4.2 Konfirmatorische Faktorenanalyse

- **S**tructural **E**quation **M**odelling: Lineare Strukturgleichungsmodelle (auch: Kovarianzstrukturanalyse)
- Geht zurück auf Modell gemeinsamer Faktoren & explorative ML-Faktorenanalyse
- CFA ist nicht implementiert in SPSS --> Spezialsoftware

Kernbereich der Anwendung:
- Prüfung zuvor spezifizierter theoretischer Modelle (wie gut passen Daten & Modell zusammen?
- Schließt traditionelle Methoden (EFA, multiple Regressionsanalyse) mit ein

4.2.1 Grundbegriffe SEM

SEM-Methoden gestatten die *Modellierung* der theoretischer lateneten Ebene, sowie die der manifesten Messebene. Gestattet wird auch die *Verbindung* der Ebenen & *Prüfung* der Modellanannahmen.

Vor Datenanalyse sind präzise Vorstellungen über *Beziehungen der Variablen* zu entwickeln / zu definieren. --> Haltbarkeit der Beziehungen soll anhand empirischer Befunde überprüft werden.

Aspekte SEM nach Kline:

Konfirmatorisch	Modellprüfend & zwingt zum Denken in Modellen
Manifeste vs. Latente V.	- Differenzierung zw. Variablentyp - Differenzierung wird auf manifester Ebene als Spezialfall eingeschlossen
Kovarianzmatrizen	- mathematische Theorie ist für Kovarianzen entwickelt worden - Korrelationsanalyse führt u.U. zu fehlerhaften Ergebnissen
Korrelative Designs	Untersuchung experimentell erhobener Daten ist dennoch möglich und sinnvoll
Spezialfälle	Die wichtigsten explorative Verfahren der multivariaten Statistik (EFA, multiple Regression, Varianzanalyse, Pfadanalyse) sind Spezialfälle der SEM
Große Stichproben	Werden in aller Regel benötigt
Pfaddiagramme	Theoretische Modelle werden in Pfaddiagrammen dargestellt

Pfaddiagramme

Konstrukte	Latente Variablen = Kreise/ Ellipsen & griechische Buchstaben
Indikatoren	Manifeste Variablen = Rechtecke & lateinische Buchstaben
Latente Variablen	- Abhängige/ endogene V. Werden innerhalb des Modells erklärt - Unabhängige/ exogene V. Werden nicht durch Modell erklärt
Residuen	Fehlerterme, griechische Buchstaben (Kreise, fallen oft weg)

Variablenklassen

η	Eta	Latent-endogene Variable
ξ	Ksi	Latent-exogene Variable
y		Indikatorvariable für latent-endogene Variable
x		Indikatorvariable für latent-exogene Variable
ε	Epsilon	Residualvariable für y
δ	Delta	Residualvariable für x
ζ	Zeta	Residualvariable für **η**

Pfeile = Pfade deren Parameter (Ausprägungen) vom Programm geschätzt werden, oder vom Forscher aufgrund theoretischer Überlegungen festgelegt werden

„einfache“ Pfeile	*Wirkrichtung* wird theoretisch kausal interpretiert, kann aber in korrelativen Querschnittdesigns nicht geprüft werden
Doppelpfeile	*Beziehungen* über die auch theoretisch nur korrelative (ungerichtete) Annahmen vorliegen
Griech Buchstaben	Beziehungen innerhalb bestimmter Variablenklassen und zwischen Klassen haben jeweils unterschiedliche Buchstaben
Doppelindizierung	(häufig); erste Ziffer steht für Ziel, zweite Ziffer steht für Ursprung

Ein vollständiges SEM besteht aus 3 Teilmodellen

ξ + x	Messmodell exogener Variablen	Einfache CFA
ξ + η	Strukturmodell	**(ξ + x) + (ξ + η) = CFA 2. Ordnung**
η + y	Messmodell endogener Variablen	

Matrizen-Grundgleichungen des Strukturgleichungsmodells:

