Faktorenanalyse und Strukturgleichungsmodelle I

Kai Arzheimer | Vorlesung Forschungsmethoden

Outline

Einführung: Faktorenanalyse

Grundlagen Strukturgleichungsmodellierung

Zwischenfazit

Karl G. Joreskog

Einführung: Faktorenanalyse

Was ist ein „Faktor“?

- „Faktor“ oder „latente Variable“
  - Common factor(s) = Gemeinsamkeit(en) zwischen verschiedenen Variablen
  - D.h. nicht direkt beobachtbare Größe
  - Die beobachtbare Variablen („Indikatoren“) beeinflußt
- Typische Beispiele: Einstellungen, Persönlichkeitsmerkmale
- Grundidee: Muster in beobachteten Korrelationsmatrizen erklären
- Psychologische Intelligenzforschung

Beispiel: Politisches Vertrauen

- ESS 4, Deutschland
- Nationales Parlament, Gerichte, Polizei, Politiker, Parteien, Europaparlament

```
. corr trstprl-trstep
(obs=2492)

             |  trstprl  trstlgl  trstplc  trstplt  trstprt   trstep
-------------+------------------------------------------------------
     trstprl |   1.0000
     trstlgl |   0.5366   1.0000
     trstplc |   0.3796   0.6029   1.0000
     trstplt |   0.6693   0.4508   0.3446   1.0000
     trstprt |   0.5912   0.4160   0.3025   0.8061   1.0000
      trstep |   0.5563   0.4146   0.3122   0.5577   0.5898   1.0000
```

Grundidee der (explorativen) Faktorenanalyse

- Konstruiere eine oder mehrere (wenige) „künstliche“ Variablen die Ausgangsmatrix „erklären“ können
- Jede gemessene Variable sollte hoch mit *einem* Faktor korrelieren
  - Korrelation mit Faktor: „erklärte“ Varianz
  - Rest: „Störvarianz“
- Regression der Ausgangsvariablen auf Faktoren
- Varianten:
  - Wie werden Faktoren konstruiert?
  - Wieviele Faktoren werden extrahiert?
  - Können Faktoren untereinander korrelieren?
- Explorativ $\approx$ atheoretisch

Beispiel Vertrauen: 1-Faktor-Lösung

```
. factor trstprl-trstep,factor(1)
(obs=2492)

Factor analysis/correlation                        Number of obs    =     2492
    Method: principal factors                      Retained factors =        1
    Rotation: (unrotated)                          Number of params =        6

    --------------------------------------------------------------------------
         Factor  |   Eigenvalue   Difference        Proportion   Cumulative
    -------------+------------------------------------------------------------
        Factor1  |      3.10719      2.62503            0.9665       0.9665
        Factor2  |      0.48216      0.47331            0.1500       1.1165
        Factor3  |      0.00885      0.04933            0.0028       1.1193
        Factor4  |     -0.04048      0.10374           -0.0126       1.1067
        Factor5  |     -0.14422      0.05452           -0.0449       1.0618
        Factor6  |     -0.19874            .           -0.0618       1.0000
    --------------------------------------------------------------------------

    LR test: independent vs. saturated:  chi2(15) = 7563.17 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances

    ---------------------------------------
        Variable |  Factor1 |   Uniqueness
    -------------+----------+--------------
         trstprl |   0.7674 |       0.4111
         trstlgl |   0.6564 |       0.5692
         trstplc |   0.5234 |       0.7261
         trstplt |   0.8411 |       0.2926
         trstprt |   0.8075 |       0.3480
          trstep |   0.6739 |       0.5459
    ---------------------------------------
```

Warum kann explorative Faktorenanalyse problematisch sein?

„factor analysis: it’s what the data get into when theory goes on holiday“

- Bezieht sich auf explorative (= böse?) Faktorenanalyse
- Standardprozedur („little jiffy“); Extraktion von Hauptkomponenten, Kaiserkriterium, VARIMAX
- Findet oft vorhandene Strukturen nicht („Tom Swift’s Electric Factor Analysis Machine“)
- Findet Strukturen, die gar nicht vorhanden sind?
- Konfirmatorische Faktorenanalyse: Prüfung, ob theoretisch sinnvolle Strukturen mit den Daten vereinbar sind
- Meßmodelle, Erweiterung möglich („LISREL“-Modell)

Grundlagen Strukturgleichungsmodellierung

Was gehört zur Spezifikation?

