

# Mehr-Ebenen-Analyse I

## Regressionsmodelle für Politikwissenschaftler

Strukturierte Daten  
Konventionelle Lösungen  
Die „echte“ Mehr-Ebenen-Analyse als Alternative  
Fazit

Was ist Struktur?  
Probleme

## Was sind strukturierte Daten?

- ▶ Bisher sind wir stets von einfachen Zufallsstichproben ausgegangen (Registerstichprobe)
- ▶ In der Realität fast immer mehrstufige Zufallsstichproben, z. B. ADM
  1. Auswahl Netz
  2. Auswahl Sampling-Point
  3. Auswahl Adresse/Haushalt
  4. Auswahl Befragungsperson
- ▶ Untersuchungsobjekte (Befragte) sind „nicht unabhängig“ voneinander

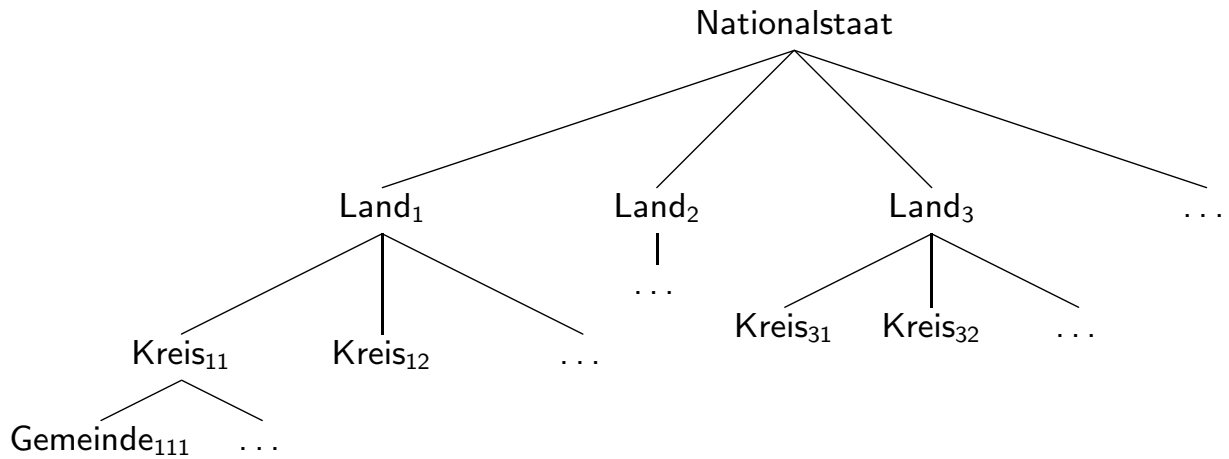
## Was heißt „nicht unabhängig voneinander“

- ▶ Objekte beeinflussen sich gegenseitig
- ▶ Sie werden in gleicher Weise beeinflusst von
  - ▶ gemessenen und nicht
  - ▶ gemessenen Eigenschaften ihres gemeinsamen **Kontextes**
- ▶ Annahme, daß  $\epsilon$  aus identischen und voneinander unabhängigen Normalverteilungen gezogen, wird hinfällig

