählen aufgrund der leichteren Interpretierbarkeit die zweite Kategorie (oft/sehr oft), d.h. die Darstellung der geschätzten Parameter $\hat{\pi}_{ig}$ (und nicht $1 - \hat{\pi}_{ig}$).

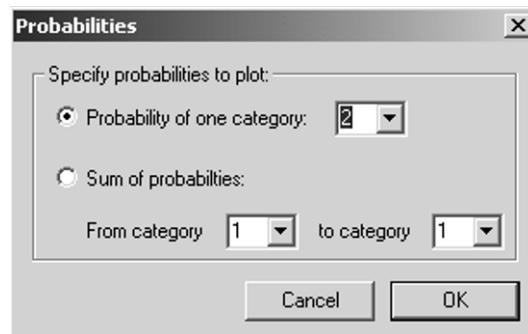


Abbildung 6.3: Spezifikation der im Klassenprofildiagramm anzuzeigenden Wahrscheinlichkeiten. Hier: Wahrscheinlichkeiten $\hat{\pi}_{ig}$ für die zweite Kategorie (oft/sehr oft).

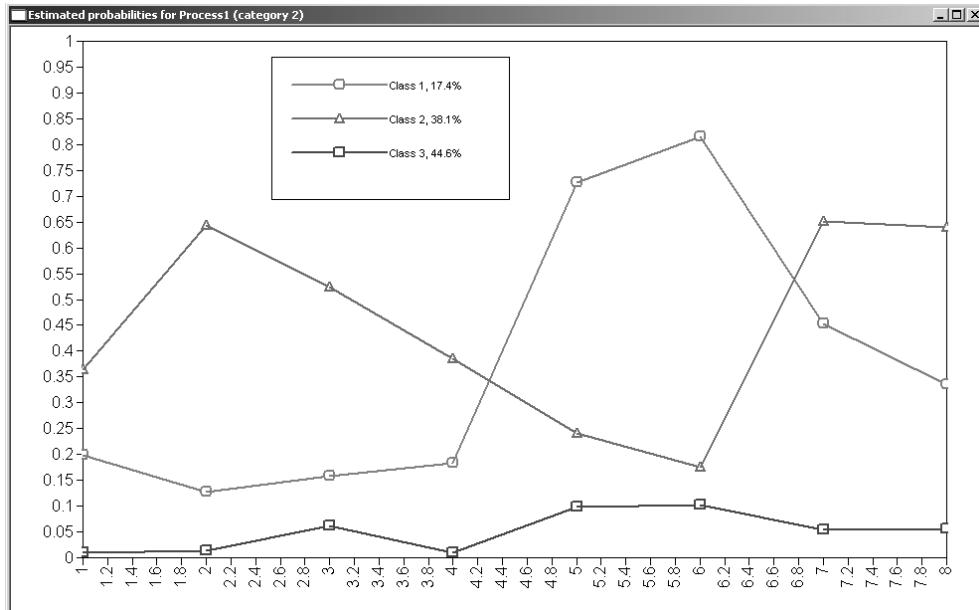


Abbildung 6.4: Mplus-Liniendiagramm zur Darstellung der geschätzten klassenbedingten Antwortwahrscheinlichkeiten $\hat{\pi}_{ig}$ in der 3-Klassenlösung für die Computerspieldaten (Klassenprofildiagramm). Auf der X-Achse sind die Items (c1-c8) dargestellt, auf der Y-Achse die geschätzte bedingte Wahrscheinlichkeit, die Kategorie „oft/sehr oft“ zu wählen.

Man erhält das in Abbildung 6.4 dargestellte Liniendiagramm, in welchem auch die geschätzten Klassengrößen nochmals angegeben werden.

Da das Mplus-Liniendiagramm beispielsweise keine Achsenbeschriftungen aufweist, ist es vor allem zur ersten Veranschaulichung der Klassenlösung nützlich, nicht jedoch zur weiteren Verwendung (z.B. in Publikationen) brauchbar. Es bietet sich an, die klassenbedingten Antwortwahrscheinlichkeiten für das letztendlich angenommene LCA-Modell unter Verwendung eines anderen Programms nochmals „schöner“ darzustellen (z.B. in Microsoft Excel etc.). Abbildung 6.5 zeigt die Lösung in einem Excel-Liniendiagramm. Hierbei wurde die Reihenfolge der Klassen gemäß Klassengröße sortiert (so dass die kleinste Klasse der Mplus-Lösung nun als Klasse 3 dargestellt wird und die größte Klasse nun als Klasse 1). Außerdem wurde die X-Achse gemäß des Iteminhalts beschriftet, so dass man die Lösung nun noch einfacher interpretieren kann.

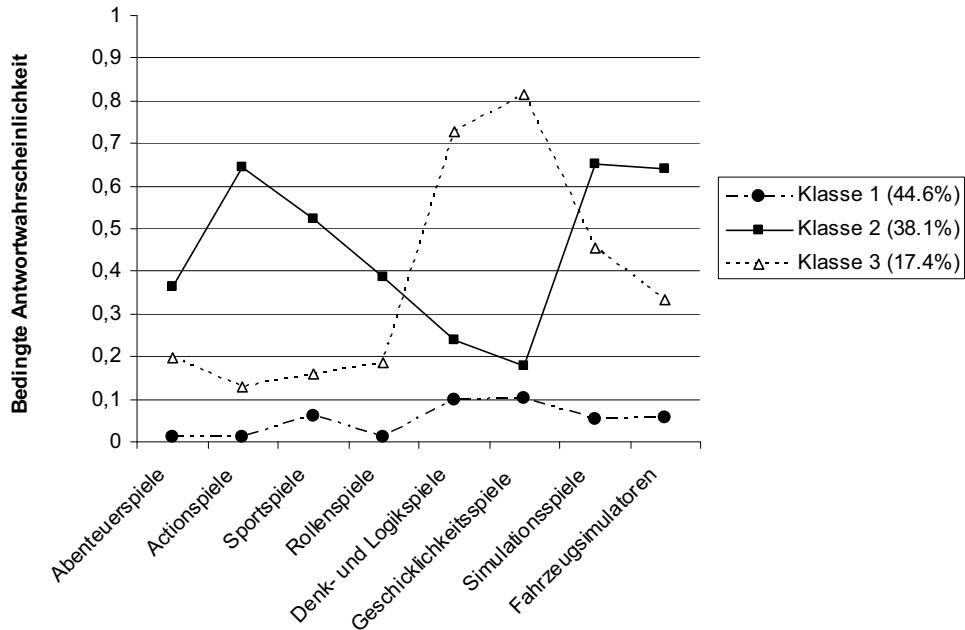


Abbildung 6.5: Excel-Liniendiagramm zur Darstellung der geschätzten klassenbedingten Antwortwahrscheinlichkeiten $\hat{\pi}_{ig}$ in der 3-Klassenlösung für die Computerspieldaten (Klassenprofildiagramm). Auf der X-Achse sind die Items (c1-c8) dargestellt, auf der Y-Achse die bedingte Wahrscheinlichkeit, die Kategorie „oft/sehr oft“ zu wählen.

Mit Hilfe der Liniendiagramme kann man die bereits oben anhand der Mplus-Parameterschätzungen vorgenommene Interpretation der latenten Klassen nochmals nachvollziehen. Die größte Klasse (hier nun Klasse 1) ist die Klasse von „Nichtspielern“, in der die Mitglieder sehr geringe Antwortwahrscheinlichkeiten für die Kategorie „oft/sehr oft“ bei allen Items haben. Die zweitgrößte Klasse (Klasse 2) repräsentiert die Gruppe von Kindern, die insgesamt häufiger spielen und eine besondere Präferenz für Action-, Sport- und Simulations-/Fahrzeugsimulatorspiele hat. In der kleinsten Klasse (nun Klasse 3) befinden sich die „Geschicklichkeitsspieler“ und „Denk- und Logikspieler“.

Außer dem Klassenprofildiagramm hatten wir im Mplus-Input mit Hilfe der `savedata`-Option auch noch die Ausgabe der individuellen Klassenzuordnungswahrscheinlichkeiten und der manifesten Klassenzuordnung aufgrund der maximalen Zuordnungswahrscheinlichkeit in einem neuen Datensatz angefordert. Am Ende des Mplus-Outputs erhalten wir eine Auflistung der Variablen, die in den mit Hilfe des `savedata`-Befehls generierten neuen Datensatz **computerspiele_3klassen.dat** abgespeichert wurden sowie Informationen zum Format dieses Datensatzes. Diese Auflistung ist u.a. wichtig, um die Reihenfolge der Variablen im neu erstellten Datensatz festzustellen, so dass die Daten in weiteren Analyse schritten (z.B. in einem anderen Statistikprogramm) korrekt verarbeitet werden können.

Beispielsweise ist die Variable **id** im vorliegenden Beispiel nun nicht mehr die erste, sondern die neunte Variable im Datensatz:

```
SAVEDATA INFORMATION

Order and format of variables

C1          F10.3
C2          F10.3
C3          F10.3
C4          F10.3
C5          F10.3
C6          F10.3
C7          F10.3
C8          F10.3
ID          F10.3
CPROB1      F10.3
CPROB2      F10.3
CPROB3      F10.3
CLASS       F10.3

Save file
computerspiele_3Klassen.dat

Save file format
13F10.3

Save file record length      5000
```

Den von Mplus neu erzeugten ASCII-Datensatz mit dem Namen **computerspiele_3Klassen.dat** können wir zum Beispiel in SPSS einlesen, um weitere Analysen durchzuführen. Beispielsweise kann man den Datensatz für diagnostische Zwecke nutzen, um für ein bestimmtes Kind herauszufinden, welchem Computerspieltyp dieses Kind am wahrscheinlichsten zuzuordnen ist. Die Begleit-CD enthält eine SPSS-Syntaxdatei zum Einlesen des Datensatzes **computerspiele_3Klassen.dat** in SPSS und zum Abspeichern als .sav-Datei. Abbildung 6.6 zeigt einen Ausschnitt aus dem hinteren Teil dieses Datensatzes, in welchem sich die vier von Mplus neu hinzugefügten Variablen **CPROB1**, **CPROB2**, **CPROB3** und **CLASS** befinden.

C1	ID	CPR0B1	CPR0B2	CPR0B3	CLASS	var
1	1	0.028	0.754	0.218	2	
2	2	0.005	0.007	0.988	3	
3	3	0.136	0.863	0.000	2	
4	4	0.826	0.006	0.168	1	
5	5	0.165	0.010	0.824	3	
6	6	0.005	0.007	0.988	3	
7	7	0.005	0.007	0.988	3	
8	8	0.112	0.017	0.871	3	
9	9	0.036	0.165	0.799	3	
10	10	0.005	0.007	0.988	3	
11	11	0.216	0.193	0.591	3	
12	12	0.970	0.016	0.013	1	
13	13	0.080	0.847	0.073	2	
14	14	0.059	0.172	0.769	3	
15	15	0.005	0.007	0.988	3	
16	16	0.005	0.007	0.988	3	

Abbildung 6.6: Ausschnitt aus dem Datensatz **computerspiele_3Klassen.sav**, welcher neben den acht Computerspielitems (weiter links im Datenblatt und hier nicht zu sehen) und der Personidentifikationsvariable **ID** auch die von Mplus geschätzten Klassenzuordnungswahrscheinlichkeiten für jedes Kind und jede Klasse (Variablen **CPROB1-CPROB3**) sowie die Klassenzuordnung anhand der maximalen Klassenzuordnungswahrscheinlichkeit (Variable **CLASS**) enthält.