1. Zahl der gemeinsamen Faktoren
2. Zahl der beobachteten Variablen
3. Varianzen/Kovarianzen der gemeinsamen Faktoren
4. Beziehungen zwischen gemeinsamen Faktoren und beobachteten Variablen
5. Beziehungen zwischen beobachteten Variablen und spezifischen Faktoren
6. Varianzen/Kovarianzen der spezifischen Faktoren (Meßfehler)

Was gehört zur Spezifikation?

1. Zahl der gemeinsamen Faktoren
2. Zahl der beobachteten Variablen
3. Varianzen/Kovarianzen der gemeinsamen Faktoren
4. Beziehungen zwischen gemeinsamen Faktoren und beobachteten Variablen
5. Beziehungen zwischen beobachteten Variablen und spezifischen Faktoren
6. Varianzen/Kovarianzen der spezifischen Faktoren (Meßfehler)

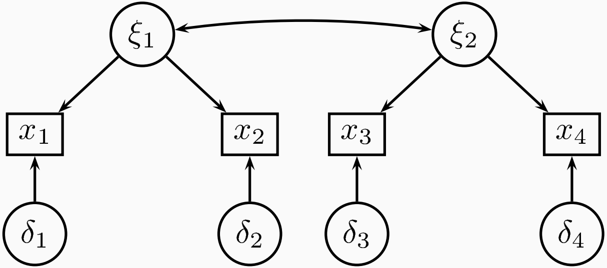
- Spezifikation ursprünglich durch eine Reihe von Matrizen

Was gehört zur Spezifikation?

1. Zahl der gemeinsamen Faktoren
2. Zahl der beobachteten Variablen
3. Varianzen/Kovarianzen der gemeinsamen Faktoren
4. Beziehungen zwischen gemeinsamen Faktoren und beobachteten Variablen
5. Beziehungen zwischen beobachteten Variablen und spezifischen Faktoren
6. Varianzen/Kovarianzen der spezifischen Faktoren (Meßfehler)

- Spezifikation ursprünglich durch eine Reihe von Matrizen
- Heute
  - Einfache Gleichungen oder graphische Eingabe
  - (Normalerweise) vernünftige Voreinstellungen

Wie werden die Beziehungen graphisch dargestellt?



- Kreise oder Ovale für latente Variablen/Meßfehler (ξ, δ)
- Rechtecke oder Quadrate für beobachtete Variablen (x)
- Gerichtete Pfeile für (kausale) Wirkungen
- Doppelpfeile für Kovarianzen/Korrelationen
- Kausalität?

Wie sieht die Terminologie aus?

- Primär interessant: Beziehungen zwischen gemeinsamen Faktoren und beobachteten Variablen
- Werden mit λ bezeichnet
- Z. B. $x_2 = \lambda_{21}\xi_1 + \delta_2$
- λ_{21} gibt an, wie sich x_2 verändert, wenn Faktor um eins zunimmt
- Analog zu $y = \beta_0 + \beta_1 x_1 + \epsilon$

Wie sieht die Terminologie aus?

- Primär interessant: Beziehungen zwischen gemeinsamen Faktoren und beobachteten Variablen
- Werden mit λ bezeichnet
- Z. B. $x_2 = \lambda_{21}\xi_1 + \delta_2$
- λ_{21} gibt an, wie sich x_2 verändert, wenn Faktor um eins zunimmt
- Analog zu $y = \beta_0 + \beta_1 x_1 + \epsilon$
- Alle (latente und manifeste) Variablen sind zentriert $\rightarrow$ Mittelwert von null, kein Achsenabschnitt notwendig

Wie sieht die Terminologie aus?

- Kovarianzen zwischen gemeinsamen Faktoren möglich, z. B. ϕ_{12}
- Kovarianzen zwischen spezifischen Fehlervarianzen möglich z. B. θ_{24}
- Unterschied zu explorativer Analyse?

Noch mehr Terminologie?