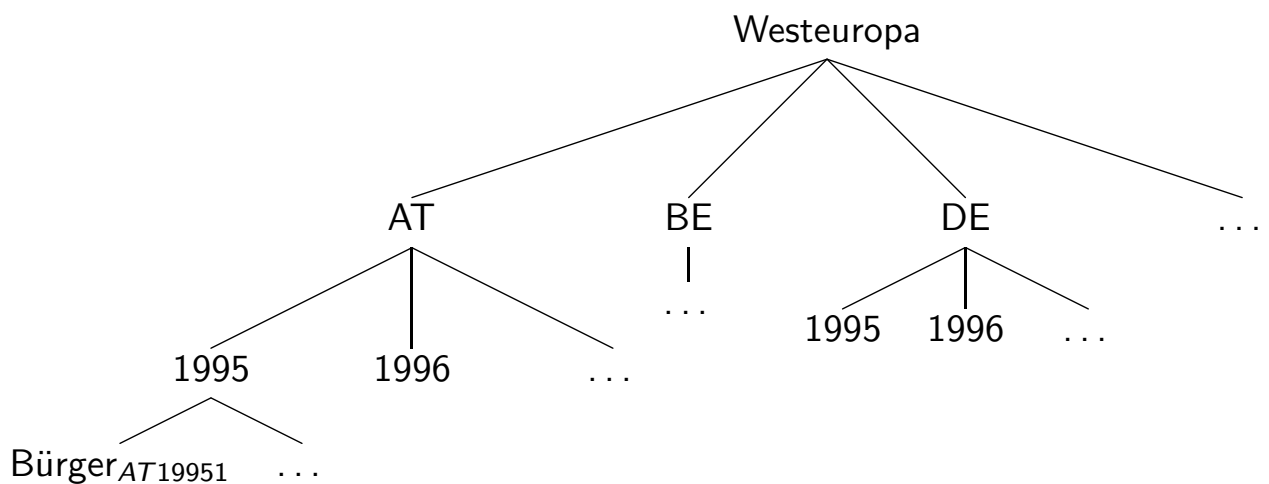
## Gibt es noch andere Quellen für „Struktur“

- ▶ Stichprobenziehung/Datenerhebung eine wichtige Quelle (Klumpenstichproben/Schulforschung)
- ▶ Selbst dann, wenn einfache Zufallsstichprobe vorliegt, häufig Strukturierung der Daten durch „Kontexte“ (Anhänger verschiedener Parteien/Leser verschiedener Zeitungen) → Kontexte müssen nicht räumlich definiert sein
- ▶ Kontexte sind überall: Wähler in Stimmbezirken in Wahlkreisen in Ländern
- ▶ Paneldaten sind auch Kontextdaten (Person als Kontext)

## Muß die Struktur einfach sein?



## Zeit innerhalb von Raum oder umgekehrt?



## Welche Komplikationen sind denkbar?

- ▶ „Gekreuzte“ (cross-classified) Strukturen sind relativ häufig
- ▶ Oft gibt es aber viel komplexere Strukturen (mehrfache/überlappende Mitgliedschaften) etc., die nicht oder nur stark vereinfacht modellierbar sind

## Welche Probleme gibt es?

1. Korrekte Standardfehler
2. Schätzungen für Effekte von Kontexteigenschaften

## Warum machen Standardfehler Probleme?

- ▶ Konventionelle Standardfehler basieren auf iid-Annahme
- ▶ Varianz von  $\epsilon$  ist nicht bekannt, sondern wird aus Residuen geschätzt
- ▶ Kontextstruktur führt zu Ähnlichkeiten zwischen Beobachtungen → Autokorrelation und möglicherweise Heteroskedastizität → Schluß von Residuen nicht gerechtfertigt
- ▶ Standardfehler für *Individualvariablen* zu klein (in Abhängigkeit von Homogenität der Kontexte) → zu schmale Konfidenzintervalle, zu laxen Signifikanztests

## Wie lassen sich Kontexteigenschaften systematisch beschreiben?

Ebene	1		2		3		...
Typ	global	→	analytisch				
	relational	→	strukturell				
	kontextuell	←	global	→	analytisch		
			relational	→	strukturell		
			kontextuell	←	global	→	
					relational	→	
					kontextuell	←	

←: Aggregation

→: Disaggregation

## Wie lassen sich Kontexteigenschaften systematisch beschreiben?

- ▶ Globale Variablen: Beziehen sich auf eine Einheit auf einer Ebene
- ▶ Analytische Variablen: Aggregation (z. B. Mittelwertbildung) über Objekte der unteren Ebene
- ▶ Relationale Variablen: Beziehen sich auf Verhältnis zwischen einem Objekt und anderen Objekten derselben Ebene (Soziometrie)
- ▶ Strukturell: Aggregation von relationalen Variablen der untergeordneten Ebene
- ▶ Kontextuell: Globale Eigenschaften der übergeordneten Ebene