Die Variablen **CPROB1**, **CPROB2**, und **CPROB3** geben für jedes Kind die geschätzte Wahrscheinlichkeit an, mit der das Kind jeder der drei Klassen angehört. Beispielsweise hat das erste Kind im Datensatz (erste Zeile) aufgrund seines Antwortmusters eine geschätzte Wahrscheinlichkeit von 0,028, der ersten Klasse anzugehören, von 0,754, der zweiten Klasse anzugehören und von 0,218, der dritten Klasse anzugehören. Da die Wahrscheinlichkeit des Kindes für Klasse 2 am höchsten ist, erhält das Kind den Wert 2 auf der Klassenvariablen **CLASS**.

Beim zweiten Kind ist die geschätzte Zuordnungswahrscheinlichkeit am größten für Klasse 3 (0,988), so dass das Kind den Wert 3 auf der Variablen **CLASS** erhält und so weiter. Auf diese Weise kann man einerseits für jedes Kind anhand der geschätzten Klassenzuordnungswahrscheinlichkeiten analysieren, wie sicher das Kind klassifiziert werden kann und andererseits erhält man eine Schätzung der Klassenzugehörigkeit aufgrund der maximalen Zuordnungswahrscheinlichkeit. Die von Mplus im Output ausgegebenen mittleren Klassenzuordnungswahrscheinlichkeiten (Average Latent Class Probabilities, s.o.) basieren auf diesen geschätzten individuellen Klassenzuordnungswahrscheinlichkeiten.

6.3 Modellgütebeurteilung und Modellvergleiche

Häufig hat man keine Theorie, die es erlaubt, a priori eine feste Annahme darüber zu treffen, wie viele Klassen extrahiert werden sollen bzw. wie viele Klassen nötig sind, um die Daten adäquat abzubilden. Da die Anzahl der benötigten Klassen kein Modellparameter des Latent-Class-Modells ist, muss indirekt über Modellttests bzw. Modellvergleiche auf die adäquate Anzahl von Klassen geschlossen werden. In der Praxis heißt dies, dass man meist eine Reihe von LCA-Modellen mit einer unterschiedlichen Anzahl von Klassen schätzen lässt (z.B. 2 bis 5 Klassen) und die Datenanpassung dieser Modelle vergleicht. Man wählt anschließend das Modell mit dem günstigsten Datenfit aus.

Zur Auswahl des am besten auf die Daten passenden Modells gibt es verschiedene Kriterien, von denen wir im Text nur diejenigen ausführlicher besprechen werden, die in Mplus verfügbar sind. Es kann hierbei zwischen Indikatoren für den *absoluten* Modellfit und Indikatoren für den *relativen* Modellfit unterschieden werden. Weitere Informationen zur Beurteilung der Modellgüte in der LCA findet man z.B. bei Collins, Fidler, Wugalter und Long (1993), Formann (2003) sowie Rost (2004).

6.3.1 Absoluter Modellfit

6.3.1.1 Likelihood-Ratio- und Pearson-X²-Statistik

Der absolute Modellfit bezieht sich darauf, inwiefern eine Klassenlösung die beobachteten Daten (Antwortmuster) adäquat reproduzieren kann. Dies kann in Mplus über den Likelihood-Ratio-Test sowie über den Pearson-X²-Test überprüft werden, welche bei einer LCA in Mplus standardmäßig ausgegeben werden. Beide Statistiken basieren auf dem Vergleich der beobachteten Häufigkeiten der Antwortmuster (Pattern) der Personen mit denjenigen Pattern, die vom Modell vorhergesagt werden. Unter günstigen Bedingungen (s.u.) sind beide Statistiken χ^2 -verteilt. Ein signifikanter Wert des Likelihood-Ratio- oder des Pearson-X²-Tests weist darauf hin, dass es eine statistisch bedeutsame Abweichung zwischen beobachteten und modellimplizierten Patternhäufigkeiten gibt. Das würde bedeuten, dass das Modell die Daten nicht perfekt abbildet (z.B. weil zu wenige Klassen extrahiert wurden).

Ungünstigerweise folgen die Likelihood-Ratio- und die Pearson-X²-Statistik nur bei großen Stichproben und relativ geringen Itemanzahlen tatsächlich der theoretischen χ^2 -Verteilung (Collins et al., 1993; Langeheine, van de Pol & Pannekoek, 1996; von Davier, 1997). Meist hat man jedoch in der Praxis relativ viele Items und eine (gemessen an der Anzahl der Items) unzureichend große Stichprobe, so dass die asymptotischen Bedingungen für die Likelihood-Ratio- und die Pearson-X²-Statistik nicht erfüllt sind. Man spricht in einem solchen Fall von „sparse data“ bzw. „sparse tables“. Die Folge ist, dass die auf der theoretischen χ^2 -Verteilung basierenden *p*-Werte für den Signifikanztest auf Abweichung des Modells von den Daten nicht korrekt sind (Rost, 2004).

Als Faustregel gilt, dass man den *p*-Werten dieser Statistiken nicht trauen sollte, wenn die Werte der Likelihood-Ratio- und der Pearson-X²-Statistik für ein und dasselbe Modell sehr unterschiedlich ausfallen. Weichen die Likelihood-Ratio- und die Pearson-X²-Statistik stark voneinander ab, so ist dies ein Hinweis darauf, dass sie in der betreffenden Anwen-

dung nicht χ^2 -verteilt sind. Collins et al. (1993) sowie Langeheine et al. (1996) schlagen für diesen Fall die Anwendung einer parametrischen Bootstrap-Methode vor, mit deren Hilfe man sich eine künstliche Prüfverteilung erzeugen kann.

Diese Methode ist allerdings in Mplus (bis Programmversion 5) nur zur Beurteilung des relativen Modelfits (Bootstrap-Likelihood-Ratio-Differenzentest, siehe Abschnitt 6.3.2.1), nicht jedoch für die Beurteilung des absoluten Modelfits direkt implementiert. Ein parametrischer Bootstrap für den absoluten Modellfit kann in Mplus derzeit nur indirekt über die Durchführung einer Monte-Carlo-Simulationsstudie (siehe dazu Muthén & Muthén, 2002 sowie den Fortsetzungsband von Geiser et al., in Vorb.) durchgeführt werden. In unserem 3-Klassenmodell fallen die Pearson- χ^2 - und die Likelihood-Ratio-Statistik folgendermaßen aus:

Chi-Square Test of Model Fit for the Binary and Ordered Categorical (Ordinal) Outcomes

Pearson Chi-Square

Value	386.883
Degrees of Freedom	229
P-Value	0.0000

Likelihood Ratio Chi-Square

Value	340.605
Degrees of Freedom	229
P-Value	0.0000

Die Werte der Statistiken sind relativ ähnlich (Likelihood-Ratio- χ^2 : 340,605; Pearson- χ^2 : 386,883), so dass wir vermuten können, dass die asymptotischen Bedingungen für diese Tests zumindest annähernd erfüllt sind und wir die *p*-Werte der Statistiken – wenn auch unter Vorbehalt – interpretieren können.

Beide Statistiken weisen signifikante *p*-Werte auf. Somit wäre das 3-Klassenmodell nach diesen Statistiken für die vorliegenden Daten zu verwerfen. Allerdings sind die Werte der Statistiken nicht sehr viel größer als die Freiheitsgrade ($df = 229$), was darauf hindeutet, dass der Grad an Misfit nicht extrem groß ist. Zudem untersuchen wir eine recht große Stichprobe ($N = 861$), so dass viel Teststärke vorliegt. Weitere Modellgüteindikatoren und Modellvergleiche sollten herangezogen werden, um ein genaueres Bild des Modelfits zu bekommen.

6.3.1.2 Residuenstatistiken (tech10-Output)

Mit Hilfe von Residuenstatistiken können wir u.a. untersuchen, welche Antwortmuster besonders zum Misfit des Modells (d.h. zur Höhe der Pearson- χ^2 - und die Likelihood-Ratio-Statistik) beitragen. Auf diese Weise lassen sich beispielsweise Outlier detektieren (Personen mit extrem abweichenden Antwortmustern, sog. „aberrant response patterns“). Die Residuen geben für jedes Pattern den Beitrag an, den das betreffende Pattern zum Pearson χ^2 - bzw.

zum Likelihood-Ratio- χ^2 -Wert beisteuert. In Mplus erhält man außerdem standardisierte Residuen in Form von z-Werten. Große absolute z-Werte weisen auf beobachtete Patternhäufigkeiten hin, deren Auftreten vom Modell besonders unter- oder überschätzt wird. Die Residuenstatistiken lassen sich anfordern, indem man die Option `tech10` über die Outputzeile anfordert:

```
output: tech10;
```

Um Zeit zu sparen, benutzen wir bei der erneuten Schätzung des Modells mit `tech10`-Output die `optseed`-Funktion im `analysis`-Befehl (siehe Box 6.6). Wir geben das Seed mit dem besten Loglikelihood-Wert aus unserer vorherigen Analyse mit multiplen Startwertesets an, so dass Mplus automatisch die beste Lösung repliziert, ohne wiederum 500 Startwertesets einzeln analysieren zu müssen:

```
analysis: optseed = 605358;
```

Der vollständige Input kann der CD zum Buch entnommen werden. Im Output erhalten wir unter den Modellparameterschätzungen nun zusätzlichen Output. Als erstes erscheinen im `TECHNICAL 10 OUTPUT` alle beobachteten Antwortmuster. Jedes beobachtete Antwortmuster über die acht Computerspielitems wird ausgegeben und nummeriert. In unserem Beispiel traten 163 (der insgesamt $2^8 = 256$ möglichen) Pattern tatsächlich auf. Das Pattern mit der Nummer 1 (00100001) ist zum Beispiel das Muster, bei welchem Kinder nur beim dritten und beim letzten Item die Kategorie „oft/sehr oft“ gewählt haben. Bei Pattern Nummer 2 wurde die Kategorie „oft/sehr oft“ bei keinem der Items gewählt und so weiter.

```
TECHNICAL 10 OUTPUT
MODEL FIT INFORMATION FOR THE LATENT CLASS INDICATOR MODEL PART
```

RESPONSE PATTERNS

No.	Pattern	No.	Pattern	No.	Pattern	No.	Pattern
1	00100001	2	00000000	3	11000101	4	00001100
5	00000100	6	00001000	7	00000001	8	00101000
9	00001110	10	10000001	11	00000010	12	00010000
13	10011111	14	01001100	15	01001010	16	11101011
17	11001100	18	01011110	19	00011110	20	00001101
21	10011100	22	10001100	23	01001111	24	00101100
25	10111111	26	11100111	27	01000011	28	00010010
29	10000000	30	00011000	31	10001110	32	10101110
33	00001010	34	00011111	35	01000110	36	00100101
37	10010111	38	00011100	39	10010100	40	11011100
41	00100011	42	01000000	43	00000110	44	00000101
45	00001001	46	00001111	47	10000110	48	00100000
49	00010001	50	11111111	51	10111011	52	00100100
53	01000010	54	11000000	55	10101111	56	11010001
57	01100010	58	01110000	59	01100000	60	00111111
61	01011001	62	00110000	63	01110111	64	01001011
65	01100011	66	00101011	67	01010001	68	11100010
69	11000111	70	11000001	71	11100001	72	01110011