$\eta = B\eta + \Gamma\xi + \zeta$	Strukturmodell
$y = \Lambda_y\eta + \varepsilon$	Messmodell latenter endogener Variablen
$x = \Lambda_x\xi + \delta$	Messmodell latenter exogener Variablen
$\eta, \xi, \zeta, y, \varepsilon, x$	Vektoren (Eta, Ksi, Zeta, y, Epsilon, x)
$B, \Gamma, \Lambda_y, \Lambda_x$	Koeffizienzmatrizen (Beta, Gamma, Lambda-y, Lambda-x)
Φ	Matrix der latenten exogenen Variablen (Phi)
$\Psi, \Theta_\varepsilon, \Theta_\delta$	Resiudual-Matrizen (Psi, Theta-Epsilon, Theta-Delta)

4.2.2 Vorgehensweise bei einer konfirmatorischen Faktorenanalyse

1. Modellspezifikation:
Der Forscher wird bei der Anwendung von SEM dazu gezwungen ein theoretisches Modell zu spezifizieren. Bei einem psychologischen Test muss also vor der SEM festgelegt werden welche Items/ Indikatoren welche latenten Variablen/ Faktoren zugeordnet sind. Weitergehend muss festgelegt werden welche Beziehungen zwischen den Faktoren bestenen und auch ob zwischen den Residuen/ Fehlertermen der Items Korrelationen zugelassen werden.

Spezifikation via SEM-Programm ist automatisch mittels Pfaddiagramm-Zeichnung. Je mehr theoretische Vorannahmen ins Modell eingehen, desto sparsamer wird es.

Arten der Spezifikation

Freie Parameter	
Fixe Parameter	Fixierung auf beliebigen Wert; häufig 0/1
Beschränkte Parameter	
Gleichheitsbeschränkung	- 2 oder mehr Koeffizienten werden vor Berechnung auf unbekannten,jedoch gleichen Wert festgelegt
Ungleichheitsbeschränkung	- Vorgabe eines bestimmten Wertebereichs < oder > 0

2. Identifizierbarkeit des Modells
Das aufgestellte Gelcihungssystem soll eine befriedigende Lösung ergeben. Es gibt unteridentifizierte Modelle, bei denen die empirische Information nicht ausreicht um einen der Parameter eindeutig zu bestimmen. Bei gerade identifizierten Modellen entspricht die Zahl der empirischen Information (6, 10) der Anzahl der Parameter (a, b). Um jedoch bestimmen zu können wie gut die Parameter mit einem theoretischen Modell übereinstimmen, benötigt man ein überidentifiziertes Modell.

Unteridentifiziert	Es gibt unendlich viele Lösungen, aber keine Gute.	$a + b = 6$
Gerade identifiziert	eindeutig lösbar, Güte: perfekt	$2a + b = 10$
Überidentifiziert	Es steht mehr empirische Information zur Verfügung, als Parameter geschätzt werden. --> Modell hat einen Freiheitsgrad	$3a + b = 12$

t-rule: notwendige, aber nicht immer hinreichende Bedingung der Identifizierbarkeit

Je mehr Freiheitsgrade vorliegen, desto strenger ist der Modelltest.

Problem der Skalierung: Metrik in der die Ladungen eines Faktors auf seine Indikatoren gemessen werden, wird durch Berechnung nur relativ zueinander festgelegt. Absolute Höhe jedoch ist willkürlich.
--> Für eine eindeutige Lösung muss ein *Indikator als Referenzvariable* ausgewählt und auf einen Wert festgelegt werden.
--> Faktor erhält die Metrik der Referenzvariable

3. Design & Datenerhebun

4.Durchführung der SEM-Analyse

4a Ausgangsmatrix & Schätzalgorithmus

Die mathematische Theorie beruht auf der Analyse von Kovarianzmatrizen (Varianzen in Diagonalen & Kovarianz ausserhalb der Diagonalen). Korrelationsmatrizen können auch verwendet werden, führen aber zu fehlerhaften Schätzungen.

Alternativen zur Kovarianzmatrix:

Dichotome Items	Matrix der tetrachorischen Korrelationen
Ratingskalen	Matrix der polychorischen Korrelationen (korrespondieren am besten im Fall 5-stufiger Ratingskalen, jedoch nicht mit dem häufigsten Schätzalgorithmus der SEM)

Steht die Ausgangsmatrix fest, besteht die Aufgabe des Schätzalgorithmus darin eine *Kovarianzmatrix $\Sigma(\Theta)$* zu finden die durch den *Vektor Θ* impliziert ist, welche jedoch auch der beobachteten Kovarianzmatrix in der *Population Σ* möglichst nahe kommt: $\boldsymbol{\Sigma = \Sigma(\Theta)}$

Populationsmatrix $\boldsymbol{\Sigma}$ kann nicht beobachtet werden, stattdessen wird die empirische Kovarianzmatrix **S** (oder Korrelationssmatrix **R**) benutzt.