## Ganz schlecht: Struktur ignorieren, Kontextvariablen verwenden

- ▶ Theoretische und praktische Relevanz von Kontexteigenschaften seit langem bekannt
- ▶ Klassische „Lösung“:
  - ▶ Standard-Regressionsverfahren
  - ▶ Kontextvariablen werden „drankumuliert“, d. h. wie Individualvariablen behandelt
- ▶ Beispiel bei Moulton (1990): Abhängigkeit individueller Löhne von relevanten und irrelevanten Eigenschaften des Staates (abhängige Variable: Individualdaten, unabhängige Variable: Aggregatdaten)
- ▶ Standardfehler für die Aggregatdaten viel zu klein, weil diese als *unabhängige* Beobachtungen betrachtet werden (Fallzahlen im Bereich von Tausenden, obwohl  $n \approx 50$ )
- ▶  $R^2$  ist zu hoch

## Warum ist das so?

- ▶ Siehe oben
- ▶ Betroffen sind nicht nur Aggregat-, sondern auch Individualmerkmale
- ▶ Entspricht dem Design-Effekt (Kish) bei mehrstufiger Zufallsauswahl
- ▶ Um so gravierender
  - ▶ je kleiner die Zahl der Kontexte in Relation zur Zahl der Befragten ist
  - ▶ je homogener die Kontexte sind
- ▶ Kontextvariablen besonders problematisch, da hier perfekte Homogenität vorliegt

## Wie läßt sich das Problem der inkorrekten Standardfehler lösen?

- ▶ Schätzungen für Parameter unverzerrt (wenn auch nicht notwendigerweise optimal)
- ▶ Spezielle Korrekturformeln für mehrstufige Zufallsauswahlen
- ▶ Generelle: „Robuste“ Standardfehler, (relativ) unempfindlich gegen Heteroskedastizität und Autokorrelation
- ▶ Gute Lösung, wenn nur zwei Ebenen und Kontextstruktur bekannt

## Was tut man (konventionell) mit den Kontextvariablen?

- ▶ Außer den Standardfehlern gibt es noch weitere Komplikationen bei Mehr-Ebenen-Designs
  1. Kontextvariablen können eine systematische Wirkung haben (wie?)
  2. Zufällige Einflüsse auf der Kontextebene/die Konstante kann über die Kontexte variieren
  3. Die Wirkung von Individualvariablen kann systematisch oder zufällig über Kontexte variieren
  4. Cross-Level-Interaktionen

## Wie schlägt sich das nieder?

1. Koeffizient (Steigung) für Kontextvariablen
2. Mehrere Achsenabschnitte (oder ein Achsenabschnitt plus Verteilung)
3. Interaktionen mit Steigungskoeffizient oder Steigung plus Verteilung
4. Interaktionsterme

## Ein Beispiel?

- ▶ Sie untersuchen rechtsextreme Einstellungen ( $y$ ) an Befragten, die in den 70 Landtagswahlkreisen Baden-Württembergs leben
- ▶  $y$  hängt (in der Realität) ab von
  1. der Individualvariable „Bildung“ ( $x_1$  beziehungsweise  $x_{ij}$ )
  2. zufälligen Einflüssen auf der Individualebene ( $\epsilon$  beziehungsweise  $\epsilon_{0ij}$ )
  3. der Kontextvariable „Ausländeranteil“ ( $x_2$  beziehungsweise  $x_j$ )
  4. sowie zufälligen Einflüssen auf der Kontextebene ( $u_{0j}$ )

## Was kann man tun? → separate Analyse (I)

- ▶ Separate Analyse der Daten aus den 70 Kontexten (70 einzelne Regressionen schätzen)
- ▶ Hier wegen großer Fallzahl unpraktisch
- ▶ Variierende Achsenabschnitte: zufällige Kontexteinflüsse; variierende Steigungen für Individualvariablen
- ▶ Bei kleiner Zahl von Kontexten ehrbare Alternative

