MA-HPSTS-7: Einführung in die Analyse linearer Strukturgleichungsmodelle

Teil 1: Lineare Strukturgleichungsmodelle mit messbaren Variablen – Pfadmodelle (Grundlagen)

Teil 2: Lineare Strukturgleichungsmodelle mit latenten Variablen

Korrelation und Kausalität

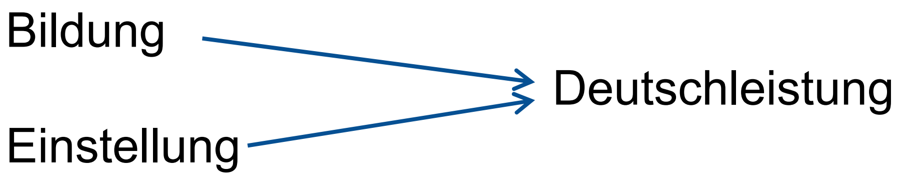
Korrelationen

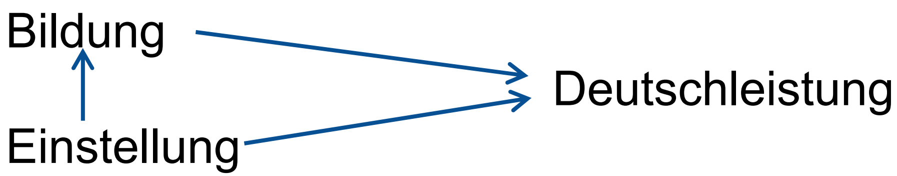
		Bildungsniveau der Eltern	Einstellung der Eltern zur Schule	Ergebnis Deutsch 4. Klasse	Ergebnis Deutsch 10. Klasse
Bildungsniveau der Eltern	Korrelation nach Pearson	1	,529**	,788**	,691**
	Signifikanz (2-seitig)		,000	,000	,000
	N	120	120	120	120
Einstellung der Eltern zur Schule	Korrelation nach Pearson	,529**	1	,764**	,688**
	Signifikanz (2-seitig)	,000		,000	,000
	N	120	120	120	120
Ergebnis Deutsch 4. Klasse	Korrelation nach Pearson	,788**	,764**	1	,904**
	Signifikanz (2-seitig)	,000	,000		,000
	N	120	120	120	120
Ergebnis Deutsch 10. Klasse	Korrelation nach Pearson	,691**	,688**	,904**	1
	Signifikanz (2-seitig)	,000	,000	,000	
	N	120	120	120	120

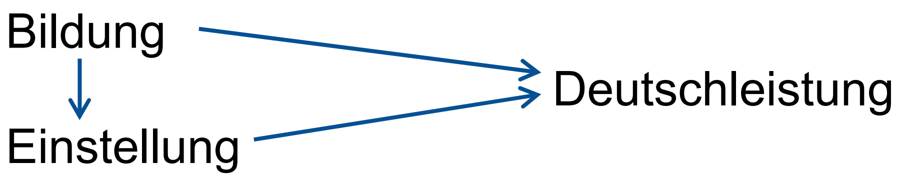
**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Korrelation und Kausalität

Auswahl an alternativen Erklärungsmöglichkeiten für die Korrelationen zwischen Bildung, Einstellung und Deutsch 4. Klasse:



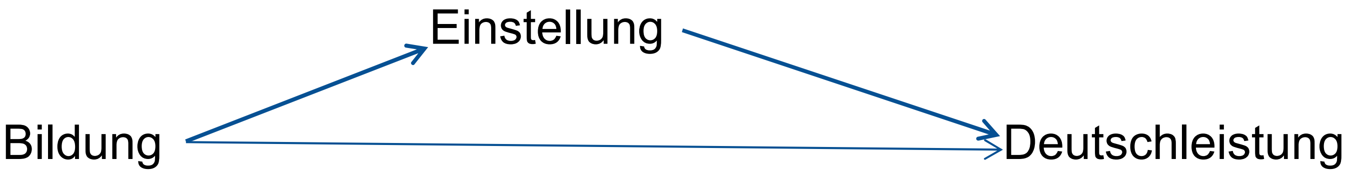




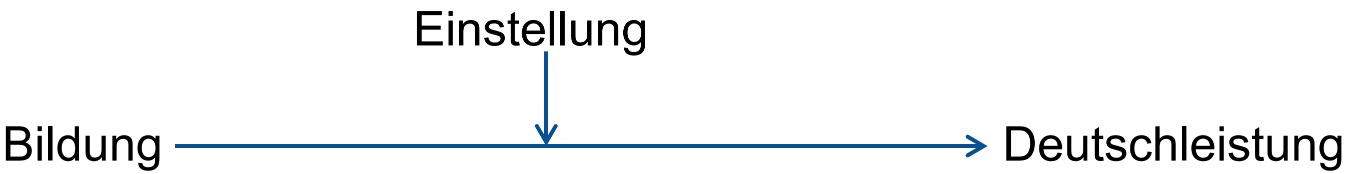
Korrelation und Kausalität

Auswahl an alternativen Erklärungsmöglichkeiten für die Korrelationen zwischen Bildung, Einstellung und Deutsch 4. Klasse:

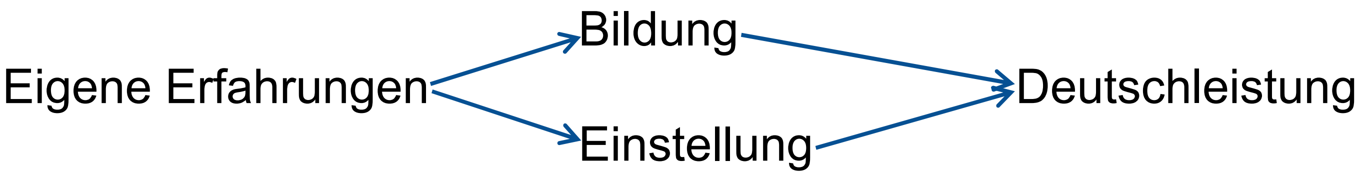
Mediation:



Moderation:



Drittvariablen:



Korrelation und Kausalität

⇨ sehr unterschiedliche Modelle zur Erklärung der Korrelationen möglich

⇨ Bewertung unterschiedlicher Modelle in einfachen Fällen über partielle Korrelationen, semipartielle Korrelationen u.ä. möglich

⇨ Bewertung unterschiedlicher Modelle in komplexeren Situationen: Pfadmodelle bzw. lineare Strukturgleichungsmodelle

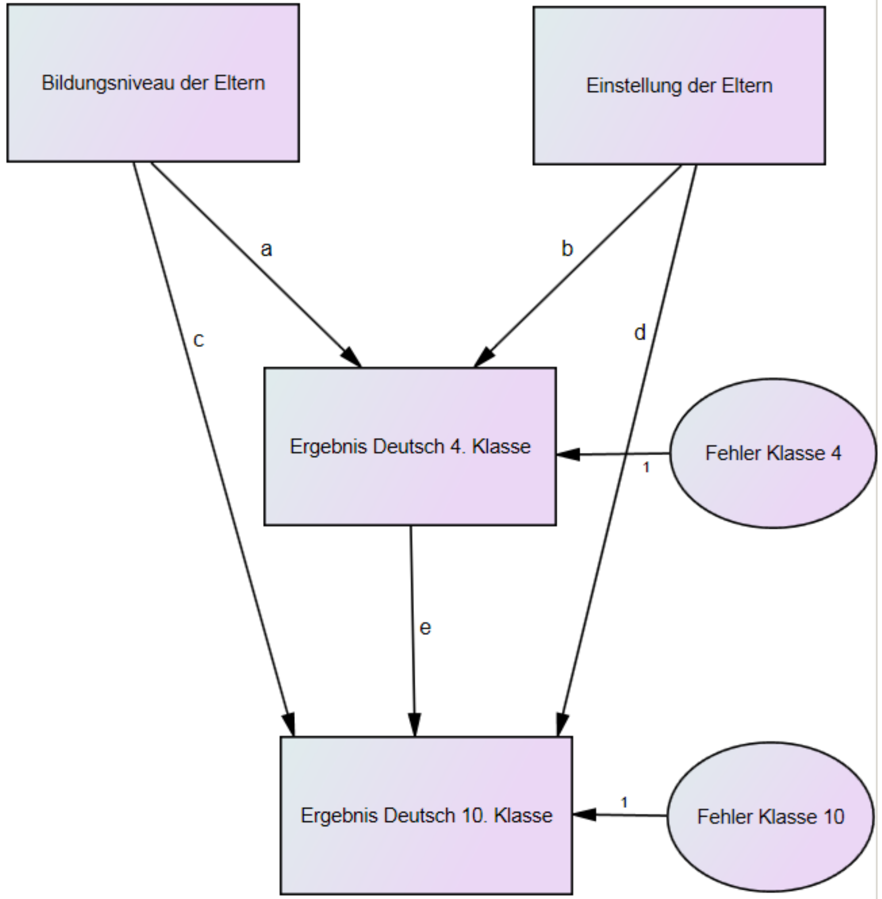
⇨ i.d.R. keine eindeutigen Ergebnisse aus der Datenanalyse möglich

⇨ Inhaltliche Modellbildung von entscheidender Bedeutung

Pfadmodelle: Grundideen

Grundlage: **Inhaltliche Hypothesen**

Umsetzung in Grafiken…



… und in Gleichungen
(Variablen verkürzt bezeichnet)

Deutsch 4 = a · Bildung + b · Einstellung + Fehler 4
Deutsch 10 = c · Bildung + d · Einstellung + e · Deutsch 4 + Fehler 10

Pfadmodelle: Grundideen

Einige wichtige zusätzliche Möglichkeiten von Pfadmodellen gegenüber multipler linearer Regressionsanalyse:

- Modellfehler werden explizit modelliert
- Kovarianzen (Korrelationen) zwischen Prädiktoren sind zugelassen und können modelliert werden (siehe folgende Folie)
- „Pfade“ können abgebildet werden, d.h. Variablen können sowohl als Prädiktoren als auch als Kriterien im Modell wirken
- Erweiterung auf Strukturgleichungsmodelle mit latenten Variablen möglich

Pfadmodelle: Pfadkoeffizienten

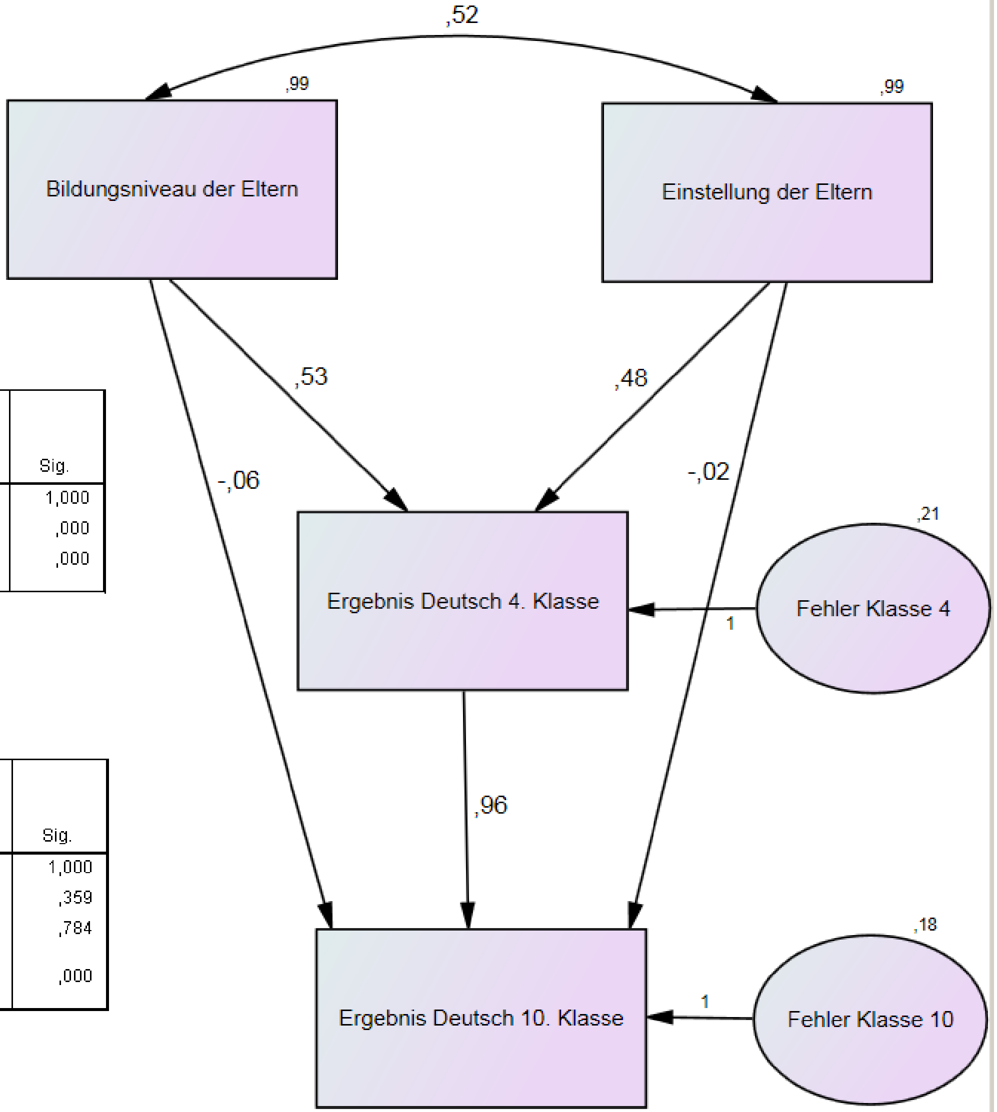
Berechnung der Pfadkoeffizienten im Ergebnis multipler Regressionsanalysen (Variablen liegen z-standardisiert vor):

Multiple Regression 1: AV: Deutsch 4. Klasse

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.
		Regressionskoeffizient B	Standardfehler	Beta		
1	(Konstante)	7,828E-17	,042		,000	1,000
	Bildungsniveau der Eltern	,533	,050	,533	10,625	,000
	Einstellung der Eltern zur Schule	,482	,050	,482	9,625	,000

Multiple Regression 2: AV: Deutsch 10. Klasse

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.
		Regressionskoeffizient B	Standardfehler	Beta		
1	(Konstante)	-5,905E-17	,039		,000	1,000
	Bildungsniveau der Eltern	-,060	,065	-,060	-,922	,359
	Einstellung der Eltern zur Schule	-,017	,062	-,017	-,275	,784
	Ergebnis Deutsch 4. Klasse	,965	,086	,965	11,251	,000