Minimierende Diskrepanzfunktion (Fit-Funktion) quantifiziert die Diskrepanz zwischen empirischer & implizierter Matrix

Fit-Funktion

Non-interative Verfahren	- keine Prüfung des Fit; keine endgültige Schätzung - Berechnung von Startwerte für Parameterschätzung
Iterative Verfahren	- Endgültige Schätzung - Zu Berücksichtigen: Stichprobengröße, Skalenniveau, multivariate Normalverteilung, Matrix positiv definit, Fitfunktion skaleninvariant, Parameter skalenfrei

Positiv Definit	- Ist Matrix wenn alle Eigenwerte >0 - Fehlende positive Definitheit entsteht durch starke lineare Abhängigkeit/ Redundanzen/ paarweise (statt listenweiser) Fallausschluss --> führt dazu, dass Matrix nicht invertiert werden kann & Schätzung mittels ML/ GLS fehlschlägt
Skaleninvarianz	Werte der Diskrepanzfunktion hängen nicht von der Skalierung der beobachteten Variable ab
Skalenfreiheit	Geschätzte Parameter lassen sich nach Veränderung der Skala einer Variable wieder in ursprüngliche Metrik transformieren

Schätzalgorithmen/ häufigste iterative Schätzalgorithmen

Maximum Likelihood	- Standardmethode & meist Voreinstellung - Schätzungen sind für $N \rightarrow \infty$ asymptotisch korrekt mit asymptotisch normalverteilten Fehlern - idr skaleninvariant & skalenfrei - Voraussetzung: Multivariate Normalverteilung --> gg Verletzung recht robust bei folgenden Grenzwerten: Schiefe <2, Exzess <7
Generalized Least Squares	- ähnlich ML, in Simulationsstudien eher etwas schlechter
Unweighted Least Squares	- entspricht PFA-Variante in EFA - Weitgehend voraussetzungsfrei hinsichtlich Verteilungseigenschaften, positiver Definitheit - Nicht skaleninvariant & nicht skalenfrei - Anwendung nur auf Korrelationsmatrix & Variablen mit einheitlicher Skalierung
Asymptotically Distribution Free	- keine/ eingeschränkte Voraussetzung an Verteilungsannahmen - Skaleninvariant & skalenfrei - Prüfstatistiken - Bieten sich an zur Analyse dichotomer & kategorialer Daten - Extreme Anforderungen an Stichprobengröße

4b Modelltest

Teilbereiche der Modellprüfung

Identifikation mgl Schätzprobleme	- Schätzalgorithmus nicht konvergiert - Unzulässige Lösungen/ Haywood cases --> unmgl einzelbefunde (negative Varianzen/ Korrelation >1) - Ursachen: geringe Stichprobengröße/ Indikatorenzahl/ Faktorenladungen oder grobe Fehlspezifikation
Fit	Prüfung der Passung zwischen impliziertem & empirischem Gesamtmodell
χ^2-Test	- prüft H_0 : Modell passt zur beobachteten Datenstruktur --> signifikanter Befund führt zur Ablehnung des Modells - Ablehnungswahrscheinlichkeit erhöht sich mit Zahl der Freiheitsgrade - Nicht signifikanter Test ist Hinweis auf absoluten/ exakten Modell-Fit
Andere Fit-Indizes	- neben χ^2-Test sollten mehrere Fit-Indizies möglichst verschiedener Klassen berichtet werden
Prüfung einzelner Modellparameter	(siehe 4d Prüfung & Interpretation)

Fit-Indizes

Absolute	Das unter Restriktionen geschätzte Modell wird mit einem Modell verglichen, bei dem die Parameter frei geschätzt wurden (letzteres entspricht saturiertem Modell/ Modell ohne Freiheitsgrade)
Badness-of-Fit-Indizes	- Root Mean Square Residual, Root Mean Error of Approximation - Höhere Werte zeigen schlechteren Fit an
Goodness-of-Fit-Indizes	- GFI & adjustierteGFI - Geben Ausmaß der Verbesserung gegenüber „No model at all“ - Hohe Werte = gute Passung
Komparative	- Fit als proportionale Verbesserung gegenüber Nullmodell/ Independence Model - Alle Kovarianzen werden auf den Wert null fixiert - Hohe Werte = guter Fit - Tucker-Lewis-Index, Comparative Fit Index

Existenz äquivalenter Modelle ist eine Prüfung anderer Art und besitzt gleichen Fit wie das geprüfte Modell.