## Welche Modellierungsmöglichkeiten gibt es?

Eigenschaften	separate Analyse(I)	dummy variable(II)	interaction(III)	ME-Analyse (IV)
systematische Kontexteinflüsse modellierbar	nein	nein	ja	ja
zufällige Kontexteinflüsse modellierbar	ja	ja	nein	ja
Variation individueller Einflüsse über Kontexte modellierbar	ja (fix)	ja (fix)	nein	ja (fix oder random)
Cross-Level Interaktionen	nein	nein	ja	ja
korrekte Standardfehler	ja	ja (robust)	ja (robust)	ja (modellbasiert oder robust)
typische Zahl der Kontexte	drei bis zehn	zehn bis 20	zehn bis 20	(sehr viel) mehr als 20

## Welche Modellierungsmöglichkeiten gibt es?

Eigenschaften	separate Analyse(I)	dummy variable(II)	interaction(III)	ME-Analyse (IV)
systematische Kontexteinflüsse modellierbar	nein	nein	ja	ja
zufällige Kontexteinflüsse modellierbar	ja	ja	nein	ja
Variation individueller Einflüsse über Kontexte modellierbar	ja (fix)	ja (fix)	nein	ja (fix oder random)
Cross-Level Interaktionen	nein	nein	ja	ja
korrekte Standardfehler	ja	ja (robust)	ja (robust)	ja (modellbasiert oder robust)
typische Zahl der Kontexte	drei bis zehn	zehn bis 20	zehn bis 20	(sehr viel) mehr als 20

## Was kann man tun? → separate Analyse (I)

- ▶ Separate Analyse der Daten aus den 70 Kontexten (70 einzelne Regressionen schätzen)
- ▶ Hier wegen großer Fallzahl unpraktisch
- ▶ Variierende Achsenabschnitte: zufällige Kontexteinflüsse; variierende Steigungen für Individualvariablen
- ▶ Bei kleiner Zahl von Kontexten ehrbare Alternative

## Was kann man tun? → dummy variable model (II)

- ▶ Gemeinsame Analyse aller Fälle + 1 Dummy pro Kontext (oder 1 Dummy weniger + Konstante)
- ▶ Hier sehr viele Dummies
- ▶ Gegebenenfalls sehr viele Interaktionen
- ▶ Bei kleinerer Zahl von Kontexten durchaus sinnvolle Alternative, wenn keine relevanten Kontextvariablen

## Welche Modellierungsmöglichkeiten gibt es?

Eigenschaften	separate Analyse(I)	dummy variable(II)	interaction(III)	ME-Analyse (IV)
systematische Kontexteinflüsse modellierbar	nein	nein	ja	ja
zufällige Kontexteinflüsse modellierbar	ja	ja	nein	ja
Variation individueller Einflüsse über Kontexte modellierbar	ja (fix)	ja (fix)	nein	ja (fix oder random)
Cross-Level Interaktionen	nein	nein	ja	ja
korrekte Standardfehler	ja	ja (robust)	ja (robust)	ja (modellbasiert oder robust)
typische Zahl der Kontexte	drei bis zehn	zehn bis 20	zehn bis 20	(sehr viel) mehr als 20

## Was kann man tun? → dummy variable model (II)

- ▶ Gemeinsame Analyse aller Fälle + 1 Dummy pro Kontext (oder 1 Dummy weniger + Konstante)
- ▶ Hier sehr viele Dummies
- ▶ Gegebenenfalls sehr viele Interaktionen
- ▶ Bei kleinerer Zahl von Kontexten durchaus sinnvolle Alternative, wenn keine relevanten Kontextvariablen

## Was kann man tun? → interaction model (II)

- ▶ Gemeinsame Analyse aller Fälle + Kontextvariablen (Zahl beschränkt)
- ▶ In der Politikwissenschaft sehr gebräuchliches Verfahren (Länderanalysen)
- ▶ Bei kleinerer Zahl von Kontexten durchaus sinnvolle Alternative, *wenn keine zufälligen Kontexteinflüsse/persistenten Kontexteinflüsse* („unit effects“ → massive Verzerrungen möglich)
- ▶ Probleme?