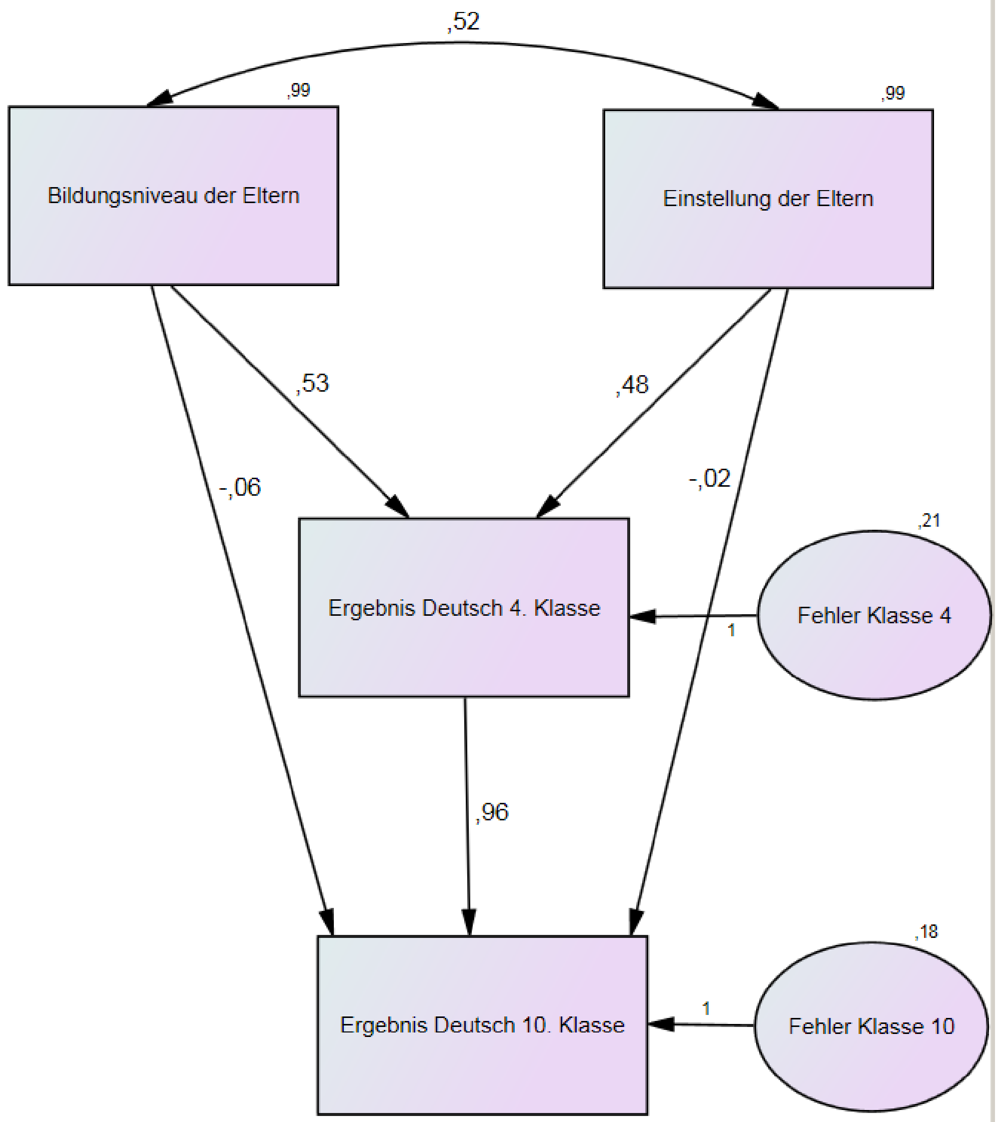
Pfadmodelle: Typen von Effekten

Direkte kausale Effekte
Zum Beispiel direkter kausaler Effekt von Bildung auf Deutsch 10: -0.06

Indirekte kausale Effekte
Zum Beispiel indirekter kausaler Effekt von Bildung auf Deutsch 10 (Mediator: Deutsch 4): $0.533 \cdot 0.965 = 0.514$

Totale kausale Effekte
Zum Beispiel totaler Effekt von Bildung auf Deutsch 10:
$-0.06 + 0.533 \cdot 0.965 = 0.454$

Indirekte korrelative Effekte
Zum Beispiel indirekter korrelativer Effekt von Bildung auf Deutsch 10:
$0.529 \cdot 0.482 \cdot 0.965 + 0.529 \cdot -0.017 = 0.237$



Pfadmodelle: Korrelationszerlegung

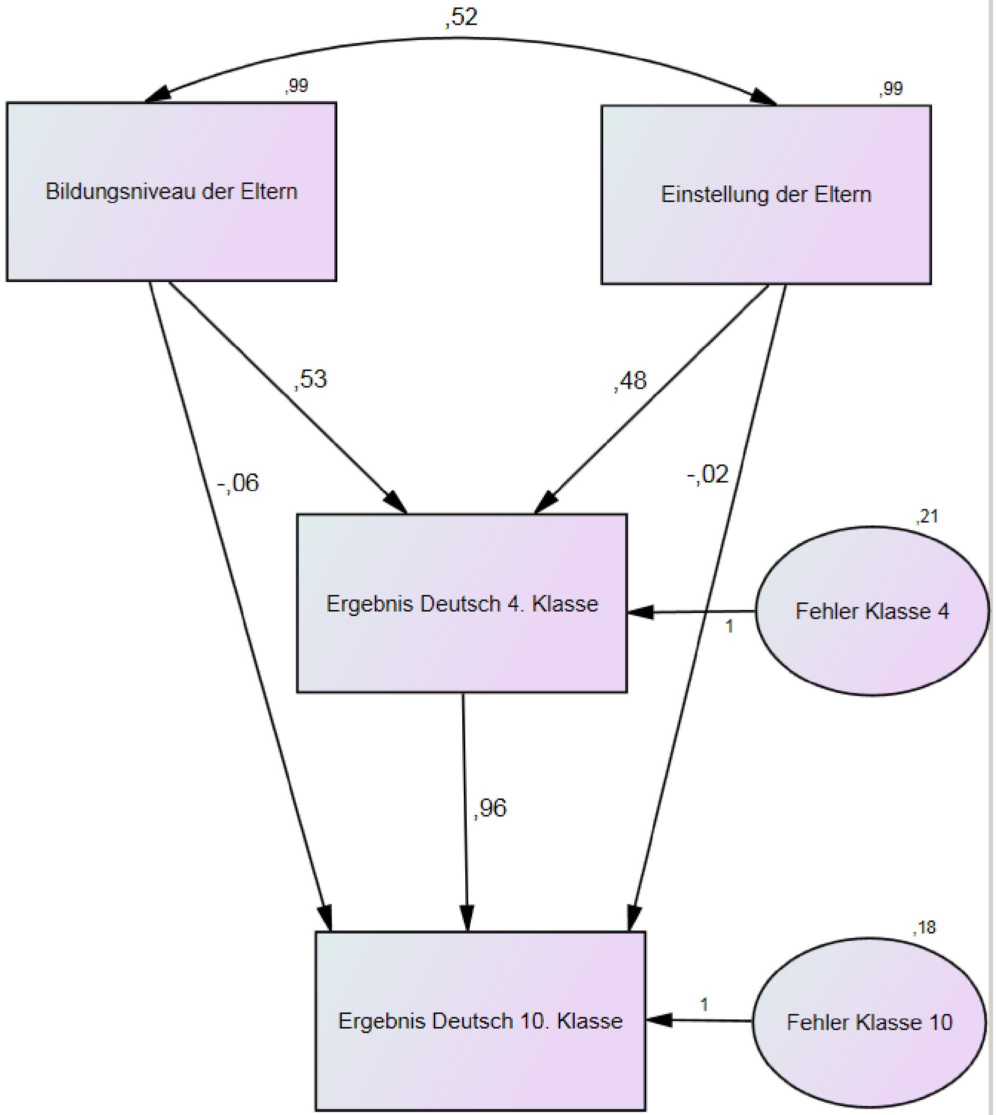
Fundamentaltheorem der Pfadanalyse:

Korrelationen zwischen unabhängigen (exogenen) und abhängigen (andogenen) Variablen ergeben sich als Summe der direkten und indirekten kausalen und der indirekten korrelativen Effekte.

Beispiel: $r_{\text{Bildung, Deutsch 10}} = 0.691$

Direkter kausaler Effekt: -0.06
Indirekter kausaler Effekt: 0.514
Indirekter korrelativer Effekt: 0.237

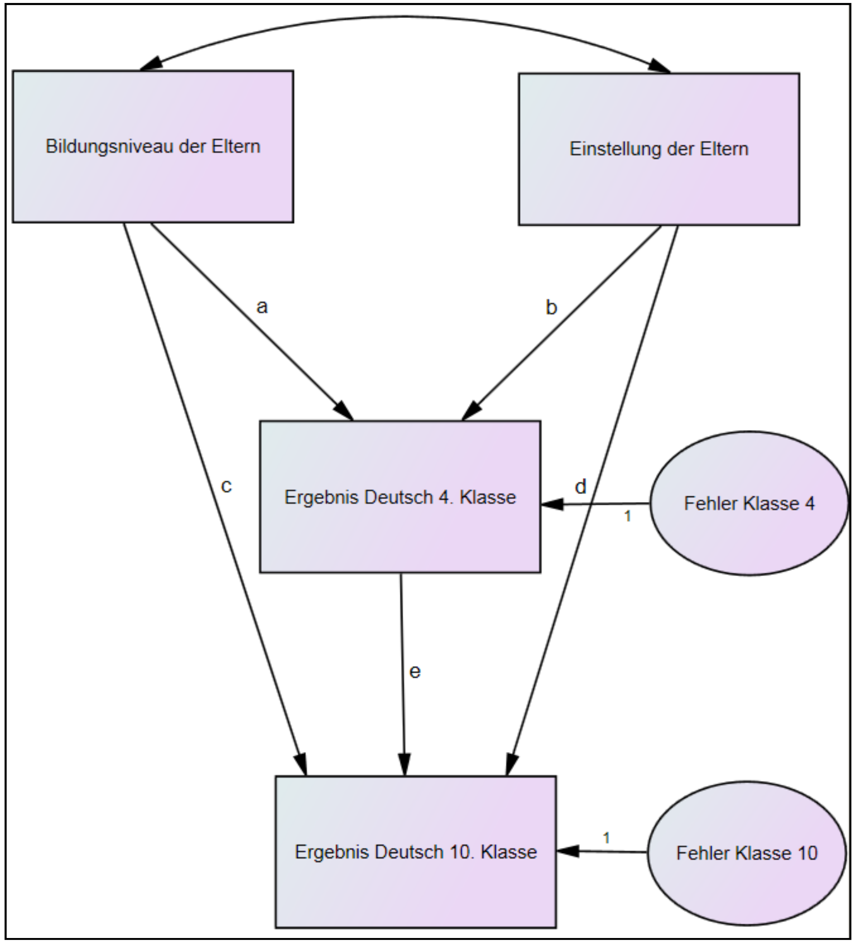
Summe der Effekte: 0.691

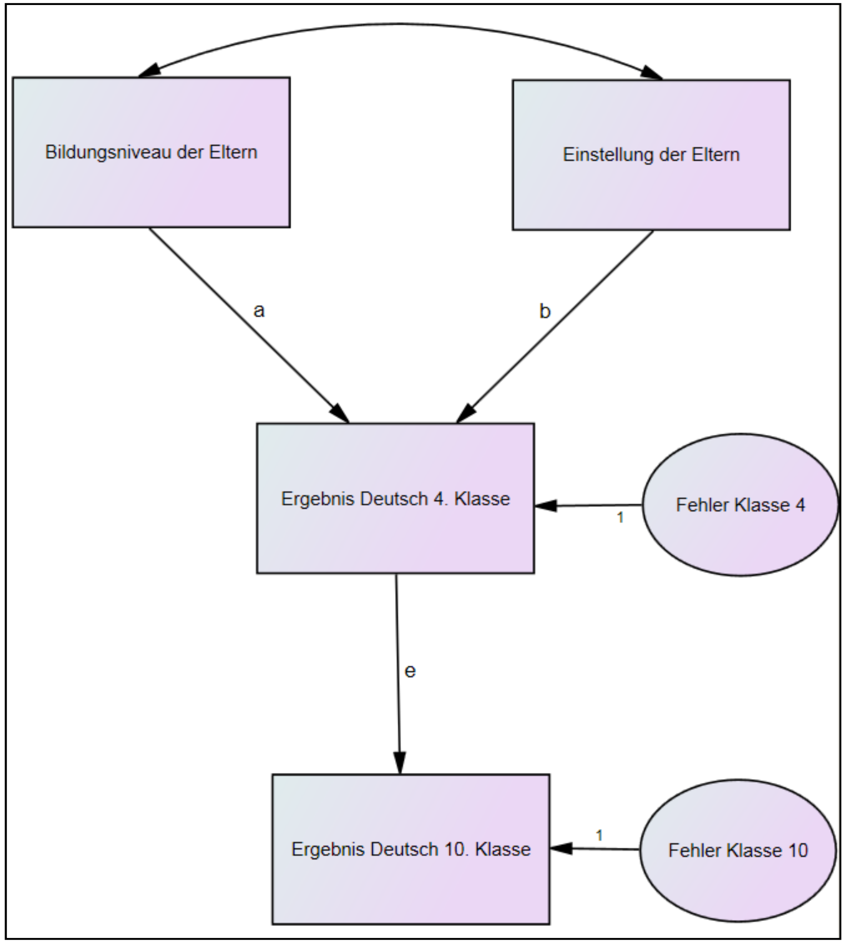


Prinzip der Parameterschätzung

Ziel pfadanalytischer Untersuchungen: Inhaltlich aussagekräftige, „sparsame“ Modelle

Modell A: Pfade zwischen allen Variablen   Modell B: Pfade entsprechend Hypothese

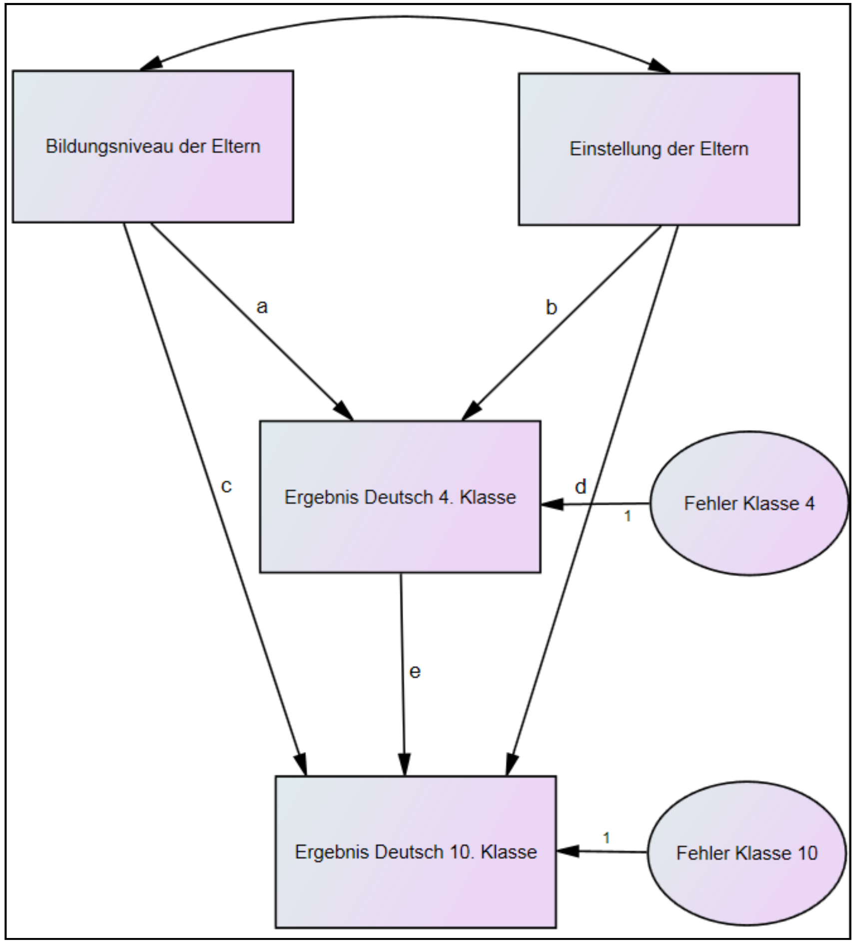


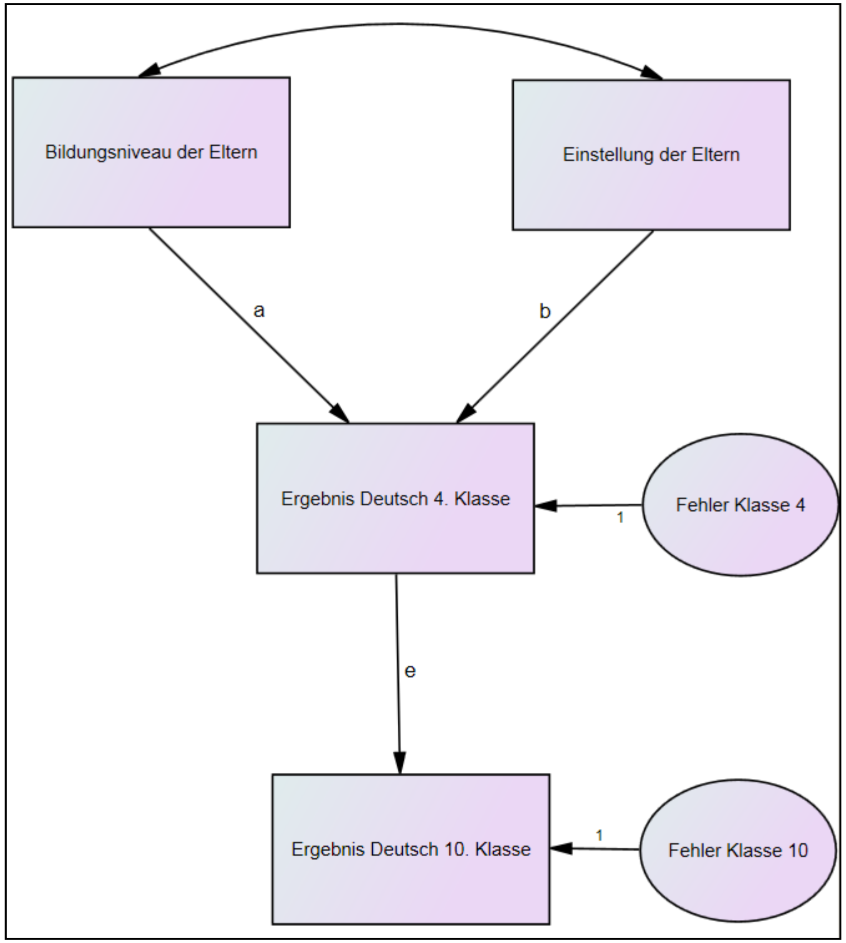


Prinzip der Parameterschätzung

Modell A: Korrelationen zwischen Variablen können vollständig repliziert werden!