Matrix	Dimension	Mittelwert	Kovarianz	Dimension	Bedeutung
ξ	$(s \times 1)$	0	$\Phi = \mathrm{E}(\xi\xi')$	$(s \times s)$	Gemeinsame Faktoren
$\boldsymbol{x}$	$(q \times 1)$	0	$\Sigma = \mathrm{E}(\boldsymbol{x}\boldsymbol{x}')$	$(q \times q)$	beobachtete Variablen
Λ	$(q \times s)$				Faktorladungen
δ	$(q \times 1)$	0	$\Theta = \mathrm{E}(\delta\delta')$	$(q \times q)$	Meßfehler

- Faktor-Gleichung: $\boldsymbol{x} = \Lambda\xi + \delta$

Noch mehr Terminologie?

Matrix	Dimension	Mittelwert	Kovarianz	Dimension	Bedeutung
ξ	$(s \times 1)$	0	$\Phi = \mathrm{E}(\xi\xi')$	$(s \times s)$	Gemeinsame Faktoren
$\boldsymbol{x}$	$(q \times 1)$	0	$\Sigma = \mathrm{E}(\boldsymbol{xx}')$	$(q \times q)$	beobachtete Variablen
Λ	$(q \times s)$				Faktorladungen
δ	$(q \times 1)$	0	$\Theta = \mathrm{E}(\delta\delta')$	$(q \times q)$	Meßfehler

- Faktor-Gleichung: $\boldsymbol{x} = \Lambda\xi + \delta$
- Kovarianz-Gleichung: $\Sigma = \Lambda\Phi\Lambda' + \Theta$

Noch mehr Terminologie?

Matrix	Dimension	Mittelwert	Kovarianz	Dimension	Bedeutung
ξ	$(s \times 1)$	0	$\Phi = \mathrm{E}(\xi\xi')$	$(s \times s)$	Gemeinsame Faktoren
$\boldsymbol{x}$	$(q \times 1)$	0	$\Sigma = \mathrm{E}(\boldsymbol{x}\boldsymbol{x}')$	$(q \times q)$	beobachtete Variablen
Λ	$(q \times s)$				Faktorladungen
δ	$(q \times 1)$	0	$\Theta = \mathrm{E}(\delta\delta')$	$(q \times q)$	Meßfehler

- Faktor-Gleichung: $\boldsymbol{x} = \Lambda\xi + \delta$
- Kovarianz-Gleichung: $\Sigma = \Lambda\Phi\Lambda' + \Theta$
- Annahmen:

Noch mehr Terminologie?

Matrix	Dimension	Mittelwert	Kovarianz	Dimension	Bedeutung
ξ	$(s \times 1)$	0	$\Phi = \mathrm{E}(\xi\xi')$	$(s \times s)$	Gemeinsame Faktoren
$\boldsymbol{x}$	$(q \times 1)$	0	$\Sigma = \mathrm{E}(\boldsymbol{x}\boldsymbol{x}')$	$(q \times q)$	beobachtete Variablen
Λ	$(q \times s)$				Faktorladungen
δ	$(q \times 1)$	0	$\Theta = \mathrm{E}(\delta\delta')$	$(q \times q)$	Meßfehler

- Faktor-Gleichung: $\boldsymbol{x} = \Lambda\xi + \delta$
- Kovarianz-Gleichung: $\Sigma = \Lambda\Phi\Lambda' + \Theta$
- Annahmen:
    - Variablen sind zentriert: $\mathrm{E}(\xi) = \mathrm{E}(x) = \mathrm{E}(\delta) = 0$

Noch mehr Terminologie?

Matrix	Dimension	Mittelwert	Kovarianz	Dimension	Bedeutung
ξ	$(s \times 1)$	0	$\Phi = \mathrm{E}(\xi\xi')$	$(s \times s)$	Gemeinsame Faktoren
$\boldsymbol{x}$	$(q \times 1)$	0	$\Sigma = \mathrm{E}(\boldsymbol{xx}')$	$(q \times q)$	beobachtete Variablen
Λ	$(q \times s)$				Faktorladungen
δ	$(q \times 1)$	0	$\Theta = \mathrm{E}(\delta\delta')$	$(q \times q)$	Meßfehler

- Faktor-Gleichung: $\boldsymbol{x} = \Lambda\xi + \delta$
- Kovarianz-Gleichung: $\Sigma = \Lambda\Phi\Lambda' + \Theta$
- Annahmen:
    - Variablen sind zentriert: $\mathrm{E}(\xi) = \mathrm{E}(x) = \mathrm{E}(\delta) = 0$
    - q: Zahl der beobachteten Variablen; s: Zahl der gemeinsamen Faktoren; $q > s$

Noch mehr Terminologie?