73	01100111	74	01101000	75	11000010	76	11111011
77	11010111	78	11110011	79	11100011	80	11101111
81	01100001	82	10010010	83	01101111	84	11101001
85	11001000	86	00101001	87	10110010	88	11110001
89	00011010	90	00000011	91	11101010	92	01010011
93	11110111	94	10001011	95	00000111	96	11000011
97	01010010	98	11010011	99	11001001	100	00100111
101	10010011	102	00110010	103	10001010	104	00011011
105	01010100	106	11110010	107	11111000	108	00101110
109	01101011	110	11011110	111	11001011	112	01100110
113	01000100	114	00001011	115	01100101	116	11011011
117	01010000	118	01001110	119	10100011	120	00010100
121	00101111	122	11010010	123	11011001	124	10000010
125	10110111	126	10010000	127	00110011	128	10010110
129	00110001	130	10100000	131	10000011	132	10011000
133	10111101	134	10100010	135	00101101	136	10001111
137	01011000	138	01110010	139	11011000	140	10100001
141	10111010	142	00010110	143	01000001	144	01101001
145	10000100	146	01010111	147	00010111	148	11110110
149	00100010	150	01111010	151	11001111	152	01011011
153	00101010	154	01011111	155	11011111	156	11010000
157	00111011	158	11011010	159	01110101	160	01110001
161	10010001	162	11010101	163	01001000		

Darunter erhalten wir die Auflistung der beobachteten (observed) sowie der unter dem Klassenmodell erwarteten (expected) Patternhäufigkeiten (frequency). Diese Liste wird aus Platzgründen hier nur für die ersten 44 und nicht für alle 163 Pattern gezeigt. Beispielsweise wurde das erste Pattern (00100001) in unseren Daten bei 13 Kindern beobachtet, aber vom 3-Klassenmodell nur bei ca. 4-5 Kindern erwartet. Diese Diskrepanz ergibt einen relativ großen standardisierten Residualwert von $z = 4,09$ für das Pattern Nr. 1. Das zeigt, dass dieses Antwortmuster bedeutsam zum Misfit des 3-Klassenmodells beiträgt.

Der größte z-Wert ergibt sich für das Pattern Nr. 40 (11011100; $z = 8,47$). Offensichtlich können Kinder mit diesem Antwortmuster in der 3-Klassenlösung nur schwer verortet werden.

Der Beitrag zur Pearson-X²-Statistik sowie zur Loglikelihood wird ebenfalls für jedes beobachtete Antwortmuster ausgegeben. Für unser Pattern Nr. 1 ergibt sich ein (relativ großer) Beitrag von 16,68 für den Pearson-X²-Wert und von 28,07 für die Loglikelihood. Man kann diese Werte ähnlich wie Modifikationsindices bei Strukturgleichungsmodellen (vgl. Kapitel 3) interpretieren. Läge beispielsweise Pattern Nr. 1 nicht vor, so würde sich der Modellfit gemäß Pearson-X²- und Loglikelihood etwa um die Beträge 16,68 bzw. 28,07 verbessern.

Die Residuenstatistiken können wertvolle Hinweise über Outlier („unskalierbare Personen“) und andere Besonderheiten in den Daten liefern. Diese Statistiken sollten allerdings niemals dazu verwendet werden, rein datengeleitet Probanden mit „ungünstigen“ Antwortmustern aus den Analysen auszuschließen.

RESPONSE PATTERN FREQUENCIES AND CHI-SQUARE CONTRIBUTIONS					
Response Pattern	Frequency Observed	Standardized Estimated Residual	Chi-square Contribution Pearson	Loglikelihood	(z-score)
1	13.00	4.42	4.09	16.68	28.07
2	261.00	255.07	0.44	0.14	12.00
3	1.00	0.78	0.25	0.06	0.51
4	27.00	18.73	1.93	3.65	19.75
5	32.00	35.06	-0.53	0.27	-5.84
6	29.00	31.54	-0.46	0.20	-4.87
7	14.00	18.34	-1.02	1.03	-7.56
8	5.00	3.05	1.12	1.24	4.93
9	11.00	13.22	-0.62	0.37	-4.04
10	2.00	2.05	-0.03	0.00	-0.10
11	19.00	18.53	0.11	0.01	0.96
12	3.00	3.94	-0.48	0.23	-1.64
13	1.00	0.50	0.71	0.50	1.39
14	1.00	2.52	-0.96	0.91	-1.85
15	3.00	2.26	0.49	0.24	1.70
16	1.00	2.05	-0.73	0.54	-1.43
17	4.00	0.68	4.03	16.22	14.18
18	2.00	0.67	1.63	2.65	4.38
19	2.00	3.03	-0.59	0.35	-1.66
20	9.00	8.18	0.29	0.08	1.71
21	2.00	0.90	1.15	1.33	3.18
22	2.00	3.92	-0.97	0.94	-2.69
23	4.00	1.64	1.85	3.40	7.14
24	5.00	3.23	0.99	0.97	4.37
25	1.00	0.22	1.67	2.77	3.03
26	2.00	1.40	0.51	0.26	1.43
27	9.00	10.32	-0.41	0.17	-2.47
28	5.00	2.40	1.68	2.83	7.35
29	4.00	3.88	0.06	0.00	0.24
30	1.00	1.41	-0.35	0.12	-0.69
31	4.00	3.30	0.39	0.15	1.55
32	2.00	0.73	1.49	2.21	4.03
33	4.00	5.48	-0.63	0.40	-2.52
34	3.00	1.70	1.00	1.00	3.41
35	2.00	1.96	0.03	0.00	0.09
36	4.00	1.37	2.25	5.04	8.57
37	2.00	0.57	1.89	3.55	5.00
38	1.00	3.59	-1.37	1.87	-2.56
39	1.00	0.46	0.80	0.64	1.56
40	4.00	0.20	8.47	71.73	23.92
41	8.00	6.40	0.64	0.40	3.57
42	7.00	6.57	0.17	0.03	0.89
43	9.00	7.12	0.71	0.49	4.21
44	8.00	5.25	1.20	1.44	6.74
...

Unter der Residuenliste wird auch noch der Gesamtbeitrag nicht beobachteter Antwortmuster („Nullzellen“) zum Pearson-X²-Wert angegeben:

```
THE TOTAL PEARSON CHI-SQUARE CONTRIBUTION FROM EMPTY CELLS IS 77.39
```

Außerdem erhält man über tech10 auch noch univariate und bivariate Modellfitstatistiken (hier nicht dargestellt), die ebenfalls zur Beurteilung des Modelfits eines LCA-Modells herangezogen werden können.

6.3.2 Relativer Modellfit

Aufgrund der Komplikationen bei der Beurteilung des absoluten Fits unter Verwendung der Pearson-X²- und Likelihood-Ratio-Statistik (Verletzung der asymptotischen Voraussetzungen; vgl. Abschnitt 6.3.1.1) betrachtet man bei LCA-Modellen häufig der Einfachheit halber eher den relativen Fit einer Klassenlösung, d.h. wie ein Modell im Vergleich zu anderen Modellen mit mehr oder weniger Klassen zu beurteilen ist. Hierbei stehen in Mplus als statistische Tests der Bootstrap-Likelihood-Ratio-Differenzentest (tech14) und der Vuong-Lo-Mendell-Rubin-Test (tech11) zur Verfügung. Außerdem lassen sich LCA-Modelle ähnlich wie andere statistische Modelle auf der Basis informationstheoretischer Maße (AIC, BIC und sample size adjusted BIC) vergleichen.

6.3.2.1 Bootstrap-Likelihood-Ratio-Differenzentest (tech14-Output)

Mit Hilfe des Bootstrap-Likelihood-Ratio-Differenzentests kann ein Modell mit g latenten Klassen statistisch gegen ein Modell mit $g - 1$ Klassen getestet werden. Hierbei wird die Differenz der Likelihood-Ratio-Testwerte zwischen dem $g - 1$ und dem g -Klassenmodell berechnet. Mit Hilfe einer parametrischen Bootstrap-Prozedur wird der approximativ korrekte p -Wert für den Likelihood-Ratio-Differenzwert aufgrund von Monte-Carlo-Bootstrap-Stichproben (siehe z.B. Langeheine et al., 1996) geschätzt. Ein signifikanter Wert zeigt an, dass das Modell mit g Klassen signifikant besser auf die Daten passt als das sparsamere Modell mit einer Klasse weniger. Bei einem nicht signifikanten Wert wird das sparsamere Modell mit einer Klasse weniger bevorzugt.

In unserem Beispiel testen wir das 3-Klassenmodell zunächst gegen ein 2-Klassenmodell, indem wir im Mplus-Input für das 3-Klassenmodell (siehe Abb. 6.1) die Option tech14 über die Outputzeile anfordern:

```
output: tech14;
```

Um die gewünschte Anzahl von Bootstrap-Stichproben (z.B. 500) zu spezifizieren, wählen wir:

```
analysis: lrtbootstrap = 500;
```

500 Bootstrap-Stichproben sollten für eine hinreichende Präzision des geschätzten approximativen Bootstrap- p -Wertes für den Differenztest ausreichen.

Um sicherzustellen, dass bei der Analyse der parametrischen Bootstrap-Stichproben keine lokalen Maxima auftreten, erhöhen wir die Anzahl der Startwertesets für das g und das $g - 1$ Klassenmodell jeweils auf 50 im ersten und 10 im zweiten Schritt der Optimierung:

ANALYSIS: LRTSTARTS = 50 10 50 10;

Bei komplexeren Modellen mit mehr Klassen empfiehlt es sich, im Zweifelsfall eine noch größere Anzahl von Startwertesets für das Bootstrapping zu verwenden, auch wenn dadurch die Rechenzeit zwangsläufig verlängert wird. Die Mplus-Voreinstellung (LRTSTARTS = 0 0 20 5) ist aufgrund der geringen Zahl an Startwerten häufig nicht ausreichend, um valide Bootstrap-Ergebnisse zu erhalten. Wir erhalten den nachfolgend dargestellten zusätzlichen Output:

TECHNICAL 14 OUTPUT

Random Starts Specifications for the k-1 Class Analysis Model		
Number of initial stage random starts		500
Number of final stage optimizations		50
Random Starts Specification for the k-1 Class Model for Generated Data		
Number of initial stage random starts		50
Number of final stage optimizations		10
Random Starts Specification for the k Class Model for Generated Data		
Number of initial stage random starts		50
Number of final stage optimizations		10
Number of bootstrap draws requested		500

PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES

H0 Loglikelihood Value	-3522.517
2 Times the Loglikelihood Difference	117.407
Difference in the Number of Parameters	9
Approximate P-Value	0.0000
Successful Bootstrap Draws	500

Der Loglikelihood-Wert für das $g - 1$ -Klassenmodell (hier das 2-Klassenmodell, welches gegen das 3-Klassenmodell getestet wird) beträgt -3522,517 (abzulesen unter H0 Loglikelihood Value). Allgemein ist es wichtig, den an dieser Stelle für das $g - 1$ -Klassenmodell angegebenen Loglikelihood-Wert mit dem Wert zu vergleichen, den man für eine separate Analyse des $g - 1$ -Klassenmodells unter Verwendung mehrerer Startwertesets erhält, um sicherzustellen, dass es sich hier nicht um eine lokale Lösung handelt. Der Output für eine separate Analyse des 2-Klassenmodells ist hier nicht gesondert dargestellt, befindet sich jedoch auf der Begleit-CD. Es zeigt sich, dass bei der separaten Schätzung des 2-Klassenmodells mit 500 Startwertesets derselbe Loglikelihood-Wert für die finale Lösung ermittelt wurde, so dass wir den vorliegenden Bootstrap-Ergebnissen trauen können.

