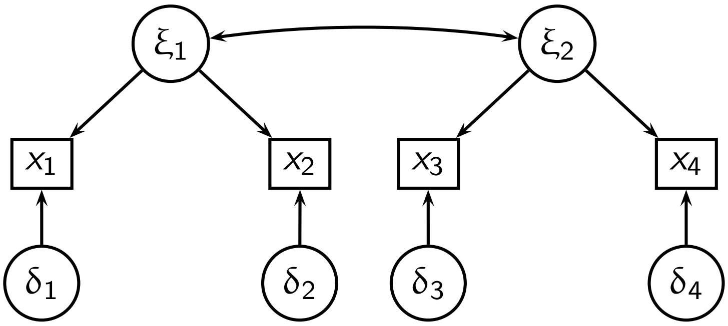
Konfirmatorische Faktorenanalyse

Regressionsmodelle für Politikwissenschaftler

Was ist ein Faktor?

- „Faktor“ oder „latente Variable“
  - nicht direkt beobachtbare Größe
  - die beobachtbare Variablen („Indikatoren“) beeinflußt
- Typische Beispiele: Einstellungen, Persönlichkeitsmerkmale
- Gemeinsame Faktoren (common factors) = Faktoren im eigentlichen Sinn
- Spezifische Faktoren (unique factors) = Störvarianzen, Meßfehler

Wie werden die Beziehungen graphisch dargstellt?



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

Wie unterscheiden sich explorative/konfirmatorische Faktorenanalyse?

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

- Bezieht sich auf explorative (= böse) Faktorenanalyse
  - Soll große Zahl von Items durch wenige Faktoren ersetzen
  - (Häufig Standardprozedur („little jiffy"); Extraktion von Hauptkomponenten, Kaiserkriterium, VARIMAX)
  - Weitere Eigenschaften im Text (vgl. auch Abbildung S. 13)
- Kann häufig selbst dann Strukturen nicht aufdecken, wenn diese in den Daten vorhanden sind („Tom Swift's Electric Factor Analysis Machine"
- **Konfirmatorische** Faktorenanalyse: Prüfung, ob theoretisch sinnvolle Strukturen mit den Daten vereinbar sind
- Meßmodelle, Erweiterung möglich („LISREL"-Modell)

Was gehört zur Spezfikation?

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

Spezfikation durch eine Reihe von Matrizen

Wie sieht die Terminologie aus?

- Betrachten Sie die Abbildung auf S. 19
- 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 ➜** Mittelwert von null, kein Achsenabschnitt notwendig

Was ist ein Erwartungswert?

- Kein Informationsverlust, primäres Interesse an Steigung, nicht an Achsenabschnitt (warum?)
- Erwartungswert, Expectation, E()
- Im Prinzip der Mittelwert einer Zufallsvariablen
  - Für diskrete Variablen die Summe aus den Produkten von Wert und jeweiliger Wahrscheinlichkeit
  - Für stetige Variablen das analoge Integral (Dichte statt Wahrscheinlichkeit)
- Mit Erwartungswerten kann man rechnen
- Zentrieren der Variablen erleichtert das sehr
  - Für zentrierte Zufallsvariable ist Erwartungswert des Produkts gleich der Kovarianz beider Variablen
  - D. h. für einen Vektor von Zufallsvariablen ist die Varianz-Kovarianz-Matrix gleich dem Erwartungswert des Vektorprodukts

Wie sieht die Terminologie aus?

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

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
$\mathbf{x}$	$(q \times 1)$	0	$\boldsymbol{\Sigma} = \mathrm{E}(\mathbf{x}\mathbf{x}')$	$(q \times q)$	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: $\mathbf{x} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta}$
- 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 gemeinsamen/spezifischen Faktoren: $\mathrm{E}(\xi\delta') = 0$

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 (vgl. S. 27-29)
  - 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 **S** als Schätzung für $\mathbf{\Sigma}$
- Schätzung setzt Identifikation voraus
- Identifikation notwendige, aber nicht hinreichende Bedingung für gültige Schätzung

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
- Voraussetzungen/constraints
  - Nicht alle Items können auf alle Faktoren laden → theoretische Überlegungen
  - Die Metrik der gemeinsamen Faktoren muß festgelegt werden

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
- Im Text Größe α ➔ unendliche Zahl von Möglichkeiten ➔ Modell nicht identifiziert
- 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
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 ➔ implizieren eine Varianz-Kovarianz-Matrix $\mathbf{\Sigma}^*$
- „Differenz“ (fitting function) zwischen $\mathbf{\Sigma}^*$ und **S** minimieren
- Verschiedene fitting functions für verschiedene (asymptotisch äquivalente) Verfahren
- Multivariate Normalverteilung der x-Variablen wird vorausgesetzt ➔ Verletzungen, Asymptotik?
- Tests wie gehabt
- Heywood Cases

Was ist das Ergebnis für heute?

- Konfirmatorische Faktorenanalyse ist mathematisches Gegenstück zu Überlegungen aus Test- und Einstellungsforschung
- Kann u. a. den wahren (=fehlerbereinigten) Zusammenhang zwischen Einstellungen schätzen
- Umsetzung komplexer Annahmen über Einstellungen/Indikatoren möglich
- Heute relativ leicht zugänglich durch moderne Programme
- Relativ robust, aber Fußangeln beachten

Übung für heute?

1. Ein Drittel der NPD=Wähler macht in einer Umfrage keine Angabe zu ihrem Wahlverhalten (1), ein weiteres Drittel gibt an nicht zur Wahl zu gehen oder die Union zu wählen (2). Liegt bei (1) beziehungsweise (2) ein missing data Mechanismus vor, und wenn ja, welcher?
2. Aus dem Grundstudium kennen Sie das Konzept der Reliablität, aus dem Aufbaukurs die Attenuation-Formel. Wo zeigt sich Reliablität bei der Faktorenanalyse? Kann diese die Attenuation-Formel ersetzen?