Matrix	Dimension	Mittelwert	Kovarianz	Dimension	Bedeutung
$\boldsymbol{\xi}$	$(s \times 1)$	0	$\boldsymbol{\Phi} = \mathrm{E}(\boldsymbol{\xi}\boldsymbol{\xi}')$	$(s \times s)$	Gemeinsame Faktoren
$\boldsymbol{x}$	$(q \times 1)$	0	$\boldsymbol{\Sigma} = \mathrm{E}(\boldsymbol{x}\boldsymbol{x}')$	$(q \times q)$	beobachtete Variablen
$\boldsymbol{\Lambda}$	$(q \times s)$				Faktorladungen
$\boldsymbol{\delta}$	$(q \times 1)$	0	$\boldsymbol{\Theta} = \mathrm{E}(\boldsymbol{\delta}\boldsymbol{\delta}')$	$(q \times q)$	Meßfehler

- Faktor-Gleichung: $\boldsymbol{x} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta}$
- Kovarianz-Gleichung: $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$
- Annahmen:
  - Variablen sind zentriert: $\mathrm{E}(\xi) = \mathrm{E}(x) = \mathrm{E}(\delta) = 0$
  - q: Zahl der beobachteten Variablen; s: Zahl der gemeinsamen Faktoren; $q > s$
  - Keine Korrelation zwischen Faktoren/Fehlervarianzen: $\mathrm{E}(\xi\delta') = 0$

Was macht man nun damit?

- Mit diesen sieben Matrizen/Vektoren kann das ganze Modell vollständig beschrieben werden
- Pfeile in graphischer Darstellung entsprechen Bedingungen (constraints) in Matrizen
  - Für Items, die *nicht* auf einen Faktor laden, wird entsprechende Zelle in Λ auf null gesetzt
  - Keine Kovarianz zwischen gemeinsamen Faktoren: (Redundante) Elemente in Φ auf null
  - Keine korrelierten Meßfehler: Alle Elemente außerhalb Diagonale in Θ auf null

Wie kommt man zu den Schätzungen?

- Konfirmatorische Faktorenanalyse geht immer von Varianz-Kovarianz-Matrix aus
- Originaldaten werden nicht benötigt
- Schlüssel ist die Kovarianz-Gleichung, die sich auf die Grundgesamtheit bezieht
- Gestattet Zerlegung der Kovarianzen in Werte für Pfade ($\mathbf{\Lambda}$, $\mathbf{\Phi}$, $\mathbf{\Theta}$)
- Beobachtete Kovarianzen $\boldsymbol{S}$ als Schätzung für $\mathbf{\Sigma}$
- Schätzung setzt Identifikation voraus
- Identifikation notwendige, aber nicht hinreichende Bedingung für gültige Schätzung

Zwischenfazit

Zusammenfassung für heute

- Faktorenanalyse mächtiges Verfahren, Vielzahl von Möglichkeiten
- Besonders adäquat für sozialwissenschaftliche Daten
- Explorative Analyse explorativ (und unzuverlässig)
- Konfirmatorische Analyse besonders adäquat für sozialwissenschaftliche Theorien
    - Meßfehler
    - Latente Variablen
    - Modellierung (kleiner) Systeme

Faktorenanalyse und Strukturgleichungsmodelle II

Kai Arzheimer | Vorlesung Forschungsmethoden

Outline

- Fortsetzung SEM
- Anwendungen
  - (Likert)-Skalierung
  - Komplexe Modelle
  - Software
- Beispiel: Einstellungen zu Migranten
- Fazit

Karl G. Joreskog

Fortsetzung SEM

Was ist Identifikation?