Modell B: Wie gut können Korrelationen zwischen Variablen repliziert werden?





Prinzip der Parameterschätzung

Beispiel: Wie gut könnte in Modell B die Korrelation zwischen Bildung und Deutsch 10 repliziert werden, wenn (theoretisch) die in Modell A berechneten Pfadkoeffizienten benutzt werden würden?

Hinweis: Die korrekte Schätzung der Koeffizienten auf der Basis von Modell B wird im praktischen Teil dieses Kurses durchgeführt.

Direkter kausaler Effekt von Bildung auf Deutsch 10: 0

Indirekter kausaler Effekt von Bildung auf Deutsch 10 (Mediator: Deutsch 4):
$0.533 \cdot 0.965 = 0.514$

Indirekter korrelativer Effekt von Bildung auf Deutsch 10:
$0.529 \cdot 0.482 \cdot 0.965 = 0.246$

$r_{\text{Bildung, Deutsch 10}} = 0.691$

Direkter kausaler Effekt: 0
Indirekter kausaler Effekt: 0.514
Indirekter korrelativer Effekt: 0.246

Summe der Effekte: 0.76 → Korrelation würde bei überschätzt

Prinzip der Parameterschätzung

Allgemeines Ziel der Parameterschätzung:

Parameter so schätzen, dass durch das Modell und die für das Modell geschätzten Parameter die Beziehungen (Korrelationen) der Variablen möglichst gut repliziert werden können.

Wichtiger Hinweis: Betrachtungen wurde hier für z-standardisierte Variablen und entsprechend für Korrelationsmatrizen vorgenommen.

Im allgemeinen Fall nicht standardisierter Variablen betrifft das Prinzip analog die Varianzen und Kovarianzen der Variablen, d.h. die Varianz-/ Kovarianzmatrix.

Prinzip der Parameterschätzung

Maximum-Likelihood-Methode ML:
Ziel in einem iterativen Verfahren: Bestmögliche Anpassung der aus den geschätzten Parametern berechnete Varianz-/Kovarianzmatrix an die empirischen Varianz-/Kovarianzmatrix; damit Maximierung der Wahrscheinlichkeit, dass die empirische Matrix auf der Grundlage der Modellparameter zustande gekommen ist.
Voraussetzung: mehrdimensionale Normalverteilung (n>100).

Generalized-Least-Squares-Verfahren GLS (mit NV)

Unweighted-Least-Squares-Verfahren ULS (ohne NV)

Asymptotically Distribution-Free ADF (ohne NV, n>500)

(Empfehlungen nach Bühner, 2006; Näheres dazu später)

Prinzip der Parameterschätzung

Beispiel: Daten

Datei Pfadmodell_Beispiel1.sav

Deskriptive Statistik

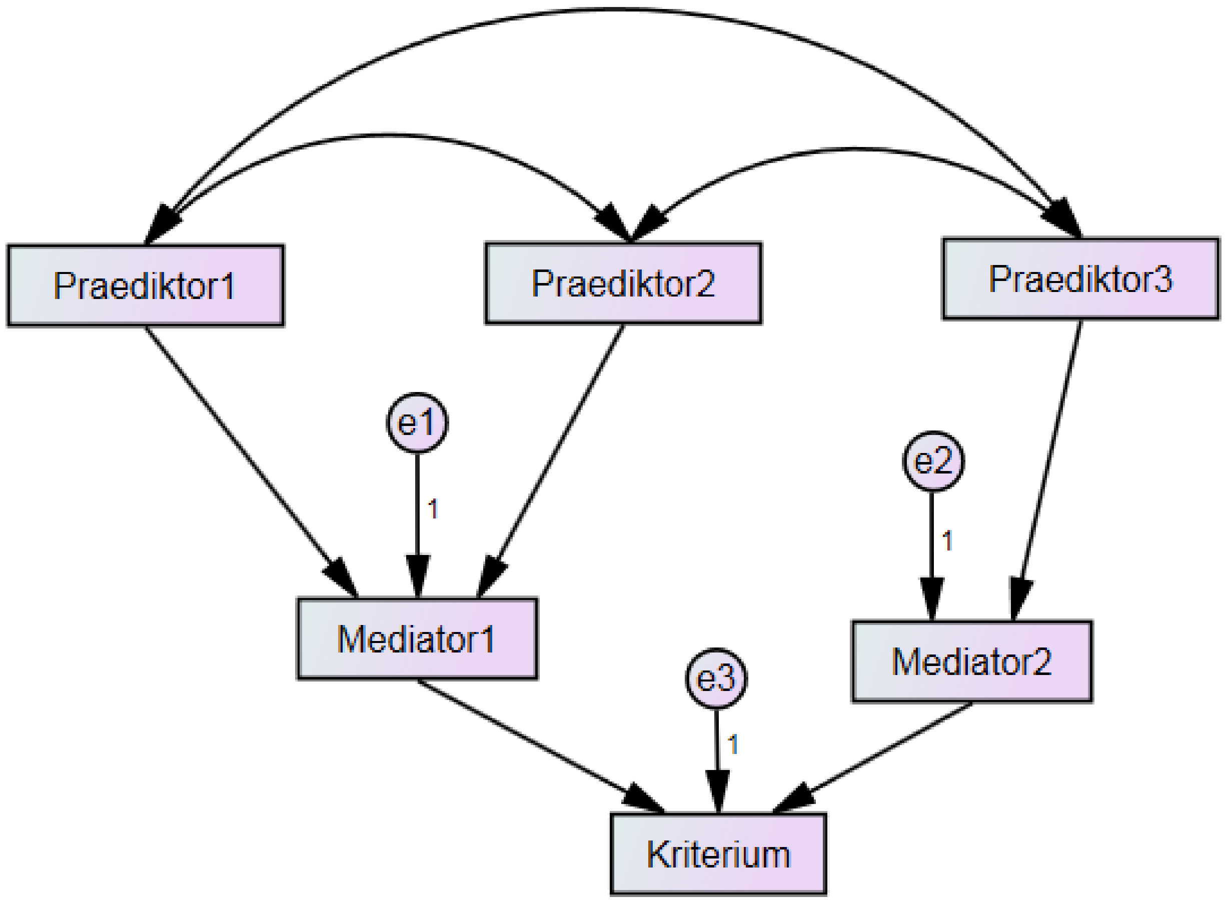
	N	Minimum	Maximum	Mittelwert	Standardabweichung
Praediktor1	100	-2,19656	2,45963	,0000000	1,00000000
Praediktor2	100	-2,91964	2,31116	,0000000	1,00000000
Praediktor3	100	-2,03713	2,78876	,0000000	1,00000000
Mediator1	100	-2,34986	2,37541	,0000000	1,00000000
Mediator2	100	-3,84415	2,20773	,0000000	1,00000000
Kriterium	100	-3,23150	2,73790	,0000000	1,00000000
Gültige Werte (Listenweise)	100				

Korrelationen

		Praediktor1	Praediktor2	Praediktor3	Mediator1	Mediator2	Kriterium
Praediktor1	Korrelation nach Pearson	1	-,035	-,128	,568**	-,028	,098
	Signifikanz (2-seitig)		,727	,205	,000	,780	,332
	N	100	100	100	100	100	100
Praediktor2	Korrelation nach Pearson	-,035	1	-,005	,632**	,631**	,575**
	Signifikanz (2-seitig)	,727		,959	,000	,000	,000
	N	100	100	100	100	100	100
Praediktor3	Korrelation nach Pearson	-,128	-,005	1	-,091	,449**	,390**
	Signifikanz (2-seitig)	,205	,959		,368	,000	,000
	N	100	100	100	100	100	100
Mediator1	Korrelation nach Pearson	,568**	,632**	-,091	1	,370**	,491**
	Signifikanz (2-seitig)	,000	,000	,368		,000	,000
	N	100	100	100	100	100	100
Mediator2	Korrelation nach Pearson	-,028	,631**	,449**	,370**	1	,797**
	Signifikanz (2-seitig)	,780	,000	,000	,000		,000
	N	100	100	100	100	100	100
Kriterium	Korrelation nach Pearson	,098	,575**	,390**	,491**	,797**	1
	Signifikanz (2-seitig)	,332	,000	,000	,000	,000	
	N	100	100	100	100	100	100

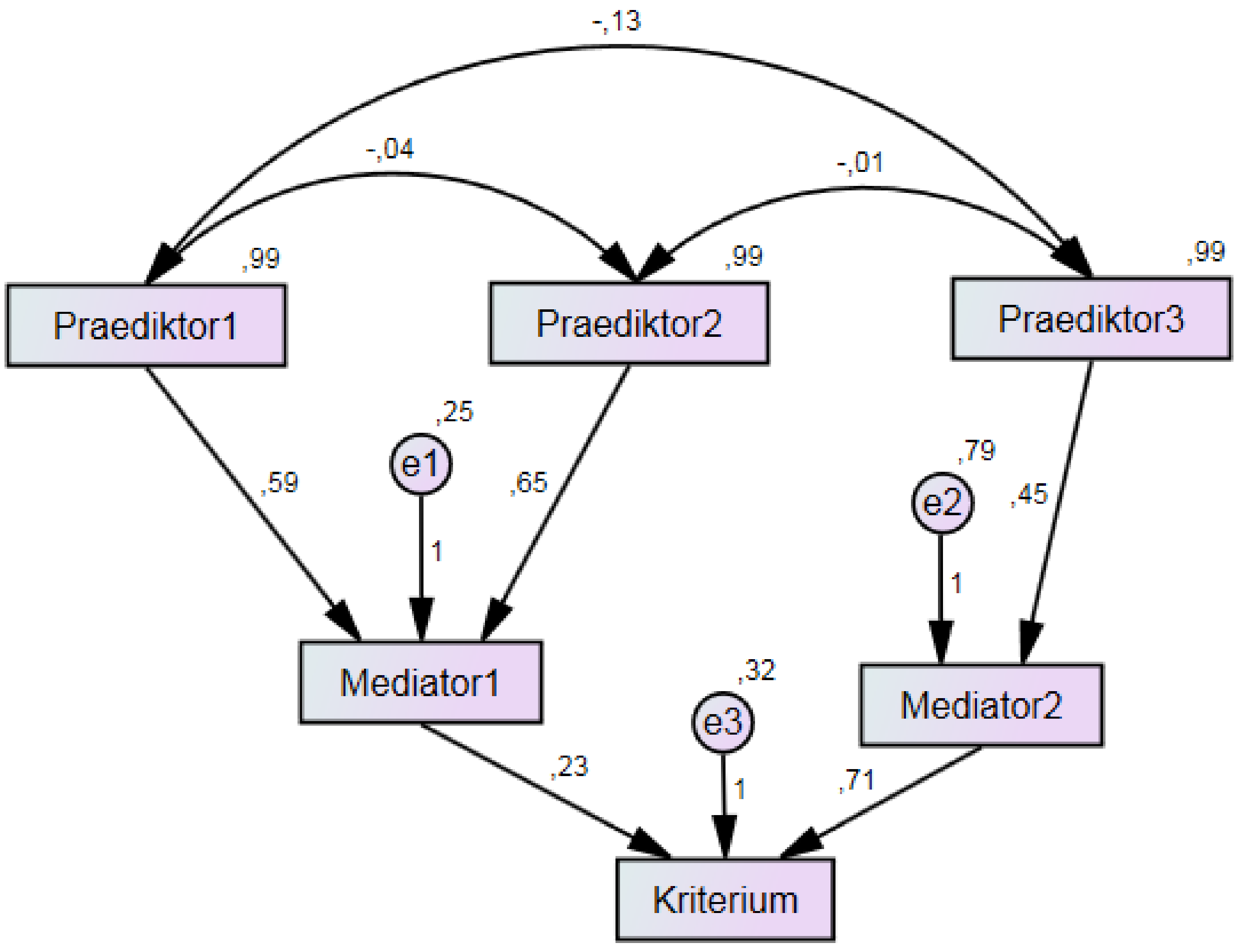
Prinzip der Parameterschätzung

Beispiel: Modell



Prinzip der Parameterschätzung

Beispiel: ML-Schätzungen



Prinzip der Modellgütebeurteilung

Allgemeine Frage zur Beurteilung der Modellgüte:

Wie gut ist es bei dem untersuchten Modell möglich, bei bestmöglicher Parameterschätzung die gegebene Korrelations- bzw. die Varianz-/ Kovarianzmatrix zu replizieren?

Prinzip der Modellgütebeurteilung

Bewertung der Gesamtstruktur: Chi-Quadrat-Test

→ Test der statistische Nullhypothese, dass die empirische Varianz-/Kovarianzmatrix der aus dem Modell ermittelten Varianz-/Kovarianzmatrix entspricht

→ Problematisch: wegen SPU-Abhängigkeit (bei großen SPU praktisch immer Ablehnung der H_0), fehlender Robustheit).

→ Trotz der Probleme werden Chi-Quadrat-Wert, Freiheitsgrade und p-Wert in Veröffentlichungen in der Regel angegeben

→ Deskriptive Empfehlung: $\chi^2 / df < 2.5$ → Hinweis auf gute Modellpassung (diese und die auf den folgenden Folien angegebenen Empfehlungen sind zitiert bei Weiber & Mühlhaus, 2010, S. 159ff)

→ Änderung des χ^2 -Wertes in Relation zu unterschiedlichen Freiheitsgraden wichtig beim Modellvergleich

Prinzip der Modellgütebeurteilung

Wichtigste Parameter der Modellgüte
(Empfehlung zit. bei Bühner, 2006, nach Beauducel & Wittman, 2005):

(1) RMSEA (Root Mean Square Error of Approximation)
RMSEA ≤ 0.06 → guter Modell-Fit
RMSEA ≤ 0.08 → akzeptabler Modell-Fit
RMSEA ≥ 0.10 → inakzeptabler Modell-Fit
(Vergleich beobachtete vs. geschätzte Varianz-/Kovarianzmatrix unter Beachtung der Modellkomplexität)

(2) CFI (Comparative Fit Index) - Guter Modell-Fit: CFI ≥ 0.95 (0.90)
(Vergleich untersuchtes Modell vs. Nullmodell)

(3) SRMR (Standardized Root Mean Residual) - Guter Modell-Fit: SRMR ≤ 0.10
(Vergleich beobachtete vs. geschätzte Varianz-/Kovarianzmatrix ohne Beachtung der Modellkomplexität)

Viele weitere Parameter der Modellgüte mit ausführlichen Beschreibungen und Angabe der Schwellenwerte siehe Weiber & Mühlhaus, 2010, S. 159 ff), z.B.