Mplus gibt uns den Likelihood-Ratio- χ^2 -Differenzenwert an, der sich aus der Differenz der Likelihood-Ratio- χ^2 -Werte beider Modelle ergibt. Für das 2-Klassenmodell beträgt der Likelihood-Ratio- χ^2 -Wert 458,012 (siehe Output für das 2-Klassenmodell auf der Begleit-CD) und für das 3-Klassenmodell 340,605 (siehe Abschnitt 6.3.1.1). Somit beträgt die Differenz 117,407. Im 3-Klassenmodell werden 9 Parameter mehr geschätzt als im 2-Klassenmodell (ein zusätzlicher Klassengrößenparameter und acht zusätzliche bedingte Antwortwahrscheinlichkeiten). Die χ^2 -Differenz von 117,407 ist gemäß Bootstrap-*p*-Wert (Approximate P-Value) signifikant ($p < 0,0001$). Somit passt nach diesem Kriterium das 3-Klassenmodell signifikant besser als das 2-Klassenmodell.

Für den Vergleich von 3 versus 4 Klassen muss zusätzlich die 4-Klassenlösung mit tech14-Output geschätzt werden (für die genaue Input-Spezifikation siehe Begleit-CD). Wir erhalten hierbei folgendes Ergebnis (verkürzt dargestellt):

TECHNICAL 14 OUTPUT

PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 3 (H0) VERSUS 4 CLASSES

H0 Loglikelihood Value	-3463.814
2 Times the Loglikelihood Difference	54.239
Difference in the Number of Parameters	9
Approximate P-Value	0.0000
Successful Bootstrap Draws	500

Wir sehen, dass unser Loglikelihood-Wert für die 3-Klassenlösung (-3463,814) wiederum exakt repliziert wurde, so dass wir den Bootstrap-Ergebnissen vertrauen können. Es zeigt sich, dass auch die Differenz der Likelihood-Ratio- χ^2 -Werte des 3- und des 4-Klassenmodells signifikant ist ($\Delta\chi^2 = 54,239, df = 9, p < 0,0001$). Nach diesem Kriterium wäre also ein Modell mit vier Klassen gegenüber der 3-Klassenlösung zu bevorzugen.

Allerdings sollte über die Anzahl der Klassen nicht nur nach statistischen sondern auch nach inhaltlichen Kriterien und Interpretierbarkeit der jeweiligen Lösungen entschieden werden. In unserem Beispiel ist die 3-Klassenlösung besser interpretierbar als die 4-Klassenlösung (siehe Abschnitt 6.3.3).

6.3.2.2 Vuong-Lo-Mendell-Rubin-Test (tech11-Output)

Der Vuong-Lo-Mendell-Rubin-(VLMR-)Test (Lo, Mendell & Rubin, 2001) basiert auf einem ähnlichen Prinzip wie der Likelihood-Ratio- χ^2 -Differenzentest. Ein signifikanter Wert des VLMR-Tests zeigt an, dass das geschätzte Modell signifikant besser passt, als ein Modell mit einer Klasse weniger. Man erhält den VLMR-Test sowie eine adjustierte Variante des Tests in Mplus, indem man folgenden Befehl in den Input mit aufnimmt:

OUTPUT: TECH11;

In unserem Beispiel liefert der VLMR-Test folgendes Ergebnis für 2 versus 3 latente Klassen:

TECHNICAL 11 OUTPUT	
Random Starts Specifications for the k-1 Class Analysis Model	
Number of initial stage random starts	500
Number of final stage optimizations	50
VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES	
H0 Loglikelihood Value	-3522.517
2 Times the Loglikelihood Difference	117.407
Difference in the Number of Parameters	9
Mean	19.334
Standard Deviation	20.559
P-Value	0.0035
LO-MENDELL-RUBIN ADJUSTED LRT TEST	
Value	115.508
P-Value	0.0038

Wiederum ist es wichtig, den unter H0 Loglikelihood Value angegebenen Loglikelihood-Wert für das $g - 1$ -Klassenmodell mit dem Wert zu vergleichen, den man für eine separate Analyse des $g - 1$ -Klassenmodells unter Verwendung multipler Startwertesets erhält (im vorliegenden Fall sind die Werte wie gewünscht identisch).

Auch gemäß VLMR-Test ist das 3-Klassenmodell für die vorliegenden Daten zu bevorzugen, da $p = 0,0035$ (beziehungsweise 0,0038 für den adjustierten Test). Der Vergleich von 3 versus 4 Klassen gemäß VLMR-Test (wiederum über die Berechnung der 4-Klassenlösung mit tech11, siehe CD) ergibt ein anderes Ergebnis als der Bootstrap-Likelihood-Ratio- χ^2 -Differenzentest:

TECHNICAL 11 OUTPUT	
VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 3 (H0) VERSUS 4 CLASSES	
H0 Loglikelihood Value	-3463.814
2 Times the Loglikelihood Difference	54.239
Difference in the Number of Parameters	9
Mean	18.229
Standard Deviation	28.354
P-Value	0.0896
LO-MENDELL-RUBIN ADJUSTED LRT TEST	
Value	53.362
P-Value	0.0925

Die p -Werte des VLMR-Tests sind hier nicht signifikant ($p = .0896$ bzw. $p = .0925$), was bedeutet, dass gemäß dieses Tests das 4-Klassenmodell nicht signifikant besser auf die Daten passt als das sparsamere 3-Klassenmodell.

Die Simulationsstudie von Nylund, Asparouhov und Muthén (2007) deutet jedoch darauf hin, dass der Bootstrap-Likelihood-Ratio- χ^2 -Differenzentest ein besserer Indikator für die adäquate Anzahl an latenten Klassen ist als der VLMR-Test, so dass für die Praxis der Bootstrap-Likelihood-Ratio- χ^2 -Differenzentest eher zu empfehlen ist.

6.3.2.3 Informationstheoretische Maße

Informationstheoretische Maße (*Information Criteria, IC*) sind deskriptive Indices für Modellvergleiche. Sie berücksichtigen sowohl die Güte der Anpassung des Modells an die Daten als auch die Modellsparsamkeit (Anzahl an geschätzten Parametern). Modelle mit mehr Klassen passen tendenziell immer besser auf die Daten, aber mit jeder zusätzlichen Klasse werden auch mehr Parameter geschätzt, und die Modellsparsamkeit nimmt ab. Das beste Modell im Sinne der Informationskriterien ist das Modell, welches einerseits gut passt und andererseits mit möglichst wenigen Parametern auskommt. In einer Serie von LCA-Modellen wird dasjenige Modell bevorzugt, das den kleinsten *AIC*, *BIC*- oder *sample size adjusted BIC* (*aBIC*)-Wert aufweist. In unserem Beispiel erhalten wir für das 3-Klassenmodell folgende *IC*-Werte im Mplus-Standardoutput:

Information Criteria		
Number of Free Parameters		26
Akaike (AIC)		6979.628
Bayesian (BIC)		7103.338
Sample-Size Adjusted BIC ($n^* = (n + 2) / 24$)		7020.769

Diese Werte sagen nur etwas aus, wenn wir sie relativ zu den Werten anderer Modelle betrachten. Beispielsweise können wir die vorliegenden Werte mit denen der 2- bzw. der 4-Klassenlösung vergleichen. Die *IC*-Werte für alle drei Lösungen sind in Tabelle 6.1 dargestellt. Es zeigt sich, dass der *BIC*-Index für die 3-Klassenlösung am kleinsten ausfällt, während *AIC* und *aBIC* auf vier (bzw. noch mehr) Klassen hindeuten.

Nylund et al. (2007) empfehlen aufgrund ihrer Simulationsstudien neben dem Bootstrap-Likelihood-Ratio-Differenzentests (siehe Abschnitt 6.3.2.1) besonders die Verwendung des *BIC*-Index zur Bestimmung der Anzahl latenter Klassen.

Tabelle 6.1: Informationstheoretische Maße für verschiedene Klassenlösungen für die Computer-spieldaten

	<i>AIC</i>	<i>BIC</i>	<i>aBIC</i>
2-Klassenlösung	7079,035	7159,923	7105,935
3-Klassenlösung	6979,628	7103,338	7020,769
4-Klassenlösung	6943,389	7109,922	6998,771

Anmerkungen. *AIC* = Akaike's Information Criterion. *BIC* = Bayesian Information Criterion. *aBIC* = sample size adjusted *BIC*. Kleinstes Werte sind fett gedruckt.

Box 6.7: Überblick über wichtige Kriterien zur Beurteilung der Güte von LCA-Modellen

- **Signifikanztests zur Bewertung des absoluten Fits:** Die Likelihood-Ratio- und Pearson-X²-Statistik; werden von Mplus standardmäßig ausgegeben. Problem: Diese Statistiken folgen nur bei geringen Itemzahlen und großer Stichprobe der theoretischen χ^2 -Verteilung. Empfehlenswert: parametrischen Bootstrap zur Bestimmung einer validen Prüfverteilung durchführen (in Mplus bis Version 5 nicht erhältlich bzw. nur über die Durchführung einer Monte-Carlo-Simulation).
- **Statistische Modellvergleiche:**
 - Informationstheoretische Maße (z.B.: AIC, BIC, aBIC) zum deskriptiven Modellvergleich; besonders empfehlenswert: BIC. Das Modell mit dem kleinsten BIC-Wert wird gewählt.
 - Statistische Tests (Vuong-Lo-Mendell-Rubin-Test, in Mplus anzufordern über output: tech11; Bootstrap-Likelihood-Ratio-Differenzen-Test; in Mplus anzufordern über output: tech14); besonders empfehlenswert: Bootstrap-Likelihood-Ratio-Differenzen-Test.
- **Mittlere Klassenzuordnungswahrscheinlichkeiten:** Für eine gute Klassenlösung sollten die mittleren Zuordnungswahrscheinlichkeiten für alle Klassen möglichst > 0,80 sein.
- **Entropy:** Globalmaß für die Zuverlässigkeit der Klassifikation in einem LCA-Modell. Werte nahe 1 sprechen für eine insgesamt hohe Sicherheit der Klassifikation, während Werte nahe 0 eine hohe Unsicherheit anzeigen.
- **Anzahl der Klassen:** Die Lösung sollte mit möglichst wenigen Klassen auskommen (Parsimonitätsprinzip). Bei Lösungen mit einer oder mehreren sehr kleinen Klassen sollte man überlegen, ob ein Modell mit weniger Klassen vorzuziehen ist.
- **Anzahl der Boundary Estimates:** Werden viele klassenbedingte Antwortwahrscheinlichkeiten auf exakt 0 oder 1 geschätzt, so kann dies ein Zeichen für die Extraktion von zu vielen Klassen oder ein lokales Maximum sein.
- **Residuenanalyse:** Residuen (output: tech10) können Aufschluss über die Ursachen einer schlechten Modellanpassung geben. Insbesondere kann analysiert werden, welche der beobachteten Antwortmuster besonders zur Fehlanpassung des Modells beitragen. So können z.B. auch Outlier (Individuen mit abweichenden Antwortmustern) identifiziert werden.
- **Interpretierbarkeit:** Jede Klasse sollte eindeutig interpretierbar sein. Dies gelingt am besten, wenn alle Items innerhalb einer Klasse entweder hohe oder niedrige (nicht aber mittlere) Antwortwahrscheinlichkeiten aufweisen. Eine interpretierbare sollte stets einer nichtinterpretierbaren Lösung vorgezogen werden.
- **Replizierbarkeit:** Günstig ist es, wenn sich eine exploratorisch ermittelte Klassenlösung mit einem neuen Datensatz replizieren lässt.
- **Zusammenhänge mit externen Validierungskriterien:** Hängen die latenten Klassen mit anderen Variablen („Kovariaten“) in theoretisch vorhergesagter Weise zusammen, so spricht dies für die Validität (oder zumindest Nützlichkeit) einer gewählten Lösung.