- Identifikation ist Eigenschaft des Modells
- Modell identifiziert: Es gibt genau *eine* Lösung für Gleichungssystem
- Modell nicht identifiziert: Es existieren mehrere (oft unendlich viele) gleichwertige Lösungen
- (Notwendig, nicht hinreichend): Zahl der zu schätzenden Parameter kleiner als Zahl der unabhängigen Informationen

- Woher weiß ich, daß Modell identifiziert ist?

Was ist Identifikation?

- Identifikation ist Eigenschaft des Modells
- Modell identifiziert: Es gibt genau *eine* Lösung für Gleichungssystem
- Modell nicht identifiziert: Es existieren mehrere (oft unendlich viele) gleichwertige Lösungen
- (Notwendig, nicht hinreichend): Zahl der zu schätzenden Parameter kleiner als Zahl der unabhängigen Informationen
- Voraussetzungen/constraints
    1. Nicht alle Items können auf alle Faktoren laden $\rightarrow$ theoretische Überlegungen
    2. Die Metrik der gemeinsamen Faktoren muß festgelegt werden
- Woher weiß ich, daß Modell identifiziert ist?

Warum ist die Skalierung der Faktoren wichtig?

- Große Varianz des Faktors/niedrige Faktorladung und kleine Varianz/hohe Faktorladung im Muster der Kovarianzen nicht zu unterscheiden

Warum ist die Skalierung der Faktoren wichtig?

- Große Varianz des Faktors/niedrige Faktorladung und kleine Varianz/hohe Faktorladung im Muster der Kovarianzen nicht zu unterscheiden
- Varianz des Faktors empirisch nicht zu bestimmen (latente Variable)

Warum ist die Skalierung der Faktoren wichtig?

- Große Varianz des Faktors/niedrige Faktorladung und kleine Varianz/hohe Faktorladung im Muster der Kovarianzen nicht zu unterscheiden
- Varianz des Faktors empirisch nicht zu bestimmen (latente Variable)
- Willkürliche Festlegung (das ist *kein* Problem)

Wie kann die Metrik eines Faktors festgelegt werden?

1. Die Ladung *eines* Items auf den Faktor wird auf eins gesetzt
    - Faktor hat gleiche Metrik wie betreffendes Item
    - Varianz des Faktors wird geschätzt

Wie kann die Metrik eines Faktors festgelegt werden?

1. Die Ladung *eines* Items auf den Faktor wird auf eins gesetzt
    - Faktor hat gleiche Metrik wie betreffendes Item
    - Varianz des Faktors wird geschätzt
2. Die Varianz des Faktors wird auf eins gesetzt
    - Alle Faktorladungen werden geschätzt
    - Faktor ist dimensionslose vollstandardisierte Variable
    - Kovarianzen zwischen solchen Faktoren sind Korrelationen

Wie werden die Parameter nun geschätzt?

- Iterative Verfahren (u. a. GLS, ML, WLS)
- Für $\mathbf{\Lambda}$, $\mathbf{\Phi}$, $\mathbf{\Theta}$ Startwerte annehmen $\rightarrow$ implizieren eine Varianz-Kovarianz-Matrix $\mathbf{\Sigma}^*$
- „Differenz“ (fitting function) zwischen $\mathbf{\Sigma}^*$ und $\boldsymbol{S}$ minimieren
- Verschiedene fitting functions für verschiedene (asymptotisch äquivalente) Verfahren
- Multivariate Normalverteilung der x-Variablen wird vorausgesetzt $\rightarrow$ Verletzungen, Asymptotik?
- Tests für Koeffizienten/Pfade: Wald oder LR, Lagrange-Multiplier
- Heywood Cases

Modelfit

- Wie gut passen Modellschätzungen zu den Daten?
- Traditionell: Test auf Basis eines χ^2-Wertes
  - Wie wahrscheinlich sind Abweichungen zwischen wahrer ($\boldsymbol{\Sigma}$) und vom Modell implizierter Kovarianzmatrix ($\boldsymbol{\Sigma}(\boldsymbol{\Theta})$) …
  - … wenn Modell in GG so zutrifft?
  - (Da $\boldsymbol{\Sigma}$ nicht bekannt, $\boldsymbol{S}$ als Schätzung)
- Probleme?