(4) TLI (=NNFI): Tucker-Lewis-Index - Guter Modell-Fit: TLI ≥ 0.95 (0.90)
(Vergleich untersuchtes Modell vs. Nullmodell unter Berücksichtigung der FG)

Bei Weiber & Mühlhaus abweichende Empfehlungen zur Kombination von Fit-Indizes auf der Basis einer Studie von Hu & Bentler. Empfehlungen zu Grenzwerten variieren im Ergebnis unterschiedlicher Studien.

Prinzip der Modellgütebeurteilung

Beispiel: Modellgüte

Notes for Model (Default model)

Computation of degrees of freedom (Default model)

Number of distinct sample moments:	21
Number of distinct parameters to be estimated:	14
Degrees of freedom (21 - 14):	7

Result (Default model)

Minimum was achieved
Chi-square = 73,890
Degrees of freedom = 7
Probability level = ,000

RMSEA = 0.31

CFI = 0.79

SRMR = 0.20

TLI = 0.57

Hypothesenüberprüfung

Bewertung von Teilstrukturen

→ Test einzelner Modellparameter: Tests auf der Grundlage der Schätzwerte und der Standardfehler

→ Tests indirekter Effekte: Bootstrap-Tests

→ Residuenanalyse

→ ggf. Vergleich alternativer Schätzverfahren

Modifikationen des Modells

Bei unbefriedigender Modellgüte
→ Modifikation der Modellstruktur auf der Grundlage der Daten und inhaltlicher Überlegungen prüfen
→ Exploratorischer (!!) Analyseschritt
→ Prüfung neuer Modelle mit neuen Daten erforderlich!

Statistische Entscheidungshilfen zur Modellmodifikation:

- Entscheidung über Hinzunahme von Pfaden, Kovarianzen u.a. auf der Basis von Modifikations-Indizes: Angabe, um welchen Betrag sich der Chi-Quadrat-Wert verändert (d.h. wie stark die Modellgüte verbessert wird), wenn ein zusätzlicher zu schätzender Parameter im Modell frei gegeben wird.
- Entscheidung über Entfernen von Pfaden, Kovarianzen u.a. auf der Basis der Ergebnisse der statistischen Tests der jeweiligen Koeffizienten (ggf. Entfernen „nicht signifikanter“ Pfade).

→ Entscheidungen immer unter vorrangiger Einbeziehung inhaltlicher Gesichtspunkte treffen

Modifikationen des Modells

Beispiel: Modification Indices

Amos Output

3 | 7 | 0

- Pfadmodell_Beispiel_1.amw
  - Analysis Summary
  - Notes for Group
  - Variable Summary
  - Parameter Summary
  - Notes for Model
  - Estimates
  - Modification Indices
  - Minimization History
  - Model Fit
  - Execution Time

Modification Indices (Group number 1 - Default model)

Covariances: (Group number 1 - Default model)

			M.I.	Par Change
e2	<-->	Praediktor2	49,923	,628

Variances: (Group number 1 - Default model)

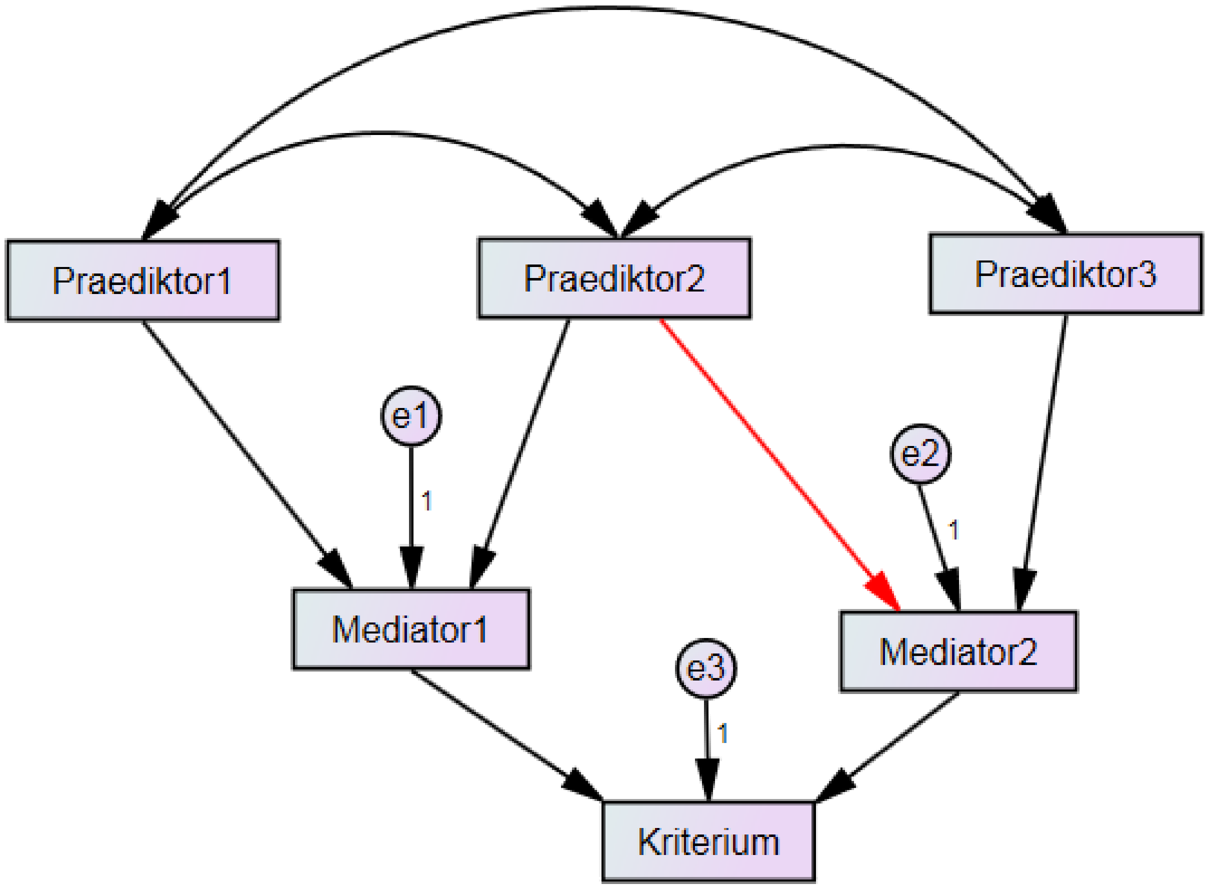
	M.I.	Par Change

Regression Weights: (Group number 1 - Default model)

			M.I.	Par Change
Mediator2	<---	Praediktor2	49,688	,633
Mediator2	<---	Mediator1	20,972	,411

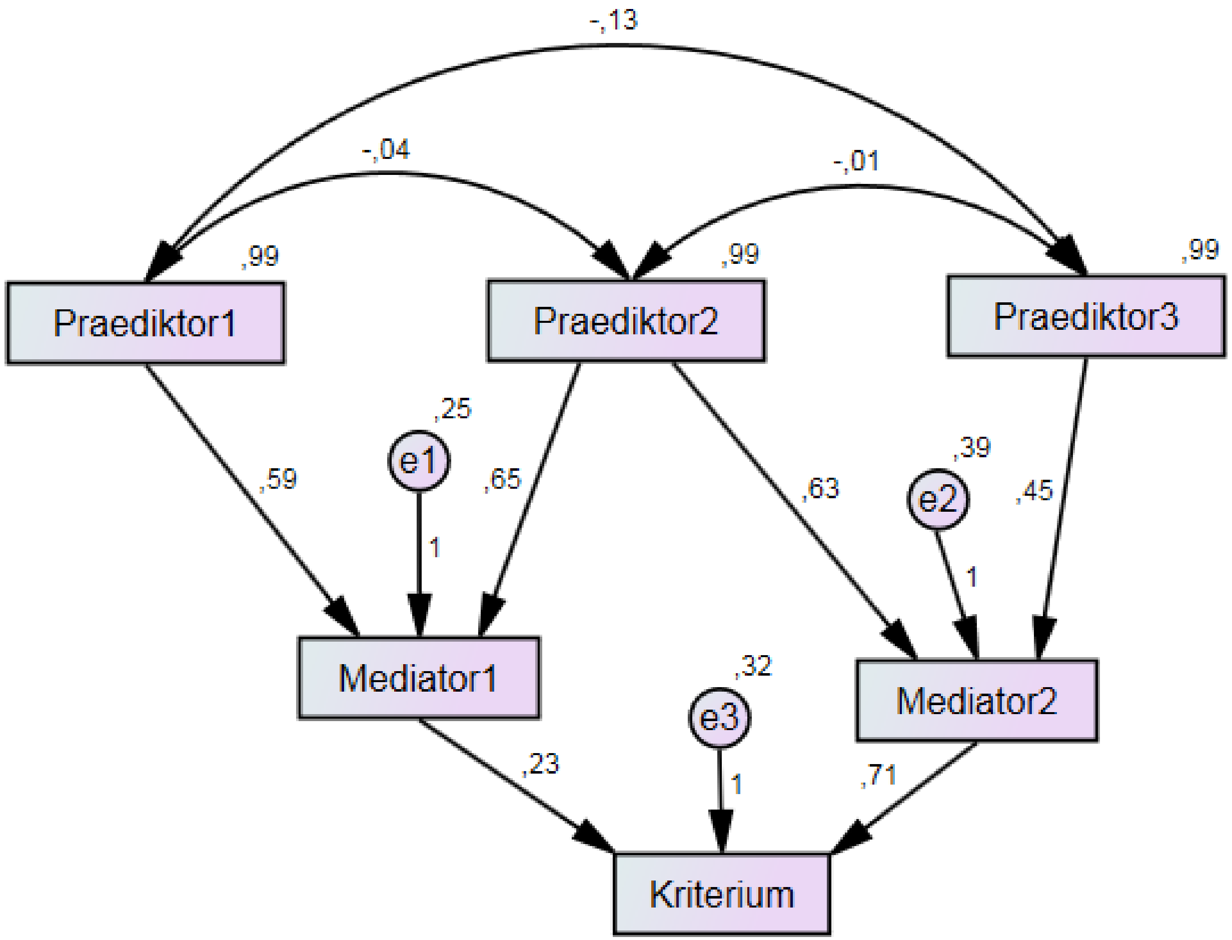
Modifikationen des Modells

Beispiel: Modifiziertes Modell



Modifikationen des Modells

Beispiel: ML-Schätzungen im modifizierten Modell



Modifikationen des Modells

Beispiel: Modellgüte im modifizierten Modell

Notes for Model (Default model)

Computation of degrees of freedom (Default model)

Number of distinct sample moments:	21
Number of distinct parameters to be estimated:	15
Degrees of freedom (21 - 15):	6

Result (Default model)

Minimum was achieved
Chi-square = 4,889
Degrees of freedom = 6
Probability level = ,558

RMSEA = 0.00

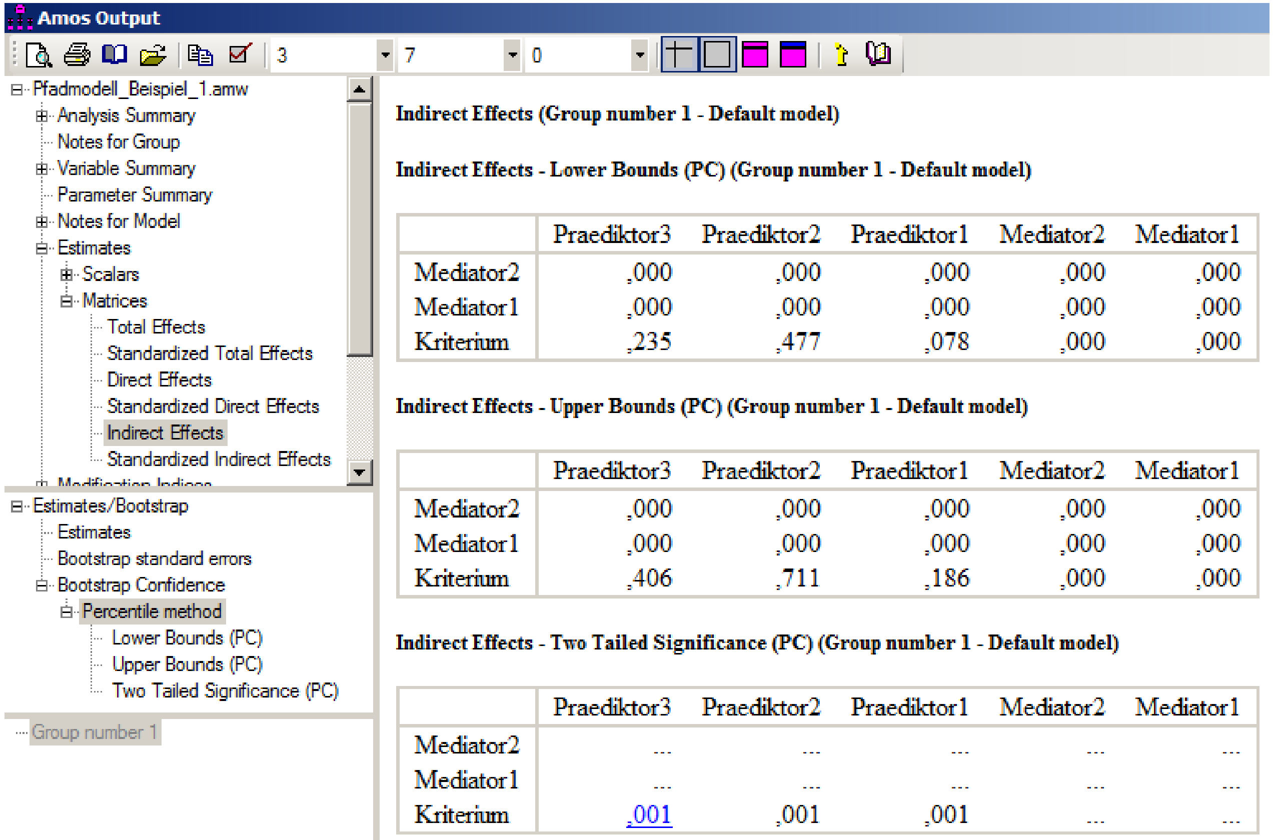
CFI = 1.00

SRMR = 0.02

TLI = 1.0

Modifikationen des Modells

Beispiel: Bootstrap-Tests des indirekten Effekts von Prädiktor 3 auf das Kriterium im modifizierten Modell



Indirect Effects (Group number 1 - Default model)

Indirect Effects - Lower Bounds (PC) (Group number 1 - Default model)

	Praediktor3	Praediktor2	Praediktor1	Mediator2	Mediator1
Mediator2	,000	,000	,000	,000	,000
Mediator1	,000	,000	,000	,000	,000
Kriterium	,235	,477	,078	,000	,000

Indirect Effects - Upper Bounds (PC) (Group number 1 - Default model)