6.3.3 Interpretierbarkeit

Bei der Beurteilung der Modellgüte und beim Vergleich verschiedener LCA-Modelle spielen nicht nur statistische Kriterien eine wichtige Rolle. Mindestens ebenso wichtig ist die Interpretierbarkeit einer Lösung. Dies können wir gut anhand unseres Beispiels illustrieren. Obwohl gemäß Bootstrap-Likelihood-Ratio- χ^2 -Differenzentest die 4-Klassenlösung der 3-Klassenlösung vorzuziehen ist, ergibt eine Inspektion der Klassenprofile der 4-Klassenlösung, dass in dieser Lösung eine der Klassen nur schwer zu interpretieren ist, da in dieser Klasse überwiegend mittlere Antwortwahrscheinlichkeiten über die Items hinweg vorliegen.

Grundsätzlich zeichnet sich eine gut interpretierbare Klassenlösung dadurch aus, dass überwiegend hohe (nahe 1) oder niedrige (nahe 0), jedoch keine mittleren bedingten Antwortwahrscheinlichkeiten in den Klassenprofilen auftreten. Klassen mit mittleren Antwortwahrscheinlichkeiten für viele Items sind meist schwer zu interpretieren, da die Personen dann eine „Fifty-fifty-Wahrscheinlichkeit“ für viele Items haben. Eine Ausnahme stellen dichotome Leistungstestitems dar, bei denen eine Wahrscheinlichkeit von 0,50 genau der Ratewahrscheinlichkeit entspricht. Hier sind „Rateklassen“ mit mittleren klassenbedingten Wahrscheinlichkeiten für einige oder alle Testitems durchaus sinnvoll und interpretierbar.

Außerdem tritt in der 4-Klassenlösung in einer Klasse ein Boundary Estimate auf (eine klassenbedingte Wahrscheinlichkeit π_{ig} wird auf exakt 0 geschätzt). Somit erscheint die 3-Klassenlösung die inhaltlich sinnvollere Repräsentation der Daten zu sein, da sie besser interpretierbar (und darüber hinaus auch sparsamer) ist. Box 6.7 fasst noch einmal die wichtigsten Kriterien zur Beurteilung der Modellgüte in der LCA zusammen.

Anhang A: Zentrale Mplus-Befehle

Befehl	Bedeutung	Bemerkungen	Kapitel
<code>title:</code>	Kommentar/Titel zur Analyse	Optionaler Befehl	2
<code>data: file = <Datensatz.dat>;</code>	Spezifikation des Datenfiles		2
<code>data: listwise = on;</code>	Anforderung der Verwendung von listenweisem Fallausschluss anstelle von FIML-Schätzung mit fehlenden Werten	Der Mplus-Default ab Version 5 ist FIML-Schätzung mit fehlenden Werten	2
<code>data: type = <Art der Summary-Daten>;</code>	Spezifikation der Art zu analysierender Summary-Daten	Beispiel 1 (Mittelwerte, Standardabweichungen und Korrelationsmatrix): <code>data: type = means std corr;</code> Beispiel 2 (Kovarianzmatrix): <code>data: type = covar;</code>	2.2
<code>data: nobservations = <Stichprobengröße>;</code>	Spezifikation der Stichprobengröße bei Verwendung von Summary-Daten	Bei Individualdaten nicht benötigt; im Buch wird auch die Abkürzung <code>nobs</code> verwendet	2.2
<code>variable: names = <Liste aller Variablennamen>;</code>	Spezifikation der Variablennamen	Jeder Variablenname darf maximal 8 Zeichen lang sein	2
<code>variable: categorical = <Liste Variablennamen>;</code>	Spezifikation von Variablen als geordnet kategorial (ordinal)		6
<code>variable: nominal = <Liste Variablennamen>;</code>	Spezifikation von Variablen als ungeordnet kategorial		6

	(nominal)		
<code>variable: usevariables = <Liste Variablennamen>;</code>	Spezifikation der im Modell verwendeten/analysierten Variablen	Nicht benötigt, wenn alle Variablen im Datensatz gleichzeitig analysiert werden sollen; im Buch wird meist die Abkürzung <code>usevar</code> verwendet	3
<code>variable: missing = <Variablenname (Missing- Value-Code)>;</code>	Spezifikation des Missing-Value-Codes	<code>missing = all <Missing- Value-Code>;</code> definiert denselben Code für alle Variablen	2
<code>variable: cluster = <Clustervariable>;</code>	Definition einer Clustervariable	Benötigt bei Mehrebenenanalysen	5
<code>variable: within = <Name Level-1-Variable>;</code>	Definition von Variablen, die in einer Mehrebenenanalyse nur auf Level-1 verwendet werden		5
<code>variable: between = <Name Level-2-Variable></code>	Definition von Variablen, die in einer Mehrebenenanalyse nur auf Level-2 verwendet werden		5
<code>variable: centering = grandmean (<Variable>);</code>	Zentrierung einer Variable am Gesamtmittelwert		3.2; 5
<code>variable: centering = groupmean (<Variable>);</code>	Zentrierung einer Variable am Clustermittelwert		5
<code>variable: classes = <Name der latenten Klassenvariable> (<Anzahl der latenten Klassen>);</code>	Spezifikation des Namens der latenten Klassenvariable und der Anzahl zu extrahierender latenter Klassen	Benötigt u.a. bei Latent-Class-Analysen	6
<code>variable: auxiliary = <Name der Auxiliary- Variable>;</code>	Spezifikation einer „Hilfsvariable“ (z.B. Versuchspersonennummer)	Relevant u.a., wenn in Mplus eine neue Datendatei abgespeichert und der Probandencode o.ä. mit abge-	6

		speichert werden soll	
analysis: type = basic;	basic-Analyse	Nur Ausgabe von Deskriktivstatistiken, keine Modellschätzung	2
analysis: type = general;	Analyseform für normale Regressions-, Pfad- und Strukturgleichungsanalysen	Mplus-Default-Analysetyp	3; 4
analysis: type = twolevel;	2-Ebenen-Analyse	Benötigt bei Mehrebenenanalysen in Verbindung mit der Spezifikation einer Clustervariable	5
analysis: type = twolevel random;	Analyse auf zwei Ebenen mit Random Slopes	Benötigt bei Mehrebenenanalysen mit Random Slopes	5.5
analysis: type = mixture;	Mischverteilungsanalyse	Benötigt u.a. bei Latent-Class-Analysen in Verbindung mit variable: classes = <Name der latenten Klassenvariablen(n)> (<Anzahl der latenten Klassen>);	6
analysis: estimator = <Schätzverfahren>;	Spezifikation des Schätzverfahrens (z.B. ml für Maximum-Likelihood)	Kann verwendet werden, um den jeweiligen Default-Estimator durch einen anderen zu ersetzen	verschiedene
model:	Grundbefehl zur Spezifikation eines Modells		4-7
model: y on x;	Regression einer Variable y auf eine unabhängige Vari-		4-6

	able x		
model: y with x;	Kovarianz bzw. Korrelation zwischen zwei Variablen y und x		4-6
model: f by y;	Faktorladung eines Indikators y auf einem latenten Faktor f	Die Ladung der ersten Variable hinter dem by-Befehl ist zur Identifizierbarkeit per Default auf 1 fixiert	3; 4
<Parameter>*	Freisetzung eines fixierten Parameters		3
<Parameter>*<Zahl>	Vergabe eines Startwertes für einen Parameter		3
<Parameter>@<Zahl>	Fixierung eines Parameters auf einen bestimmten Wert		3
<Parameter> (Zahl oder Label)	Gleichsetzung von Parametern		4.1.4
	Spezifikation von Wachstumsfaktoren oder Random Slopes	Benötigt bei Wachstumskurvenmodellen und Mehrebenenmodellen mit Random Slopes	4.5; 5.5
model: %within% <Modellspezifikation>	Spezifikation des Level-1-Teils eines Modells in einer Mehrebenenanalyse		5
model: %between% <Modellspezifikation>	Spezifikation des Level-2-Teils eines Modells in einer Mehrebenenanalyse		5
<Name Random Slope> y on x;	Spezifikation eines Random Slopes für die Regression von y auf x in einem Mehrebenenmodell	Benötigt bei Mehrebenenmodellen mit Random Slopes	5.5
model indirect:	Befehl zur Ausgabe von indirekten Effekten	Vor allem in Pfadanalysen (Mediatormodellen) relevant	3.5
model indirect: y ind x;	Ausgabe aller indirekten Effekte von x auf y		3.5
output:	Grundbefehl zur Anforderung von Nichtstandard-		verschiedene

	output		
output: sampstat;	Deskriptivstatistiken		verschiedene
output: patterns;	Ausfallmuster bei fehlenden Werten (in Verbindung mit FIML-Schätzung)	Kann nicht angefordert werden, wenn data: listwise = on; gewählt wurde	2
output: standardized;	Alle standardisierten Lösungen		verschiedene
output: stdyx;	Nur vollstandardisierte Lösung		verschiedene
output: cinterval;	Konfidenzintervalle für alle Parameterschätzungen		3.5
output: residual;	Modellimplizierte Kovarianz- und Mittelwertsstruktur sowie Residuenstatistiken		
output: modindices;	Modellmodifikationsindizes		
plot:	Grundbefehl zur Ausgabe von Diagrammen		4.5; 6
savedata: file = <Name neuer Datensatz>;	Befehl zum Abspeichern von neuen Datensätzen		6
savedata: save = <Zu speichernde Information>;	Spezifikation der zu speichernden Schätzwerte	z.B. Klassenzuordnungswahrscheinlichkeiten in der LCA	6

Anhang B: Häufige Fehler bei der Mplus-Spezifikation und „Troubleshooting“

Die folgende Tabelle fasst einige häufige Fehlerquellen bei der Mplus-Input-Spezifikation sowie Fehlerbehebungsmöglichkeiten zusammen.