Modelfit

- Wie gut passen Modellschätzungen zu den Daten?
- Traditionell: Test auf Basis eines χ^2-Wertes
  - Wie wahrscheinlich sind Abweichungen zwischen wahrer ($\boldsymbol{\Sigma}$) und vom Modell implizierter Kovarianzmatrix ($\boldsymbol{\Sigma}(\boldsymbol{\Theta})$) …
  - … wenn Modell in GG so zutrifft?
  - (Da $\boldsymbol{\Sigma}$ nicht bekannt, $\boldsymbol{S}$ als Schätzung)
- Probleme?
  - Annahme $\boldsymbol{\Sigma} = \boldsymbol{S}$ unrealistisch
  - Annahme perfekte Reproduktion (Nullhypothese) unrealistisch
  - Signifikanztests und Fallzahlen

Modelfit

- Wie gut passen Modellschätzungen zu den Daten?
- Traditionell: Test auf Basis eines χ^2-Wertes
    - Wie wahrscheinlich sind Abweichungen zwischen wahrer ($\mathbf{\Sigma}$) und vom Modell implizierter Kovarianzmatrix ($\mathbf{\Sigma}(\mathbf{\Theta})$) ...
    - ... wenn Modell in GG so zutrifft?
    - (Da $\mathbf{\Sigma}$ nicht bekannt, $\boldsymbol{S}$ als Schätzung)
- Probleme?
    - Annahme $\mathbf{\Sigma} = \boldsymbol{S}$ unrealistisch
    - Annahme perfekte Reproduktion (Nullhypothese) unrealistisch
    - Signifikanztests und Fallzahlen
- $\rightarrow$ Anpassungsindizes

Anpassungsindizes

- Vielzahl von Anpassungsindizes („shot gun approach“)
- Momentan populär: RMSEA
  - Weniger abhängig von Stichprobenumfang
  - Relativiert Abweichungen an der Zahl der Freiheitsgrade $\rightarrow$ Strafe für overfitting
  - Geht nicht von perfekter Anpassung aus; Konfidenzintervall
- Wert möglichst klein; Daumenregel: $\leqslant 0.1$, besser noch $\leqslant 0.08$

Anwendungen

Reliabilität und Dimensionalität

Reliabilität: Zuverlässigkeit, Konsistenz einer Messung

Skala: Instrument mit mehreren äquivalenten Einzelindikatoren

Interne Konsistenz: Korrelation der Indikatoren untereinander

- Einfache SEM (konfirmatorische Faktorenanalyse) $\leftrightarrows$ viele Probleme der Skalierung
  - Konsistenz und schwache Items
  - Eine oder mehrere Dimensionen
  - Meßfehlerbereinigte Korrelation zwischen Konstrukten
  - Stabilität von (wahren) Meßwerten über die Zeit

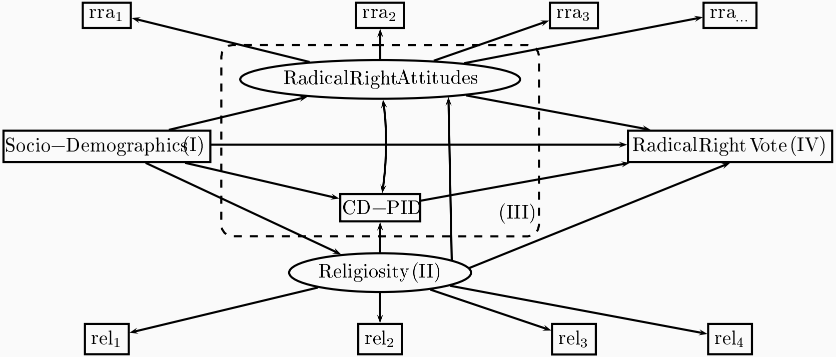
Gruppenvergleiche und interkulturelle Einstellungsforschung