4c Modellvergleich

Multi-Gruppen-Vergleich	Es lässt sich die Übertragbarkeit eines Modells zwischen Teilgruppen im Datensatz prüfen
Nestung	Beschreibt den Fall ineinander verschachtelter Spezifikationen. Ein Modell ist mit einem anderen genestet, wenn es sich vom ersten nur durch die Einführung zusätzlicher Restriktionen unterscheidet.

Der Signifkanztest lässt sich nur bei genesteten Modellen anwenden. Anwendung findet der $\Delta\chi^2$-Test (Chi-Quadrat-Differenz-Test).

Exkurs/ Grade der Äquivalenz

Äquivalente Messung	Einzelne Teile/ Items eines Test, welche dasselben Konstruktes messen --> Die Messungen werden untereinander als austauschbar angesehen (sofern Messgelegenheiten frei von Fehlern wären)
Streng parallel	Für jede Person ist in beiden Messungen der wahre Wert & Fehler gleich hoch --> impliziert für alle Items gleiche Schwierigkeit & Trennschärfe
Essentiell parallel	Annahme gleicher Mittelwerte/ Schwierigkeiten wird aufgegeben --> wahre Wert kann um additive Komponente verschoben sein
Tau-äquivalent	Gleiche Mittelwerte bei unterschiedlichen Reliabilitäten/ Messfehlern --> Korrelationen der Testteile, die um Messfehler bereinigt wurden, sollten gleich sein --> Feststellung durch gleich hohe Ladungen je Faktor
Essentiell tau-äquivalent	Mittelwert kann um eine Konstante verschoben sein
Tau-kongerisch	Messwerte können um multiplikative Konstante verschoben sein --> Faktorladungen & Fehlerterme zwischen Items dürfen in Höhe variieren, Items dürfen jedoch nur auf einen Faktor laden --> eindimensionale Messung - häufigste Voraussetzung bei CFA

Unstandardisierte Lösung	Die Ladungen entsprechen z.B. Unstandardisierten Regressionsgewichten
Teilstandardisierte Lösung	Nur die latenten Variablen sind standardisiert, die Indikatoren nicht.
Vollständig standardisierte Lösung	Ladungsparameter lassen sich wie Musterkoeffizienten der EFA als standardisierte Regressionskoeefizienten/ Korrelationen interpretieren

Bei Analyse einer Korrelationsmatrix sind alle Werte standardisiert. Die Parameter der Pfade sind als Beziehungen auf der Konstrukteuren zu interpretieren.

Signifikanzprüfung der Unstandardisierten Lösung	- Standardfehler: Berechnung von Konfidenzintervallen; nicht auf standardisierte Lösung übertragbar - t-Werte: es gelten Signifikanzgrenzen
Modifikationsindizes (MI)	- Beurteilung der Parameter - Programme geben für jeden auf Null fixierten Parameter an, welche Verbesserung für χ^2 zu erwarten sind, wenn der betreffende Parameter freigesetzt werden würde - Sehr hohe MI: Hinweis auf grobe Fehlspezifikation - Veränderungen aufgrund MI sind rein explorativ

Bewertung SEM/ CFA:
- mathematisch sehr komplex
- Häufig unqualifizierte Anwendung, da Programme leicht bedienbar
- Zahlreiche subjektive Entscheidungen notwendig
- Wesentliche Ergänzung der deskriptiven Analysen im Rahmen der KTT
- Unverzichtbarer Bestandteil bei rational konstruierten Tests.

4.3 Aggregation zu Skalenwerten/ Normierung/ Interpretation

Nach Entscheidung welche Items eliminiert & welche Items zu Skalen zusammengefasst werden sollen, liegt nun die *Endform des Tests* inhaltlich fest.

Nächster Konstruktionsschritt: Entscheidung über die *Berechnungsvorschrift für Skalenrohwert.* Meistens berechnet sich dieser als Summe/ Mittelwert der Rohwerte aller Items einer Skala.