## Welche Modellierungsmöglichkeiten gibt es?

Eigenschaften	separate Analyse(I)	dummy variable(II)	interaction(III)	ME-Analyse (IV)
systematische Kontexteinflüsse modellierbar	nein	nein	ja	ja
zufällige Kontexteinflüsse modellierbar	ja	ja	nein	ja
Variation individueller Einflüsse über Kontexte modellierbar	ja (fix)	ja (fix)	nein	ja (fix oder random)
Cross-Level Interaktionen	nein	nein	ja	ja
korrekte Standardfehler	ja	ja (robust)	ja (robust)	ja (modellbasiert oder robust)
typische Zahl der Kontexte	drei bis zehn	zehn bis 20	zehn bis 20	(sehr viel) mehr als 20

## Was kann man tun? → interaction model (II)

- ▶ Gemeinsame Analyse aller Fälle + Kontextvariablen (Zahl beschränkt)
- ▶ In der Politikwissenschaft sehr gebräuchliches Verfahren (Länderanalysen)
- ▶ Bei kleinerer Zahl von Kontexten durchaus sinnvolle Alternative, *wenn keine zufälligen Kontexteinflüsse/persistenten Kontexteinflüsse* („unit effects“ → massive Verzerrungen möglich)
- ▶ Probleme?

## Unter welchen Voraussetzungen ist Anwendung möglich/sinnvoll?

- ▶ Große Zahl von Kontexten (Mindestwert 30 bis 50, für Varianzkomponenten nach Möglichkeit mehr)
- ▶ Typischerweise 2 bis 100 Objekte pro Kontext (ansonsten oft kein Unterschied zu konventionellen Analysen)
- ▶ **Kontexte sind als Zufallsstichprobe aus einer großen Grundgesamtheit von Kontexten zu interpretieren**
- ▶ Der letzte Punkt wird häufig ignoriert

## Was sind die Grundannahmen des Modells?

- ▶ Variation von Achsenabschnitten/Steigungen über Kontexte ...
- ▶ ... kann durch Zufallsvariable(n) beschrieben werden
- ▶ Statt 70 separater Modellschätzungen Schätzung von Mittelwert und Varianz *der Modellparameter* („random effects“) → **sehr** effizient
- ▶ Korrekte Standardfehler, Kontexteffekte und Cross-Level-Interaktionen machbar
- ▶ Shrinkage Estimates

## Welche Modellierungsmöglichkeiten gibt es?

Eigenschaften	separate Analyse(I)	dummy variable(II)	interaction(III)	ME-Analyse (IV)
systematische Kontexteinflüsse modellierbar	nein	nein	ja	ja
zufällige Kontexteinflüsse modellierbar	ja	ja	nein	ja
Variation individueller Einflüsse über Kontexte modellierbar	ja (fix)	ja (fix)	nein	ja (fix oder random)
Cross-Level Interaktionen	nein	nein	ja	ja
korrekte Standardfehler	ja	ja (robust)	ja (robust)	ja (modellbasiert oder robust)
typische Zahl der Kontexte	drei bis zehn	zehn bis 20	zehn bis 20	(sehr viel) mehr als 20

## Wie sieht das Modell aus?

- ▶ Viele Varianten mit unterschiedlicher Komplexität
- ▶ Relativ einfach: „Random Intercept“
- ▶ In unserem Beispiel: Rechtsextremismus hängt ab von individuellen Eigenschaften, das mittlere Niveau variiert aber über die Kreise (z. B. wegen politischer Traditionen)
- ▶ Dummy variable model mit 69/70 Dummies?
- ▶ Oder lieber so?