- SEM ermöglicht systematischen Vergleich von Gruppen (z. B. Test von Gleichheitsrestriktionen)
- Funktionieren Instrumente in verschiedenen Ländern in gleicher Weise (z. B. ESS, ISSP); bspw. verschiedene Indikatoren/Aspekte von Nationalismus/Patriotismus (Davidov 2009)?
- Definitionen von Äquivalenz
  1. Configural Invariance: Indikatoren laden überall auf selbe Faktoren (gleiche Grafik)
  2. Metric Invariance: Faktorladungen in allen Ländern gleich (mittleres Niveau kann sich trotzdem unterscheiden)
  3. Scalar Invariance: Zusätzlich Achsenabschnitte der Indikatoren über Länder identisch $\rightarrow$ Absolute Werte von Indikatoren/Faktoren vergleichbar, Fehlervarianzen können sich unterscheiden

Komplexes Beispiel: Arzheimer/Carter 2009

- Komplexe SEM bilden Systeme von Hypothesen ab
- Z. B. Arzheimer/Carter 2009: Negative Einstellungen zu Migranten, Religiosität, Parteiidentifikation und Rechtswahl
- In verschiedenen europäischen Ländern (mehrere Gruppen)

Mega-Monster-Modell (vereinfacht)



LISREL

- Für viele Synonym für Strukturgleichungsmodelle
- Lange Zeit unzugänglich wg Syntax
- Seit 90er Jahre graphische Oberfläche und vereinfachte Syntax (SIMPLIS)
- Zahlreiche neue Erweiterungen
- Test- und Studierendenversionen hier: `http://www.ssicentral.com/lisrel/`

EQS

- Peter Bentler
- Kleine Unterschiede im Modell, äquivalent zu LISREL
- In Deutschland wenig verbreitet
- Entwicklung seit Jahren stockend
- `http://www.mvsoft.com/eqs60.htm`

AMOS

- James Arbuckle
- Technische Besonderheiten und graphische Oberfläche (ca. 1988)
- Später Modul für SPSS
- Inzwischen Übernahme durch IBM
- `http://www.amosdevelopment.com/`

MPlus

- Bengt Muthén
- Strukturgleichungsmodelle als Spezialfall eines generelleren Modells
- Transparente Behandlung von kategorialen Variablen (Indikatoren und Faktoren)
- Relativ altmodische Oberfläche, aber sehr schnell und leistungsfähig
- `http://www.statmodel.com/demo.shtml`

Weitere

- Mehrere Implementationen in R, vor allem lavaan, `http://lavaan.ugent.be/`
- SmartPLS, `http://www.smartpls.de`
- Ox (*nur für hardcore user*, `http://www.doornik.com/ox/`)
- *Sehr gute Implementation ab Stata 12*

Beispiel: Einstellungen zu Migranten

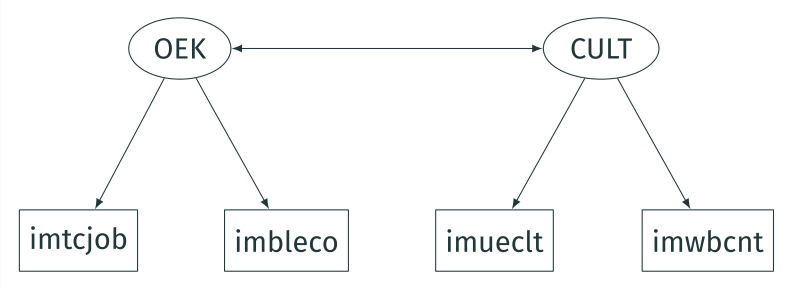
Dimensionalität

- Theorie: Unterscheidung zwischen kulturellen und ökonomischen Bedrohungsgefühlen
- Aber: Findet sich das so in der Realität? $\rightarrow$ Forschungsfrage
- Kann zum Beispiel mit ESS 1 untersucht werden