Fehler/Problem	Mögliche Fehlerquelle	Mögliche Lösung
Fehlermeldung(en) im Output, kein Modell berechnet	<ul style="list-style-type: none"> Datensatz wurde von Mplus nicht gefunden Datensatz konnte von Mplus nicht korrekt gelesen werden Inputzeile länger als 80 bzw. 90 Zeichen Semikolon als Befehlsabschluss in einem Befehl vergessen usevar-Befehl inkorrekt spezifiziert (zu viele oder zu wenige Variablen aufgelistet) oder usevar-Befehl vergessen Variablennamen falsch geschrieben (bei usevar oder im model-Befehl) 	<ul style="list-style-type: none"> Überprüfen, ob Dateipfad und -name korrekt spezifiziert wurden Überprüfen, ob die Anzahl der Variablen korrekt spezifiziert wurde Zu lange Zeilen im Input mit Return umbrechen Semikolon im Input am Ende jedes Unterbefehls hinzufügen usevar-Befehl korrekt spezifizieren model-Befehl überprüfen
Inkorrekte Berechnungen/Schätzergebnisse; ggf. nicht positiv definite Matrizen/Konvergenzprobleme;	<ul style="list-style-type: none"> Fehlende Werte nicht oder nicht korrekt definiert Inkorrektes Einlesen der Daten Reihenfolge der Variablen im Datensatz stimmt nicht mit der Reihenfolge überein, die unter variable = names angegeben wurde 	<ul style="list-style-type: none"> Im variable-Befehl Missing Value Code definieren mit Hilfe des Unterbefehls missing = Datensatz überprüfen, ggf. neu abspeichern und basic-Analyse durchführen Reihenfolge der Variablennamen im Input überprüfen

<p>Fehlermeldung im Output: „At least one variable is uncorrelated with all other variables in the model. Check that this is what is intended.“</p>	<ul style="list-style-type: none"> • usevar-Befehl vergessen (Achtung: dann benutzt Mplus <i>alle</i> Variablen im Datensatz!) • usevar-Befehl inkorrekt spezifiziert (zu viele Variablen aufgelistet) • usevar-Befehl einfügen sofern nicht <i>alle</i> Variablen im Datensatz simultan benutzt werden sollen. • Nur tatsächlich benutzte Variablen im usevar-Befehl auflisten
<p>Daten konnten von Mplus nicht gelesen werden oder wurden inkorrekt gelesen; aus diesem Grund gar keine oder inkorrekte Berechnungen/ Schätzergebnisse</p>	<ul style="list-style-type: none"> • Datensatz enthält Kommas statt Punkte als Dezimaltrennzeichen • Datensatz enthält Stringvariablen • Fehlende Werte wurden nicht kodiert, sondern sind als leere Zellen im Datensatz vorhanden • Variablennamen wurden versehentlich im Arbeitsblatt mit abgespeichert • Daten wurden in SPSS nicht als Tabulatorgetrennt (*.dat) sondern als festes ASCII (*.dat) abgespeichert • Datensatz wurde nicht gefunden, da der Name oder der Pfad im Input unter <code>data</code> falsch spezifiziert wurde • Reihenfolge oder Anzahl der Variablen im <code>variable</code>-Befehl unter <code>names =</code> stimmt nicht mit der Reihenfolge/Anzahl im Datenfile überein • Eine Input-Zeile ist länger als 80 bzw. 90 Zeichen • Sicherstellen, dass in dem von Mplus verwendeten Datensatz Punkte als Dezimaltrennzeichen verwendet werden (s. Kapitel 1) • Stringvariablen nach Möglichkeit vermeiden bzw. vor dem Einlesen in numerische Variablen umkodieren • Fehlende Werte in SPSS kodieren (siehe Abschnitt 1.2), dann Datensatz erneut als *.dat-Datei abspeichern (siehe Abschnitt 1.3) • Datensatz erneut in SPSS abspeichern, dabei Option „Variablennamen im Arbeitsblatt speichern“ deaktivieren (siehe Abb. 1.14); Alternative: erste Zeile in der *.dat-Datei manuell löschen und dann STRG + S drücken • Pfad korrekt spezifizieren • Reihenfolge der Variablen im <code>names</code>-Befehl korrekt spezifizieren • Zu lange Input-Zeilen mit Return umbrechen

Mplus schätzt kein Modell, obwohl alle Input-Befehle korrekt sind und die Daten richtig gelesen werden

Im Input ist unter analysis:
`type = basic;` spezifiziert.
Unter dieser Option werden keine Modellparameter geschätzt.

Lösung: `type = basic;` im Input ersetzen, z.B. durch
`type = general;`

Anhang C: Literaturhinweise zu Mplus

Das Originalhandbuch zu Mplus kann über die Mplus-Homepage (<http://www.statmodel.com>) als PDF-Datei heruntergeladen werden:

Muthén, L. K. & Muthén, B. O. (1998-2007). *Mplus User's Guide. Fifth Edition.* Los Angeles, CA: Muthén & Muthén. <http://www.statmodel.com/ugexcerpts.shtml>

Der Fortsetzungsband zum vorliegenden Buch befasst sich mit komplexeren statistischen Verfahren und Auswertungen in Mplus:

Geiser, C., Crayen, C. & Enders, C.K. (in Vorbereitung). *Datenanalyse mit Mplus für Fortgeschrittene.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Auf Spezialliteratur zu den statistischen Grundlagen der einzelnen besprochenen Analyseverfahren wird in den entsprechenden Kapiteln hingewiesen.

Literatur

- Aiken, L. S. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K. A. & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. New York: Wiley.
- Bollen, K. A. & Long, S. (Eds.) (1993). *Testing structural equation models*. Newbury Park: Sage Publications.
- Bryk, A. S. & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 101, 147-158.
- Byrne, B. M., Shavelson, R. J. & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456-466.
- Chen, F., Bollen, K. A., Paxton, P., Curran, P. & Kirby, J. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research*, 29, 468-508.
- Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Cole, D. A., Martin, J. M. & Powers, B. (1997). A competency-based model of child depression: A longitudinal study of peer, parent, teacher, and self-evaluations. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 38, 505-514.
- Cole, D. A., Martin, J. M., Powers, B. & Truglio, R. (1996). Modeling causal relations between academic and social competence and depression: A multitrait-multimethod longitudinal study of children. *Journal of Abnormal Psychology*, 105, 258-270.
- Cole, D. A. & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, 112, 558-577.
- Collins, L. M., Fidler, P. L., Wugalter, S. E. & Long, J. D. (1993). Goodness-of-fit testing for latent class models. *Multivariate Behavioral Research*, 28, 375-389.
- Collins, L. M. & Lanza, S. T. (2010). *Latent Class and latent transition analysis with applications in the social, behavioral, and health sciences*. New York: Wiley.
- Clogg, C. C. (1995). Latent class models: Recent developments and prospects for the future. In G. Arminger, C. C. Clogg & M. E. Sobel (eds.), *Handbook of statistical modeling in the social sciences* (pp. 311-359). New York: Plenum.
- Crayen, C. (2010). *Chi-square distributions calculator Version 3* [Computer software]. Berlin: Freie Universität.
- Duncan, T. E., Duncan, S. C. & Strycker, L. A. (2006). An introduction to latent variable growth curve modeling: Concepts, issues, and applications (2nd Ed.). Psychology Press.
- Eid, M. (2003). Veränderungsmessung und Kausalanalyse. In M. Jerusalem & H. Weber (Hrsg.), *Psychologische Gesundheitsförderung – Diagnostik und Prävention* (S. 105-120). Göttingen: Hogrefe.
- Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, 65, 241-261.

- Eid, M. (1999). Lineare Strukturgleichungsmodelle. In B. Strauß, H. Haag & M. Kolb (Hrsg.), *Datenanalyse in der Sportwissenschaft*. Hermeneutische und statistische Verfahren (S. 427-454). Schorndorf: Hofmann.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Forschungsmethoden und Statistik*. Weinheim: Beltz.
- Eid, M., Langeheine, R. & Diener, E. (2003). Comparing typological structures across cultures by multi-group latent class analysis. *Journal of Cross-Cultural Psychology*, 34, 195-210.
- Eid, M., Lischetzke, T. & Nussbeck, F. W. (2006). Structural equation models for multitrait-multimethod data. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 283-299). Washington, DC: American Psychological Association.
- Eid, M., Schneider, C. & Schwerkmezger, P. (1999). Do you feel better or worse? The validity of perceived deviations of mood states from mood traits. *European Journal of Personality*, 13, 283-306.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford.
- Enders, C. K. & Tofghi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12, 121-138.
- Ferrer, E., Balluerka, N. & Widaman, K. F. (2008). Factorial invariance and the specification of second-order growth models. *Methodology*, 4, 22-36.
- Finney, S. J. & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course*. Greenwich, CT: Information Age Publishing.
- Formann, A. K. (2003). Latent class model diagnostics—a review and some proposals. *Computational Statistics & Data Analysis*, 41, 549-559.
- Geiser, C., Crayen, C. & Enders, C. K. (in Vorb.). *Datenanalyse mit Mplus für Fortgeschrittene*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Geiser, C., Eid, M. & Nussbeck, F. W. (2008). On the meaning of the latent variables in the CT-C(M-1) model: A comment on Maydeu-Olivares & Coffman (2006). *Psychological Methods*, 13, 49-57.
- Geiser, C., Lehmann, W. & Eid, M. (2006). Separating “Rotators” from “Non-Rotators” in the Mental Rotations Test: A multigroup latent class analysis. *Multivariate Behavioral Research*, 41, 261-293.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.
- Hagenaars, J. A. (1993). *Loglinear models with latent variables*. Beverly Hills: Sage Publications.
- Hagenaars, J. A. & McCutcheon, A. (Eds.) (2002). *Applied latent class analysis*. Cambridge: Cambridge University Press.
- Hancock, G. R., Kuo, W. & Lawrence, F. R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling*, 8, 470-489.
- Heller, K., Gaedicke, A.-K. & Weinläder, H. (1976). *Kognitiver Fähigkeitstest (KFT 4-13)*. Weinheim: Beltz.
- Hertzog, C. & Nesselroade, J. R. (1987). Beyond autoregressive models: Some implications of the state-trait distinction for the structural modeling of developmental change. *Child Development*, 58, 93-109.
- Hofmann, D. A. & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, 24, 623-641.
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jöreskog, K. G. (1979a). Statistical models and methods for analysis of longitudinal data. In K. G. Jöreskog & D. Sörbom (Eds.), *Advances in factor analysis and structural equation models* (pp. 129-169). Cambridge, MA: Abt.

- Jöreskog, K. G. (1979b). Statistical estimation of structural models in longitudinal–developmental investigations. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 303–351). New York, NY: Academic Press.
- Kaplan, D. (2009). *Structural Equation Modeling: Foundations and Extensions* (2nd Edition). Newbury Park, CA: Sage Publications.
- Klein, A. (2000). *Moderatormodelle – Verfahren zur Analyse von Moderatoreffekten in Strukturgleichungsmodellen*. Hamburg: Verlag Dr. Kovač.
- Kline, R. B. (Ed.) (2004). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.
- Kreft, I. & de Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage.
- Kreft, I. G. G., de Leeuw, J. & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, 30, 1-21.
- Lance, C. E., Noble, C. L. & Scullen, S. E. (2002). A critique of the correlated trait–correlated method and correlated uniqueness models for multitrait–multimethod data. *Psychological Methods*, 7, 228–244.
- Langeheine, R. & Rost, J. (1993). Latent Class Analyse. *Psychologische Beiträge*, 35, 177-198.
- Langeheine, R. & Rost, J. (Eds.) (1988). *Latent trait and latent class models*. New York: Plenum.
- Langeheine, R., Pannekoek, J. & van de Pol, F. (1996). Bootstrapping goodness-of-fit measures in categorical data analysis. *Sociological Methods & Research*, 24, 249–264.
- Lazarsfeld, P. F. & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Lo, Y., Mendell, N. R. & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88, 767-778.
- Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis*. Mahwah: Erlbaum.
- Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks: Sage.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Mahwah: Erlbaum.
- MacKinnon, D. P., Lockwood, C. M. & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39, 99–128.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G. & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7, 83–104.
- Marsh, H. W., Wen, Z. & Hau, K. T. (2006). Structural equation models of latent interaction and quadratic effects. In G. Hancock & R. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 225–265). Greenwich, CT: Information Age.
- McArdle, J. J. (1988). *Dynamic but structural equation modeling of repeated measures data*. In R. B. Cattell & J. Nesselroade (Eds.), *Handbook of multivariate experimental psychology* (pp. 561–614). New York: Plenum Press.
- McArdle, J. J. & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analysis with incomplete longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 137–175). Washington, DC: American Psychological Association.
- McCutcheon, A. C. (1987). *Latent Class Analysis*. Beverly Hills: Sage.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525–543.
- Meredith, W. & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins & A. G. Sayer (eds.), *New methods for the analysis of change* (pp. 203–240). Washington, DC: American Psychological Association.
- Meredith, W. & Tisak, J. (1984). On “Tuckerizing” curves. Paper presented at the annual meeting of the Psychometric Society, Santa Barbara, CA.
- Meredith, W. & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107–122.