	Praediktor3	Praediktor2	Praediktor1	Mediator2	Mediator1
Mediator2	,000	,000	,000	,000	,000
Mediator1	,000	,000	,000	,000	,000
Kriterium	,406	,711	,186	,000	,000

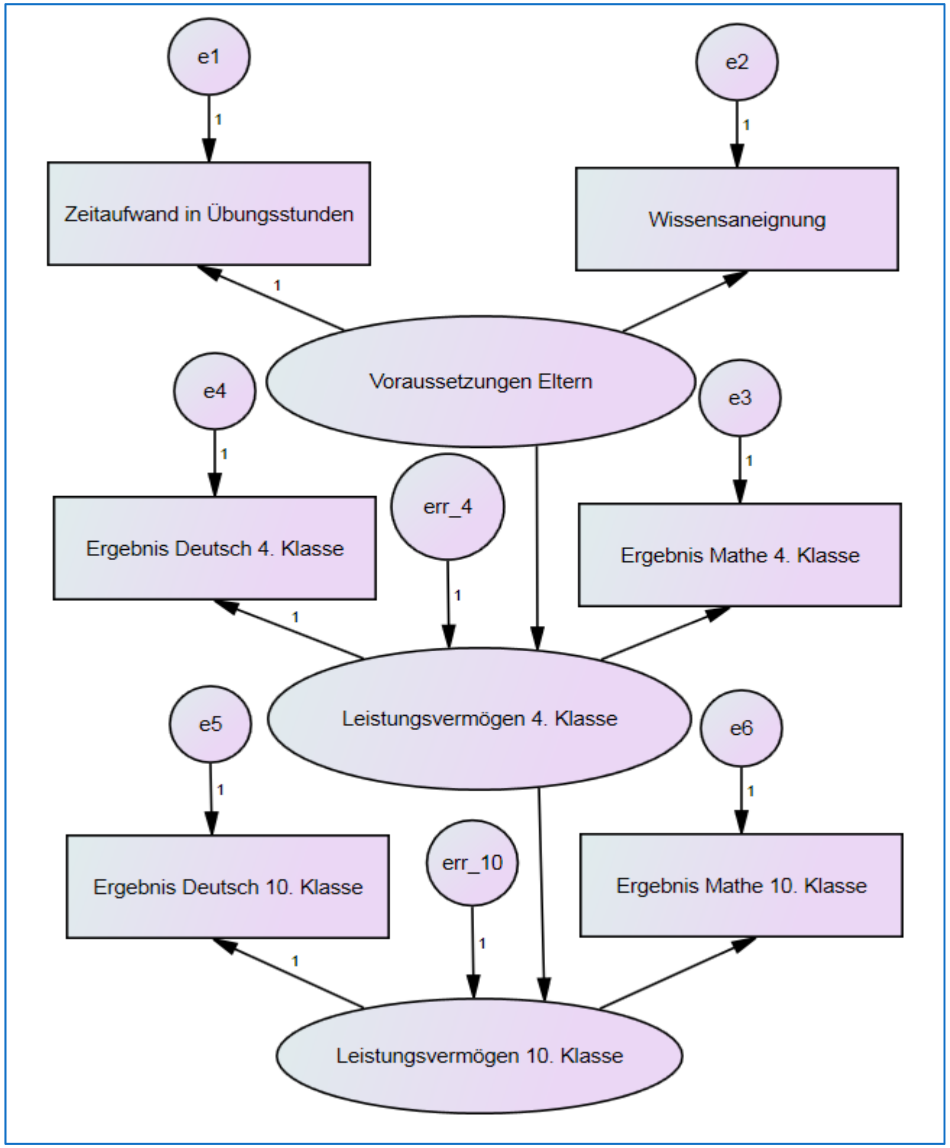
Indirect Effects - Two Tailed Significance (PC) (Group number 1 - Default model)

	Praediktor3	Praediktor2	Praediktor1	Mediator2	Mediator1
Mediator2	...	...	...	...	...
Mediator1	...	...	...	...	...
Kriterium	,001	,001	,001	...	...

LSM mit latenten Variablen

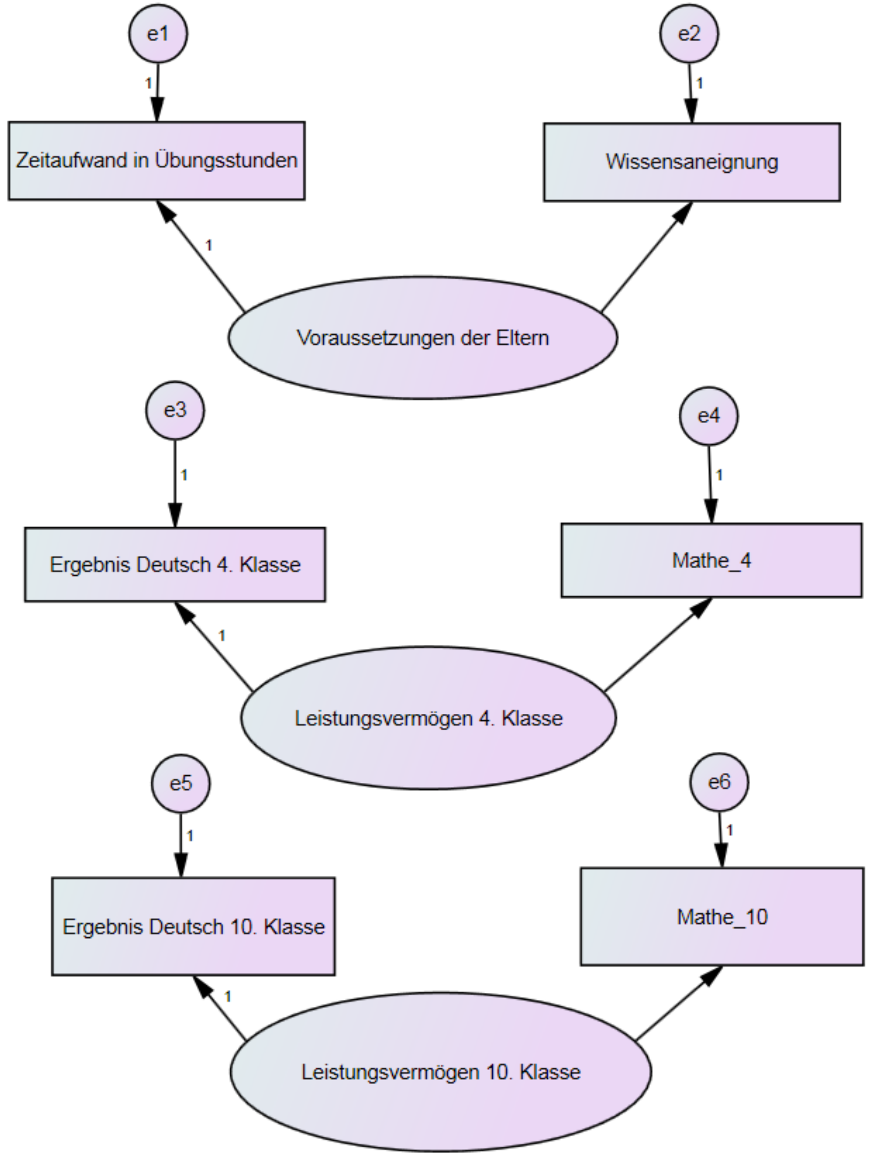
Verallgemeinerung der Pfadanalyse, Verbindung mir der (konfirmatorischen) Faktorenanalyse

→ Pfadmodelle mit latenten Variablen



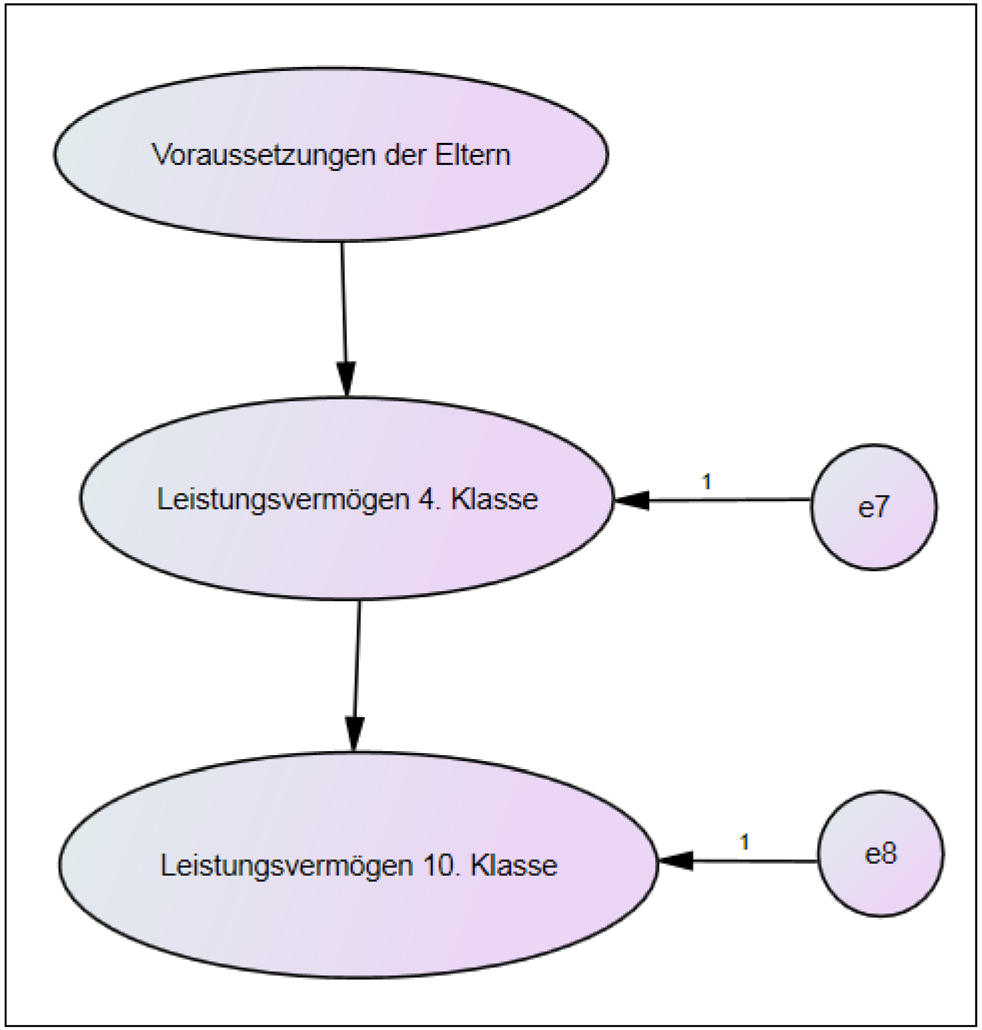
Messmodell

Messmodelle beschreiben, welche manifesten Variablen Indikatoren von latenten Variablen sind.



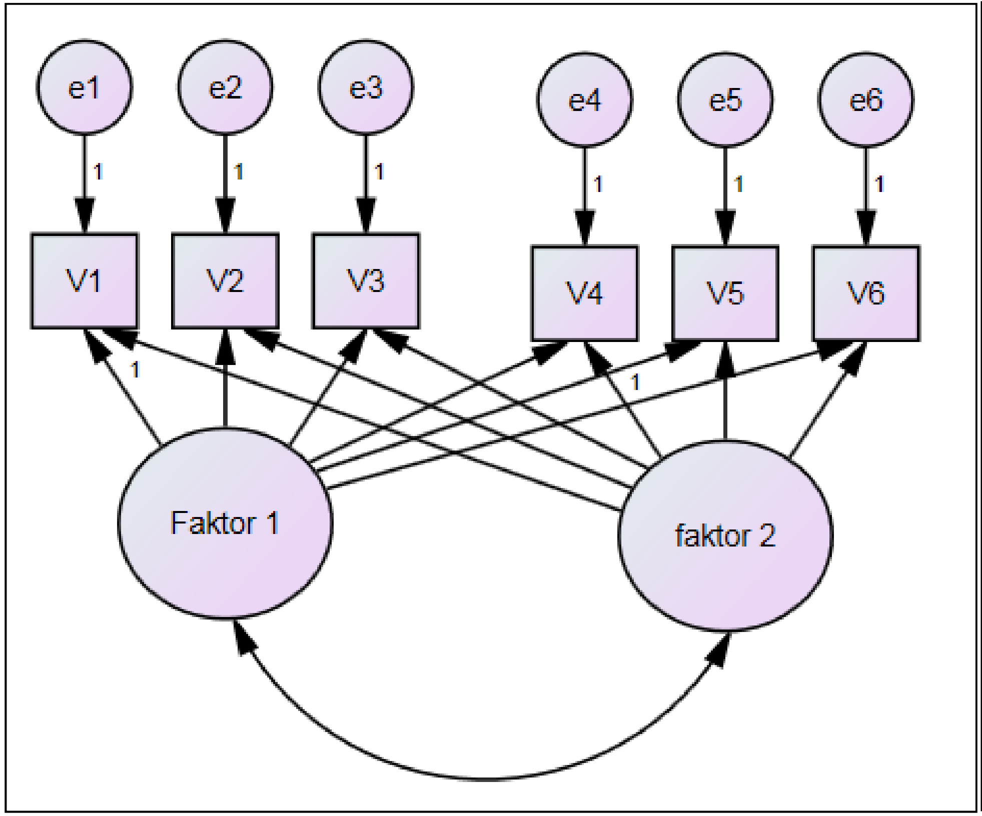
Strukturmodell

Strukturmodelle beschreiben die Beziehungen der latenten Variablen im untersuchten Modell.

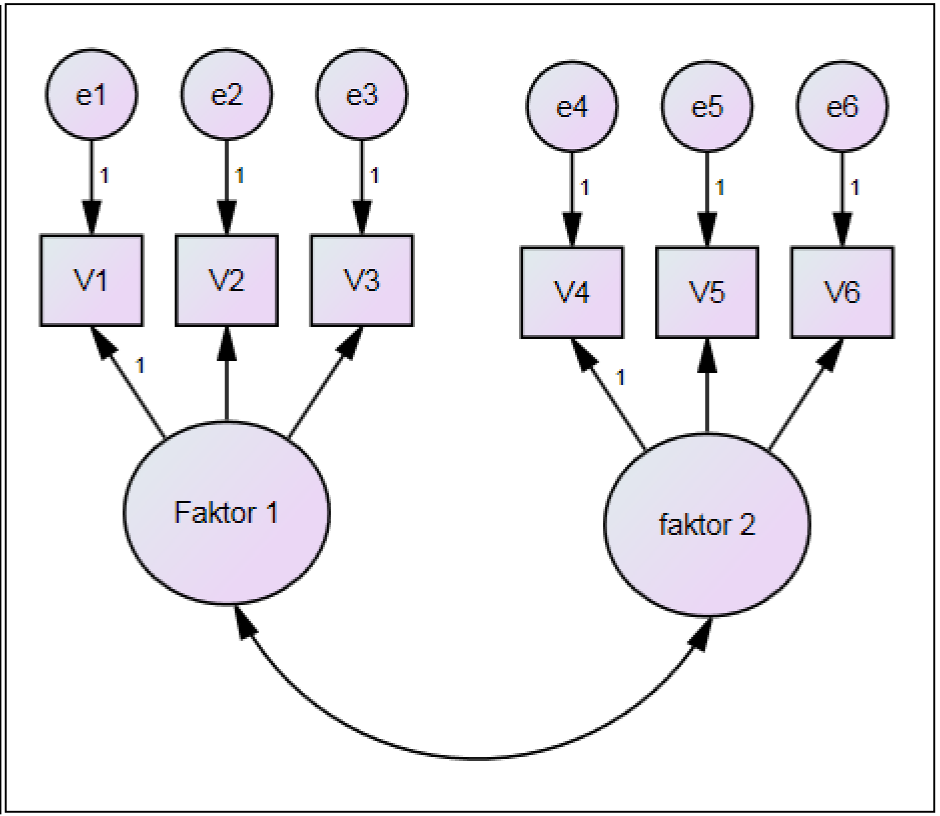


Konfirmatorische Faktorenanalyse

Exploratorische Faktorenanalyse



Konfirmatorische Faktorenanalyse



(Abbildung nach Bühner, 2006, S. 260)

Analysevorbereitung

- Ausreißerprüfung
- Ersetzung fehlender Werte
- Prüfung der Multinormalverteilung

Behandlung fehlender Werte

(1) Not missing at random (NMAR)

→ Ausfallmechanismus muss bekannt sein, sonst keine adäquate Behandlung möglich (Unterstützung durch SPSS unter Analysieren → Analyse fehlender Werte → Muster)

(2) Missing completely at random (MCAR)
(3) Missing at random (MAR)

→ FIML-Technik in AMOS (Full Information Maximum Likelihood). Ersetzung fehlender Werte direkt im Prozess der Parameterschätzungen. Empfohlenes Vorgehen bei der Analyse von Strukturgleichungsmodellen.
→ Vorgehen in AMOS: Anklicken von View und Analysis Properties unter der Registerkarte Estimation, danach Haken bei Estimate means and intercepts.
→ Nachteil: Keine Berechnung von Modification indices möglich.