Mittelwert	Erlaubt Vergleich zwischen Skalen unterschiedlicher Länge, verschleiert aber gerade diesen Längenunterschied
Toleranz	In welchem Umfang sollen fehlende Werte einzelner Items bei der Berechnung von Skalenwerten toleriert werden?

Nachteil	Mögliche Verzerrung von Skalenwerten
Vorteil	Datensätze, bei denen nur ein Bruchteil der Information fehlt, bleiben erhalten

- Je mehr Items zu einer Skala zusammengefasst werden...
- Je homogener die Items einer Skala sind...
- Je geringer die Folgen einzelner Verzerrungen sind....
- Je geringer der Anteil tolerierter fehlender Werte...

... Desto mehr spricht für die Tolerierung fehlender Werte.

Normorientierte Testinterpretation

Normstichprobe	Vergleich mit Referenzstichprobe
Normierung/ Eichung	Aus Daten der Stichprobe werden Normwerte berechnet und entstehen durch Transformation der Skalenrohwerte
Relative Interpretation	Testwerte sind meist nur als quantifizierte Abweichung von einem Referenzpunkt sinnvoll zu interpretieren
Lineare Transformation	entsteht entweder unmittelbar durch z-Standardisierung, oder wird durch weitere lineare Transformation daraus abgeleitet
z-Wert	Entsteht durch Abweichungsrelativierung des individuellen Testwertes X vom Stichprobenmittelwert M an der Standardabweichung SD der Stichprobenwerte --> $z = (X-M) / SD$ - Intervallskalenniveau; Normalverteilung = Vorteile für Interpretation, aber keine Voraussetzung

Weitere gebräuchliche Skalen:
- z-Skala
- IQ-Skala
- Stanine-Skala/ Centil-Skala (9/ 11-polige Skala bei Persönlichkeitstest)
- Prozentrangnormen: Prozentsatz der Normstichprobe, welche einen geringeren/ höchstens gleich hohen Testwert erreicht. Nicht-linear, relativ verteilungsfrei, Ordinalskala

Berechnung von Skalenrohwerten	- TRANSFORMIEREN --> VARIABLE BERECHNEN --> FUNKTIONSGRUPPE: ALLE --> MEAN/ SUM - Verschieben der Funktion MEAN/ SUM in Fenster NUMERISCHER AUSDRUCK - Auswahl der Items die zur Skala gehören --> in Klammer verschieben & durch Kommata trennen - ZIELVARIABLE: Zuweisung eines Variablennamens zum Skalenwert (Erläuterung via TYP&LABEL --> BESCHRIFTUNG

Berechnung von Skalenrohwerten	- TRANSFORMIEREN --> VARIABLE BERECHNEN --> FUNKTIONSGRUPPE: ALLE --> MEAN/ SUM - Verschieben der Funktion MEAN/ SUM in Fenster NUMERISCHER AUSDRUCK - Auswahl der Items die zur Skala gehören --> in Klammer verschieben & durch Kommata trennen - ZIELVARIABLE: Zuweisung eines Variablennamens zum Skalenwert (Erläuterung via TYP&LABEL --> BESCHRIFTUNG

Kriterienorientierte Testinterpretation

Cut-off	Testergebnis und daran geknüpfte Konsequenzen ändern sich grundlegend bei über-/ Unterschreitung eines Schwellenwertes
Dichotome Entscheidungen	Liegt ein bestimmter Sachverhalt vor: ja/ nein --> Vier-Felder-Schema der Fehlerrisiken

Mögliche „Diagnosen“ beim Vier-Felder-Schema der Fehlerrisiken:
- RP/ Hit: Kriterium erfüllt, Klassifikation durch Test
- FP/ False Alarm: Kriterium nicht erfüllt, Klassifikation durch Test
- FN/ Misses: Kriterium erfüllt, keine Klassifikation durch Test
- RN/ Correct Rejection: Kriterium nicht erfüllt, keine Klassifikation durch Test

--> Formale Definitionen für bestimmte Anteile richtiger/ falscher Entscheidungen

Sensitivität	Trefferquote = RP/ (FN+RP)
1- Sensitivität	Verpasserquote = FN/ (FN+RP)
Spezifität	Quote korrekter Ablehnungen = RN / (FP+RN)
1- Spezifität	Quote falscher Alarme = FP / (FP+RN)