- Millsap, R. E. & Meredith, W. (2007). Factorial invariance: Historical perspectives and new problems. In R. Cudeck & R. MacCallum (Eds.), *Factor Analysis at 100*. (pp. 131-152) Mahwah, NJ: Lawrence Erlbaum.
- Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, 29, 81-117.
- Muthén, L. K. & Muthén, B. O. (2007). *Mplus 5* [Computer software]. Los Angeles, CA: Muthén & Muthén. <http://statmodel.com>
- Muthén, L. K. & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and power. *Structural Equation Modeling*, 9, 599-620.
- Muthén, L. K. & Muthén, B. O. (1998-2007). *Mplus User's Guide. Fifth Edition*. Los Angeles, CA: Muthén & Muthén. <http://statmodel.com/ugexcerpts.shtml>
- Nylund, K., Asparouhov, T. & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14, 535-569.
- Pinquart, M. (2001). Correlates of subjective health in older adults: A meta-analysis. *Psychology and Aging*, 16, 414-426.
- Quaiser-Pohl, C., Geiser, C. & Lehmann, W. (2006). The relationship between computer-game preference, gender, and mental rotation ability. *Personality and Individual Differences*, 40, 609-619.
- Raffalovich, L. E. & Bohrnstedt, G. W. (1987). Common, specific, and error variance components of factor models: Estimation with longitudinal data. *Sociological Methods & Research*, 15, 385-405.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks: Sage.
- Raykov, T. & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd Ed.). Mahwah: Erlbaum.
- Reinecke, J. (2005). *Strukturgleichungsmodelle in den Sozialwissenschaften*. München: Oldenbourg.
- Reuter, T., Ziegelmann, J. P., Wiedemann, A. U., Geiser, C., Lippke, S., Schüz, B. & Schwarzer, R. (2010). Dynamics in health behavior change: An application of latent true change modeling. *Journal of Health Psychology*. doi: 10.1177/1359105309360071
- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion* (2. Aufl.). Bern: Huber.
- Rost, J. (2006). Latent-Class-Analyse. In F. Petermann & M. Eid (Hrsg.), *Handbuch der Psychologischen Diagnostik* (S. 275-287). Göttingen: Hogrefe.
- Rost, J. & Langeheine, R. (Eds.) (1997). *Applications of latent trait and latent class models in the social sciences*. Münster: Waxmann.
- Sayer, A. G. & Cumsille, P. E. (2001). Second-order latent growth models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 177-200). Washington, D. C.: APA.
- Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147-177.
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research - Online*, 8, 23-74. <http://www.dgps.de/fachgruppen/methoden/mpr-online>
- Schumacker, R. E. & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah: Erlbaum.
- Singer, J. D. & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
- Snijders, T. A. B. & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological Methodology 1982* (pp. 290-312). Washington DC: American Sociological Association.
- Sörbom, D. (1975). Detection of correlated errors in longitudinal data. *British Journal of Mathematical and Statistical Psychology*, 28, 138-151.

- Steyer, R. (1988). *Experiment, Regression und Kausalität. Die logische Struktur kausaler Regressionsmodelle.* Unveröffentlichte Habilitationsschrift, Universität Trier.
- Steyer, R., Eid, M. & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research - Online*, 2, 21–33. <http://www.dgps.de/fachgruppen/methoden/mpr-online>
- Steyer, R., Ferring, D. & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment*, 8, 79–98.
- Steyer, R., Partchev, I. & Shanahan, M. (2000). Modeling true intra-individual change in structural equation models: The case of poverty and children's psychosocial adjustment. In T. D. Little, K. U. Schnabel & J. Baumert (Eds.), *Modeling longitudinal and multiple-group data: Practical issues, applied approaches, and specific examples* (pp. 109–126). Hillsdale, NJ: Erlbaum.
- Steyer, R., Schmitt, M. J. & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, 13, 389–408.
- Uebersax, J. (2000). *A brief study of local maximum solutions in latent class analysis.* <http://www.john-uebersax.com/stat/local.htm>
- von Davier, M. (1997). Bootstrapping goodness-of-fit statistics for sparse categorical data: Results of a Monte Carlo study. *Methods of Psychological Research-Online*, 2, 29–48.
- Ware, J. E. & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care*, 30, 473–483.
- Whitelaw, N. A. & Liang, J. (1991). The structure of the OARS physical health measures. *Medical Care*, 29, 332–347.
- Widaman, K. F. & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Register

- aberrant response patterns 261
Alpha-Fehler 200, 208
Antwortmuster, Pattern 31, 235ff., 260ff., 270
ASCII-Daten 11, 21, 23ff., 27, 258
Autoregressive Modelle 8, 93, 95f., 131, 132f., 138ff.
Autoregressive Struktur 95, 139
Autokorrelierte Fehlervariablen 99f., 103
- Baseline-Change-Version 153, 163
Baselinemodell 59f.
Basic-Analyse 25ff., 29, 45, 48, 203ff., 209ff., 275, 278
Benutzerdefiniert fehlend 16
Bias-korrigierter Bootstrap 77, 84f.
Bivariate Regressionsmodell 50
Bootstrap-Likelihood-Ratio-Differenzentest 261, 265
Bootstrap-Methode 77, 83, 261
Bootstrap-Stichproben 84, 265f.
Boundary Estimate 241, 243, 270f.
- CFI, Comparative fit index 51, 59ff., 68f., 80, 88, 98f., 101, 103, 111, 117, 120f., 127, 134f., 142, 144, 158, 174f., 192
Chi-Quadrat-Test, Chi-Quadrat-Differenzentest, χ^2 -Differenzentests, χ^2 -Differenzwert 50, 60, 111f., 117, 121
Class proportions 237
Clusterstichprobe 199
Clusterstruktur 199, 203, 205, 208, 213, 223, 226f.
Clusterung 199, 200, 223
Conditional response probabilities 237
Correlated-uniqueness-Modell 99
Covariance coverage 32
Cross-Level-Interactions 200
Curve-of-factors model 188
- Data-Befehl 27, 33
Datenlesefehler 13, 16
- Deskriptivstatistiken 25, 28, 32f., 37ff., 48f., 206, 275, 277
Determinationskoeffizienten 54
Dezimaltrennzeichen 12ff., 21, 24, 37, 279
Direkte Effekte 76f., 79
Diskriminante Validität 66, 70
Dummyvariable 201
- Einfache Regressionsanalyse 44
Einlesen von Rohdaten 25
Einlesen von Summary-Daten 11, 37
Empirische Prüfgröße 52
Entropy-Maß 249
Erwartungswert 98
Estimated sample statistics 34, 207
- Faktorenanalyse 39, 44, 65, 75, 85, 233
Faktoren 7, 8, 41, 43f., 46, 55f., 63, 65ff., 70ff., 85, 87f., 95ff., 101ff., 108, 116, 120, 122f., 125, 127f., 140, 142ff., 157, 170, 188ff., 194, 235
Faktorkorrelationen 58, 66, 70, 104
Faktorkovarianzen 44, 68
Faktorladung 41f., 44, 56, 58, 63, 65, 70, 86, 94ff., 102f., 108f., 111, 116, 123f., 128, 143, 155, 171, 173, 184, 188f., 276
Faktormetrik 58, 68
Faktormittelwert 44, 125
Faktorvarianz 44, 58, 68f., 104, 108, 122
Faktorielle Invarianz 108, 125, 152, 188
Fehlende Werte 12, 15f., 273, 277ff.
Fehlermeldung 24, 29, 48, 50, 218, 278f.
Fehlervariable 42f., 53, 96, 99ff., 107, 188
Fehlervarianz 44, 64, 92, 99, 109, 114, 121, 158, 163
Fitindices 50, 59
Fixed effect 209, 217
Fixieren 58, 110, 116, 158, 174
Fixierung 58, 74, 109, 117, 276
Formative Messmodelle 41
Freiheitsgrade, Degrees of Freedom 51, 59ff., 79f., 88, 95, 98, 111, 127, 248, 261

- Full Information Maximum Likelihood Schätzung 16
- Gerichtete Zusammenhänge 41, 42
- Gesamtmittelwert 53, 206, 209f., 213f., 232, 274
- Gleichheitsrestriktionen 108f., 112, 117, 153
- Grand Mean 53, 203, 209ff., 213f., 225, 230, 232
- Group Mean centering 214
- Heterogenität der Indikatoren 103
- Heywood Case 72ff.
- Hierarchische lineare Modelle 199
- Hilfsvariable 240, 274
- homogene Subgruppen 235
- Homogenität der Indikatoren 96
- Incremental Fit Indices 60f.
- Indikatorspezifische Effekte 96f., 99, 101, 103ff., 142, 191
- Indikatorspezifische Varianz 65f., 91f.
- Indirekte Effekte 76
- Individualdaten, individual data 5, 11, 25, 37f., 170, 179f., 205, 273
- Informationstheoretische Maße, AIC, BIC 51, 61, 68f., 98, 103, 111f., 117, 120ff., 127, 135, 144, 175, 248, 265, 269f.
- Interaktion 77, 199f., 228
- Intercept-Faktor 170f., 173ff.
- Intercepts, Achsenabschnitte 44, 52, 54, 62ff., 69, 71, 81, 89f., 98, 104f., 107ff., 112, 114ff., 121, 125, 127, 129ff., 136f., 146, 149, 152, 155, 158f., 161, 163, 165, 167, 171, 173f., 176f., 187f., 190f., 193, 196, 212f., 219, 222, 224, 227ff.
- Intercepts-and-Slopes-as-Outcomes-Modelle 222
- Intraklassenkorrelation 204, 208, 212
- Iteration 241, 246
- Klassengröße 200, 236, 248f., 251, 256
- Klassenprofildiagramm 244, 255ff.
- Klassenzugehörigkeit 236, 243, 259
- Klassenzuordnungswahrscheinlichkeit 238, 240, 243, 249f.0, 257, 259, 270, 277
- Klassifikation 236, 249f., 270
- Kodierschema 218
- Kommentarzeile 26, 29, 30
- Konfidenzintervall 61, 69, 77, 85
- Konfidenzintervalle 77, 83f., 92, 200, 224, 277 - asymmetrisch 77, 83
- Konfidenzintervallschätzungen 200
- Konfigurale Invarianz 108
- Konfirmatorische Faktorenanalyse 5, 65, 71, 233
- Konsistenz, consistency 94
- Konstante 53, 76, 86, 209, 213, 233
- Konvergenzkriterium 241
- Konvergenzprobleme 50, 278
- Korrelationsmatrix 25, 37f., 67, 94, 273
- Kovarianzmatrix 25, 32, 38, 53, 60, 185, 273
- Kreuzverzögerte Effekte, cross-lagged effects 132f., 135, 138ff., 143
- Kriteriumsvariable 45, 213, 228
- Kurvenlinear 168, 184
- Ländereinstellungen 5, 12
- Längsschnittanalyse 60, 93ff., 99, 150
- Längsschnittstrukturgleichungsmodelle 93
- Latent-Change-Modelle 93, 96, 151ff.
- Latent-Class-Analyse, LCA, Latent-Structure-Analysis 8f., 235ff., 246, 249ff., 256, 260, 265, 269ff., 277, 286
- Latent-Difference-Modelle 151
- Latente autoregressive Modelle 132
- Latente Differenzvariable 151f., 162
- Latente Mittelwertsunterschiede 120
- Latente Mittelwertsvergleiche 108
- Latente Moderatormodelle 77
- Latente Regressionsanalyse 42
- Latente Veränderung 162
- Latente Wachstumskurven 8, 93, 168f., 187
- latente Wachstumskurvenmodelle 1. Ordnung 169
- latente Wachstumskurvenmodelle 2. Ordnung 96, 187
- Latent-State-Trait-Modelle, LST 93ff., 122ff., 131, 188
- Likelihood-Ratio 260f., 265, 267ff.
- Likert-Skala 235
- Lineare Strukturgleichungsmodelle 7f., 25, 41, 236
- Linearer Slope-Faktor, Slopefaktor 170f., 184, 191
- Liniendiagramm 242, 254, 256f.
- Listenweiser Fallausschluss, listwise deletion 16, 33, 37
- Local solution 240
- Logistische Regression, Probitregression 48
- Logit-Parametrisierung 237, 243, 251
- Loglikelihood-Wert 240f., 245ff., 262, 266ff.
- Lokales Likelihood-Maximum 241
- Lösungsstrategie 9, 235f.