Vgl. Weiber & Mühlhaus (2010) S. 142ff

Umgang mit Ausreißern

(1) Verfahrenstechnische Fehler (Eingabefehler)
(2) Ungewöhnliche Werte (sachlogisch erklärbar)
(3) „echte“ Ausreißer

Hilfe in SPSS: Boxplots

Identifikation von Ausreißern in AMOS:

→ Grundlage: Mahalanobis-Distanz (multivariates Distanzmaß).

→ Mahalanobis-Distanz beschreibt, wie weit die Werte einzelner Probanden vom multivariaten Mittelwert abweichen.

→ Ordnung der Mahalanobis-Distanzen der Probanden.

→ Deutliche Knicke in der Abfolge der Mahalanobis-Distanzen weisen auf Ausreißer hin.

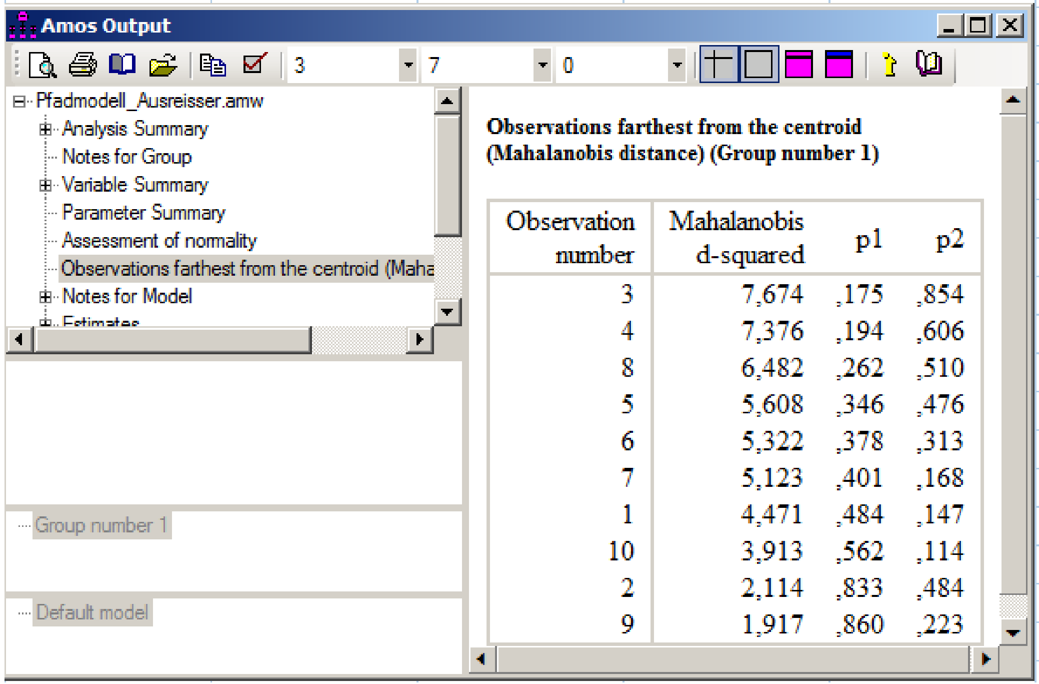
→ Umsetzung in AMOS: siehe praktischer Teil

Vgl. Weiber & Mühlhaus (2010) S. 145ff

Umgang mit Ausreißern

Beispiel 1:

variable_1	variable_2	variable_3	variable_4	variable_5
74,49	28,71	-2,63	-2,06	895,83
118,14	16,49	,79	1,30	805,38
106,22	13,59	-2,05	,14	887,83
101,98	22,16	1,18	-3,09	1108,83
27,20	21,19	,32	,58	738,80
31,65	14,96	2,39	,88	1095,19
104,14	25,30	-,18	,40	987,01
201,69	19,09	2,01	1,52	863,63
110,67	17,20	,94	,31	1078,77
104,15	27,26	-1,34	-1,87	740,09



Observations farthest from the centroid (Mahalanobis distance) (Group number 1)

Observation number	Mahalanobis d-squared	p1	p2
3	7,674	,175	,854
4	7,376	,194	,606
8	6,482	,262	,510
5	5,608	,346	,476
6	5,322	,378	,313
7	5,123	,401	,168
1	4,471	,484	,147
10	3,913	,562	,114
2	2,114	,833	,484
9	1,917	,860	,223

Umgang mit Ausreißern

Beispiel 2:

variable_1	variable_2	variable_3	variable_4	variable_5
74,49	28,71	-2,63	-2,06	895,83
118,14	16,49	,79	1,30	805,38
106,22	13,59	-2,05	,14	887,83
101,98	22,16	1,18	-3,09	1108,83
27,20	21,19	,32	,58	738,80
31,65	14,96	2,39	,88	1095,19
104,14	25,30	-,18	,40	987,01
201,69	19,09	2,01	1,52	863,63
110,67	17,20	,94	,31	1078,77
304,15	37,26	5,34	-4,87	1500,09

Amos Output

3 7 0

- Pfadmodell_Ausreisser.amw
  - Analysis Summary
  - Notes for Group
  - Variable Summary
  - Parameter Summary
  - Assessment of normality
  - Observations farthest from the centroid (Maha
  - Notes for Model
  - Estimates

Group number 1

Default model

Observations farthest from the centroid (Mahalanobis distance) (Group number 1)

Observation number	Mahalanobis d-squared	p1	p2
10	7,875	,163	,832
6	6,127	,294	,841
1	5,644	,342	,722
3	5,629	,344	,470
5	5,473	,361	,273
4	5,277	,383	,139
8	5,118	,402	,056
7	4,621	,464	,033
2	2,127	,831	,478
9	2,110	,834	,162

Umgang mit Ausreißern

Beispiel 3:

	variable_1	variable_2	variable_3	variable_4	variable_5
19	156,75	17,58	-,99	,80	1293,37
20	149,20	12,28	-,56	,42	1049,80
21	120,71	17,65	2,69	1,26	815,70
22	75,99	21,09	-,65	-1,70	1245,55
23	195,84	13,72	3,14	1,47	1364,63
24	106,68	17,42	-,05	,92	782,38
25	119,26	22,14	-1,22	,79	1116,20
26	188,18	18,90	-,11	1,54	851,23
27	109,30	21,35	4,26	-,46	961,89
28	104,07	22,66	1,08	-1,03	662,81
29	154,77	6,92	1,88	,79	651,71
30	70,96	18,19	2,33	-2,35	896,61
31	119,83	21,19	1,83	-1,62	1077,09
32	172,89	20,61	,96	-,01	1097,16
33	50,76	24,35	,38	1,12	1094,11
34	105,97	12,19	2,24	,39	1073,05
35	230,02	31,19	5,09	5,09	1500,86
36	104,27	21,43	,27	-1,78	1085,75
37	63,04	22,03	1,17	-,03	1198,23
38	177,37	26,99	3,52	1,77	1685,85
39	187,73	14,22	2,03	-,86	1101,04
40	58,89	22,80	,78	-,31	916,40
41	118,32	21,13	-1,53	-,64	1055,45
42	139,99	20,20	1,48	3,02	1068,73
43	121,32	27,60	-1,72	,83	1386,09
44	140,33	19,31	,66	1,81	1204,44
45	126,36	17,66	-,10	,29	1132,67
46	104,10	12,51	,49	,77	1061,53
47	161,32	21,67	-1,18	-4,89	1024,58

*Ausgabe1 [Dokument1] - IBM SPSS Statistics Viewer

Datei Bearbeiten Ansicht Daten Transformiere Einfügen Format Analysierer Direktmarketin Grafik Extras Fenster Hilfe

	N	Minimum	Maximum	Mittelwert	Standardabweichung
variable_1	100	,37	237,32	113,6450	52,13479
variable_2	100	4,68	31,81	19,9431	5,30752
variable_3	100	-3,15	5,09	,5145	1,88123
variable_4	100	-4,89	6,30	,2496	1,85931
variable_5	100	397,01	1685,85	1010,3148	219,68318
Gültige Werte (Listenweise)	100				

IBM SPSS Statistics -Prozessor ist bereit | Fälle: 100 | Unicode:ON

Umgang mit Ausreißern

Beispiel 3:

	variable_1	variable_2	variable_3	variable_4	variable_5
19	156,75	17,58	-,99	,80	1293,37
20	149,20	12,28	-,56	,42	1049,80
21	120,71	17,65	2,69	1,26	815,70
22	75,99	21,09	-,65	-1,70	1245,55
23	195,84	13,72	3,14	1,47	1364,63
24	106,68	17,42	-,05	,92	782,38
25	119,26	22,14	-1,22	,79	1116,20
26	188,18	18,90	-,11	1,54	851,23
27	109,30	21,35	4,26	-,46	961,89
28	104,07	22,66	1,08	-1,03	662,81
29	154,77	6,92	1,88	,79	651,71
30	70,96	18,19	2,33	-2,35	896,61
31	119,83	21,19	1,83	-1,62	1077,09
32	172,89	20,61	,96	-,01	1097,16
33	50,76	24,35	,38	1,12	1094,11
34	105,97	12,19	2,24	,39	1073,05
35	230,02	31,19	5,09	5,09	1500,86
36	104,27	21,43	,27	-1,78	1085,75
37	63,04	22,03	1,17	-,03	1198,23
38	177,37	26,99	3,52	1,77	1685,85
39	187,73	14,22	2,03	-,86	1101,04
40	58,89	22,80	,78	-,31	916,40
41	118,32	21,13	-1,53	-,64	1055,45
42	139,99	20,20	1,48	3,02	1068,73
43	121,32	27,60	-1,72	,83	1386,09
44	140,33	19,31	,66	1,81	1204,44
45	126,36	17,66	-,10	,29	1132,67
46	104,10	12,51	,49	,77	1061,53
47	161,32	21,67	-1,18	-4,89	1024,58

Amos Output

Pfadmodell_Ausreisser.amw
- Analysis Summary
- Notes for Group
- Variable Summary
- Parameter Summary
- Assessment of normality
- Observations farthest from the centroid (Mahalan
- Notes for Model
- Estimates
- Minimization History
- Model Fit
- Execution Time

Group number 1

Default model

Observation number	Mahalanobis d-squared	p1	p2
35	23,405	,000	,028
91	16,194	,006	,132
66	15,280	,009	,066
38	14,002	,016	,072
48	11,170	,048	,530
11	9,850	,080	,816
47	9,256	,099	,878
89	9,184	,102	,811
92	9,167	,103	,708
29	9,059	,107	,634
97	8,956	,111	,558
79	8,638	,124	,598
96	8,614	,125	,491
23	8,171	,147	,622
84	7,864	,164	,687
86	7,806	,167	,618
95	7,733	,172	,558
80	7,646	,177	,508
63	7,645	,177	,406
83	7,335	,197	,509
75	7,017	,219	,628
16	6,923	,226	,600
43	6,764	,239	,619

Prüfung auf Multinormalverteilung

Univariate Prüfung einzelner Items.
→Signifikante Abweichungen von der NV sprechen noch nicht gegen Anwendung der ML-Methode
→Empfehlung: Schiefe (skewness) sollte kleiner 2 sein, Exzess (kurtosis) kleiner 7 für jedes Item: Exzess für die Beurteilung bedeutsamer

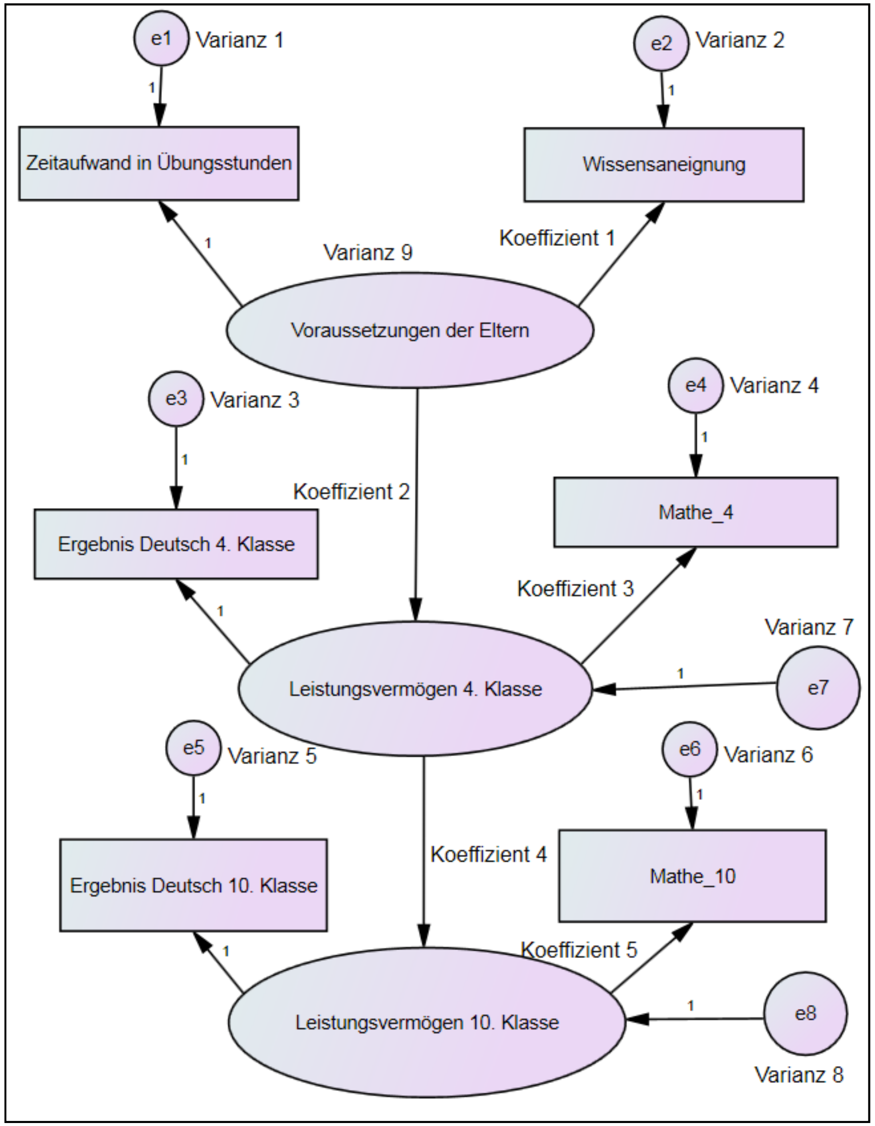
Multivariate Prüfung:
→Mardia's Koeffizient der multivariaten Wölbung incl. Test
→Empfehlung (vgl. Bühner)
- bei Verletzung der NV und kleiner SP: Bollen-Stine-Bootstrap-Verfahren bei der ML-Schätzung (geänderter p-Wert bei Chi-Quadrat-Test)
- bei Verletzung der NV und großer SP: ADF-Methode

→ Mögliche Auswege bei Nicht-NV:
- Ausreißer-Eliminierung
- Transformationen der Variablen

Umsetzung in AMOS: siehe praktischer Teil

Vgl. Weiber & Mühlhaus (2010) S. 146ff, Bühner (2006) 261 ff, siehe auch Byrne (2009)

Konstruktion des LSM



Parametertypen

→Feste Parameter (z.B. je ein abgehender Pfadkoeffizient für jede latente Variable gleich 1 gesetzt)

→ Restringierte Parameter (Parameter mit bestimmten Einschränkungen, z.B. sollen bestimmte Pfadkoeffizienten den gleichen (zu schätzenden) Wert bekommen (im Beispiel zunächst nicht enthalten)

→ Freie Parameter, im Beispiel 14: Pfadkoeffizienten 1-5, Varianzen der Fehlervariablen E1 bis E8 und Varianz der latenten Variablen Voraussetzungen der Eltern.