## Wie sieht das random intercept model aus?

$$y_{ij} = \beta_{0ij} + \beta_1 x_{1ij} + \beta_2 x_{2j} \quad (1)$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij} \quad (2)$$

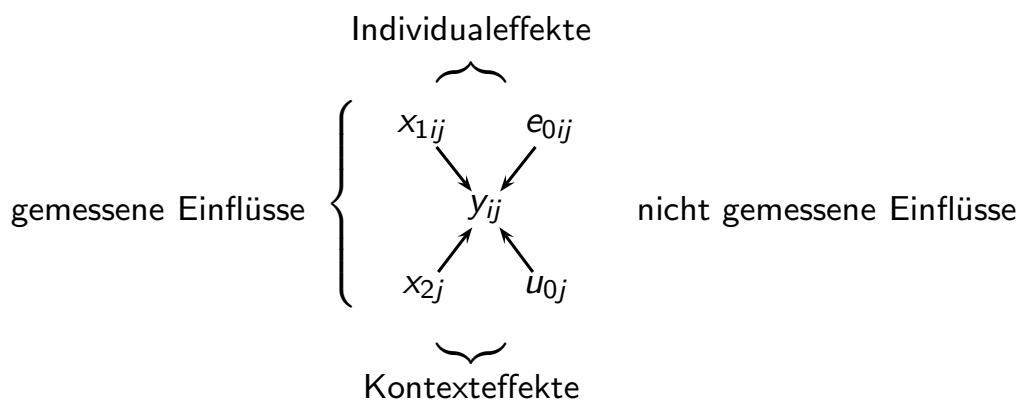
mit

$$(u_{0j}) \sim (0, \mathbf{\Omega}_u) : \mathbf{\Omega}_u = (\sigma_{u0}^2) \quad (3)$$

$$(e_{0ij}) \sim (0, \mathbf{\Omega}_e) : \mathbf{\Omega}_e = (\sigma_{e0}^2) \quad (4)$$

## Was bedeutet das alles?

- ▶  $y_{ij}$ : Wert der abhängigen Variablen beim Befragten mit der Ordnungsnummer  $i$  im Kontext  $j$ , hängt ab von fixen Regressionsparametern ( $\beta_1, \beta_2$ )
- ▶  $u$  beziehungsweise  $e$ : Zufällige Einflüsse auf Kontext- und Individualebene
- ▶  $\Omega$ : Varianz-Kovarianz-Matrizen für zufällige Einflüsse
- ▶ Zufällige Einflüsse auf Kontextebene sorgen für Varianz von  $\beta_{0ij}$  (random intercept)



## Wie lassen sich die Ergebnisse für heute zusammenfassen?

- ▶ Strukturierte Daten sind allgegenwärtig
- ▶ Wenn Struktur ignoriert wird, sind Standardfehler meist zu optimistisch
- ▶ Gilt besonders (aber nicht nur) für Kontextvariablen
- ▶ In vielen Standardsituation ist Einsatz konventioneller Methoden sinnvoll/geboten (z. B. Analyse von ESS- oder vergleichbaren Daten)
- ▶ Mehr-Ebenen-Analyse ist sehr flexibel, aber ebenfalls mit vielen Annahmen behaftet
- ▶ Kein Wundermittel und nicht immer angebracht/notwendig

## Was ist die Übung für heute?

Sie untersuchen den Zusammenhang zwischen der individuellen formalen Bildung ( $x$ ), dem Ausländeranteil auf Kreisebene ( $z$ ) und der individuellen Ausländerfeindlichkeit ( $y$ ) in Ost- und Westdeutschland

1. Was spricht dagegen, ein Drei-Ebenen-Modell (Region – Kreis – Person) zu konstruieren? Welche einfachere Lösung ist sinnvoller?
2. Bitte schreiben Sie diese einfachere Lösung als Formel nieder. Gehen Sie dabei von fixen Effekten für Bildung und Ausländeranteil aus.
3. Wären random effects für diese beiden Variablen möglich/sinnvoll?
4. Eine Kommilitonin macht Sie darauf aufmerksam, daß der Ausländeranteil in Ost- beziehungsweise Westdeutschland einen unterschiedlichen Effekt haben dürfte (warum?). Wie läßt sich dies im Modell berücksichtigen?

Bitte schicken Sie mir die Lösung wie üblich bis nächsten Mittwoch per email.