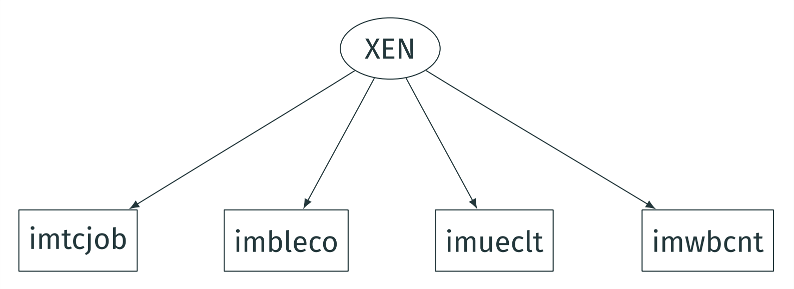
Indikatoren

1. „Was würden Sie sagen, nehmen Zuwanderer, die hierher kommen, im Allgemeinen Arbeitnehmern in Deutschland die Arbeitsplätze weg oder helfen sie im Allgemeinen, neue Arbeitsplätze zu schaffen?“ (nehmen Arbeitsplätze weg (10) – schaffen neue Arbeitsplätze (0), *imtcjob*)
2. „Die meisten Zuwanderer, die hierher kommen, arbeiten und zahlen Steuern. Sie nehmen außerdem das Gesundheitssystem und Sozialleistungen in Anspruch. Wenn Sie abwägen, denken Sie, dass Zuwanderer mehr bekommen als sie geben, oder mehr geben, als sie bekommen?“ (bekommen mehr (10) – geben mehr (0) *imbleco*)
3. „Würden Sie sagen, dass das kulturelle Leben in Deutschland im Allgemeinen durch Zuwanderer untergraben oder bereichert wird?“ (untergraben (10) – bereichert (0) *imueclt*)
4. „Wird Deutschland durch Zuwanderer zu einem schlechteren oder besseren Ort zum Leben?“ (schlechterer Ort (10) – besserer Ort (0) *imwbcnt*)

Zweidimensionales Modell



Eindimensionales Modell



In Stata

```
* Zweidimensionales Modell, ML-Schaetzung, standardisierte Faktoren
sem (imtcjob<-OEK) (imbleco<-OEK) (imueclt<-CULT) ///
  (imwbcnt<-CULT) ,   means(OEK@0 CULT@0) var(OEK@1 CULT@1)
* Schaetzungen speichern
est store zweiml
estat eqgof
estat gof, stats(ic rmsea)
* Eindimensionales Modell, ML-Schaetzung, standardisierte Faktoren
sem (imtcjob<-XEN) (imbleco<-XEN) (imueclt<-XEN) ///
  (imwbcnt<-XEN) , means(XEN@0) var(XEN@1)
est store einml
estat eqgof
estat gof, stats(ic rmsea)
```

Ergebnisse: ein/zweidimensionale Lösung

Ladung/Achsenabschnitt	
imtcjob: OEK	1.321***
	(0.0522)
imtcjob: Konstante	4.372***
	(0.0472)
imbleco: OEK	1.296***
	(0.0542)
imbleco: Konstante	4.019***
	(0.0496)
imueclt: CULT	1.603***
	(0.0556)
imueclt: Konstante	6.308***
	(0.0548)
imwbcnt: CULT	1.809***
	(0.0532)
imwbcnt: Konstante	4.997***
	(0.0521)
cov(OEK,CULT)	0.826***
	(0.0239)

Ladung/Achsenabschnitt	
imtcjob: XEN	1.168***
	(0.0487)
imtcjob: Konstante	4.372***
	(0.0472)
imbleco: XEN	1.148***
	(0.0511)
imbleco: Konstante	4.019***
	(0.0496)
imueclt: XEN	1.603***
	(0.0533)
imueclt: Konstante	6.308***
	(0.0548)
imwbcnt: XEN	1.749***
	(0.0503)
imwbcnt: Konstante	4.997***
	(0.0521)

Modellvergleich

Anpassungsmaße	ein Faktor	zwei Faktoren
Parameter	12	13
df_M	2	1
χ^2_M	63.9	12.8
RMSEA	.137	.0843
TLI	.897	.961
BIC	26842	26799
LL	-13377	-13351
N	1659	1659

Fazit

Zusammenfassung

- SEM: Extrem mächtiges Verfahren (mit diversen Fallstricken)
- Kein Wundermittel
- Aber ausgezeichnetes Werkzeug zur Bearbeitung vieler sozialwissenschaftlicher Fragen

$$y = \frac{\log_e\left(\frac{x}{m} - sa\right)}{r^2}$$

$$yr^2 = \log_e\left(\frac{x}{m} - sa\right)$$

$$e^{yr^2} = \frac{x}{m} - sa$$

$$me^{yr^2} = x - msa$$

$$me^{rry} = x - mas$$