Modellspezifikation: Freiheitsgrade

→ Voraussetzung: explizit formulierte Hypothesen über die Zusammenhangsstruktur

→ Grundlage: Varianzen und Kovarianzen der messbaren Variablen (im Beispiel $6 \cdot (6-1) / 2 + 6 = 21$

→ Zahl der zu schätzenden Parameter darf diese Anzahl nicht überschreiten (eine notwendige Bedingung für die Identifizierbarkeit des Modells)

→ Freie Parameter, im Beispiel 14: Pfadkoeffizienten 1-5, Varianzen der Fehlervariablen E1 bis E8 und Varianz der latenten Variablen Voraussetzungen der Eltern.

→Freiheitsgrade im Beispiel: FG = 21 - 14 = 7

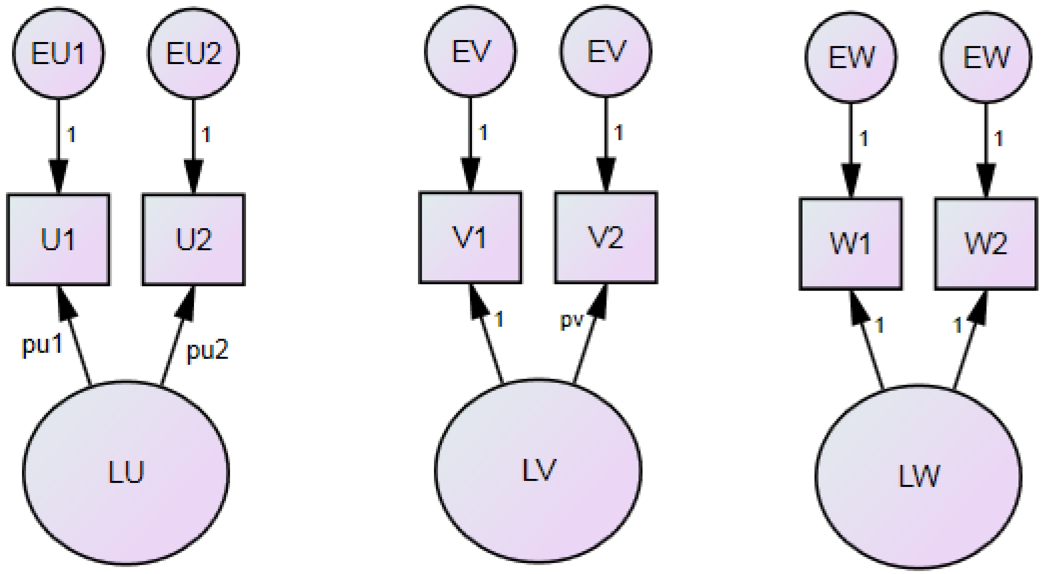
Problem der Identifizierbarkeit

→ Notwendige Bedingung: df $\geq$ 0
(in Praxis i.d.R. großer df-Wert anzustreben)

→ Keine lineare Abhängigkeit der Modellgleichungen

→ Identifizierbarkeit von Teilstrukturen

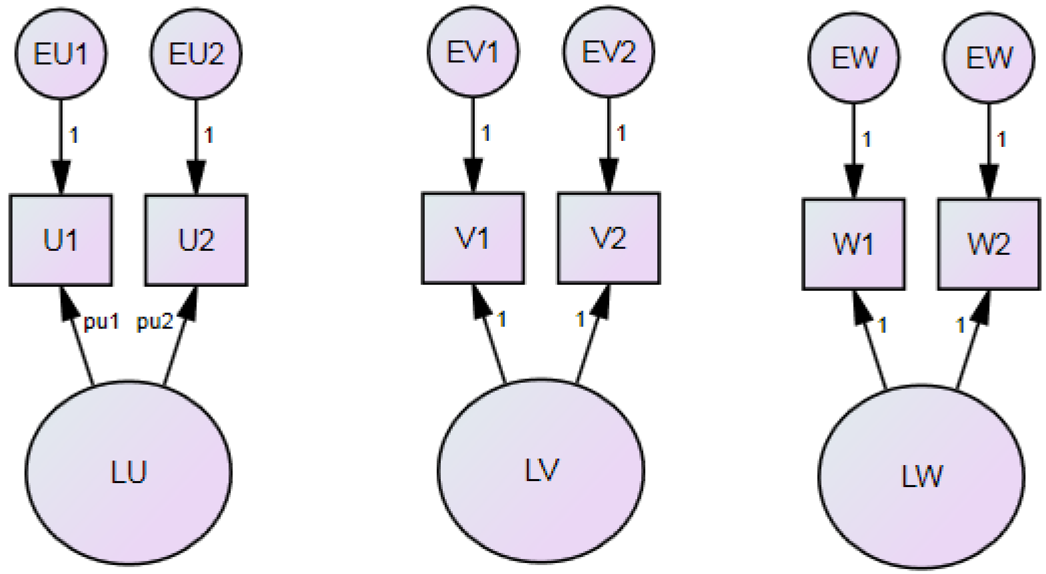
Problem der Identifizierbarkeit



→ Anzahl der empirischen Parameter: 3 (2 Varianzen, 1 Kovarianz)

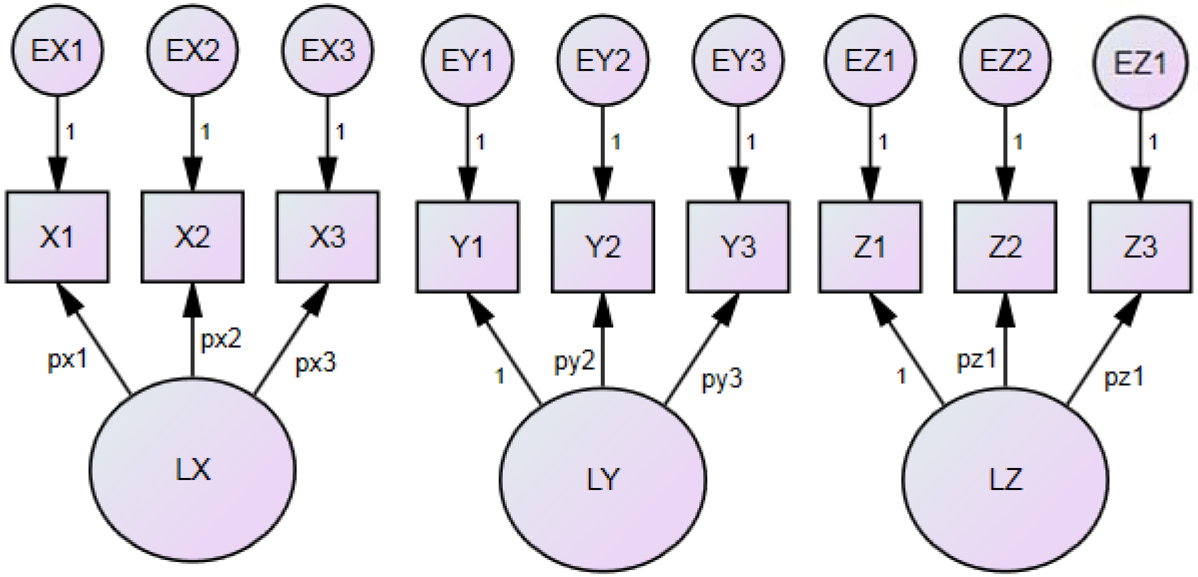
→ Zu schätzende Parameter:
→ Modell U: Var (LU), pu1, pu2, Var (EU1), Var (EU2) → df<0
→ Modell V: Var (LV), pv, Var (EV)→ df = 0
→ Modell W: Var (LV), Var (EW) → df = 1

Problem der Identifizierbarkeit



→ Modell U: unteridentifiziert →keine Schätzungen der Parameter möglich
→ Modell U: genau identifiziert → Schätzungen der Parameter möglich, aber keine stat. Tests
→ Modell W: überidentifiziert → Schätzungen der Parameter und statistische Tests möglich → praktisch relevante Modelle

Problem der Identifizierbarkeit



→ Anzahl der empirischen Parameter: ?

→ Zu schätzende Parameter:
→ Modell X: ?
→ Modell Y: ?
→ Modell Z: ?

→ Identifizierbarkeit der Teilmodelle?

Problem der Identifizierbarkeit

→ Praktisch bedeutsam: Überidentifizierte Gesamtmodelle, die auf der Grundlage inhaltlicher Überlegungen formuliert werden

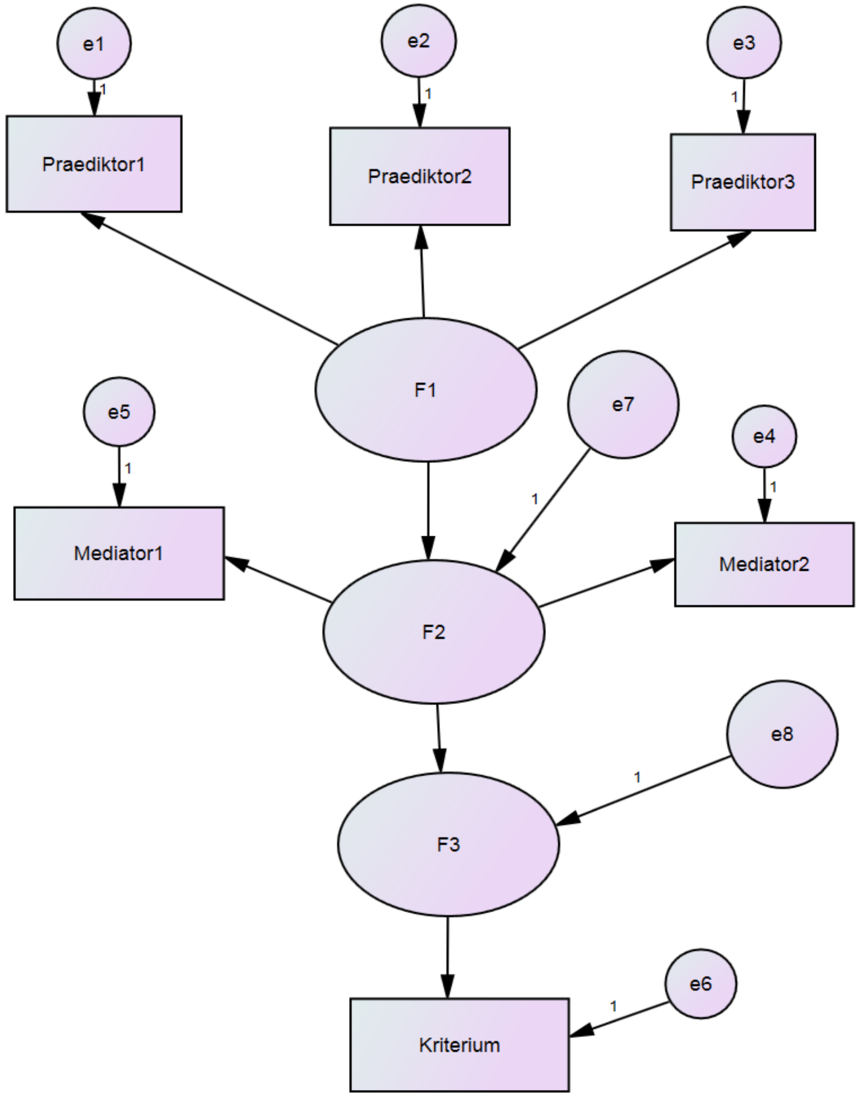
(oft in komplexeren Modellen angestrebt: df ≥ Anzahl der zu schätzenden Parameter).

→ Wichtige Regel zur Erzeugung genau identifizierter bzw. überidentifizierter Teilstrukturen: Von jeder latenten Variablen mit unbekannter Varianz MUSS der Pfadkoeffizient eines abgehenden Pfeiles fixiert werden (i.d.R. mit 1), um der latenten Variablen eine „Referenzmetrik" zuzuweisen.

→ Weitere Möglichkeiten zur Erzeugung identifizierter bzw. überidentifizierter Teilstrukturen: Parameterrestriktionen (Fixierung, Gleichsetzung von Pfadkoeffizienten, Varianzen, Kovarianzen)

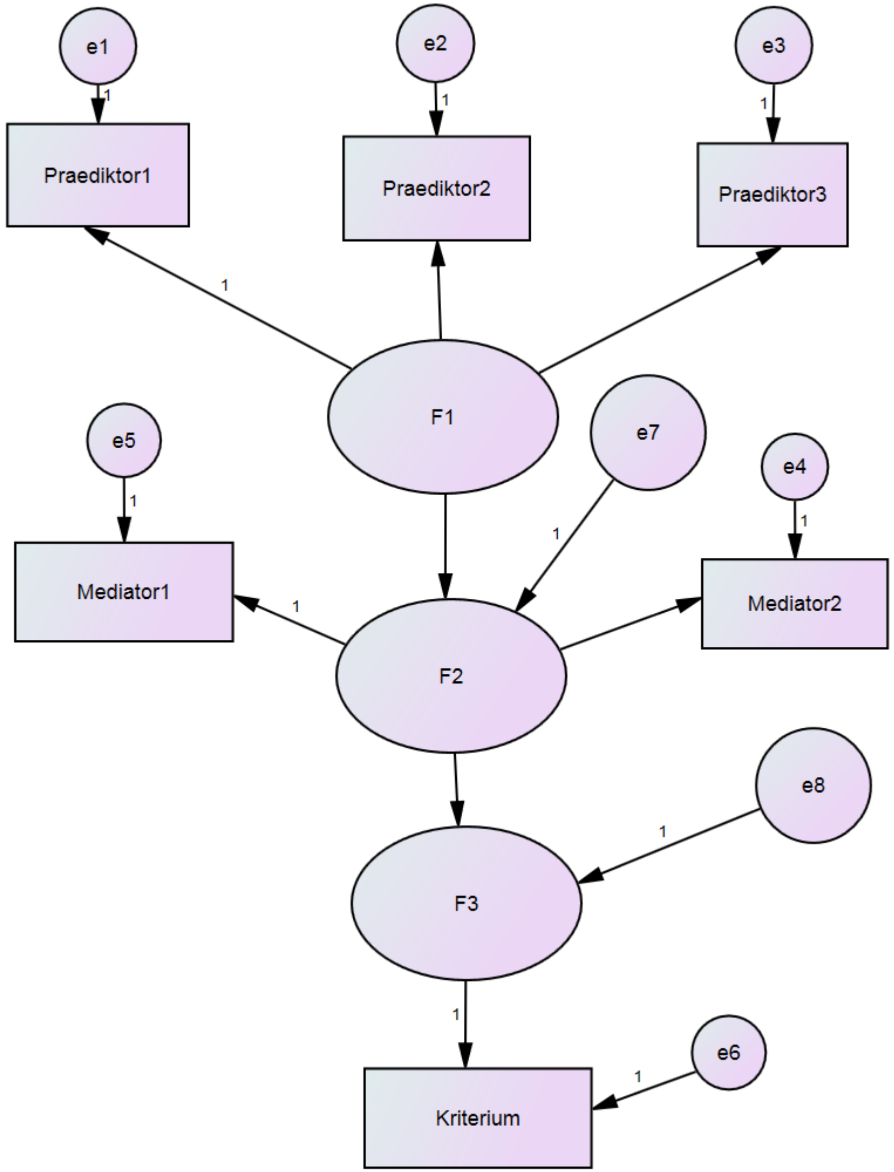
Problem der Identifizierbarkeit

Beispiel:



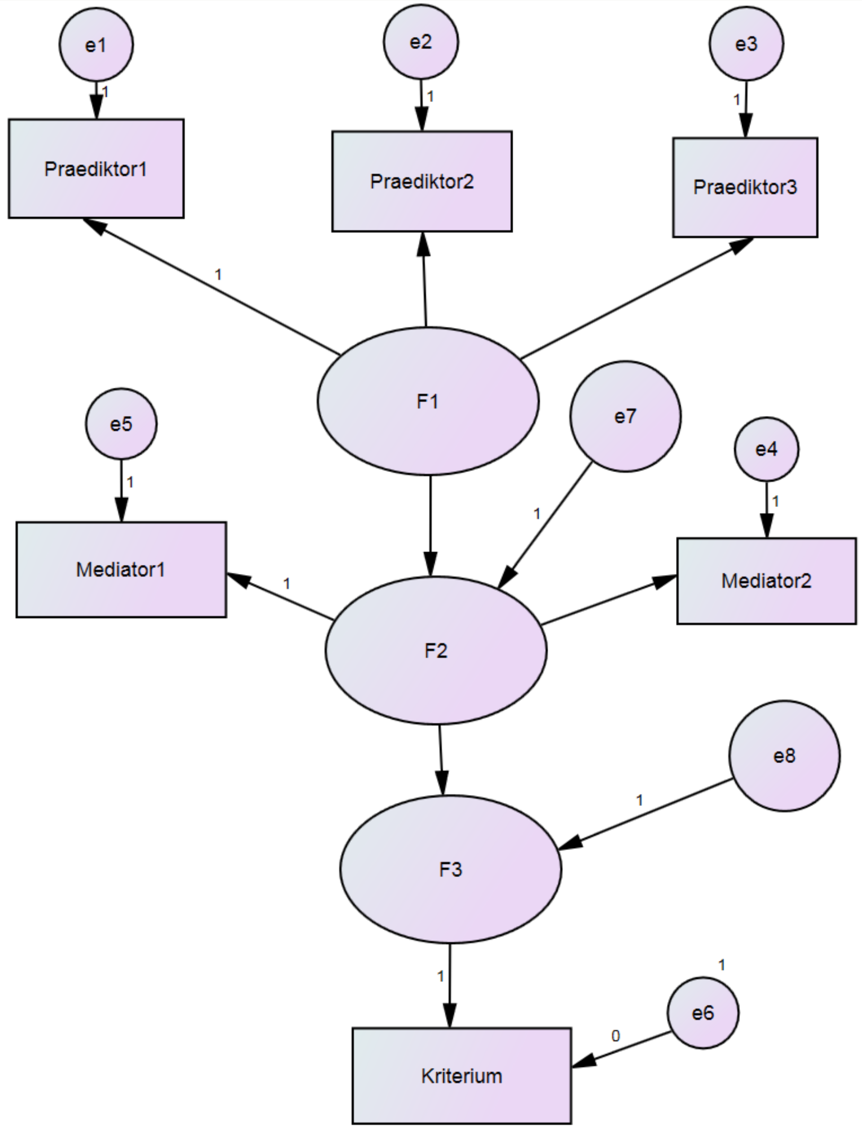
Problem der Identifizierbarkeit

Beispiel:



Problem der Identifizierbarkeit

Beispiel:

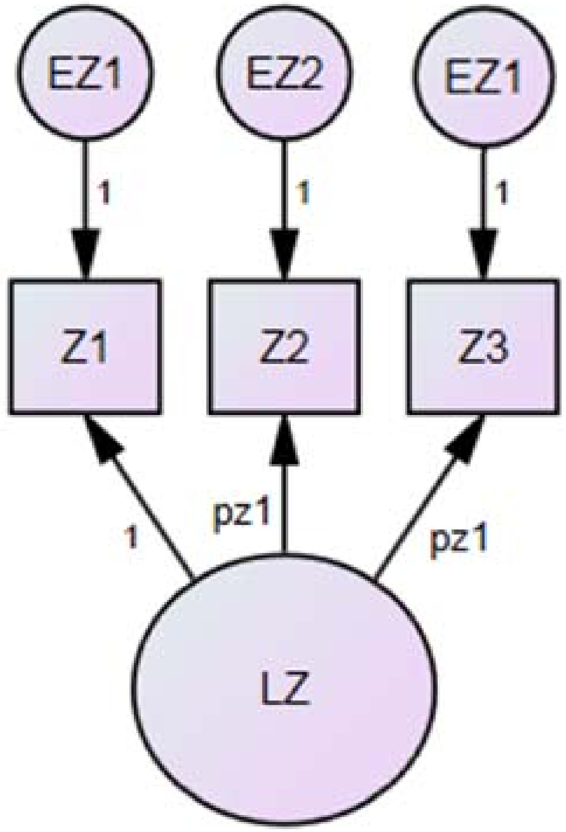


Modell-Varianz-/ Kovarianzmatrix

- Berechnung der Modell-Varianz-/ Kovarianzmatrix im Modell Z:

- Var (Z1) = Var (LZ) + Var (EZ1)
- Var (Z2) = $pz1^2 \cdot$ Var (LZ) + Var (EZ2)
- Var (Z3) = $pz1^2 \cdot$ Var (LZ) + Var (EZ1)

- Cov (Z1, Z2) = 1 · pz1 · Var (LZ)
- Cov (Z1, Z3) = 1 · pz1 · Var (LZ)
  → Cov (Z1, Z3) = Cov (Z1, Z2)
- Cov (Z2, Z3) = $pz1^2 \cdot$ Var (LZ)
  → Cov (Z2, Z3) = pz1 · Cov (Z1, Z2)



Modell-Varianz-/ Kovarianzmatrix

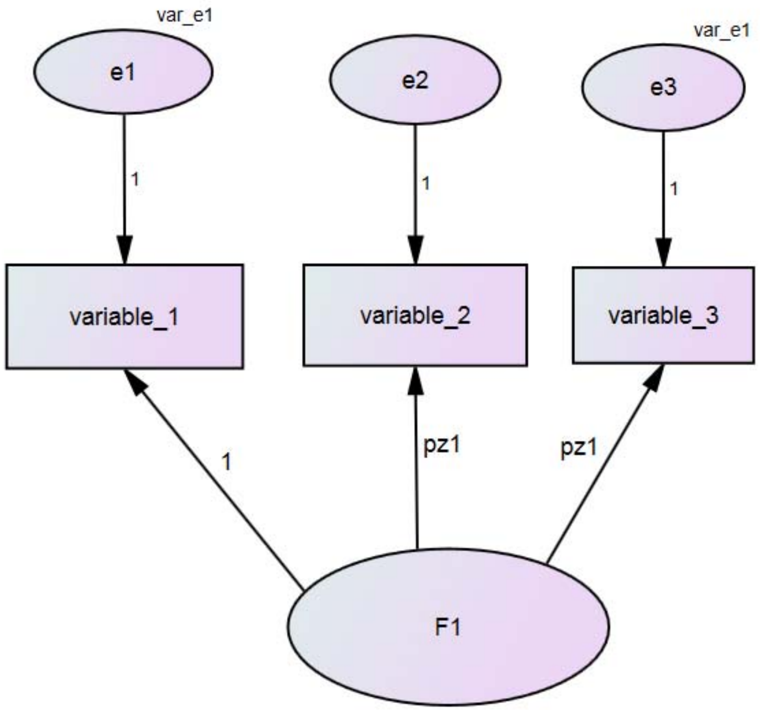
Beispiel:

Empirische Parameter:

Korrelationen

		variable_1	variable_2	variable_3
variable_1	Korrelation nach Pearson	1	,628**	,476**
	Signifikanz (2-seitig)		,000	,001
	Kovarianz	3,558	3,717	3,775
	N	42	42	42
variable_2	Korrelation nach Pearson	,628**	1	,724**
	Signifikanz (2-seitig)	,000		,000
	Kovarianz	3,717	9,859	9,558
	N	42	42	42
variable_3	Korrelation nach Pearson	,476**	,724**	1
	Signifikanz (2-seitig)	,001	,000	
	Kovarianz	3,775	9,558	17,683
	N	42	42	42

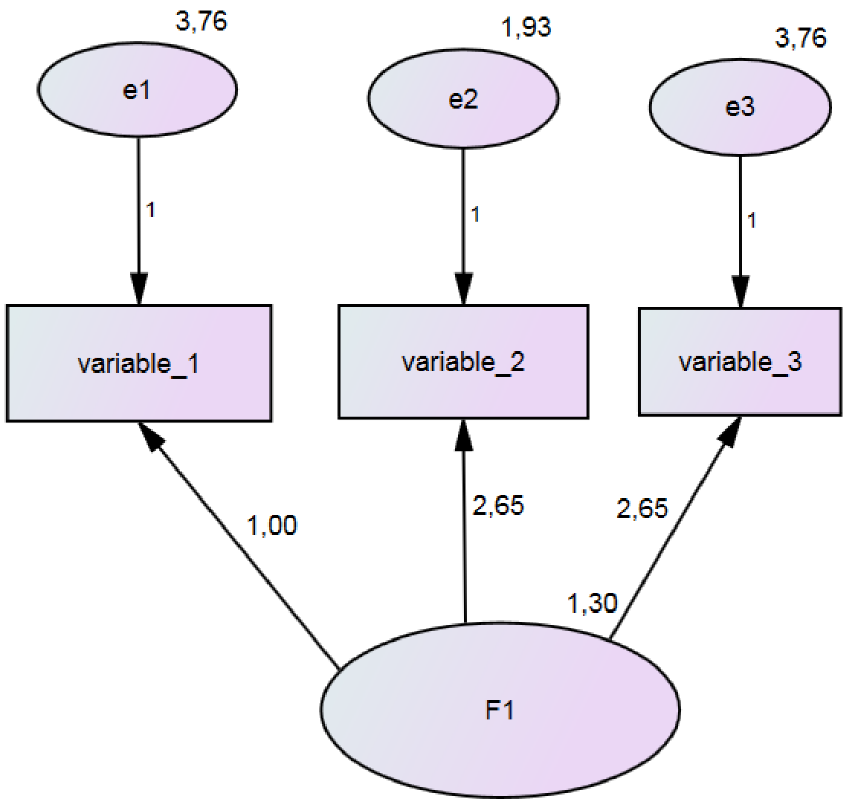
**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.



Modell-Varianz-/ Kovarianzmatrix

Beispiel:

- **Aus dem Modell geschätzte Parameter:**
- Var (V1) = 1.30 + 3.76 = **5.06**
- Var (V2) = $2.65^2 \cdot 1.30 + 1.93$ = **11.06**
- Var (V3) = $2.65^2 \cdot 1.30 + 3.76$ = **12.89**
- Cov (V1, V2) = $2.65 \cdot 1.30$ = **3.44**
- Cov (V1, V3) = $2.65 \cdot 1.30$ = **3.44**
- Cov (V2, V3) = $2.65^2 \cdot 1.30$ = **9.12**



Modell-Varianz-/ Kovarianzmatrix

Beispiel:

Vergleich Empirische Parameter - Aus dem Modell geschätzte Parameter:

- Var (V1): **3.56** **5.06**
- Var (V2): **9.86** **11.06**
- Var (V3). **17.7** **12.89**
- Cov (V1, V2): **3.72** **3.44**
- Cov (V1, V3): **3.78** **3.44**
- Cov (V2, V3): **9.56** **9.12**

Notes for Model (Default model)

Computation of degrees of freedom (Default model)

Number of distinct sample moments:	6
Number of distinct parameters to be estimated:	4
Degrees of freedom (6 - 4):	2

Result (Default model)

Minimum was achieved
Chi-square = 13,775
Degrees of freedom = 2
Probability level = ,001

Vorbereitungsschritte

- Wahl des Schätzverfahrens
- Wahl der Gütekriterien (Gesamtmodell; siehe nächste Folie)
- Wahl der auszuwertenden Teile des Modells; Wahl der Gütekriterien
- Entscheidung über Kriterien zur Modellmodifikation

Vergleich alternativer Modelle

- χ^2-Test auf Basis $\Delta\chi^2$ und Δdf beim Vergleich hierarchisch strukturierter Modelle (Modelle, in denen einzelne Beziehungen weggelassen bzw. hinzugefügt werden)

- AIC beim Vergleich echter Modellalternativen → Modell mit dem kleinsten AIC ist zu favorisieren

$$AIC = \chi^2 + 2t \text{ (t: Anzahl zu schätzender Parameter)}$$

Mehrgruppen-Kausalanalyse

Einige Typische Fragestellungen der MGKA:
(vgl. Weiber & Mühlhaus, 2010, S. 225)

- Vergleich der Basisstruktur latenter Variablen: Können in verschiedenen Gruppen die gleichen Indikatoren zur Messung der Konstrukte verwendet werden?
- Vergleich der Strukturbeziehungen zwischen latenten Variablen: Besitzen die Strukturbeziehungen in unterschiedlichen Gruppen Gültigkeit, ist die Stärke der Effekte vergleichbar?
- Vergleich der Mittelwerte von latenten Variablen: Gibt es Unterschiede in den Mittelwerten von latenten Variablen zwischen den Gruppen

Mehrgruppen-Kausalanalyse

Typisches Vorgehen bei einer MGKA:
(vgl. Weiber & Mühlhaus, 2010, S. 228ff)

- Schätzung des „unrestringierten“ (unconstrained) Modells (alle Parameter werden unabhängig für die Gruppen separat geschätzt)
- Schätzung des „restringierten“ (measurement residual) Modells (alle Parameter werden für die Gruppen gleich gesetzt)
- Vergleich der Modell-Fit-Indizes
- → falls Übereinstimmung → keinerlei gruppenspezifische Unterschiede
- → falls keine Übereinstimmung → Suche nach Unterschieden durch schrittweise Identitätsrestriktionen (in AMOS automatisiert möglich), Modellmodifikationen, spezifische Restriktionen

Einige ausgewählte Literaturhinweise

- Arbuckle, J. L. & Wothke, W. (1999). *Amos 4.0 User's Guide*. Chicago, IL.: Small Waters Corporation.
- Blunch, N.J. (2013). Introduction to Structural Equation Modeling Using IBM SPSS and AMOS(2. ed.). London: Sage.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2. Aufl.). München: Pearson-Studium.
- Byrne, B. M. (2009). *Structural Equation Modelling With Amos: Basic Concepts, Applications, and Programming* (2. ed.). London: Chapman & Hall/CRC.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden*. Weinheim: Beltz.
- Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (Methodology in the Social Sciences) (3rd ed.). New York: Guilford Press.
- Rudolf, M. & Müller, J. (2012). *Multivariate Verfahren. Eine praxisorientierte Einführung mit Anwendungsbeispielen in SPSS* (2. Aufl.). Göttingen: Hogrefe.
- Schumacker, R. E. & Lomax, R. G. (2009). *A Beginner's Guide to Structural Equation Modelling.* London: Taylor & Francis.
- Weiber, R. & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung. Eine anwendungs-orientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS* (2. Aufl.). Heidelberg: Springer.