Faktorenanalyse und Strukturgleichungsmodelle I

Kai Arzheimer | Vorlesung Forschungsmethoden

Outline

Karl G. Joreskog

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
```

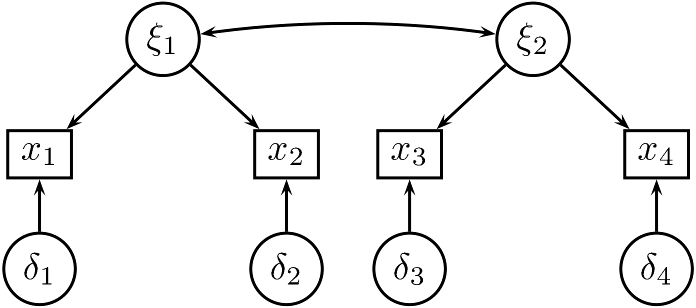
Variable	Factor1	Uniqueness
trstprl	0.7674	0.4111
trstlgl	0.6564	0.5692
trstplc	0.5234	0.7261
trstplt	0.8411	0.2926
trstprt	0.8075	0.3480
trstep	0.6739	0.5459

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)

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
- Kausalität?

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 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	$\mathbf{\Phi} = \mathrm{E}(\xi\xi')$	$(s \times s)$	Gemeinsame Faktoren
$\boldsymbol{x}$	$(q \times 1)$	0	$\mathbf{\Sigma} = \mathrm{E}(\boldsymbol{xx}')$	$(q \times q)$	beobachtete Variablen
$\mathbf{\Lambda}$	$(q \times s)$				Faktorladungen
δ	$(q \times 1)$	0	$\mathbf{\Theta} = \mathrm{E}(\delta\delta')$	$(q \times q)$	Meßfehler

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

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: $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$

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:

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: $\mathrm{E}(\xi) = \mathrm{E}(x) = \mathrm{E}(\delta) = 0$

Noch mehr Terminologie?

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

- Faktor-Gleichung: $\boldsymbol{x} = \mathbf{\Lambda}\xi + \delta$
- Kovarianz-Gleichung: $\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Theta}$
- 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 ($\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Theta}$)
- Beobachtete Kovarianzen $\boldsymbol{S}$ als Schätzung für $\boldsymbol{\Sigma}$
- Schätzung setzt Identifikation voraus
- Identifikation notwendige, aber nicht hinreichende Bedingung für gültige Schätzung

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

Zusammenfassung für die nächste Woche

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

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

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

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

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