- Lösungswahrscheinlichkeit 237
- Macro-Ebene 199
- Manifeste Modelle 65, 79, 88f., 91f., 140, 150f.
- Maximum-Likelihood-Schätzung 58, 60, 211
- Means-as-Outcomes-Modelle 212
- Mediatoreffekte 9, 41, 76f.
- Mediatorvariablen 41, 76, 168
- Mehrebenenanalyse, Mehrebenenmodelle, Multilevelmodelle, Hierarchische Lineare Modelle, Random Coefficient Models, Mixed Models 199ff., 203, 208, 213, 233, 274ff.
- Mehrebenenregressionsanalyse 53, 200, 201
- Mehrebenenregressionsmodelle 8, 199, 222
- Mehrstufig-kategoriale Variablen 236
- Messfehler 7, 42, 55, 65, 86, 91, 93, 138f., 170
- Messfehlereinflüsse 7, 65, 85, 93, 96, 122, 138
- Messgenauigkeit 106, 122
- Messinvarianz, Messkonstanz, Messäquivalenz 107ff., 122, 131, 139f., 150ff., 155, 158, 163, 187ff., 191
- Messmodell, measurement model 41, 56, 74, 85, 87f., 92, 95, 122, 140f., 156, 163, 188
- Metrik 56, 58, 88, 109
- Micro-Ebene 199
- Mischverteilungsanalyse 240, 275
- Missing data patterns, Muster fehlender Werte 31
- Missing-Value-Code 16, 18, 20f., 28, 274
- Mittelwert 25, 32, 37f., 53, 60f., 79, 94, 98, 101, 116f., 120, 125, 127, 131, 133, 152, 158, 162f., 168, 171ff., 192, 206, 209, 213, 218, 220, 273
- Mittelwertsstruktur 98, 104, 124f., 127, 131, 158, 191, 277
- Modellanpassung 50, 60, 109, 112, 184, 270
- Modellfit 59, 60ff., 72, 99, 112, 120, 142, 163, 249, 260f., 263, 265
- absoluter 260
- Modellgütebeurteilung 61, 260
- Modellimplizierte Wachstumskurve 179, 183, 197
- Modellmisspezifikation 74
- Modelltestung 42
- Modellvergleich 109, 236f., 260f., 269f.
- Moderatoreffekt 77
- Moderierte Regression 77
- Modifikationsindices 61f., 263
- Monte-Carlo-Simulationsstudie 261
- Mplus Language Generator 26
- Mplus-Input-Datei 26f., 39, 45f., 67f., 73, 78, 87, 97, 133, 141, 172, 175, 185, 190, 204, 210, 215, 219, 225, 230, 239
- Multigruppen-LCA 238
- Multiple Indikatoren 55, 65, 122, 188
- Multitrait-Multimethod 43, 99
- Negative Varianz 50, 74
- Neighbor-Change-Version 153, 162
- Nested models 60f., 111
- Nichtlinear 184
- Nullmodell, Intercept-Only-Modell, Unconditional Cell Means Model, Unconstrained Model, Empty Model, Unabhängigkeitsmodell 38f., 50, 60, 175, 181, 203, 208ff., 216f., 220, 226f.
- Nullpunkt 53, 213, 232
- Occasion-specificity, Situationsspezifität 94
- Optseed 247, 262
- Ordinary-Least-Squares, OLS 199, 204, 209, 213f., 223, 226, 231f.
- Outlier 261, 263, 270
- output: residual 277
- Paarweiser Fallausschluss 33
- Parameterrestriktionen 236
- Parameterschätzungen 51, 56, 58, 62, 68, 72, 74, 81, 84, 88, 103, 113, 163, 186, 192, 194, 211, 226, 230, 237, 241, 251, 257, 277
- Parametrische Bootstrap-Methode 261
- Pearson-X²-Statistik 260, 263, 270
- Personenidentifikationsvariable 243, 259
- Pfadanalyse 8, 9, 39, 41, 42, 44, 75ff., 85ff., 131f., 134, 233, 276
- Pfaddiagramm 8, 43, 45, 55, 134
- Pfadkoeffizient 44, 56, 76, 80, 86, 131f., 134f., 137ff., 142f., 150, 157
- Plot 174, 178, 181, 197, 242f., 254, 277
- Posterior probabilities 249
- Pseudomessmodelle 156
- p-Wert 52, 60f., 63, 80, 111, 162, 200, 211, 224, 260f., 265, 269, 267
- Random Coefficient 173, 225, 233
- Random Effect 209
- Random Intercept 173, 203, 212, 214, 218, 225f.
- Random-Coefficient-Regressionsmodelle 222
- Random-Effects-ANCOVA 212, 215, 223f., 227

- Random-Intercept-and-Slope-Modell 212, 222
 Random-Intercept-Modell 208f., 212, 217
 Referenzindikator 101, 103f., 191
 Reflektives Messmodell 41
 Regressionsgerade 223f., 228, 231, 233
 Regressionskoeffizient 7, 41f., 44, 52, 54, 63, 77, 83, 199, 208, 216f., 221, 224, 231
 Regressionsmodell 39, 41, 44f., 47, 50, 55f., 62, 75, 77, 87, 203, 209, 214, 223f., 227, 287
 Relativer Modellfit 265
 Reliabilität 55, 65, 88, 93, 99, 101, 106, 122, 150, 188, 194, 236, 250
 Residualfaktor 101f., 104, 191
 Residualkorrelation 134
 Residualkovarianz 44, 229f., 233
 Residuen 61f., 98, 134, 151, 261, 270
 Residuenstatistik 261, 263, 277
 Restriktionen 60f., 72, 79f., 88, 95, 98, 104, 108ff., 112, 114, 117, 120, 127, 134
 Root-Mean-Square-Error-of-Approximation, RMSEA 51, 59, 61, 68f., 80, 88, 98f., 101, 103, 111, 117, 120f., 127, 134f., 142, 144, 158, 174f., 192
 Rohdatensätze 11
 Sampstat 38f., 48, 277
 Saturiertes Modell 50
 SAVE TRANSLATE 16, 21
 Schwellenparameter, Thresholds 237, 251
 Second order growth curve models 188
 Seed 241, 245, 247, 262
 Semikolon 28f., 278
 Signifikanztestung 52, 77, 83
 Sparse data, sparse tables 260
 Spezifische Varianzanteile 65, 91
 SPSS 11f., 14ff., 25ff., 33, 37, 201, 258, 279
 Standardized-Root-Mean-Square-Residual, SRMR 51, 59, 61, 68f., 80, 88, 98f., 101, 103, 111, 117, 120f., 127, 134f., 142, 144, 158, 174f., 192
 Standardabweichung 25, 37f., 54, 94, 133, 220, 273
 Standardfehler 52, 63, 83, 113, 117, 199, 208, 211, 224, 227, 251
 Standardfehlerbias 200
 Standardisierte Residuen 61, 262
 Standardisierter Intercept-Term 54
 Standardisiertes Regressionsgewicht 54, 63
 Startwertesets 241, 245ff., 262, 266, 268
 Statistische Äquivalenz 72
 Steigungskoeffizient, Slope 45, 52, 168, 170ff., 184, 189, 192, 194, 203, 212ff., 217f., 220, 222ff., 276
 Strukturmodell, structural model, latent variable model 42, 55, 58, 65, 72, 87f., 95, 127, 140f., 150, 163
 Summary-Daten 25, 37ff., 133, 140, 174, 273
 Systemdefiniert fehlende Daten 12, 16, 18
 Tabulator-getrennte Formate 11, 15, 23
 Textoptions decimal = dot 16
 Title-Befehl 26
 Tucker-Lewis-Index, TLI 51, 59, 61, 68f., 88, 98, 127
 Totaler Effekt 77, 82
 Totaler indirekter Effekt 82f.
 Trait-Veränderungen, Trait-Veränderungsmodelle 93, 131
 True-Change-Modelle 151
 Twolevel basic 203ff., 209ff.
 Typologie 236
 Überfaktorisierung 186, 197
 Ungerichtete Zusammenhänge 41ff., 65, 85
 Unskalierbare Personen 263
 Unzulässige Schätzergebnisse 50, 74
 Usevar 47f., 56, 67, 205, 274, 278f.
 Variabilität, Variabilitätsmodelle 93, 131, 173, 176, 210
 Variable-Befehl 27, 28, 46, 48, 53, 278f.
 Variablen
 - abhängige 48, 201, 208, 217, 227
 - unabhängige 41, 43, 53, 55, 76, 116, 168, 173, 178, 213f., 275
 - beobachtete, manifeste 43f., 123f., 138, 154, 189
 - Clustervariable 201f., 205, 274, 275
 - dichotome 218, 236
 - endogene 41f.
 - exogene 41, 63
 - intervenierende 41
 - latente 41f., 44, 49, 55, 63, 85f., 88, 91f., 98, 107ff., 116, 125, 127f., 139, 151f., 162, 173, 185, 225f., 235
 Variablennamen 23, 27ff., 38, 43, 76, 79, 86, 95, 116, 123f., 132, 135, 139, 143, 153f., 170, 189, 239, 273f., 278f.
 Varianzkomponenten 94, 96, 131

- Veränderungsmessung 5, 42, 93, 131, 283
Vollstandardisierte Lösung, standardized stdyx
48, 63, 70, 215, 218, 277
Vuong-Lo-Mendell-Rubin-Test, VLMR-Tests 265,
267, 269f.
Wachstumsfaktor, Growth-Faktor 169, 171, 173,
176ff., 185, 188, 191, 276
Wachstumskurve 96, 131, 168ff., 173, 184, 189,
191, 233
Wachstumskurvenmodellen 1. Ordnung 169
Warnmeldung 30, 73, 185, 211
Zentrierung, Centering 53, 213f., 225, 230, 274,
285
Zweiseitiger Test 162
z-Wert 52, 61, 103, 127, 162, 262f.